Aus dem Max-Delbrück-Centrum für molekulare Medizin
der Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

A complex interplay of regulatory domains controls cell cycle dependent subnuclear localization of DNMT1 and is required for the maintenance of epigenetic information

DISSERTATION

Zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

im Fach BIOLOGIE

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von

Hariharan P. Easwaran

geb. am 16.09.1974 in Mumbai (Indien)

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Michael Linscheid

Gutachter:

Prof. Dr. Harald Saumweber

Prof. Dr. Heinrich Leonhardt

Prof. Dr. Peter B. Becker

Eingereicht: 23. April 2003

Tag der mündlichen Prüfung: 1. October 2003

Table of contents

Abbreviations

A

Ab	antibody
AdoHCy	S-adenosyl-L-homocysteine
AdoMet	S-adenosyl-L-methionine
ADP	adenosine diphosphate
Ar	argon
5AzaC	5-azacytidine

B

BAH	bromo adjacent homology domain (also called PBHD)
BDGP	Berkeley Drosophila Genome Project
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BrdU	5-bromodeoxyuridine

C

C2C12	mouse myoblast cell line
$CaPO_4$	calcium phosphate
CCD	charge coupled device
cdk2	cyclin dependent kinase-2
CENP-B	centromere protein B
CMT	chromomethyltransferase
COS-7	african green monkey kidney fibroblast-like cells transformed with SV40 T antigen
CpA	cytosine-adenine doublet
CpG	cytosine-guanine doublet
Cy5	indodicarbocyanine

D

DABCO	1,4-diazabicyclo[2,2,2]-octane
DIC	differential interference contrast
DMAP	DNMT1-associated protein
DmDNA Ligase I	*Drosophila melanogaster* DNA Ligase I
DMEM	Dulbecco's Modified Eagle Medium
DmPCNA	*Drosophila melanogaster* PCNA
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNA MTase	DNA methyltransferase
DNase	deoxyribonuclease
DsRed	red fluorescent protein from *Discosoma* species

E

EC	embryonic carcinoma
EDTA	ethylenediaminetetraacetic acid
ES	embryonic stem cells
EST	expressed sequence tag
E-value	expectation value

F

FCS	fetal calf serum
FITC	fluorescein isothiocyanate

G

G1	gap 1 phase
G2	gap 2 phase
G418	geneticin
GFP	green fluorescent protein

H

H167V	mutation of histidine at position 167 to valine in DNMT1
H2A	histone H2A
H2B	histone H2B
H3	histone H3
H3K9Me	histone H3 methylated at lysine at position 9
H4	histone H4
HDAC	histone deacetylase
HeNe	helium neon
HP-1	heterochromatin protein 1
hr	hours
HsDNA Ligase I	human DNA Ligase I
HsPCNA	human PCNA

I

ICF	immunodeficiency, centromeric instability and facial anomalies

K

kbp	kilobase pairs
kDa	kilo Dalton

M

M	mitosis
MBD	methyl-CpG-binding domain
Mbp	megabase pairs
5mC	C5-methylcytosine
MeCP	methyl-CpG-binding protein
mM	millimolar

N

neo^r	neomycin resistance gene
NLS	nuclear localization signal
NuRD	nucleosome remodelling and histone deacetylation complex

P

PAGE	polyacrylamide gel electrophoresis

PBD	PCNA binding domain
PBHD	polybromo homology domain (also called BAH domain)
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
pol	DNA polymerase
PSI-BLAST	position-specific iterated BLAST
PVDF	polyvinylidene fluoride
R	
RF	replication foci
RFP	red fluorescent protein
RFTS	replication factory targeting sequence
RPA-70	70 kDa subunit of replication protein A
S	
S	Synthesis phase
S2	*Drosophila* Schneider's line 2 cells
S514A	mutation of serine at position 514 to alanine in DNMT1
S514D	mutation of serine at position 514 to aspartate in DNMT1
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
ScPCNA	*Saccharomyces cerevisiae* PCNA
SDS	sodium dodecylsulphate
SV40	simian virus 40
T	
293T	human embryonic kidney cells transformed with SV40 T antigen
TR	texas red
Tris	tris (hydroxymethyl) aminomethane
TS	targeting sequence
W	
WSTF	Williams syndrome transcription factor
Y	
YFP	yellow fluorescent protein
Z	
Zn	zinc
ß	
ß-gal	betagalactosidase epitope
µ	

µm	micrometer
µM	micromolar

Summary

DNA methylation constitutes an essential epigenetic mark controlling chromatin organization and gene regulation in higher eucaryotes, which has to be duplicated together with the genetic information at every cell division cycle. In mammals duplication of DNA methylation is mediated by DNA methyltransferase-1 (DNMT1). It associates with sites of nuclear DNA replication, called replication foci (RF), and thereby couples maintenance of DNA methylation to DNA duplication. In this work, we have analyzed the role of regulatory sequences in the N-terminal domain of DNMT1 in controlling its subnuclear localization throughout the cell cycle, and the evolutionary conservation of these sequences and of the mechanisms that mediate association of proteins with RF.

We provide evidence that DNMT1 shows dynamic subnuclear distribution that is controlled by the regulatory sequences depending on the cell cycle stage. To determine the subnuclear distribution of DNMT1 throughout the cell cycle, an RFP-Ligase fusion protein was developed as a marker that allows identification of the cell cycle stage in live cells. Various DNMT1 mutants fused to GFP were coexpressed with RFP-Ligase and imaged by 4-dimensional live cell microscopy during an entire cell cycle. The PBD (PCNA binding domain) drives the localization of DNMT1 at RF throughout S phase and the TS (targeting sequence) mediates retention of DNMT1 only at the late replicating pericentric heterochromatin from late-S phase until early-G1. In contrast, the PBHD (polybromo homology domain) seems to be required for unloading DNMT1 from the pericentric regions in G1. Overexpression of the TS to interfere with this association lowers cell viability and induces the formation of micronuclei and coalescence of centromeric heterochromatin. These results bring forth a novel function of the TS in mediating association of DNMT1 with pericentric heterochromatin from late-S phase through G2 until mitosis, which is important for maintenance of DNA methylation, and heterochromatin structure and function. Database searches indicate that the TS is a domain unique to the DNMT1 family of proteins. Amongst the DNMT1 family, only the metazoan DNMT1 proteins have the PBD. This suggests that coupling of maintenance of DNA methylation with DNA replication occurs only in metazoans, while plants and fungi have alternative mechanisms that maintain DNA methylation patterns, probably mediated by the TS.

The evolutionary conservation of the mechanisms by which proteins associate with RF in mammalian cells was directly tested by analyzing the ability of mammalian replication proteins PCNA and DNA Ligase I as well as DNMT1 to associate with RF in *Drosophila* cells. Of all the proteins tested, only PCNA associated with RF while the others showed diffused nuclear distribution although they contain a functional PBD. Surprisingly, *Drosophila* DNA Ligase I associates with RF in mammalian but not in *Drosophila* cells. These results suggest differences in the dynamics and organization of the replication machinery in these distantly related organisms, which correlates with the increased size and complexity of mammalian genomes.

Zusammenfassung

DNA-Methylierung spielt eine wichtige Rolle bei der Kontrolle der Chromatinorganisation und Genregulation in höheren Eukaryoten und muss zusammen mit der genetischen Information in jedem Zellzyklus dupliziert werden. Bei Mammalia wird DNA durch die DNA-Methyltransferase 1 (DNMT1) methyliert, die dabei mit nukleären Replikationsstellen (RF) assoziiert und so die Erhaltung des Methylierungsmusters mit der Duplikation der DNA verbindet. In dieser Arbeit wurden die Funktion der regulatorischen Sequenzen in der N-terminalen Domäne von DNMT1 bei der Kontrolle ihrer subnukleären Lokalisierung während des Zellzyklus und die evolutionäre Konservierung dieser Sequenzen, sowie die Mechanismen die eine Assoziation von Proteinen mit RF vermitteln, untersucht.

Es konnte gezeigt werden, dass DNMT1 eine dynamische Verteilung im Kern aufweist, die durch regulatorische Sequenzen zellzyklusabhängig gesteuert wird. Um die subnukleäre Verteilung von DNMT1 während des Zellzyklus zu untersuchen, wurden RFP-Ligase Fusionsproteine hergestellt, die als Marker für die Identifikation von Zellzyklusstadien in lebenden Zellen dienen. Verschiedene, mit GFP fusionierte DNMT1 Mutanten wurden zusammen mit RFP-Ligase exprimiert und über einen ganzen Zellzyklus hinweg mit 4-dimensionaler Lebendzellmikroskopie verfolgt. Die PBD (PCNA-Bindungsdomäne) bewirkt die Lokalisierung von DNMT1 an RF während der S-Phase, und die TS (*targeting sequence*) vermittelt die Retention von DNMT1 an spät replizierendem Heterochromatin von der späten S- bis zur frühen G1-Phase. Im Gegensatz dazu scheint die PBHD (Polybromohomologiedomäne) für die Freisetzung von DNMT1 von perizentrischen Regionen während der G1-Phase notwendig zu sein. Eine Überexpression der TS zu Störung dieser Assoziation, senkt die Überlebensrate der Zellen und fördert die Bildung von Mikronuklei sowie die Verschmelzung von zentromerem Heterochromatin. Diese Ergebnisse zeigen eine neue Funktion für die TS bei der Assoziation von DNMT1 mit perizentrischem Heterochromatin von der später S- über die G2-Phase bis hin zur Mitose, die eine wichtige Voraussetzung für die Erhaltung der DNA-Methylierung und Heterochromatinstruktur und –funktion ist. Datenbankanalysen zeigten, dass es sich bei der TS um eine einzigartige Domäne innerhalb der DNMT1 Proteinfamilie handelt. Innerhalb der DNMT1 Familie besitzen nur die DNMT1 Proteine der Metazoen die PBD. Das lässt vermuten, dass die Verknüpfung von Beibehaltung der DNA Methylierung mit der DNA Replikation nur in Metazoen auftritt, während in Pflanzen und Pilzen alternative Mechanismen zur Aufrechterhaltung des Methylierungsmusters, wahrscheinlich vermittelt durch die TS, zur Anwendung kommen.

Die evolutionäre Konservierung von Mechanismen, zur Assoziation von Proteine mit RF in Säugerzellen, wurde durch die Analyse der Säugerproteine PCNA, DNA Ligase I und DNMT1 in *Drosophila*-zellen direkt getestet. Von allen untersuchten Proteinen assoziiert nur PCNA mit RF, während die anderen nur eine diffuse Verteilung innerhalb des Kerns zeigten, obwohl sie eine funktionale PBD enthalten. Überraschenderweise assoziierte auch die *Drosophila* DNA Ligase I in Säugerzellen nicht aber in *Drosophila*-zellen mit RF. Diese Ergebnisse weisen auf Unterschiede in der Dynamik und dem Aufbau der Replikationsmaschinerie in diesen entfernt verwandten Organismen hin, was mit der Vergrösserung und höheren Komplexität des Säugergenoms korreliert.

1. Introduction

1.1. Genetic and Epigenetic information

The hereditary material in all known biological systems (except prions) consists of nucleic acids, wherein the information is encoded in the sequence of four nitrogenous bases, viz. adenine, guanine, cytosine, thymidine/uridine. This forms the genetic information which codes for the complete repertoire of functional molecules. However, not all functions are required at all times and only a subset of genes is expressed at any given time. Thus expression of genes is regulated in such a way that some genes are kept repressed and called upon for action by appropriate environmental cues. As we move from procaryotes to eucaryotes and lower eucaryotes to higher eucaryotes, the genome size and gene number increases, as does complexity of the organisms. In higher eucaryotes, complex functions are performed by specialized cellular systems, for example the nervous system and muscular system, each requiring expression of unique sets of genes which make up the distinct phenotypes. Thus even though cells from different tissues within the same organism have the same genetic information, they have different gene expression programs. Once a specific gene expression program is established for a tissue, all progenitor cells in that tissue faithfully inherit both the genetic information and the gene expression program to maintain the integrity of the tissue. To fulfill these requirements, organisms have evolved novel mechanisms to 'mark' genes as transcriptionally on or off in such a way that the sequence of the four bases is not changed. Such 'mark' forms the epigenetic information, which along with the genetic information, is passed on to the daughter cells when a cell divides.

Here we examine some of the mechanisms by which transfer of genetic and epigenetic information occurs during cell division, which are essential for stable inheritance of phenotype by daughter cells.

1.2. Replication of genetic information

1.2.1. DNA replication origins

Studies on procaryotic DNA replication have led to the replicon model which provided a framework for understanding the process of initiation of DNA replication and its regulation. According to this model, a replicon is a genetic element that is replicated from a single origin of replication (the replicator), which is recognized by a specific positive regulatory protein (the initiator) (Jacob et al., 1964). In procaryotes, the whole genome makes up one replicon in that replication initiates at a single defined origin from where the whole genome is replicated. Identification of origins in eucaryotes extended this model to the eucaryotic system. However, in contrast to procaryotes, the eucaryotic genome contains multiple origin sites where DNA replication initiates. In the unicellular eucaryote *S. cerevisiae*, origins of replication are composed of conserved DNA sequences spread over 100 – 150 bp called autonomously replicating sequence (ARS) (Stinchcomb et al., 1979). The *S. cerevisiae* genome consists of numerous such elements and DNA replication initiates at a subset of them (Newlon and Theis, 1993). Metazoan origins of replication are

very complex as they cannot be defined in terms of DNA sequence and virtually any sequence could behave as an origin in *Xenopus* egg extracts and *Drosophila* cells (Cox and Laskey, 1991) (Smith and Calos, 1995) (Gilbert, 1998). Even though DNA replication initiates at defined sites in the genome during normal cellular DNA replication (Giacca et al., 1994) (Abdurashidova et al., 2000), these sites do not show any sequence conservation. Moreover, during early development there seems to be no sequence requirement for defining origin function, as observed in the early embryos of *D. melanogaster* and *X. laevis* that can virtually replicate any DNA sequence (Blow, 2001). Thus, it seems that even though the replicon model operates in eucaryotes, the requirement for specific sequence elements has been replaced by as yet unidentified feature(s). There have been various hypotheses to explain the specificity of replication origins in higher eucaryotes, including role of nuclear matrix, transcriptional activity of genes, chromatin structure, DNA sequence and DNA methylation (reviewed in (DePamphilis, 1999). By whatever mechanism origins are selected in higher eucaryotes, DNA replication consistently initiates at these sites, and exact copies of the whole genome are synthesized during normal cell proliferation.

1.2.2. DNA replicates in discrete sites in the nucleus called replication foci

Cells grown in the presence of halogenated nucleotide analogues (e.g. BrdU) incorporate the analogue into genomic DNA during the process of DNA replication. The halogenated nucleotides incorporated into DNA can be detected by immunocytochemical techniques that allow investigation of DNA replication at the cellular level (reviewed in (Leonhardt and Cardoso, 1995). Using such an approach it was observed that in cells grown in the presence of BrdU for a short period of time (pulse labeling), the BrdU signal was present at discrete sites in the nucleus (Nakamura et al., 1986). Since these sites correspond to regions of the chromosome which had incorporated BrdU by the process of DNA replication, these sites are called replication foci (RF). Extraction of nuclei with physiological salt concentrations, non-ionic detergents and endonucleases retains the labeled replicated DNA as discrete sites in the nuclear matrix indicating that the RF is stably tethered to the nuclear matrix (Jackson and Cook, 1986). It is suggested that each replication focus is constituted of a subchromosomal domain of adjacent replicons from the same chromosomal region that replicate together. These domains are stably maintained as single units through multiple rounds of cell division, and they occupy stable positions in the nucleus (Sparvoli et al., 1994) (Jackson and Pombo, 1998) (Ma et al., 1998) (Sadoni et al., 1999).

The sites of DNA replication can also be visualized as discrete foci where proteins involved in DNA replication and associated activities assemble during S phase (reviewed in (Leonhardt and Cardoso, 1995). The first protein to be identified at these foci during S phase is PCNA (Celis and Celis, 1985) (Bravo and Macdonald-Bravo, 1987). Many other proteins have since been shown to be associated with these sites, including the maintenance DNA methyltransferase (DNMT1) (Leonhardt et al., 1992), DNA polymerase α (Hozak et al., 1993), DNA Ligase I (Cardoso et al., 1997), RPA-70 and cell cycle regulators cyclin A and cdk2 (Cardoso et al., 1993), and DNA repair factors like uracil-DNA-glycosylase (Otterlei et al., 1999). Nuclear matrix preparations retain some replication factors, like PCNA, and polymerizing activity at these sites in the nuclear matrix indicating that the replication factors form insoluble complexes at the sites of DNA replication during S phase (reviewed in (Leonhardt

and Cardoso, 1995). Based on these findings, these sites were called replication factories or replication foci (we use the latter term throughout this work and abbreviate it as RF). Thus, RF can be defined as microscopically visible subchromosomal domains in the nucleus tethered to the nuclear matrix where DNA is being replicated and replication factors are concentrated.

It is estimated that each replication focus comprises on average 1Mbp of DNA. Most of the mammalian replicons are in a size range of 75-150 Kbp. Thus, it is estimated that each replication focus consists of at least 10 replicons (reviewed in (Berezney et al., 2000). Since DNA replication occurs bidirectionally in each replicon, on average it is expected that each replication focus consists of 20 replication forks. However, the RF are heterogenous in their size (0.25 µm to several µm) and lifetimes (30 min to over 3 h) (Leonhardt et al., 2000a), and the number of replication forks might vary greatly. Based on studies on mouse 3T3 cells, it is estimated that the whole genome is replicated in ~10,000 RF (Ma et al., 1998). These RF are all not active at the same time and rather follow a defined pattern of activation throughout S phase as discussed below.

1.2.3. Temporal and spatial order of DNA replication

Specific regions of eucaryotic chromosomes replicate at defined times during S phase and this timing is correlated to the transcriptional competence of the DNA elements (Goldman et al., 1984) (Hatton et al., 1988). In general, it was observed that transcriptionally active regions replicate during early-S phase and the transcriptionally inactive regions replicate at any interval during S phase. The more condensed heterochromatin, which is typically found at centromeric regions, has been shown to replicate during late-S phase (Ten Hagen et al., 1990). At the cellular level, this temporal program of replication is reflected in the number, size and location of RF throughout S phase (Nakayasu and Berezney, 1989) (O'Keefe et al., 1992). This results in formation of distinct spatial patterns of RF in the nucleus during S phase (Fig. 1.1). Typically, during early-S phase, when most of the euchromatin replicates, the RF are distributed throughout the nucleus. During early to mid-S phase (called mid-S phase throughout this work), the RF can be observed as discrete perinuclear and perinucleolar sites. During mid to late-S phase (called late-S phase throughout this work), the highly condensed heterochromatin replicates and RF form large "donut" shaped structures. Existence of the various patterns of RF has been shown in living mammalian cells by labeling the RF with GFP fused to core replication factors, like DNA Ligase I or PCNA (Cardoso et al., 1997) (Leonhardt et al., 2000a). Such studies have shown that replication proteins continuously assemble at many subchromosomal domains to form RF, and continuously disassemble from sites that have completed replication (reviewed in (Cardoso et al., 1999). This cycle of assembly and disassembly progresses throughout S phase giving rise to the distinct patterns of RF.

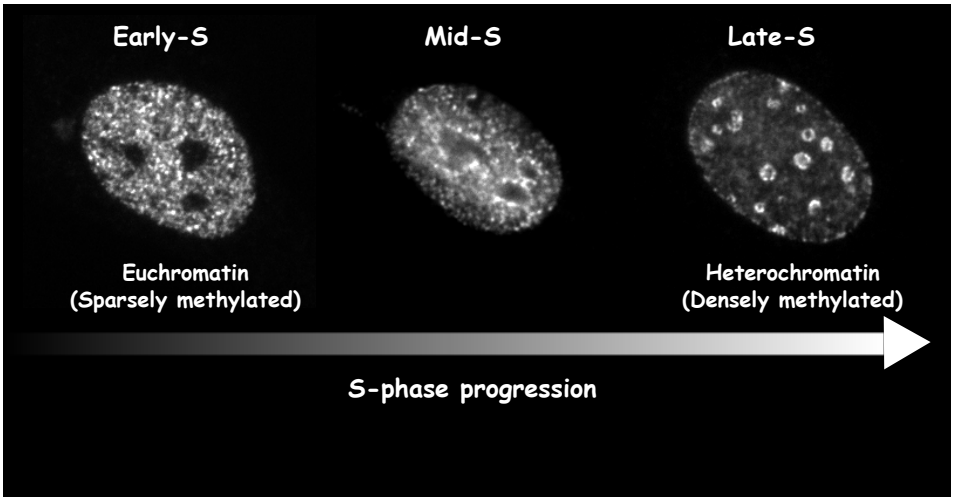


Fig. 1.1. Distinct spatio-temporal patterns of RF in S phase mammalian nuclei. *Mouse cells in S phase displaying the distinct patterns of RF. Euchromatin replicates in early-S phase and heterochromatin replicates in late-S phase The RF are labeled with a GFP-PCNA fusion protein.*

The discovery of RF indicated that replication is not an event occuring randomly throughout the nucleus, but an event organized in discrete domains in time and space. Formation of such functional domains could have consequences on regulation of replication and formation/maintenance of chromatin states. For an understanding of the process of replication as it occurs *in vivo*, it is essential to define the mechanisms by which various proteins assemble at the RF (Leonhardt et al., 2000b).

1.2.4. Specific protein sequences mediate assembly of replication factors at RF

Replication factors undergo dynamic re-distribution in the nucleus during S phase. In the G1 phase, most replication proteins are diffused in the nucleus. On entry into S phase, replication proteins form punctate patterns that co-localize with RF (reviewed in (Leonhardt and Cardoso, 1995). This association with RF is mediated by specific peptide sequences in the replication proteins called replication foci targeting sequence (RFTS). The first such domain identified was in the maintenance methyltransferase (DNMT1) and was called targeting sequence (TS) (Leonhardt et al., 1992). Fusion of TS with a heterologous protein like the β-gal epitope mediated association of the fusion protein with RF. It is interesting to note that DNMT1 is an enzyme responsible for catalyzing transfer of a methyl group to cytosine residues in DNA and not involved in the process of DNA replication *per se*. The association of DNMT1 with RF is best interpreted as a feature suited to its role in maintaining methylation patterns whereby it can methylate newly synthesized DNA at the site of synthesis. Later on, two more domains were identified in DNMT1 that mediated association with RF. One of these domains interacts with PCNA, called PCNA binding domain (PBD), and site directed mutagenesis of this domain abolishing PCNA binding also abolished association with RF (Chuang et al., 1997). The other domain identified to mediate association with RF is the PBHD/BAH domain (Liu et al., 1998).

The second protein in which an RFTS was mapped is DNA ligase I (Cardoso et al., 1997). Subsequently, it was shown that the region in DNA ligase I that mediates association with RF binds to PCNA via a domain that is related to the PBD

in DNMT1. Also, this interaction of DNA ligase I with PCNA was shown to be essential for association with RF (Montecucco et al., 1998). Another protein shown to associate with RF mediated by a similar PBD includes DNA polymerase η (Kannouche et al., 2001).

Over the years many proteins involved in DNA metabolism and cell cycle regulation have been shown to interact with PCNA via the canonical PBD found in DNMT1 and DNA Ligase I (reviewed in (Leonhardt et al., 1998), some of which have been shown to be associated with RF. Considering the presence of a PBD in various DNA replication and repair factors and the requirement of the PBD to mediate association of DNA Ligase I with RF , it has been suggested that the mechanistic basis for the association of PBD containing proteins with RF is an interaction of the PBD with PCNA (Montecucco et al., 1998). In the scheme shown in Fig. 1.2, the PCNA trimer forms the central core encircling DNA and various replication factors associate with PCNA via the PBD. The PBD is conserved in homologous proteins from archaebacteria, yeast, worms, flies, amphibians and mammals (Warbrick et al., 1998). Such conservation across different classes of organisms suggests that the recruitment of proteins to RF mediated by PBD-PCNA interaction is a mechanism conserved throughout evolution. However, like DNMT1, there could be other proteins with a PBD that have additional domains involved in mediating association of replication factors with RF. In the case of DNMT1 all the three RFTS are conserved in the DNMT1 homologues from metazoans (this study) indicating that they are essential. It is suggested that assembly of a replication focus might involve a web of unique interactions among replication factors (Fig. 1.2) (Leonhardt et al., 1992). To better understand the assembly of replication proteins to RF, it is essential to learn the roles played by these additional RFTS (Leonhardt et al., 2000b).

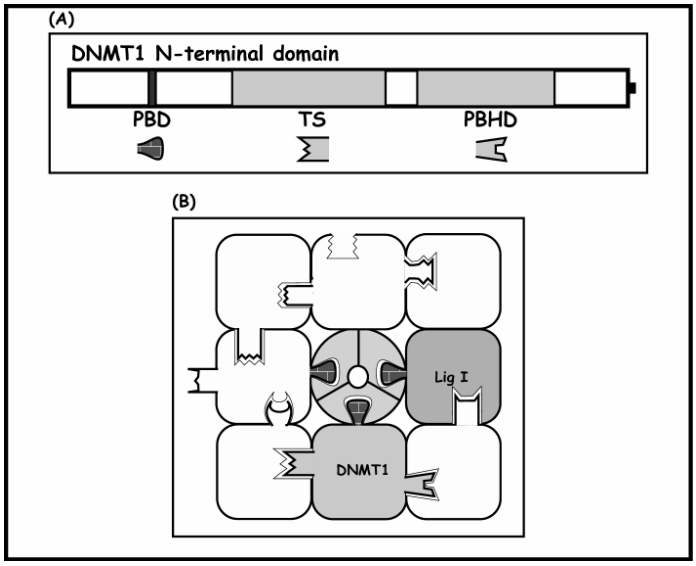


Fig. 1.2. Schematic representation of RF. (A) *The three regions in the N-terminal domain that have been shown to independently associate with RF.* ***(B)*** *Schematic of the RF. Various proteins are shown as boxes with protrusions to depict the protein domains that interact with other members of the RF by protein-protein interactions. The different shapes of the protrusions are meant to depict different domains that may play a role in targeting proteins to the RF. The central green ring is the PCNA trimer that encircles DNA (not shown). Association of proteins with the RF is principally mediated by interaction with PCNA. DNMT1 and DNA Lig I are two typical proteins which are targeted to the RF by interaction of their PBD with PCNA. The TS and PBHD are depicted to interact with other unidentified proteins in the complex.*

1.3. Epigenetic information

Epigenetic information is defined as a heritable mark on DNA that does not change the DNA sequence but that modulates gene activity, and that is stably inherited through mitotic/meiotic divisions (Wu and Morris, 2001).

1.3.1. Types of epigenetic information

Eucaryotic genomes contain two types of epigenetic marks:

a) Histone modifications: The DNA in eucaryotes is wound around a protein core called nucleosome that consists of histones. Four types of histones make up the octameric nucleosome core, a H3-H4 tetramer and two H2A-H2B dimers. All four histones are small basic proteins closely related to each other in that they share a globular motif called the histone fold. In addition to the histone fold, each of the core histones has a long N-terminal tail, which is rich in basic amino acid and extends out from the histone core. These histone tails are subject to several types of post-translational modifications, viz. acetylation, phosphorylation, methylation, ubiquitination and ADP-ribosylation (Berger, 2002). Covalent modifications in the globular domain have also been described (reviewed in (Varga-Weisz and Dalgaard, 2002). These modifications were purported to play a role in chromatin structure by influencing histone-DNA and histone-histone contacts, and thereby influencing transcription. However, observations made in the past two to three years have led to a "histone code hypothesis" that proposes an active and decisive role of histone modifications in chromatin function (Strahl and Allis, 2000). According to this, distinct histone tail modifications, individually or in combinations, would create specific binding sites for various chromatin modifiers with distinct functions thereby inducing formation of specialized chromatin domains. Such domains would have far-reaching consequences on processes like transcription, replication, recombination, mitosis etc. An example of the histone code hypothesis is the opposing effects of methylation of histone H3 tail at lysine-4 (H3K4Me) and lysine-9 (H3K9Me) on transcriptional activity (reviewed in (Lachner and Jenuwein, 2002). Presence of H3K9Me is correlated with transcriptionally silenced chromatin while H3K4Me marks transcriptionally active regions (Noma et al., 2001) (Litt et al., 2001). Mechanistically, H3K9Me attracts a transcriptionally repressive protein HP-1 that induces formation of silent chromatin domain (Lachner et al., 2001) (Bannister et al., 2001). In contrast, H3K4Me prevents association of the negatively acting nucleosome remodelling and histone deacetylation (NuRD) complex, and induces formation of a transcriptionally active domain (Nishioka et al., 2002). Even though very little is known about how regions in the genome are identified for establishing specific histone modifications, it is now clear that histone modifications function as epigenetic marks that can stably establish gene expression states.

b) DNA methylation: The DNA of most organisms is modified by a post-replicative process which results in three types of methylated bases in DNA: C5-methylcytosine (5mC), N4-methylcytosine and N6-methyladenine. The latter two

are more widespread in procaryotes, while the former is the major class of methylated base in all organisms. Formation of 5mC is accomplished by an enzyme called DNA methyltransferase (DNA MTase), which transfers a methyl group from S-adenosyl-L-methionine (SAM) to carbon-5 in the pyrimidine ring of cytosine (Fig. 1.3) (Wu and Santi, 1987). Briefly, in this process a cysteine thiol of the enzyme attacks carbon-6 of cytosine and forms a covalent DNA-protein intermediate. The addition of the cysteine thiol activates the carbon-5 allowing transfer of the methyl group from SAM and release of S-adenosyl-L-homocysteine (SAH). This reaction mechanism is conserved in all organisms and the enzyme involved is conserved across the whole spectrum of organisms. DNA methylation is a covalent modification of DNA that does not change the DNA sequence, but has an influence on gene activity. Although in procaryotes one of the major role of DNA methylation is to protect host DNA from the restriction-modification system, in eucaryotes the role of DNA methylation as an epigenetic mark has gained great importance. In vertebrates, DNA methylation is distributed throughout the genome and primarily occurs at CpG sequences, producing methyl-CpG symmetrically on both strands of the DNA (reviewed in (Bird, 2002). In human somatic cells, 5mC constitutes about 1% of total DNA bases and therefore 70-80% of all CpG dinucleotides in the genome are methylated (Ehrlich and Wang, 1981). Some of the remaining unmethylated CpG dinucleotides constitute the CpG islands and are found at promoter segments of genes. Some of these CpG islands become methylated during development and this results in stable silencing of the gene, for example genes silenced in the inactive X chromosome and silenced alleles of imprinted regions. In general, it is an accepted view that promoter methylation is one of the regulatory mechanisms employed in gene silencing (reviewed in (Cardoso and Leonhardt, 1999a). However, this is not a an absolute requirement as there are examples where a CpG island in a promoter is unmethylated while the gene is still kept silent, for example the CpG island in human α-globin gene promoter is unmethylated in both erythroid and non-erythroid tissues (Bird et al., 1987). Such cases might now be explained by the role of histone modifications in gene silencing. However, the importance of DNA methylation in mammalian development and in regulation of gene expression is well established and is known to be essential. This is best emphasized by the fact that disruption of the gene(s) encoding the enzyme that catalyzes DNA methylation is lethal in mice early in development (Li et al., 1992) (Lei et al., 1996) (Okano et al., 1999)

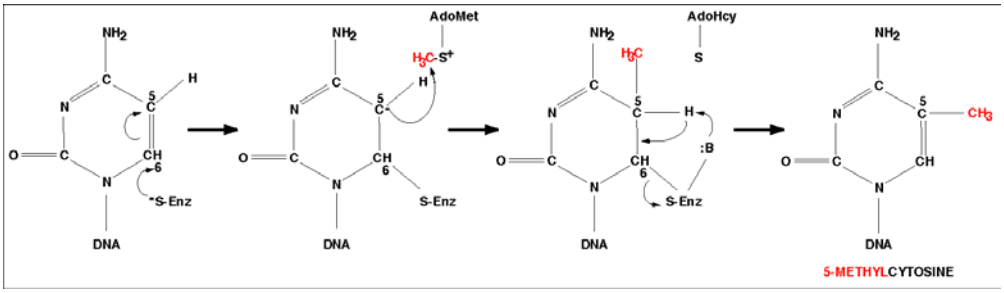


Fig. 1.3. Mechanism of transfer of methyl group to C5-cytosine based on the mechanism proposed by Wu and Santi (Wu and Santi, 1987) for thymidylate synthase and tRNA-(uracil-5)methyltransferase. *A cysteine thiol of the enzyme attacks the 6-carbon of cytosine and forms a covalent DNA-enzyme intermediate. The resulting carbanion at 5-carbon of cytosine then attacks the methyl group of SAM (AdoMet) forming a covalent bond with the methyl group and SAH (AdoHcy) is released. Elimination of the conjugate occurs through abstraction of the proton from carbon-5 by a base (B:) to yield the product 5-methylcytosine.*

1.3.2. Role of DNA methylation

DNA methylation has been demonstrated to play important roles during development, differentiation, aging, X-chromosome inactivation, genomic imprinting, tumourigenesis and transposon inactivation (reviewed in (Cardoso and Leonhardt, 1999a) (Leonhardt and Cardoso, 2000) (Bird, 2002) (Ehrlich, 2002) (Yoder et al., 1997b)). Most of these roles arise as a consequence of the effect of DNA methylation on transcription and chromatin structure as discussed below.

Inhibitory effect on transcription: There is a strong correlation between DNA methylation and gene silencing. For example, the CpG islands that span promoter regions are heavily methylated in the inactive X chromosome while the corresponding regions in the active X chromosome are not (reviewed in (Bird, 2002). DNA methylation can inhibit transcription in three ways (Fig. 1.4A) (reviewed in (Leonhardt and Cardoso, 2000). Firstly, DNA methylation can directly block transcription factor binding, which has been shown to be the case for some transcription factors (AP-2, c-Myc/Myn, E2F and Nf-κB) (Becker et al., 1987). However, other transcription factors are not sensitive to methylation (Sp1, CTF and YY1) (Tate and Bird, 1993). Secondly, DNA methylation represses promoter activity indirectly by attracting factors that specifically recognize and bind methylated cytosines thereby blocking access of transcription factors to promoters (Fig. 1.4B). These factors share a domain called methyl-CpG-binding domain (MBD). Out of five MBD containing proteins identified, four (MBD1, MBD2, MBD3, and MeCP2) have been implicated in DNA methylation dependent transcriptional silencing. Thirdly, DNA methylation represses transcription by altering chromatin structure through the MBD proteins that can function as a complex containing nucleosome remodeling factors (for example, the MeCP1 complex consists of MBD2 + Mi2/NuRD; MeCP2 binds Sin3/HDAC) (Fig. 1.4C). In this case DNA methylation will lead to deacetylation of histones thereby suppressing transcription. Thus, DNA methylation can modulate the histone code and lead to repression.

Effect on chromatin structure: DNA methylation is known to have a profound influence on chromatin structure (reviewed in (Leonhardt and Cardoso, 2000). For example, mutation of a gene encoding a DNA MTase (DNMT3B, discussed later) is linked to a hereditary disorder called ICF syndrome (Xu et al., 1999) (Okano et al.,

1999). Cells from these patients show deletions or duplications of entire chromosomal arms, isochromosomes and centromere breakage (Franceschini et al., 1995). DNA methylation studies in ICF patients showed hypomethylation of classical satellites II and III, which are major components of constitutive heterochromatin (Jeanpierre et al., 1993). These regions are normally highly methylated indicating that DNA methylation is essential for proper centromere structure and stability (reviewed in (Robertson and Wolffe, 2000).

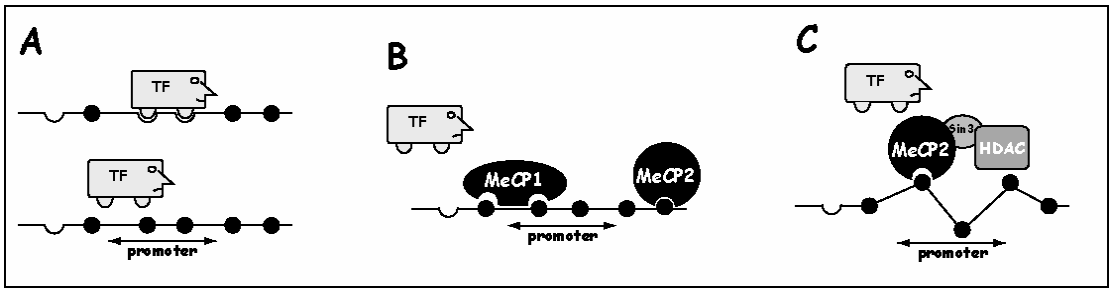


Fig. 1.4. Models for effect of methylation on gene activity. *Unmethylated DNA is depicted as half circles and methylated DNA as filled circles.* ***(A)*** *Methylation directly prevents binding of transcription factor (TF) thereby inhibiting transcription.* ***(B)*** *Methyl DNA binding proteins bound to the methylated promoter prevent binding of TF and inhibit transcription.* ***(C)*** *MeCP2 complex containing HDAC binds to methylated*

1.3.3. Regulation of DNA Methylation

The mammalian genome undergoes sweeping changes in its methylation pattern, the most dynamic being observed during development. In mice, just after fertilization the male pronucleus is rapidly demethylated while the maternal genome progressively loses methylation until the blastocyst stage. The methylation levels decrease to ~30% of that in the adult somatic cells but return to higher levels during implantation (Monk et al., 1987) (Oswald et al., 2000) (Mayer et al., 2000) (Reik et al., 2001). In effect, these changes cause an erasure of the existing DNA methylation patterns (except some regions like imprinted loci) followed by establishment of new DNA methylation patterns. Such reprogramming of DNA methylation patterns is essential for setting up tissue specific gene expression, X chromosome inactivation in female mammals and genomic imprinting. For normal functioning, once methylation patterns are established, these patterns have to be faithfully inherited by daughter cells. Changes in methylation patterns sometimes occur in adult tissues with harmful effects. For example, in somatic cells, some CpG islands get methylated during aging and tumourigenesis (reviewed in (Cardoso and Leonhardt, 1999a) (Bird, 2002)). Formation of many tumours have been correlated with hypomethylation and/or hypermethylation of specific regions in the genome resulting in activation of oncogenes or suppression of tumour suppressors (reviewed in (Leonhardt and Cardoso, 2000) (Ehrlich, 2002)). The changes in DNA methylation that occur during all biological processes, both normal and diseased, are mediated by three processes, viz. *de novo* methylation, maintenance methylation and demethylation (Fig. 1.5). An understanding of how these processes function is central to our understanding of development and disease.

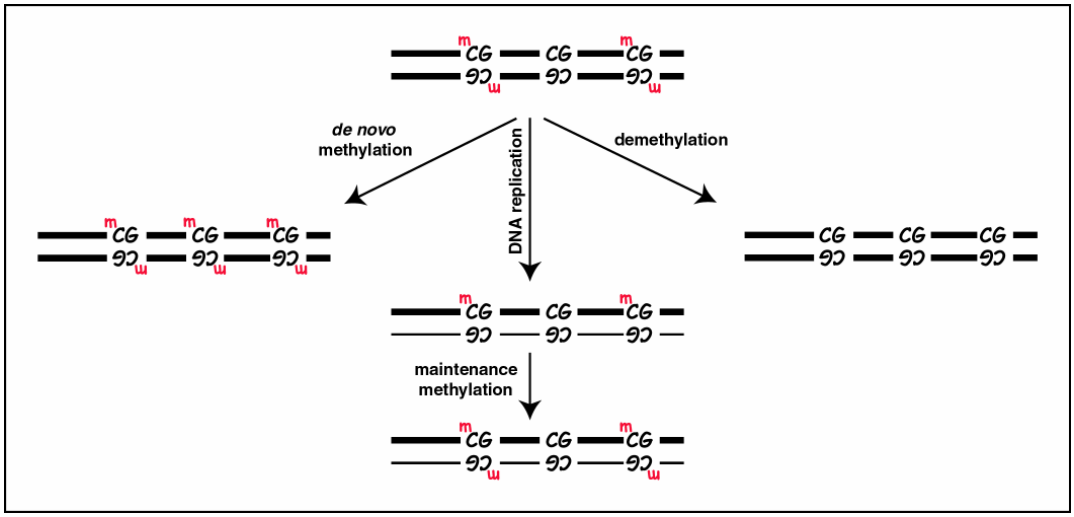


Fig. 1.5. Processes that change or maintain DNA methylation pattern. *In vertebrates DNA methylation occurs mainly in CpG dinucleotides depicted here as CG. Methyl residues are depicted as 'm'. New methylation patterns are established by the process of de novo methylation (left). Existing methylation pattern can be erased by demethylation (right). During DNA replication (centre), the newly synthesized DNA strand (thin line) is unmethylated while the parent strand (thick line). retains its methylation pattern. The methylation pattern from the parent strand is copied on to the daughter strand by maintenance methyltransferase.*

A) *De novo* methylation: It is the process in which unmethylated sites in DNA are methylated resulting in formation of new methylation patterns (Fig. 1.5). The highest *de novo* MTase activity is detected in embryonal carcinoma (EC) and embryonic stem (ES) cells (Stewart et al., 1982) (Lei et al., 1996) and specific MTases have been shown to be involved in *de novo* methylation (discussed in the next section). In mammals, most of the *de novo* methylation occurs during development when both paternally and maternally derived genomes undergo gross DNA methylation (Monk et al., 1987). Other examples of *de novo* methylation of sites in the genome where viral DNA integrates (Toth et al., 1990), age related hypermethylation in the c-myc gene in liver of mice (Ono et al., 1989) and methylation of the estrogen receptor (ER) gene in ageing colorectal mucosa resulting in predisposition to sporadic colorectal tumorigenesis (Issa et al., 1994). Even though there exist many examples of *de novo* methylation, little is known about how specific DNA sequences are selected for DNA methylation. Many observations indicate that the DNA methylation machinery is targeted to transcriptionally inactive regions. In the case of the X-linked Hprt gene, DNA methylation occurs after chromosome inactivation and transcriptional silencing (Lock et al., 1987). Where ever examined, DNA methylation occurs after methylation of histone at lysine9 (H3K9Me), the other epigenetic mark characteristic of silent chromatin (Heard et al., 2001) (Bachman et al., 2003). Notably, silencing occurs prior to DNA methylation and concomitant with formation of H3K9Me. These studies have suggested that DNA methylation plays an important role in stabilizing the silenced state established by histone modifications.

B) Maintenance methylation: It is the process by which DNA methylation patterns are maintained after each round of DNA replication. Since each round of DNA replication results in a newly synthesized strand that is unmethylated while the parent strand is methylated, a mechanism is required to methylate the newly synthesized strand (Fig. 1.5). It was proposed that once a methylation pattern has

been set by *de novo* methylation, this would be clonally inherited by the action of a maintenance DNA methyltransferase specific for hemi-methylated CpG sites (Riggs, 1975) (Holliday and Pugh, 1975). Biochemical experiments have shown that mammalian DNA methyltransferases purified from somatic cells prefer hemimethylated DNA as substrate (Gruenbaum et al., 1982) (Bestor and Ingram, 1983). Such an activity has also been demonstrated *in vivo* wherein it was observed that *in vitro* methylated DNA introduced into mouse cells by transfection retains the methylation pattern after several rounds of replication. In contrast, cells transfected with unmethylated DNA showed no methylation of the DNA suggesting that the cell had some MTases that specifically "replicate" methylation patterns (Wigler et al., 1981). An MTase, DNMT1 (discussed in the next section), with a preference to methylate DNA containing hemi-methylated CpG dinucleotide was later cloned and characterized (Bestor et al., 1988) (Bestor, 1992). Several biochemical, cell biological and genetic studies have shown that DNMT1 is the key enzyme involved in maintaining DNA methylation patterns.

C) Demethylation: It is the process by which DNA methylation patterns are erased (Fig. 1.5). Demethylation mainly occurs during preimplantation development, but also occurs throughout development as a prelude to transcriptional activation. The demethylation process is not as clear as DNA methylation and two possible mechanisms have been reported for demethylation:

Passive demethylation: This results in the gradual decrease of DNA methylation levels due to the absence of maintenance methylation at each round of DNA replication. For example, passive demethylation occurs after inactivation of cellular MTases with 5AzaC (Jones and Taylor, 1980). Analysis of the kinetics of demethylation that occur in the maternally inherited genome during preimplantation has been correlated with successive loss of methylation at each chromosome replication cycle (Rougier et al., 1998). This is explained to occur through active retention of DNMT1 in the cytoplasm during development from the oocyte to the blastocyst stage (Cardoso and Leonhardt, 1999b) (Carlson et al., 1992).

Active demethylation: Active demethylation occurs independent of DNA replication, and is mediated by enzymes. Cases where this is observed include the global demethylation that occurs in the zygotic paternal genome (Mayer et al., 2000), demethylation of the vitellogenin gene in chick liver upon induction of transcription (Wilks et al., 1984), demethylation of globin gene stimulated in erythroleukemia cells (Razin et al., 1986) and the genome wide demethylation in differentiating myoblasts (Jost and Jost, 1994). Three main biochemical mechanisms have been proposed to carry out active demethylation: excision of the methylated base by a glycosylase, excision of the methylated nucleotide, or direct replacement of the methyl group by a hydrogen atom (Kress et al., 2001). However, the molecular mechanism behind these processes is still not known.

1.3.4. Enzymes involved in methylating DNA

Numerous DNA methyltransferases (MTases) have been identified and cloned from both procaryotes and eucaryotes, and have been shown to share a conserved catalytic domain in the form of 10 small sequence motifs. Based on a phylogenetic

comparison of the catalytic domains, the eucaryotic MTases are grouped into 5 families, DNMT1, DNMT2, DNMT3, Masc 1 (only one member from a fungus), and CMT (chromomethylase, only in plants) (Fig. 1.6) (Colot and Rossignol, 1999). All eucaryotic MTase families, except the DNMT2 family, consist of proteins which have an additional N-terminal domain with various functional motifs. DNA methylation being an important epigenetic mark, presence of many different MTases suggest special functions in regulating DNA methylation. Here we examine the known mammalian DNA methyltransferases.

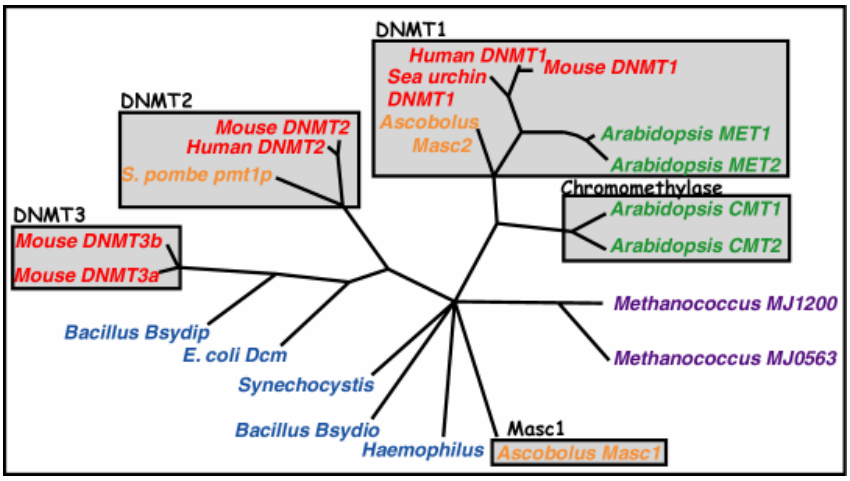


Fig. 1.6. Five families of eucaryotic MTases. *Phylogenetic relationship between known MTases based on comparison of the conserved motifs in the catalytic domains (adapted from (Colot and Rossignol, 1999)). The eucaryotic MTases group into five families (boxed): DNMT1, DNMT2, DNMT3, Chromomethylase (CMT), Masc1 (only one member). Some of the recently identified proteins (CMT3 and DNMT3L) are not shown. MTases from eubacteria and archaebacteria are divergent and lie scattered.*

DNMT1: DNMT1 is the first mammalian MTase that was cloned, and is referred to as maintenance MTase because it is responsible for maintaining DNA methylation patterns. Based on functional and structural data it is suggested to result from fusion of three genes, one of them being an ancestral procaryotic DNA MTase (Margot et al., 2000). The enzyme consists of two main domains - the C-terminal catalytic domain (570 amino acids) and the N-terminal regulatory domain (1051 amino acids) - linked by a stretch of repeating Gly-Lys dipeptide (linker) (Fig. 1.7). DNMT1 homologues have been identified in a wide range of organisms including fungi, plants, sea urchin, amphibians, fish, birds and mammals, all having a similar structure with a long N-terminal domain and shorter C-terminal catalytic domain. In mammalian DNMT1, the N-terminal domain has various motifs with specific functions (Fig. 1.7) (reviewed in (Leonhardt and Cardoso, 2000) (Bestor, 2000)). Two properties of DNMT1 distinguishes it as maintenance MTase. Firstly, the enzyme shows a preference for hemimethylated DNA (Bestor, 1992), and secondly it is specifically relocated to RF when the cell enters S phase (Leonhardt et al., 1992). These observations paved the way for an understanding of the mechanism by which cells maintain their methylation pattern. At every round of DNA replication in organisms with methylated genome, the product is double stranded DNA wherein the parent strand retains the methylation pattern while the newly synthesized daughter strand is unmethylated (Fig. 1.5). The preference of DNMT1 for hemimethylated CpG sequences enables DNMT1 to copy the methylation pattern into the newly

synthesized strand. By virtue of being targeted to RF, DNMT1 would be positioned exactly at the site where its substrate, hemi-methylated DNA, is synthesized (Fig. 1.8). Three regions in the N-terminal domain of DNMT1, viz. TS, PBD and PBHD/BAH, have been reported to mediate association with RF (Leonhardt et al., 1992) (Chuang et al., 1997) (Liu et al., 1998).

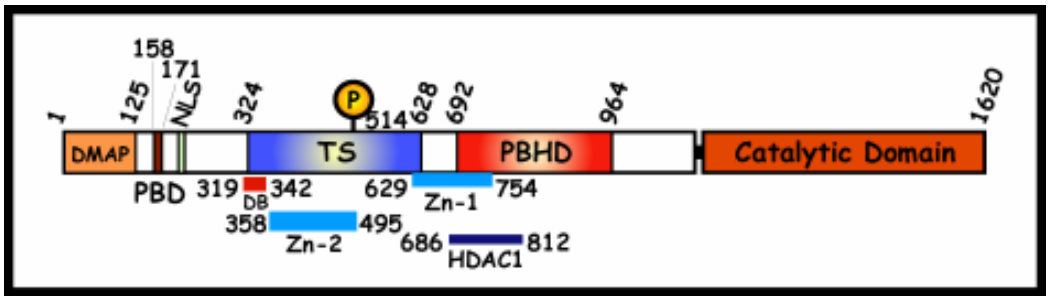


Fig. 1.7. Domain structure of DNMT1. *The somatic long isoform of DNMT1 is shown here. DMAP corresponds to the region in DNMT1 that binds DMAP (Rountree et al., 2000). PBD (PCNA binding domain) (Chuang et al., 1997), TS (targeting sequence) (Leonhardt et al., 1992) and PBHD (polybromo homology domain) (Liu et al., 1998) are reported to target to replication foci. P marks an identified phosphorylation site (Glickman et al., 1997). NLS is the nuclear localization signal (Cardoso and Leonhardt, 1999b). Zn-1 (Bestor, 1992) and Zn-2 (Chuang et al., 1996) are the two Zn binding domains. DB is a DNA binding domain just preceding the TS (Chuang et al., 1996). HDAC1 corresponds to the region in DNMT1 that binds to HDAC1 (Fuks et al., 2000).*

In addition, other regions in the N-terminus have been demonstrated to play special functions (Fig. 1.7). At least three nuclear localization signals (NLS) have been identified that mediate nuclear import (Cardoso and Leonhardt, 1999b). A Cysteine-rich region and two Zn-binding regions have been identified (Bestor, 1992) (Chuang et al., 1996). Using biochemical approaches, it was observed that the N-terminal domain can mediate transcriptional repression that is partially mediated by interaction with histone deacetylases (HDAC1 and HDAC2) and a novel trancriptional repressor (DMAP1) (Fuks et al., 2000) (Robertson et al., 2000) (Rountree et al., 2000). The N-terminal domain has also been observed to mediate transcriptional repression directly through a region related to the trithorax-related protein HRX (Fuks et al., 2000). It is suggested that DNMT1 mediates recruitment of HDAC2 to late-RF and that this serves as a mechanism to deacetylate the acetylated histones that are assembled on newly replicated DNA (Rountree et al., 2000). Thus, in addition to its role in maintaining DNA methylation pattern, DNMT1 is proposed also to play a role in maintaining heterochromatin structure.

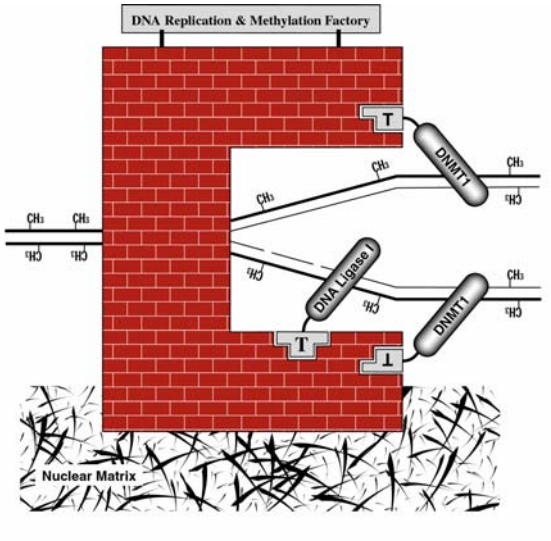


Fig. 1.8. Coupling of DNA methylation with DNA replication. *Association of DNMT1 with the replication machinery mediated by RFTS (marked as T) couples maintenance of DNA methylation with DNA replication. The replication machinery is shown tethered to the nuclear matrix. RFTS mediated association of DNA Ligase I with the replication machinery is also shown.*

Observations made on DNMT1 mutant mice created by targeted mutation of the Dnmt1 gene have strongly supported the role of DNMT1 as a maintenance methyltransferase (Li et al., 1992) (Lei et al., 1996). Homozygous Dnmt1 mutation causes a severe reduction of 5mC in ES cells and embryos. These mutant embryos die very early in development and were reported to be extensively demethylated at all sites in the genome that were examined. Loss of DNMT1 results in activation of silenced alleles of imprinted genes due to inability to maintain the methylation pattern at the imprinted loci (Li et al., 1993). These observations made on mice lacking DNMT1 have established DNMT1 as an enzyme responsible for maintaining DNA methylation patterns.

It is also suggested that DNMT1 might function as a *de novo* methyltransferase. This is based on *in vitro* experiments using extracts from various tissues and cell types in which DNMT1 had a significant activity on unmethylated DNA substrates (Yoder et al., 1997a). Other studies have shown that the *de novo* methyltransferase activity of mouse DNMT1 is higher than that of the known *de novo* methyltransferases (DNMT3 proteins, see below) (reviewed in (Bestor, 2000). In conclusion, although DNMT1 could potentially methylate DNA *de novo*, genetic studies show that it seems to be mainly involved in maintenance of methylation patterns. It is possible that *in vivo* it performs both functions.

DNMT2: DNMT2 forms a family of proteins that is related to pmt1p of *S. pombe*, an organism that does not show any 5mC in its DNA. In *S. pombe*, disruption of the $pmt1^+$ gene results in no discernible phenotype, and purified pmt1p does not show any methylation activity *in vitro* (Wilkinson et al., 1995). However, deletion of a single amino acid in pmt1p restores catalytic activity (Pinarbasi et al., 1996). Disruption of the Dnmt2 gene in mouse ES cells did not yield any detectable effect on DNA methylation, nor did purified DNMT2 show any methylation activity in biochemical assays (Okano et al., 1998b). Homologues of DNMT2 have been identified in other vertebrates, *D. melanogaster*, plants and *S. pombe* (reviewed in

(Bestor, 2000) but it is not clear whether DNMT2 has any function in these organisms.

DNMT3: The DNMT3 family consists mainly of three enzymes, DNMT3A, DNMT3B and DNMT3L, that were identified from EST database searches. The catalytic domain of DNMT3A and DNMT3B proteins is more similar to the bacteriophage MTases than to DNMT1 and DNMT2. The murine Dnmt3a and Dnmt3b genes are highly expressed in undifferentiated ES cells but downregulated after differentiation and expressed at low levels in adult somatic tissue (Okano et al., 1998a). In contrast, DNMT1 is expressed at high levels both in ES cells and somatic cells. Biochemical experiments showed that DNMT3A and DNMT3B could methylate both unmethylated and hemi-methylated DNA with equal activity (Okano et al., 1998a). *In vivo* studies showed that DNMT3 proteins indeed have the ability to catalyze *de novo* methylation. This was shown by an assay wherein retroviral DNA is introduced into wild type and mutant ES cells and the methylation state of the retroviral DNA is tested after several days. In such an assay it was observed that ES cells deficient in both DNMT3A and DNMT3B (double mutant), completely lacked the ability to methylate the retroviral DNA (Okano et al., 1999). These observations have heralded DNMT3A and DNMT3B as enzymes that mediate *de novo* methylation.

Studies on the sub-nuclear localization of epitope tagged DNMT3A and DNMT3B have shown that both specifically associate with pericentric heterochromain in embryonic stem cells, while in embryonic fibroblasts only DNMT3A is associated with these sites (Bachman et al., 2001). This corroborates some of the observations made in mice mutant for Dnmt3 genes, which show that DNMT3B is important in methylating the centromeric repeats during early development, and not in differentiated cells (Okano et al., 1999). It is not clear what role DNMT3A has at the pericentric regions in differentiated cells. An alternative form of DNMT3A, DNMT3A2, produced from an alternative promoter in the Dnmt3a gene exhibited localization to euchromatin (Chen et al., 2002). In contrast to these reports, it has been demonstrated that both Dnmt3a and Dnmt3b are not associated with pericentric heterochromatin in mouse myoblast cells (C2C12) (Margot et al., 2001). These studies indicate that DNMT3 proteins are recruited to their target sites by non-overlapping mechanisms, and that these mechanisms might be specific to the developmental stage and cell type. It follows from these studies that the targeting mechanisms could be controlled to regulate *de novo* methylation. Further, such studies on the localization of the various methyltransferases should shed light on the mechanisms involved in targeting and the *in vivo* targets of these enzymes, which are essential for our understanding of regulation of DNA methylation.

DNMT3L: DNMT3L is the most recent addition to the list of DNA MTases. It is closely related to DNMT3A and DNMT3B, but lacks the conserved residues in the catalytic domain that are essential for enzymatic activity. Consistent with this, it does not show any catalytic activity *in vitro* (Hata et al., 2002). DNMT3L is shown to interact with DNMT3A and DNMT3B, and also co-localizes with these enzymes in nuclei of transfected COS cells (Hata et al., 2002). DNMT3L is shown to be specifically expressed in undifferentiated ES cells and mice lacking DNMT3L are defective in establishing maternal genomic imprints (Hata et al., 2002) (Bourc'his et al., 2001). It is suggested that DNMT3L co-operates with the DNMT3 family proteins

to carry out imprinting of genes during oogenesis and early mouse development (Hata et al., 2002) (Bourc'his et al., 2001).

1.3.5. Organisms that lack DNA methylation

Even though DNA methylation is an essential epigenetic mark in vertebrate system, many lower eucaryotes, like yeast, *C. elegans* and *D. melanogaster* lack this modification. Although *D. melanogaster* was long considered to lack DNA methylation, there are studies showing the presence of low levels of 5mC in its genome (Achwal et al., 1984) (Gowher et al., 2000) (Lyko et al., 2000). Unlike in mammalian cells, most of the methylation here is found in CpA dinucleotides and not CpG. In *D. melanogaster*, 5mC accounts for just about 0.1% of the total cytosines (Gowher et al., 2000) as compared to ~2-10% of cytosines in mammmalian cells (Ehrlich and Wang, 1981). The only DNA methyltransferase found in *D. melanogaster* is the DNMT2 homologue which does not show methyltransferase activity *in vitro* (Tweedie et al., 1999) (Lyko et al., 2000). It could be possible that this enzyme is active *in vivo* and is responsible for the traces of 5mC. No homologues of the functional DNA MTases, present in all other organisms with a methylated genome, have been observed in *Drosophila*. Even though there are insects whose genome is methylated, for example the cricket *Acheta domesticus* (Tweedie et al., 1999), it is not known whether they have a DNMT1 homologue. However, after complete sequencing of the *Drosophila* genome, it is now established that *Drosophila* does not have any homologue of DNMT1. Absence of a homologue of DNMT1 might mean that *Drosophila* never had a DNMT1 homologue or has lost it during evolution.

Homologues of downstream effectors of methylation, the methyl DNA binding proteins, identified in *D. melanogaster* totally lack the highly conserved methyl DNA binding domain (MBD) found in other organisms whose genome is known to be methylated (Tweedie et al., 1999). Also, in some fungi (*N. crassa* and *A. immersus*) whose genome is methylated, the MTases are not essential. Absence of DNA methylation in some of these eucaryotes indicates that other epigenetic mechanisms, like histone modifications, are sufficient for gene regulation in these organisms.

1.4. Questions addressed in this work

To study mechanisms by which proteins associate with RF in diverse eucaryotes and the role of the regulatory sequences of DNMT1 in this association, we addressed the following questions:

Are the mechanisms that mediate association of proteins with RF conserved in evolution?

Many features of DNA replication are conserved in higher eucaryotes. Firstly, general features of the replication process itself, like bi-directional replication fork movement, continuous leading and discontinuous lagging strand synthesis, requirement of RNA primers to start DNA synthesis (Baker and Bell, 1998) are all conserved. Secondly, the proteins involved in controlling DNA replication and

catalyzing the process of DNA replication are conserved in diverse eucaryotes (Leipe et al., 1999). Thirdly, organization of replication into RF that follow a spatio-temporal pattern is a conserved feature in various eucaryotes (Samaniego et al., 2002) (Ahmad and Henikoff, 2001). However, it is not known whether the mechanisms by which replication factors associate with RF are conserved in evolution. To determine this, we have analyzed the ability of the various domains in DNMT1 to associate with RF in *Drosophila* cells. Moreover, as shown in Fig 1.8, maintenance of epigenetic information (DNA methylation) is coupled to DNA replication by association of DNMT1 with the replication machinery. In this regard, *Drosophila* is interesting as its genome is scarcely methylated and it lacks the DNMT1 homologue (see section 1.3.5; when this work was planned *Drosophila* was still known as completely lacking DNA methylation). An understanding of the ability of the three targeting sequences in DNMT1 to associate with RF and subnuclear structures in *Drosophila* cells would tell us whether the extra targeting domains have specifically evolved to function in organisms with methylated and complex genomes coupling DNA replication with maintenance of DNA methylation.

What is the function of the various regulatory sequences in controlling the subnuclear distribution of DNMT1 during the cell cycle?

As discussed earlier, it is suggested that PBD-containing replication proteins associate with RF via the PBD. DNMT1 has three targeting sequences (TS, PBD and PBHD) that have been reported to independently mediate association with RF. Here we sought to understand whether the two extra sequences, viz TS and PBHD, have any specialized functions. Replication in eukaryotes follows a defined spatial and temporal order that reflects the state of transcriptional activity of the chromatin. In general, the sparsely methylated euchromatin replicates early and the densely methylated heterochromatin replicates late, which can be easily discerned by microscopic examination of the pattern of RF (Fig. 1.1). DNMT1 associates with the RF and maintains DNA methylation patterns. Considering this function of DNMT1, and the differences in methylation density at early-replicating euchromatin and late-replicating heterochromatin, we were prompted to investigate whether the three targeting sequences in DNMT1 have any preference for early or late replication foci. We directly addressed this question by analyzing the subnuclear localization of the three targeting sequences, each fused individually and in combinations to GFP/YFP, during different stages in S phase and throughout the entire cell cycle.

Are the regulatory sequences of the mammalian DNMT1 present in other proteins?

As mentioned earlier, there are five classes of DNA Mtases (see Fig 1.6). Only the DNMT1 family of proteins are purported to play a role in maintaining DNA methylation patterns. One evidence that strongly supports such a role for DNMT1 is its association with RF thereby coupling DNA replication with maintenance methylation. Such an analysis has been performed only for mammalian DNMT1 and it is not known whether plant and fungal DNMT1 proteins also localize to RF or other subnuclear structures. Here all the MTase family members and other proteins in the database were analyzed for the presence of sequences similar to the targeting domains in DNMT1. This would provide insight into whether and how other MTases and other nuclear proteins could associate with RF. This analysis may thus contribute to our

knowledge of the evolution of nuclear architecture and the introduction of epigenetic information.

2. Materials and Methods

2.1. Construction of plasmids encoding various fusion proteins

The various plasmid constructs encoding translational fusions of DNMT1 (Fig. 3.4 and Fig. 3.14) , HsDNA Ligase I (Fig. 3.18) and DmDNA Ligase I (Fig. 3.10) with GFP/YFP/RFP were derived from the following vectors: pEMT (Czank et al., 1991), pEGFP-N2 (Clontech), pEGFP-C1 (Clontech), pEVRF0 (Matthias et al., 1989), pDsRed1-C1 (Clontech). Eucaryotic expression in all cases is driven by the cytomegalovirus immediate-early enhancer-promoter. All plasmid constructs were made using standard cloning techniques and transformation of *Escherichia coli* (Sambrook and Russell, 2001). Plasmid DNA was isolated from transformed *E. coli* using the alkaline lysis method (Birnboim and Doly, 1979) and subsequently verified by restriction enzyme analysis. An SV40-NLS was included for efficient nuclear targeting of deletion proteins which lost the endogenous NLS. Wherever PCR was used to generate the required insert, the sequence was verified by sequencing (GATC). The PBD inserts for MTPBD-GFP, HsLigPBD-GFP and DmLigPBD-GFP fusion proteins were generated by synthesizing sense and antisense oligonucleotides corresponding to both strands of the PBD coding region (Table 2.1). The oligonucleotides were flanked with overhangs corresponding to products of EcoR I and Xma I digestions. The sense and anti-sense oligonucleotides were annealed in Sequenase buffer [0.2 M Tris-Cl (pH 7.5), 0.2 M $MgCl_2$, 1 M NaCl] and subcloned into pEGFP-N2 digested with EcoR I and Xma I. The inserts were verified by sequencing. Plasmid DNA was further purified using anion-exchange columns (Qiagen) according to the instructions of the manufacturer and used for transfection of cultured *Drosophila* and mammalian cells.

Table 2.1. Oligonucleotides used for generating PBD-GFP fusions

Origin of PBD	Sequence of Oligonucleotides*
DNMT1	Sense strand 5'AATTCTGCAGTGCGACCATGGGACGCACCACCAGGCAGACCACCATCACGGCTCAC TTCACGAAGGGCCCCACTAAACGGAAACCAGGAC 3'
	Antisense strand 5'CCGGGTCCTGGTTTCCGTTTAGTGGGGCCCTTCGTGAAGTGAGCCGTGATGGTGGTC TGCCTGGTGGTGCGTCCCATGGTCGCACTGCAG 3'
HsDNA Ligase I	Sense strand 5'AATTCTGCAGTGCGACCATGCAGCGAAGTATCATGTCATTTTTCCACCCCAAGAAA GAGGGTAAAGCAAAGAAGCCAGGGC 3'
	Antisense strand 5'CCGGGCCCTGGCTTCTTTGCTTTACCCTCTTTCTTGGGGTGGAAAAATGACATGATA CTTCGCTGCATGGTCGCACTGCAG 3'
DmDNA Ligase I	Sense strand 5'AATTCTGCAGTGCGACCATGCAAAAGTCTATAACCTCCTTCTTCAAGAAGAAATCC GATGCCACCGACAGCCCCTCGC 3'
	Antisense strand 5'CCGGGCGAGGGGCTGTCGGTGGCATCGGATTTCTTCTTGAAGAAGGAGGTTATAGA CTTTTGCATGGTCGCACTGCAG 3'

*Letters in bold flanking the sequences are the overhangs corresponding to products of EcoR I and Xma I digestions. The start codon is underlined.

2.2. Cell culture and transfection

2.2.1. Drosophila cells

The ability of various proteins and their mutants to associate with RF in *Drosophila* cells was analyzed in Schneider's line 2 (S2 cells). These cells have been generated from primary cells prepared from minced or enzymatically disaggregated late embryos of *Drosophila melanogaster* (Schneider, 1972) and are likely to have originated from lymphoid cells (Roberts, 1998). The cells were grown in a humidified atmosphere at a temperature of 25˚C and maintained in Shields and Sang M3 medium (Sigma) with 10% FCS at a density of 0.5 x 10^6 to 2 x 10^6 cells/ ml in tissue culture flasks. The cells grow partially attached and were detached by mechanical shaking and pipetting to harvest or to subculture them. Frozen cells were thawed by warming the cryovials in a 37ºC water bath and pipetting the cell suspension into a plate containing medium at 25ºC. For long term storage, the cells were harvested and resuspended at a density of 1-2 x 10^7 cells/ml in 10% FCS supplemented Shields and Sang M3 medium containing 45% conditioned medium (medium in which the cells were growing). The cell suspension was distributed into cryovials and DMSO was added to a concentration of 10% and the cells were immediately frozen and stored in liquid nitrogen.

For immunostaining and microscopic analysis, S2 cells were cultured on poly-lysine coated glass coverslips to enhance attachment of cells to the coverslip. Transient transfections were done with the eucaryotic expression plasmid DNA using the $CaPO_4$-DNA coprecipitation method (Graham and van der Eb, 1973) (Parker and Stark, 1979). Cells were transfected at a density of 1 x 10^6 cells/ ml. After 48 hrs of transfection, the cells were washed gently with phosphate buffered saline (PBS) and fixed with 3.7% formaldehyde in PBS and immunostained.

2.2.2. Mammalian cells

Mouse C2C12 myoblasts (Yaffe and Saxel, 1977) were used for all immunostaining and microscopic analyses as replication foci are well characterized in these cells (Cardoso et al., 1993). COS-7 (African green monkey kidney fibroblast-like cells transformed with SV40 T antigen; (Gluzman, 1981)) or 293T (human embryonic kidney cells transformed with SV40 T antigen; gift from Isao Suetake and Shoji Tajima) cells were used to test the proteins produced from the various plasmid constructs by western blot analysis.

The cells were grown in a humidified atmosphere of 5% CO_2 at 37ºC in DMEM with 10% and 20% FCS for COS-7 or 293T cells and C2C12 cells respectively. The cells were grown to 70% confluency, and were subcultured every 48-72 hrs. To harvest or to subculture the cells, trypsin (0.25% trypsin, 0.02% EDTA in PBS) was used to detach C2C12 and COS-7 cells, while 293T cells were detached using 0.02% EDTA in PBS. Freezing and thawing of cells were done as described in 2.2.1, except that cells were frozen at a density of 10^6 cells/ml.

C2C12 cells were grown on coverslips and transfected using the $CaPO_4$-DNA coprecipitation method. After 24 hrs of transfection C2C12 cells were washed in PBS and fixed either with ice cold absolute methanol or 3.7% formaldehyde followed by immunostaining. COS-7 cells and 293T cells were transfected using Polyfect reagent

following the instructions of the manufacturer (Qiagen). Polyfect is a synthetic polymer (dendrimer) built up from branched units that form a spherical architecture. The branches terminate in charged amino groups which interact with the negatively charged phosphate groups of nucleic acids. The Polyfect-DNA complex has a net positive charge and interacts with negatively charged moieties (like glycoproteins) at the cell surface and is taken into the cell by nonspecific endocytosis. The reagent buffers the pH of the endosome, leading to pH inhibition of endosomal nucleases, which ensures stability of PolyFect–DNA complexes. After 48-72 hrs of transfection COS-7 or 293T cells were washed in PBS, harvested and extracted for western blot analysis.

2.3. Cell extracts and western blot analysis

Transfected cells were harvested by centrifugation at 228 g for 5 min at 4˚C and cell pellets were resuspended in Laemmli sample buffer (2% SDS, 20% glycerol, 250 mM Tris-HCl pH 6.8, 10% β-mercaptoethanol, 0.1% bromophenol blue) and denatured by boiling at 100ºC for 5 min. Proteins were separated by SDS-PAGE. A 12% gel was used in most cases, but proteins with expected molecular weight of 71-212 kDa (Fig. 3.14B) were separated on a 5-20% gradient gel. Separated proteins were transferred to a PVDF membrane (BIO-RAD) with transfer buffer (48 mM Tris-Cl, 39 mM glycine, 0.037 % SDS and 20 % absolute methanol) using a semidry blotter (Hoefer) for 1 hr at 2 mA/cm^2. Non-specific binding to the membrane was blocked by incubating it in blocking buffer (5% nonfat milk powder in PBS) for 30 min at room temperature followed by incubation in a solution of the primary antibody diluted with blocking buffer and 0.2% Tween-20 for 2 hrs at room temperature. The blot was washed three times 20 min each in PBS containing 0.2% Tween-20, incubated for 1 hr with HRP-conjugated secondary antibody diluted with the blocking buffer. The HRP signals were detected by ECL+ reagent (Amersham Pharmacia) following the instructions of the manufacturer. This is based on the principle that HRP/hydrogen peroxide catalyses oxidation of chemiluminescent substrates, like luminol, in alkaline conditions. The product is in an excited state and decays to ground state by emitting light. The emitted light signals were recorded in a luminescent image reader (Fuji). Images were assembled and annotated using Adobe Photoshop 5.5 and Adobe Illustrator 8.0.1. The following primary antibodies were used: rabbit anti-GFP (1:500; abcam), rabbit anti-human DNA Ligase I [1:5000, (Cardoso et al., 1997)], anti-PATH5Δ (1:5000 ; Nowak, D., Leonhardt, H. and Cardoso, M.C.). The secondary antibody used was anti-rabbit-IgG-HRP (1:10,000; Sigma).

2.4. Cell cycle and immunofluorescence analysis

2.4.1. BrdU labeling of replication foci and immunostaining

For detection of RF, cells grown on coverslips were incubated in medium with 100 μM BrdU for 15 min (pulse labeling). The cells were then washed two times with PBS and fixed with 3.7 % formaldehyde in PBS. Cells were permeabilized with

0.25% Triton-X-100 for 10 min followed by washing three times with PBS. Non-specific binding of antibodies was prevented by blocking in 0.2% fish skin gelatin (FSG) for 30 min. The cells were then incubated with mouse monoclonal anti-BrdU antibody (Beckton-Dickinson) or rat monoclonal anti-BrdU antibody (Harlan Sera-lab) along with other desired primary antibodies against replication proteins (DNA Ligase I, DNMT1, PCNA) for 1 hr at 37˚C. The primary antibodies were diluted in buffer containing 0.2% FSG, 20U/ml DNase I (Boehringer Mannheim), 0.5 mM ß-mercaptoethanol, 0.33 mM $MgCl_2$, 33 mM Tris-Cl pH 8.1. The DNA in the chromatin is partially digested by DNase I exposing the incorporated BrdU. Following incubation in primary antibody, the cells were washed three times with 0.1% NP40 in PBS followed by incubation with fluorophore conjugated secondary antibodies at room temperature. The following primary antibodies against replication and methylation proteins were used in the various experiments: rabbit anti-PATH52 (against DNMT1) [1: 2000; (Bestor, 1992)], mouse monoclonal anti-PCNA (1:1000 with methanol fixed cells, Clone PC10, Dako), rabbit anti-PCNA (1:100, FL-261, Santa Cruz), rabbit anti-human DNA Ligase I [1:250, (Cardoso et al., 1997)]. Secondary antibodies (Jackson Immuno Research and Molecular Probes) conjugated to the following fluorophores were used: fluorescein isothiocyanate (FITC), texas red (TR), Cy5, Alexa Fluor 568 and Alexa Fluor 647. The DNA was counterstained with Hoechst 33258 or TOPRO-3 and cells were mounted in mowiol with 2.5% DABCO as an anti-fading agent. In the same immunostaining, the fluorophores were selected so that there is no overlap in their excitation-emission maxima. For example, a combination of FITC, TR, Cy5 and Hoechst 33258 would be used for a quadruple staining. The excitation-emission maxima of the various fluorescent proteins used and the fluorophores are listed in Table 2.2.

2.4.2. BrdU pulse-chase for identification of cells in G2 phase

C2C12 cells were pulse labeled with 10 µM BrdU for 15 min (pulse labeling) and washed twice with pre-warmed DMEM containing 100 µM thymidine. BrdU incorporation was chased by incubating the cells in conditioned medium containing 100 µM thymidine for 2-3 hrs. The cells were then fixed with 3.7 % formaldehyde and immunostained as described in 2.4.1.

Table 2.2. Excitation and emission maxima of various fluorophores used.

Fluorescent proteins/ Fluorophore	Excitation maxima (nm)	Emission maxima (nm)
GFP	489	508
YFP	514	527
DsRed (RFP)	558	583
FITC	494*	518*
Texas red	595*	615*
Cy5	650*	670*
Alexa Fluor 568	578*	603*
Alexa Fluor 647	650*	668*
Hoechst 33258	346	460
TOPRO-3	642	661

* Approximate excitation and fluorescence emission maxima for conjugates.

2.4.3. Localization of DNMT1 at mitotic chromatin

Association of DNMT1 with mitotic chromatin was detected by fixing cells with ice cold absolute methanol for 10 min followed by immunostaining with anti-PATH52 Ab in buffer containing DNase I as described in 2.4.1. DNMT1 was detected in mitotic chromatin with anti-PATH52 Ab only in methanol fixed cells and when DNase I was included in the buffer. This suggests the epitope recognized by anti-PATH52 Ab [amino acid 255-753 (Bestor, 1992)] which lies within the TS region) is hidden and gets unmasked by DNase I treatment only in methanol fixed cells. However, association of DNMT1 with mitotic chromatin was observed with other antibodies against DNMT1 [anti-DNMT1 N-term Ab against the first 118 amino acids of DNMT1 which was a gift from Suetake, I., (Suetake et al., 2001); anti-DNMT1 C-term Ab against the C-terminal domain of DNMT1 which was a gift from Gaudet, F., (Gaudet et al., 1998)] in formaldehyde fixed cells.

2.4.4. Obtaining cells in G1

G1 cells were obtained by isolating mitotic cells by mechanical shake off. C2C12 cells (50% confluent) growing on a 100 mm dish with 10 ml medium were vigorously agitated without spilling the medium. Mitotic cells being loosely attached get detached from the surface. The mitotic cells were harvested by centrifugation at 228 g for 5 min and resuspended in 2 ml medium. The cells were then laid on coverslips in a 35 mm dish and incubated for 2 hrs. The cells were pulse labeled with BrdU and immunostained as described in 2.4.1. None of the cells stained BrdU positive indicating that they were still in G1.

2.5. Microscopy

Immunostained cells were examined on a Zeiss LSM 510 microscope with 63x or 100x NA 1.4 Plan-Apochromat oil immersion objective with Nomarsky optics. Ar-laser (488, 514 nm), HeNe-laser 1 (543 nm) and HeNe-laser 2 (633 nm) were used to excite the fluorophores. The excitation lasers and detection filters used for the different flurophores are listed in Table 2.3. Images were acquired using the Zeiss LSM510 software and processed, assembled and annotated using Adobe Photoshop 5.5 and Adobe Illustrator 8.0.1.

Hoechst 33258 counterstained nuclei were imaged in an Axioplan 2 microscope equipped with phase-contrast and epifluorescence optics, using a 63x NA 1.4 Plan-Apochromat oil immersion objective. A Hg lamp was used as the light source in combination with excitation and emission filters listed in Table 2.3. Images were acquired with a cooled CCD camera (SensiCam) using Zeiss Axiovision software.

Table 2.3. Excitation and emission filters used for detecting the signals from the different fluorophores and fluorescent proteins.

Fluorophore/ Fluorescent protein	Zeiss LSM 510			Zeiss Axioplan 2		
	Excitation Laser (nm)	Beam splitter (nm)	Emission filters (nm)	Excitation Filters (nm)	Beam splitter (nm)	Emission filters (nm)
FITC, GFP, YFP*	488	HFT UV/488/ 543/633	500-530 BP	450-490 BP	510	515-565 BP
TR, Alexa Fluor 548, DsRed (RFP)	543	HFT UV/488/ 543/633	565-615 BP	530-585 BP	600	615 LP
Cy5, Alexa Fluor 647, TOPRO-3	633	HFT UV/488/ 543/633	650 LP	575-625 BP	645	660-770 BP
Hoechst 33258	-	-	-	365/12 BP	395	397 LP

* YFP can be excited also by 488nm laser. **Note**: BP = Band pass; LP = Long pass; HFT = Hauptfarbteiler (Main beam splitter)

2.6. Live cell microscopy

For live cell microscopy, cells were grown on 40 mm diameter glass coverslips and cotransfected with plasmids encoding F-TS-GFP or GMT and RFP-Ligase (cell cycle marker for live cells). 24 hrs after transfection, the coverslip was assembled in a FCS2 live cell microscopy chamber (Bioptechs). The chamber was mounted onto the stage of the microscope and the temperature of the coverslip was maintained at 37˚C using a temperature controller (Bioptechs). Care was taken to prevent photo-damage of cells by screening the cells with minimum intensity of excitation light. Images were acquired using the 488 nm and 543 nm laser lines at low power (1-5%). Four Z-stacks of 1 μm distance were imaged every hour and the cells were followed throughout the cell cycle. During the progress of imaging over a long period, the fluorescence decreased because of bleaching. In such cases, the detector gain was increased to detect weak signals.

After image acquisition, the different Z scans at each time point were analyzed visually to correct for movement of the cell in the Z plane by arranging them sequentially so as to include the same structures in each sequential image. A sequence of images were selected and processed in Adobe Photoshop 5.5 and Adobe Illustrator 8.0.1 for assembly and annotation in the figures. Movies were created from these sequential images using Adobe Premiere 5.1.

2.7. Sequence analysis

2.7.1. Search for DNA Ligase I homologue in Drosophila

DmDNA Ligase I cDNA was identified by searching the *Drosophila* genome database and the EST database (BDGP) with DNA Ligase I sequence from different organisms (mouse, human, yeast) as query using the BLAST program (Altschul et al., 1990). The sequences obtained were aligned pairwise using BLAST2 (at BCM Search Launcher http://searchlauncher.bcm.tmc.edu/) for generating the schematic of alignments shown in Fig. 3.8. Phylogenetic comparison of the human DNA Ligases and the putative *Drosophila* Ligases were done using the Jotun-Hein method (Hein, 1990) in Lasergene program. The DmDNA Ligase I cDNA clone (LD41868) was obtained from ResGen.

2.7.2. Multiple sequence alignments, profiles and profile search

Multiple sequence alignments were generated using programs available with the Lasergene software or the Heidelberg Unix Sequence Analysis Resources (HUSAR) (DKFZ, Heidelberg). The specific programs used in each case are mentioned in the figure legends in the Results section. The multiple sequence alignment in Fig. 3.29 is displayed using PrettyBox, which shades regions that agree with a calculated consensus sequence. In some cases the alignment obtained were edited by visual inspection to get maximal alignment.

A ‘profile’ is a quantitative representation of the occurrence of residue at a given position in a group of aligned sequences. Profiles were generated in cases

where the set of related sequences have small sequence lengths (10-30 amino acids; the PBD and TS motif). Generation of profiles and searching of databases with the profiles were carried out with the ProfileMake and ProfileSearch programs (Gribskov et al., 1987) available at HUSAR. Profiles of the aligned sequences were made with default settings. ProfileMake creates the profile which is a position-specific scoring table that quantitatively represents the information from a group of aligned sequences. The profile generated was used to search for similar sequences in a given sequence(s) or the Swissprot and GenBank databases using ProfileSearch with default settings.

2.7.3. PSI-BLAST searches

PSI-BLAST (position-specific iterated-BLAST) is a database search program that automatically combines three distinct operations: it constructs a multiple sequence alignment from BLAST output data; it calculates a position-specific score matrix from this alignment; and it uses this matrix to search the database for similar sequences (Altschul et al., 1997). PSI-BLAST (NCBI) was used for searching the non-redundant peptide database for sequences similar to the TS. After the first round of iteration, only the statistically significant sequences (E-value better than threshold) in the BLAST output data were selected for calculating the position-specific score matrix for successive iterations. E-value is the expectation value - the number of different alignments that is expected to occur in a database search by chance with scores equivalent to or better than the one obtained. After four iterations, no more new sequences were obtained.

3. Results

3.1. Association of proteins with RF in *Drosophila*

3.1.1. PCNA is highly conserved in S. cerevisiae, D. melanogaster and mammals

The mechanistic basis for the association of replication factors with RF has been proposed to be mediated via interaction of PBD with PCNA. PCNA functions as a processivity factor for the replicative DNA polymerases (Bravo et al., 1987) (Prelich et al., 1987). To understand whether the PBD-mediated association of proteins with RF is conserved in evolution, we first analyzed the conservation of PCNA protein sequence from divergent eucaryotes, especially the regions shown to interact with PBD of $p21^{Cip1/Waf1}$ (Gulbis et al., 1996). PCNA is highly conserved in organisms as divergent as yeast, flies and mammals (Fig. 3.1). The residues of PCNA that interact with PBD of $p21^{Cip1/Waf1}$ are more than 75% conserved in human, *Drosophila* and yeast. Also the PBD is conserved in homologous proteins from archaebacteria, yeast, worms, flies, amphibians and mammals (Fig. 3.2) (Warbrick et al., 1998) suggesting that the association of proteins to RF mediated by PCNA is a mechanism conserved through evolution. We put this hypothesis to test by analyzing whether the PBD of DNMT1 can associate with RF in *Drosophila* cells (S2 cells). Further, nothing is known about the mechanism underlying the association of the TS and PBHD of DNMT1 with RF. To understand whether these mechanisms are conserved in divergent eucaryotes, we have analyzed the ability of the TS and PBHD to target to RF in *Drosophila* cells (S2 cells) as it lacks DNA methylation. (see Section 1.3.5).

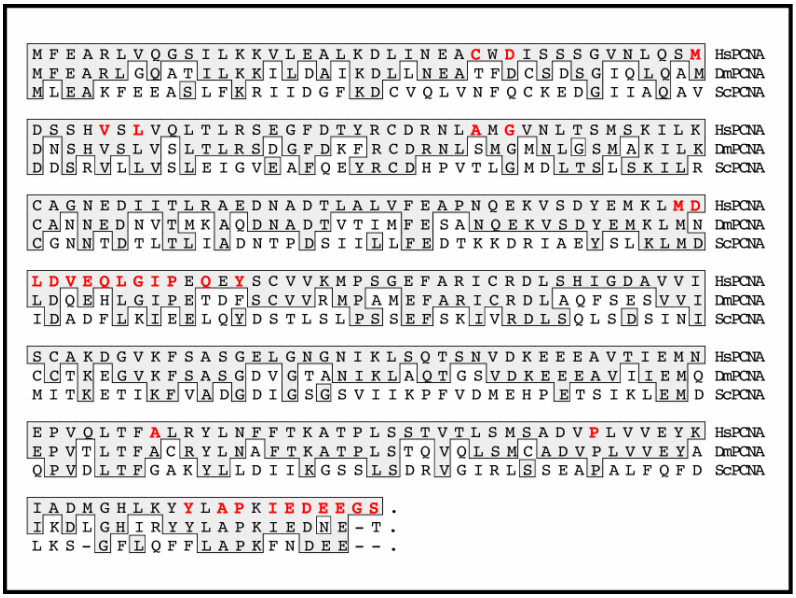


Fig. 3.1. Drosophila and human PCNA are highly similar. *Sequence alignment of HsPCNA (human), DmPCNA (Drosophila) and ScPCNA (budding yeast). The protein sequences were aligned using the Jotun-Hein method (in Lasergene software). Residues identical to HsPCNA are boxed. HsPCNA is about 35% identical to ScPCNA and 70% identical to DmPCNA. Regions in HsPCNA that interacts with the PBD of p21 (Gulbis et al., 1996) are in red.*

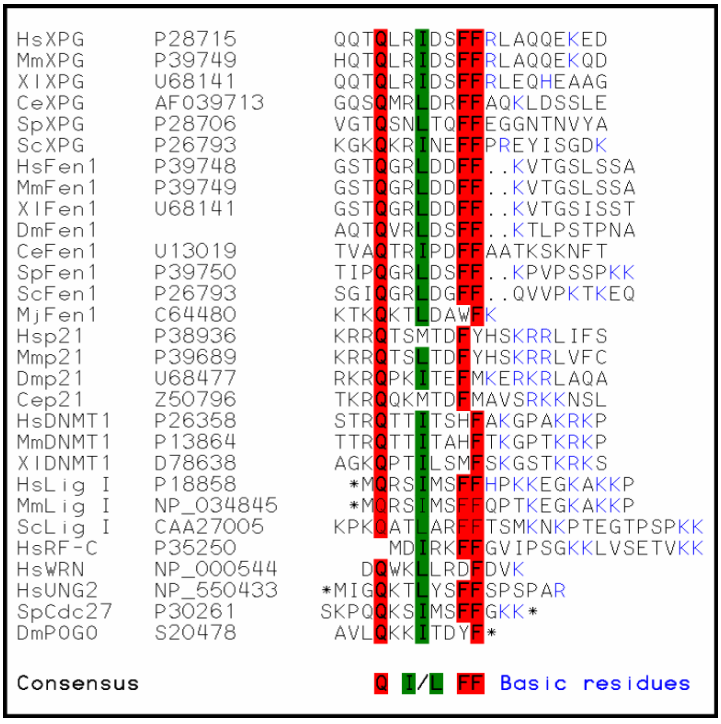


Fig. 3.2. A conserved PBD in proteins involved in DNA metabolism. *Alignment of the PBD from various homologous proteins from different organisms. Accession numbers are indicated in the second column where available. Asterisk denotes the beginning or end of the protein. Identical residues are highlighted in red and conserved substitution is highlighted in green. Association with replication foci requires both the conserved residues and the stretch of peptide rich in basic amino acids (in blue) (Montecucco 1998). (Hs: Homo sapiens; Mm: Mus musculus; Xl: Xenopus laevis; Dm: Drosophila melanogaster; Ce: Caenorhabditis elegans; Sp: Schizosaccharomyces pombe; Sc: Saccharomyces cerevisiae; Mj: Methanococcus jannaschii).*

3.1.2. Human PCNA and Drosophila PCNA can associate with RF across the two organisms

To answer our main question whether the three RFTS from mouse DNMT1 can function in *Drosophila*, we had to first establish whether the replication machinery in the two systems were similar in that a conserved core replication factor from one organism would be targeted to the RF in the heterologous system. To this end, we analyzed the association of human PCNA (HsPCNA) with RF in *Drosophila* cells (S2 cells) and *Drosophila* PCNA (DmPCNA) in mouse cells (C2C12 cells). Plasmids encoding HsPCNA, GFP-HsPCNA and GFP-DmPCNA were constructed and introduced into cells by calcium phosphate transfection. Subnuclear localization of HsPCNA in transfected S2 cells was accomplished by indirect immunostaining with anti-PCNA antibody (FL261, Santacruz) that reacted specifically with HsPCNA (Fig. 3.3A). Transfected cells in S-phase showed punctate BrdU staining representing RF and HsPCNA colocalized with these foci (Fig. 3.3B). Also GFP-HsPCNA is targeted to RF in S2 cells (Fig. 3.3B). Similarly, C2C12 cells expressing GFP-DmPCNA showed complete colocalization of GFP-DmPCNA with RF (Fig. 3.3C). Since the available antibody against DmPCNA cross reacts with mouse PCNA, localization of untagged DmPCNA in mouse cells was not possible. Owing to the high sequence similarity of HsPCNA with DmPCNA (Fig. 3.3A) and the conserved

structure of PCNA from divergent eucaryotes (Krishna et al., 1994), efficient association of HsPCNA with RF in S2 cells indicates that HsPCNA must be replacing endogenous PCNA in S2 cells. Thus, these results indicate that the basic replication machinery in the two organisms is conserved and core replication factors are inter-changeable.

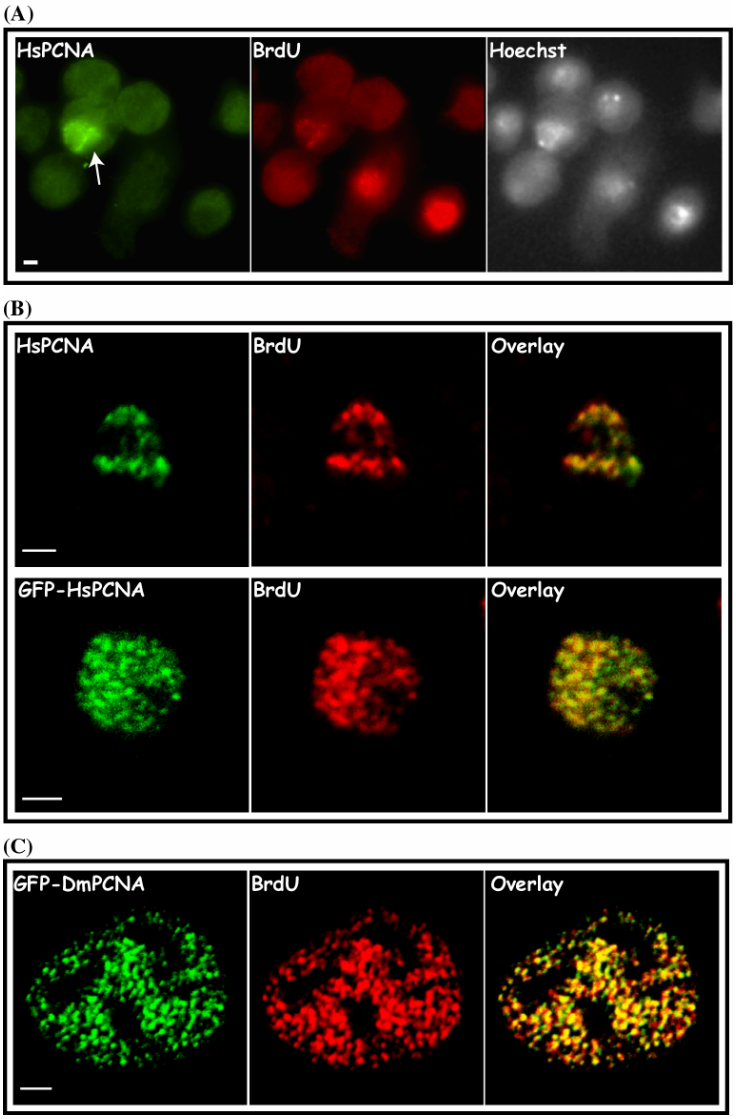


Fig. 3.3. Mammalian and Drosophila PCNA associate with RF interchangeably between these divergent organisms. *Drosophila S2 and mammalian C2C12 cells were transfected, pulse labeled with BrdU to label sites of active DNA replication, and fixed with formaldehyde followed by immunostaining.* ***(A)*** *S2 cells transfected with plasmid encoding HsPCNA and coimmunostained with anti-PCNA (FL261; Santa Cruz) and anti-BrdU antibodies. Cells positively staining with anti-BrdU were in S phase at the time of BrdU labeling. Anti-PCNA antibody (FL261) specifically stains one transfected cell (arrow) expressing HsPCNA that is in S phase and does not stain endogenous PCNA in the other non-transfected S phase and non-S phase cells.* ***(B)*** *S2 cells transfected with plasmid encoding HsPCNA (top) or GFP-HsPCNA (bottom) and immunostained with anti-BrdU. HsPCNA was detected by coimmunostaining with anti-PCNA antibody (FL261) as in (A). Overlay shows colocalization of HsPCNA and GFP-HsPCNA with BrdU foci. (C) C2C12 cells transfected with plasmid encoding GFP-DmPCNA and immunostained with anti-BrdU. GFP-DmPCNA colocalizes with BrdU foci. Scale bar = 2 µm.*

3.1.3. Subnuclear localization of DNMT1 in Drosophila cells

In order to determine whether the mechanisms by which PBD, TS and PBHD mediate association with RF in mammalian cells (methylated genome) are conserved in *Drosophila* (unmethylated genome), we evaluated the ability of these domains to associate with RF in S2 cells. S2 cells were transfected with plasmids encoding chimeric fusions of the N-terminal domain of DNMT1 or the PBD/TS/PBHD with GFP/YFP/ß-gal epitope (summarized in Fig. 3.4). As shown in Fig. 3.5A, NMT^2-GFP associates with RF in mouse C2C12 cells. NMT^2-GFP has a PBD and therefore it is predicted that this fusion protein would be targeted to RF in *Drosophila* S2 cells. Surprisingly, NMT^2-GFP does not colocalize with RF in S2 cells (Fig. 3.5B). Rather, in S2 cells, NMT^2-GFP seems to be diffused throughout the nucleus as well as aggregating in some regions to form large structures. To test whether this is an artifact of the fusion protein in S2 cells, S2 cells were transfected with a plasmid encoding full length untagged DNMT1(s) and the association of DNMT1(s) with RF was analyzed. Like NMT^2-GFP, DNMT1(s) is not associated with RF (Fig. 3.5C top panel) in the majority of cells, though in some cells in S-phase DNMT1(s) seemed to partially colocalize with RF (Fig. 3.5C; bottom panel), or with a subset of *Drosophila* chromatin.

It is surprising that a PBD-containing protein that efficiently associates with RF in mammalian cells does not behave similarly in *Drosophila* cells even though their basic replication machinery is similar. Reduced efficiency in the targeting of DNMT1 to RF in *Drosophila* cells could be an artifact of a protein which is very foreign to *Drosophila* as it lacks a DNMT1 homologue. So we tested the targeting of human DNA Ligase I (HsDNA Lig I), which is a core replication factor involved in ligating Okazaki fragments during synthesis of lagging strand, to RF in S2 cells. DNA Ligase I is conserved in evolutionarily distant organisms and the *Drosophila* DNA Ligase I (DmDNA Lig I) homologue is 50% identical to HsDNA Lig I (Fig. 3.8). S2 cells were transfected with a plasmid encoding a GFP fusion to HsDNA Lig I (GFP-HsLig) and the subnuclear localization of GFP-HsLig with respect to RF was analyzed. GFP-HsLig showed a diffused nuclear distribution in S2 cells that are in S phase and did not show any discernible colocalization with RF (Fig. 3.6A). In contrast, consistent with previous studies demonstrating association of human DNA Ligase I with RF (Cardoso et al., 1997), GFP-HsLig is efficiently targeted to RF in mouse C2C12 cells (Fig. 3.6B). Unlike NMT^2-GFP and DNMT1(s) that aggregates in structures, GFP-HsLig showed a diffused nuclear distribution in all S2 cells observed. This aggregation could be due to the various other functional domains present in the N-terminus of DNMT1. Thus, taking together the subnuclear localization of NMT^2-GFP, DNMT1(s) and GFP-HsLig in S2 cells, it appear with plasmid encoding HsPCNA aefficiently targeted to sites of DNA replication in *Drosophila* cells. Strikingly, the inability of GFP-HsLig to associate with RF in any of the S2 cells in S phase indicates that PBD is unable to function as an RFTS in *Drosophila*. Partial association of DNMT1(s) with RF in some S2 cells in S phase might mean that the other two RFTS in DNMT1 might be able to target the protein to RF, although at a low efficiency.

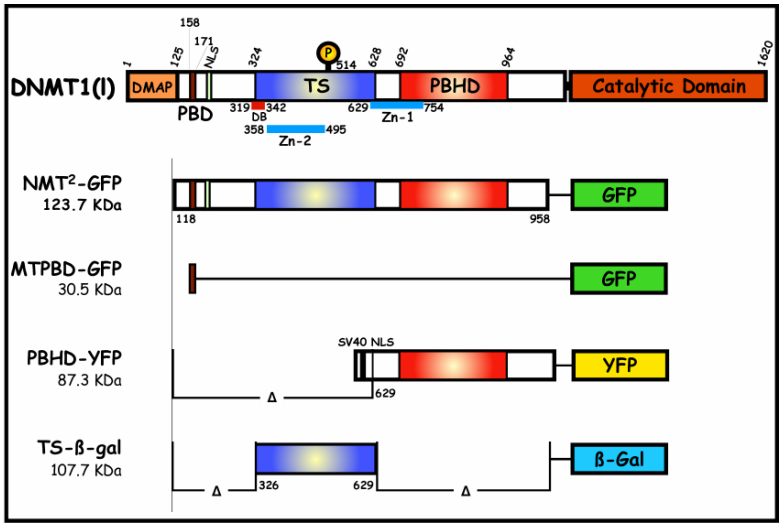


Fig. 3.4. Deletions of DNMT1 used for evaluating the ability of PBD, TS and PBHD to associate with RF in S2 cells. *The somatic isoform, DNMT1(l), is illustrated at the top. Various deletions in the N-terminus fused to GFP/YFP/ß-gal epitope are shown. The predicted molecular weights of the fusion proteins are indicated. See Fig 3.12 for Western blot analysis of the fusion proteins. The boundaries of the different domains and deletions are indicated by the amino acid numbers corresponding to the DNMT(l) isoform. See Fig. 1.7 for description of the domains. An SV40 NLS was added in PBHD-YFP for efficient nuclear targeting.*

Next, we asked whether the weak targeting of DNMT1(s) to RF in S2 cells can be attributed to the TS and PBHD. In order to answer this, we determined the ability of each of the RFTS from DNMT1 to associate with RF in S2 cells. The plasmid constructs used are summarized in Fig. 3.4. The PBD from DNMT1 and HsDNA Lig I were fused individually to GFP to test the ability of the minimal PBD to associate with RF in *Drosophila* cells. As shown in Fig. 3.7A, MTPBD-GFP and HsLigPBD-GFP both are diffusely distributed in the nucleus and did not show anyidiscernible localization or exclusion from the RF. The same constructs when expressed in mouse cells show a similar diffused pattern in many cells but are clearly targeted to RF in other cells (see Section 3.2). In S2 cells, none of the transfected cells in S phase showed targeting of both PBDs fused to GFP. Thus, even though the 10 aa PBD plus the 10 aa region rich in basic residues are enough for targeting to RF as observed in mouse cells, it is unable to do so in *Drosophila* cells. Ability of the PBD to assemble at replication sites requires interaction with PCNA (Chuang et al., 1997) (Montecucco et al., 1998). Inability to target might also be because the PBD from mammalian origin is not able to interact with the endogenous DmPCNA in S2 cells. Although this is unlikely as the PCNA is highly conserved between human and *Drosophila* (Fig. 3.1), and the PBD from *Drosophila* proteins, like $p21^{Cip1/Waf1}$ and POGO transposase, are similar to mammalian PBD (Fig. 3.2) and can efficiently interact with DmPCNA *in vitro* (Warbrick et al., 1998). We tested whether HsPCNA would aid targeting of mammalian PBD to RF in *Drosophila* by co-transfecting plasmids encoding MTPBD-GFP and HsPCNA in S2 cells. As can be seen in Fig. 3.7B, MTPBD-GFP still remained diffused in the nucleus while HsPCNA was associated with RF. Thus, the inability of mammalian PBD to target to RF in *Drosophila* is not due to the absence of its natural interacting partner (HsPCNA).

The previously described TS-ß-gal epitope tagged fusion used for the mapping of the RFTS in DNMT1 (Leonhardt et al., 1992) (Fig. 3.4) was used to test the ability of the TS to associate with RF in *Drosophila* cells. Although, as previously shown

(Leonhardt et al, 1992), this fusion protein associates with RF in mammalian cells, in S2 cells it did not associate with RF (Fig. 3.7C). This fusion protein is not efficiently imported into the nucleus as can be seen from the strong cytoplasmic staining. However, inability to associate with RF cannot be attributed to this because the TS-ß-gal fusion protein can be clearly detected in the nucleus as well. Moreover, in spite of its weak nuclear localization, it efficiently associates with RF in mammalian cells, whereas none of the S2 cells in S-phase showed association of TS-ß-gal with RF. Similarly, expression of PBHD-YFP in S2 cells showed that the fusion protein is unable to associate with RF in S2 cells (Fig. 3.7D). However, PBHD-YFP did not colocalize with RF even in mouse cells (Section 3.2).

Thus, none of the three RFTS from DNMT1 is able to associate with RF in *Drosophila* cells. Most often these fusion proteins showed a diffused nuclear distribution. Inability of these domains to associate with RF could be due to the following reasons: (a) Possibly these ectopically expressed proteins are produced in high amounts and as a result, the excess protein remains diffused throughout the nucleus making it difficult to discern the population which associates with RF and that which is unbound and present in the nucleoplasm. However, this seems unlikely because the amount of protein produced is quite low as judged by the weak fluorescence intensity in transfected cells. Also, it is difficult to detect protein expression in transfected S2 cells by western blotting as compared to mammalian cells. Moreover, in S2 cells ectopically expressed HsPCNA and GFP-HsPCNA associates with RF (Fig. 3.3B). (b) Alternatively, in *Drosophila*, the mechanism behind recruitment of proteins at RF may not be dependent on the PBD, but rather some other unique protein motif might function as an RFTS.

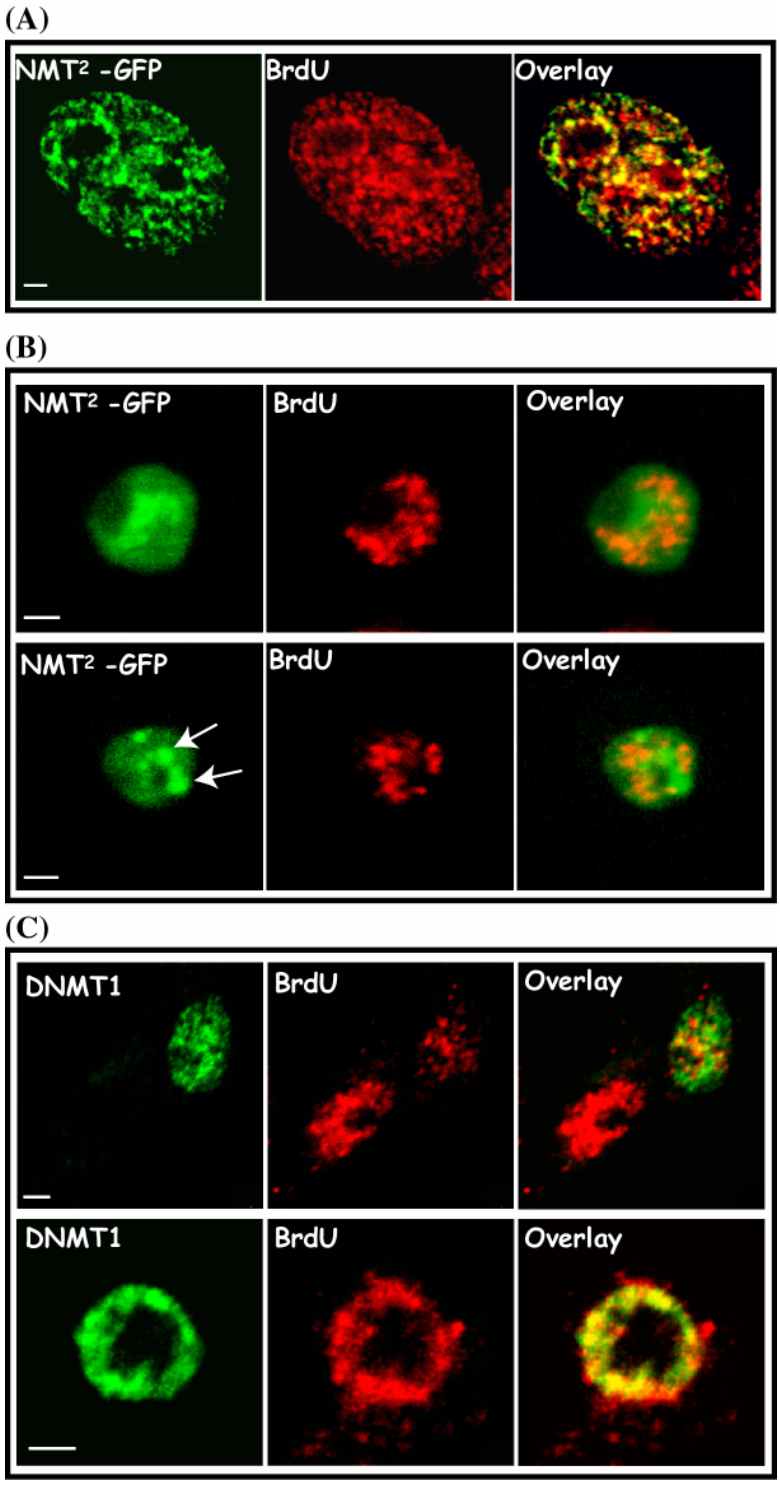


Fig. 3.5. Subnuclear localization of NMT^2-GFP and DNMT1(s) with RF in S2 cells. *Cells were transfected and pulse labeled with BrdU 24 hrs later followed by indirect immunostaining for BrdU.* ***(A)*** *C2C12 cells transfected with plasmid encoding NMT^2-GFP. Colocalization of NMT^2-GFP with BrdU is seen in the overlay as yellow.* ***(B)*** *S2 cells transfected with plasmid encoding NMT^2-GFP. NMT^2-GFP does not colocalize with BrdU. In most cells it aggregates to form structures (arrows in bottom panel) that are excluded from BrdU incorporation sites.* ***(C)*** *S2 cells transfected with a plasmid encoding DNMT1(s). Cells were immunostained with anti-pATH52 to detect DNMT1. In 80-90 % of cells, DNMT1 is excluded from BrdU sites (top panel). In about 10% of S2 cells, DNMT1 partially colocalizes with BrdU incorporated sites (bottom panel). Scale bar = 2 µm*

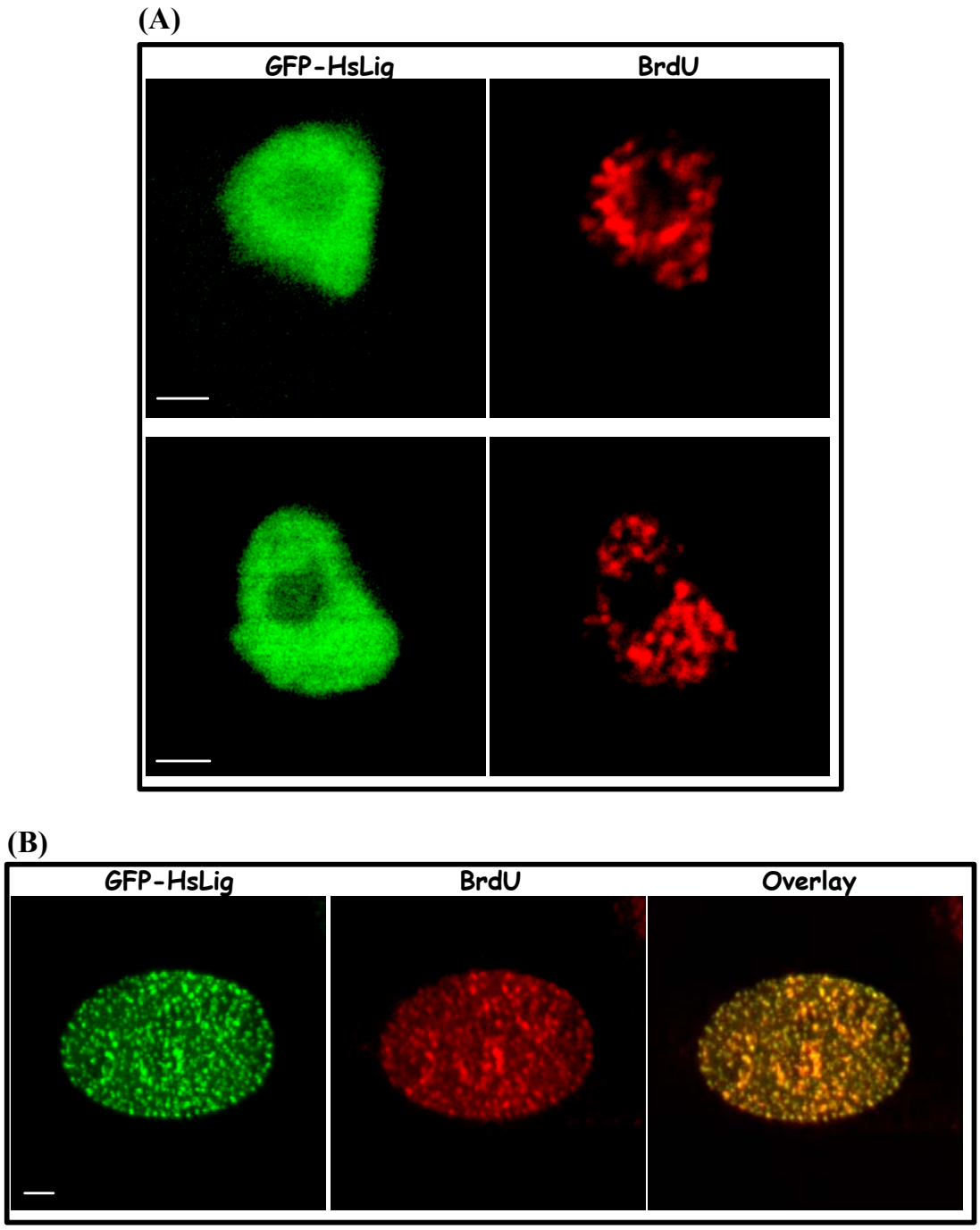


Fig. 3.6. GFP-HsLig does not associate with RF in Drosophila S2 cells. *Cells were transfected with plasmid encoding GFP-HsLig and pulse labeled with BrdU 24 hrs later followed by indirect immunostaining for BrdU.* ***(A)*** *In S2 cells, GFP-HsLig shows diffused nuclear localization in all cells observed.* ***(B)*** *GFP-HsLig colocalizes with BrdU labeled RF in C2C12 cells seen as yellow colour in overlay. Scale bar = 2 µm*

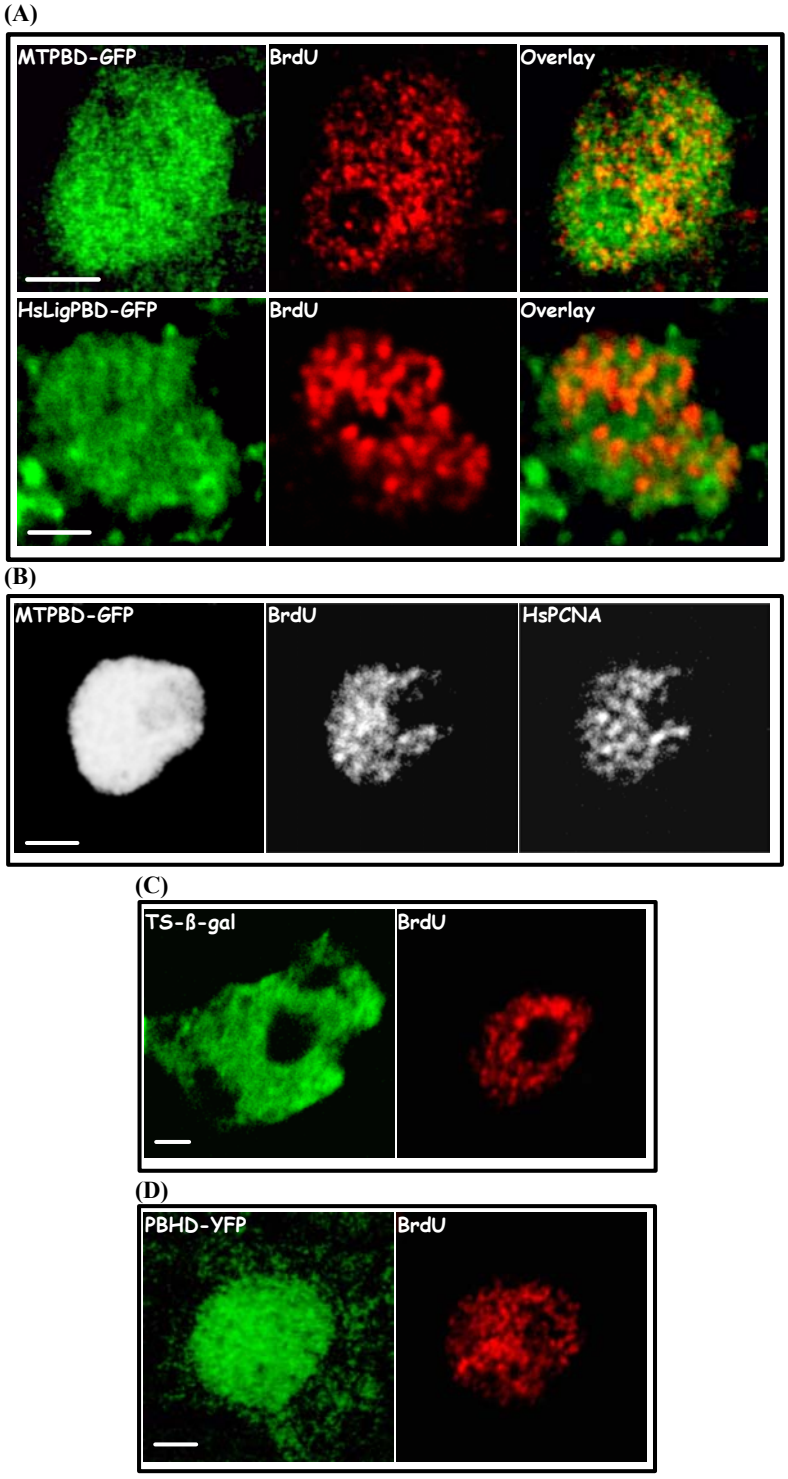
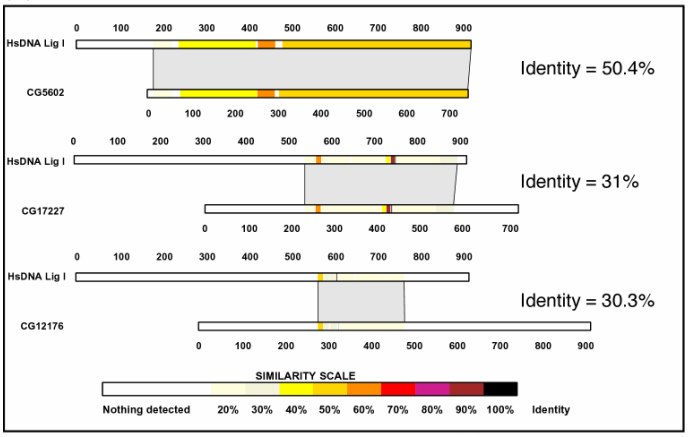


Fig. 3.7. PBD, PBHD and TS do not mediate association with RF in S2 cells. *S2 cells were transfected and pulse labeled with BrdU 24 hrs later followed by indirect immunostaining for BrdU.* ***(A)*** *S2 cells transfected with plasmid encoding MTPBD-GFP (top panel) or HsLigPBD-GFP (bottom panel). Both PBD-GFP fusions show diffused nuclear distribution.* ***(B)*** *S2 cells cotransfected with plasmids encoding MTPBD-GFP and HsPCNA. Presence of HsPCNA does not aid association of MTPBD-GFP with RF as MTPBD-GFP still shows diffused nuclear distribution.* ***(C)*** *S2 cells transfected with plasmid encoding TS-ß-gal. ß-gal epitope was detected by indirect immunostaining with anti-ß-gal antibody. (D) S2 cells transfected with plasmid encoding PBHD-YFP. Both TS-ß-gal and PBHD-YFP show diffused nuclear distribution. Scale Bar = 2 µM.*

3.1.4. Search for an RFTS in Drosophila

In order to determine whether another protein sequence is functioning as an RFTS in *Drosophila*, we sought to clone a replication protein from *Drosophila* and identify the RFTS in it. Since DNA Ligase I is a well conserved protein from yeast to humans, the *Drosophila* genome database and the EST database (BDGP) were searched for putative DNA Ligase I homologues with DNA Ligase I sequence from different organisms (mouse, human, yeast) as query. Three putative homologues were identified in the EST database, viz. LD41868. AT15112, and LD06019 with corresponding predicted proteins CG5602, CG17227 and CG12176 respectively (Fig. 3.8). Comparison of the protein sequences of each of these putative homologues with HsDNA Lig I showed that CG5602 has maximum identity (53.7 %) to HsDNA Ligase I spanning over a large region of the protein as compared to CG17227 (31% identical) and CG12176 (30.3% identical) (Fig. 3.8A). Moreover, comparison of each of the three putative *Drosophila* homologues with the three mammalian DNA ligases, viz. Ligase I, Ligase III and Ligase IV, indicated that CG17227 and CG12176 are homologues of Ligase III and Ligase IV respectively (Fig. 3.8B). Also, a phylogenetic analysis of the human and putative *Drosophila* ligases showed that CG5602 is closer to HsDNA Lig I (Fig. 3.8C). Thus, these sequence analyses indicate that the putative DNA ligase I in *Drosophila* is CG5602 (henceforth called DmDNA Lig I). The EST clone (LD41868) coding for CG5602 was obtained and the sequence was verified. An 81 bp segment in the coding region of the predicted gene was found to be absent in the cDNA clone (Fig 3.9A). The DmDNA Lig I protein sequence was analyzed in order to identify the presence of sequences similar to the PBD or the other RFTS from DNMT1. This revealed the presence of a PBD sequence (designated as DmDNA Lig I-PBD) that had all the features of HsDNA Lig I-PBD required for association with replication foci, viz. the 10 amino acid region interacting with PCNA immediately followed by the 10 amino acid sequence rich in basic residues (Fig. 3.9B). Additionally, two sequences with weak similarity to the conserved PBD were identified and called as PBD-1 and -2 . No sequence with similarity to PBHD and TS were observed. The DmDNA Lig I protein sequence was searched for potential NLS using signature NLS sequences like Caudal NLS, Engrailed NLS, SV40 Large T NLS, Nucleoplasmin NLS, and as well the NLS from HsDNA Ligase I (Montecucco et al., 1995) (Cardoso et al., 1997). Two regions in DmDNA Lig I showed similarity to the nucleoplasmin and T antigen NLS (labeled as NLS-1 and NLS-2, Fig. 3.10A).

(A)



(B)

	CG5602	CG17227	CG12176
HsDNA Lig I	**50.4 %**	31%	30.3 %
HsDNA Lig III	30.3 %	**51.4 %**	24.7 %
HsDNA Lig IV	30 %	28.4 %	**33.2 %**

(C)

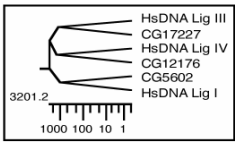


Fig. 3.8. Basis for selection of the EST clone LD41868(CG5602) as the putative DNA ligase I homologue in Drosophila. *BLAST search of the D. melanogaster genome and EST database with DNA Ligase I (from different species) identified three similar sequences CG5602, CG17227 and CG12176.* ***(A)*** *Each of the putative DNA ligase protein sequences identified in Drosophila is aligned pairwise with HsDNA Lig I. The percentage identity obtained from the alignment is indicated at the right. CG5602 shows maximum identity to HsDNA Lig I compared to the other two sequences.* ***(B)*** *Table showing the percentage identity obtained from pairwise alignment of the three putative DNA ligase homologues in Drosophila with each of HsDNA Ligase I, III and IV. CG17227 and CG12176 show more identity with HsDNA Lig III and IV respectively.* ***(C)*** *Phylogenetic tree from the comparison of HsDNA ligases and the Drosophila homologues created in Lasergene program (J. Hein method with PAM250 residue weight table. Scale represents the amino acid substitution).*

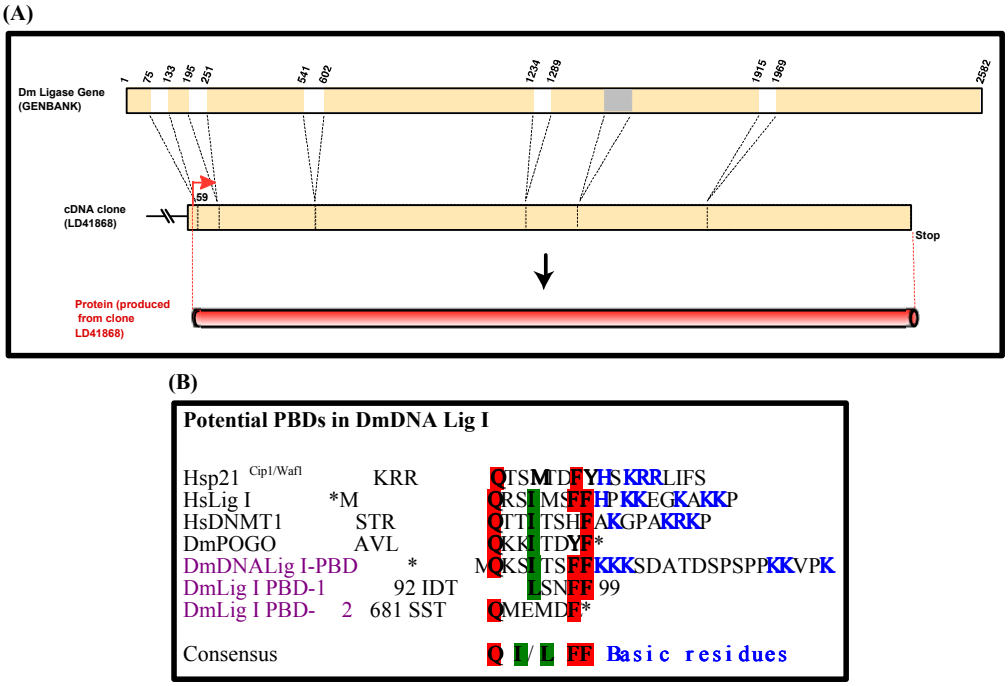


Fig. 3.9. DmDNA Lig I gene, putative protein and potential PBD. (A) *Structure of DmDNA Lig I gene and the predicted protein. The gene structure is illustrated at the top and position of exons are marked by nucleotide numbers. The regions shown to be spliced out are the introns (white boxes). A region in the 5-th exon (marked gray) was found to be absent in the cDNA clone. This region was not detected as an intron using programs to detect introns. Middle shows the LD41868 cDNA clone obtained from ResGen. The translational start is shown by the bent arrow and the protein produced from this cDNA clone is illustrated at the bottom.* ***(B)*** *The DmDNA Lig I protein sequence was searched for presence of a PBD (including the deleted region) using ProfileSearch. The sequence which showed all the features of a PBD essential for association with RF is labeled as DmDNA Lig I-PBD. The other two hits show weak similarity. Asterisk indicates the beginning or end of the protein sequence.*

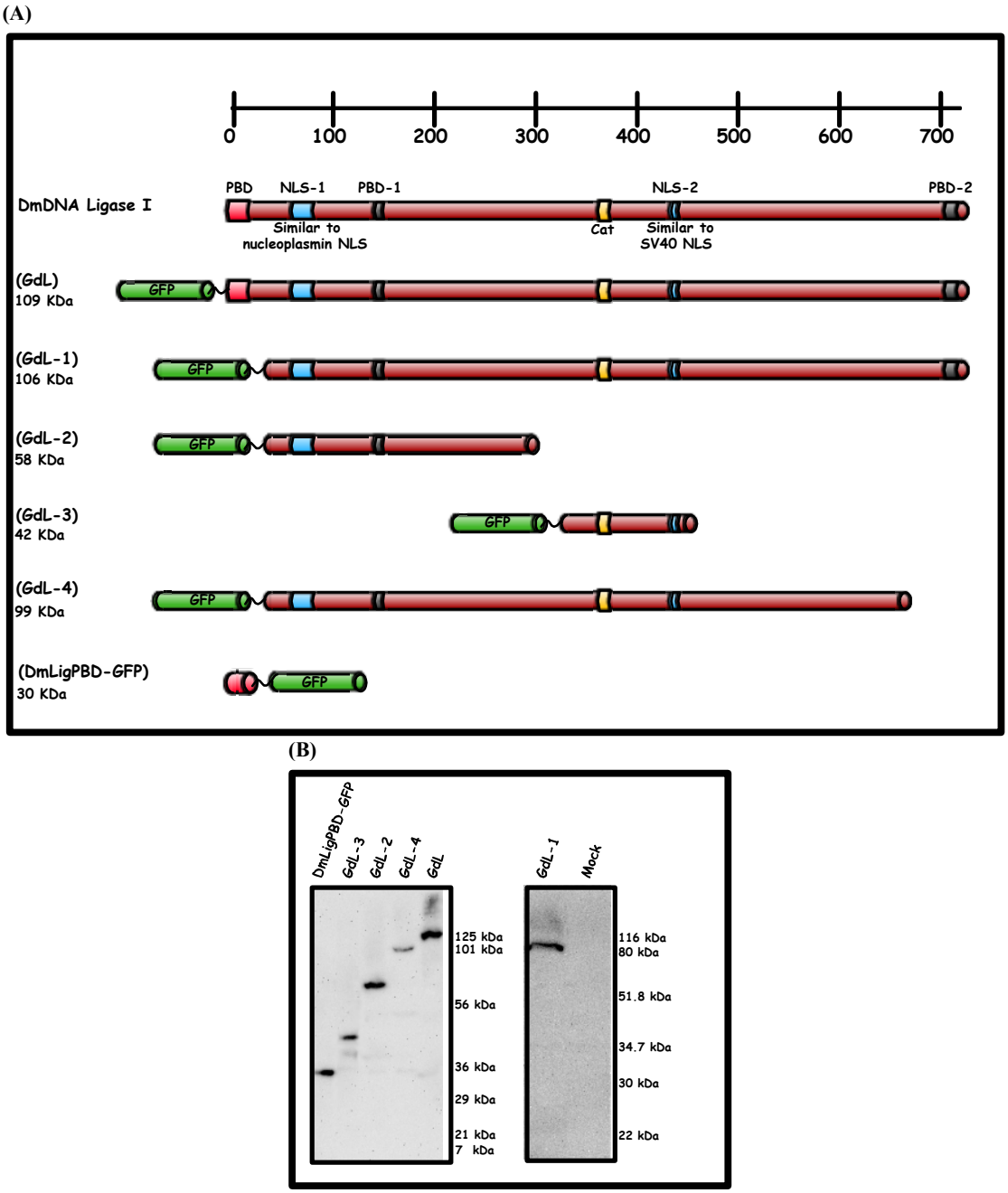


Fig. 3.10. Deletion constructs used for mapping of the RFTS in DmDNA Lig I. (A) *Organization of domains in DmDNA Lig I is shown at the top. Various deletions of DmDNA Ligase I fused to GFP at the N-terminus were made as described in Methods. The predicted molecular weight of the fusion proteins is indicated on the left.* ***(B)*** *Western blot analysis of the chimeric proteins confirmed their protein sizes. COS7 or 293T cells were transfected with the fusion constructs shown in (A) and after 24 hrs of incubation the cells were lysed by boiling in Laemmli's loading buffer and analyzed by immunoblotting with rabbit polyclonal antibody against GFP.*

In order to determine whether DmDNA Lig I employs some other protein sequence as an RFTS instead of the conserved PBD, a series of deletion mutants of DmDNA Lig I fused to GFP were made (Fig. 3.10A) and their subnuclear localization with respect to RF was analyzed in both mouse and *Drosophila* cells. Immunoblotting analysis showed that proteins of the right size were expressed from these plasmids

(Fig. 3.10B). The full length protein fused to GFP (GdL) did not show association with RF in S2 cells (Fig. 3.11A), whereas it colocalized with RF in C2C12 cells (Fig. 3.11B). None of the deletions, GdL-1, -2, -3 and –4, showed colocalization with RF in both S2 and C2C12 cells (Fig. 3.11A and B). Although GdL-2 and GdL-4 are not efficiently targeted to the nucleus, lack of association with RF cannot be attributed to this as these fusion proteins are not completely excluded from the nucleus. Thus, neither the full length DmDNA Lig I nor any of the deletions fused to GFP associated with RF in *Drosophila* cells. The fact that only GdL, but none of GdL-1 to 4, associates with RF in mouse cells indicates that this association is mediated by the DmDNA Lig I-PBD. Most importantly, it shows that GdL is folded properly and inability of GdL to associate with RF in *Drosophila* cells cannot be due to improper folding of the protein.

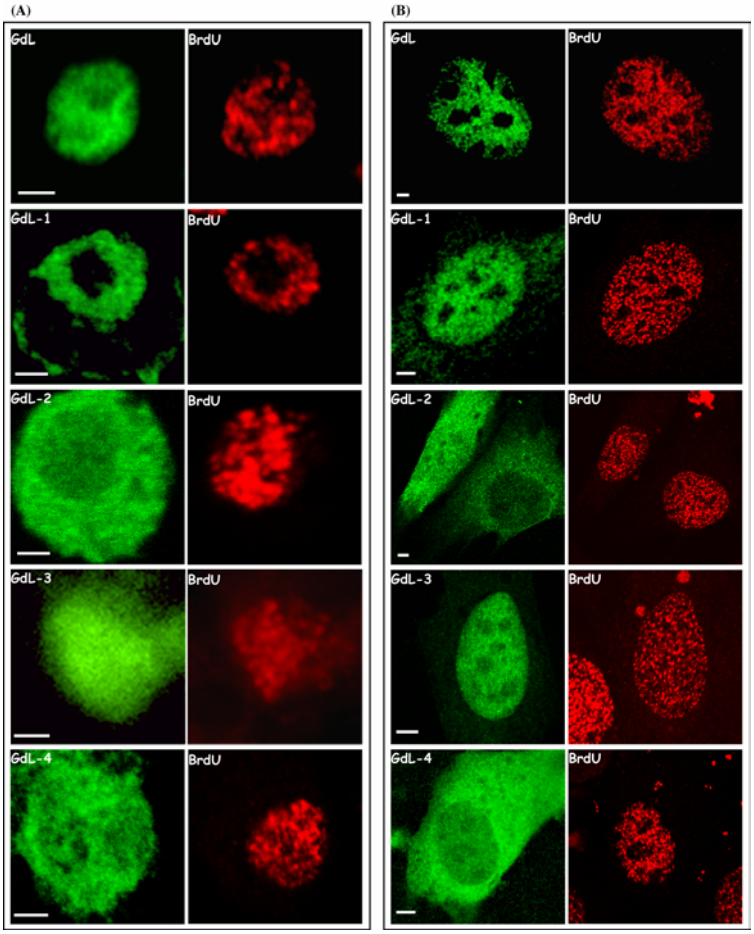


Fig. 3.11. Subcellular localization of the DmDNA Lig I deletions fused to GFP. *Cells were transfected with plasmids encoding the various deletions of DmDNA Lig I fused to GFP and pulse labeled with BrdU after 24 hrs followed by indirect immunostaining for BrdU.* ***(A)*** *None of the fusion proteins associate with RF in S2 cells.* ***(B)*** *Only the full length DmDNA Ligase I fused to GFP (GdL) associates with RF in C2C12 cells. GdL-2 and GdL-4 are not efficiently taken up into the nucleus. Scale Bar = 2 µm.*

The ability of the the minimal DmDNA Lig I-PBD to associate with RF in *Drosophila* and mouse cells was analyzed. As observed with HsLigPBD-GFP and MTPBD-GFP, DmLigPBD-GFP colocalized with RF in C2C12 cells but showed diffused nuclear distribution in S2 cells (Fig. 3.12A and B). Taken together, these results show that: (i) Dm DNA Lig I associates with RF in mouse cells via the DmDNA Lig I-PBD. (ii) Interestingly, Dm DNA Lig I does not associate with RF in *Drosophila* cells.

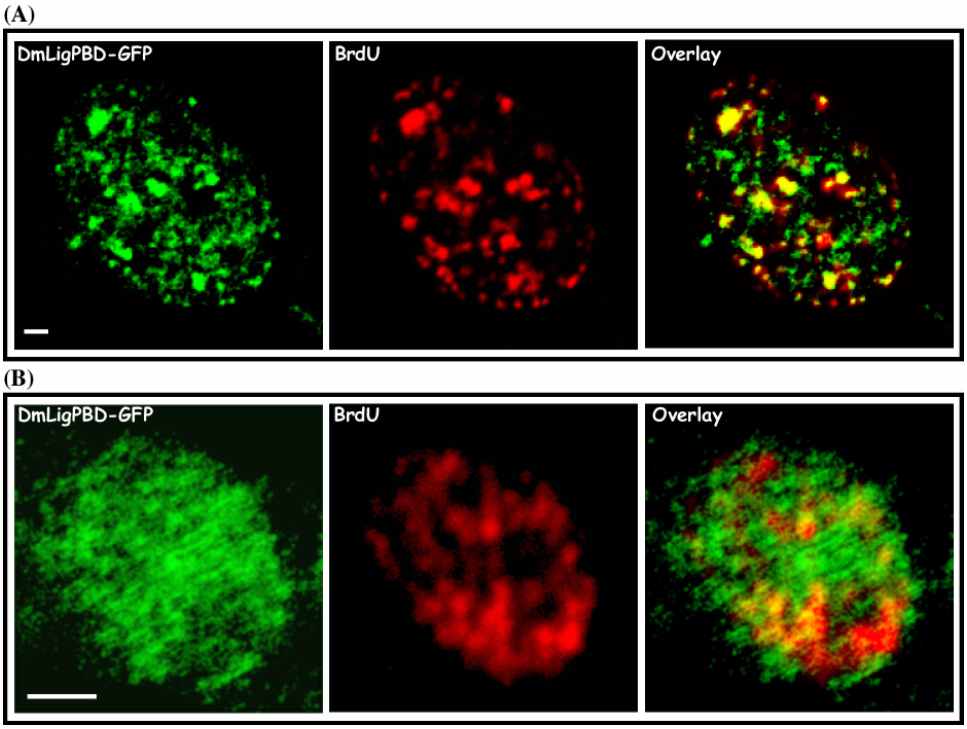


Fig. 3.12. DmLigPBD-GFP does not associate with RF in S2 cells. *Cells were transfected with a plasmid encoding DmLigPBD-GFP and pulse labeled with BrdU 24 hrs later followed by indirect immunostaining for BrdU.* ***(A)*** *Colocalization of DmLigPBD-GFP with BrdU in C2C12 cells.* ***(B)*** *DmLigPBD-GFP does not colocalize with BrdU in S2 cells. Scale Bar = 2 µm*

Table 3.1 summarizes the ability of the various fusion proteins to associate with RF in *Drosophila* and mammalian cells. Except for PCNA none of the other replication proteins associate with RF in *Drosophila* cells. Most notably DmDNA Lig I does not associate with RF in *Drosophila* cells while it does so in mouse cells. This probably indicates a fundamental difference in the kinetics of association of proteins with RF in *Drosophila* S2 cells and mammalian cells (see Discussion.

Table 3.1. Efficiency of association of various fusion proteins with replication foci in *Drosophila* and mammalian cells

Fusion protein	Percentage cells showing association with RF[a]	
	***Drosophila S2* cells**	***Mammalian C2C12 cells***
HsPCNA	80% (n = 25)	ND
GFP-HsPCNA	90% (n =20)	100% (n = 20)
GFP-DmPCNA	82% (n = 23)	100% (n = 15)
NMT^2-GFP	0% (n = 40)	75% (n = 20)
DNMT1	13%[b,] (n = 16)	100% (n = 15)
DNMT1 + HsPCNA	15%[b, c] (n = 27)	ND
GFP-HsDNA Lig I	0% (n = 38)	70% (n = 20)
GFP-HsDNA Lig I + HsPCNA	0%[c] (n = 19)	ND
MTPBD-GFP [d]	0% (n = 22)	40% (n = 32)
MTPBD-GFP + HsPCNA	0%[c] (n = 19)	ND
TS-ß-gal	0% (n =25)	ND
PBHD-YFP	0% (n = 25)	0% (n = 45)
GFP-DmDNA Lig I	0% (n = 40)	62% (n = 24)

[a]Number in parentheses indicates the number of S phase cells counted.
[b]Only partial colocalisation with BrdU foci was observed.
[c]Cotransfection efficiency was about 93%.
[d]HsLigPBD-GFP and DmLigPBD-GFP behave similar to MTPBD-GFP.

3.2. Regulation of subcellular localization of DNMT1 in mammalian cells throughout the cell cycle

3.2.1. Targeting preference of the three targeting domains of DNMT1

Previous studies with synchronized cells have shown that RF form ordered patterns at different stages during S phase (Nakamura et al., 1986) (Nakayasu and Berezney, 1989) (Fox et al., 1991) (O'Keefe et al., 1992). Formation of these patterns has been demonstrated in live cells to involve a gradual and asynchronous assembly and disassembly of the replication machinery (Leonhardt et al., 2000). The spatial pattern of RF indicates the temporal position during S phase (Leonhardt et al., 2000) and this provides a strategy to identify cells at different stages of S phase in an asynchronously growing population. Fig. 3.13 shows the typical patterns of RF in an asynchronous population of C2C12 cells diagnostic of the temporal position in S phase: RF in early-S are evenly distributed throughout the nucleus; in mid-S they are concentrated more around the nucleoli and nuclear periphery; and in late-S they form large toroidal structures at pericentric heterochromatin that is densely stained with Hoechst 33258 (Leonhardt et al., 1992).

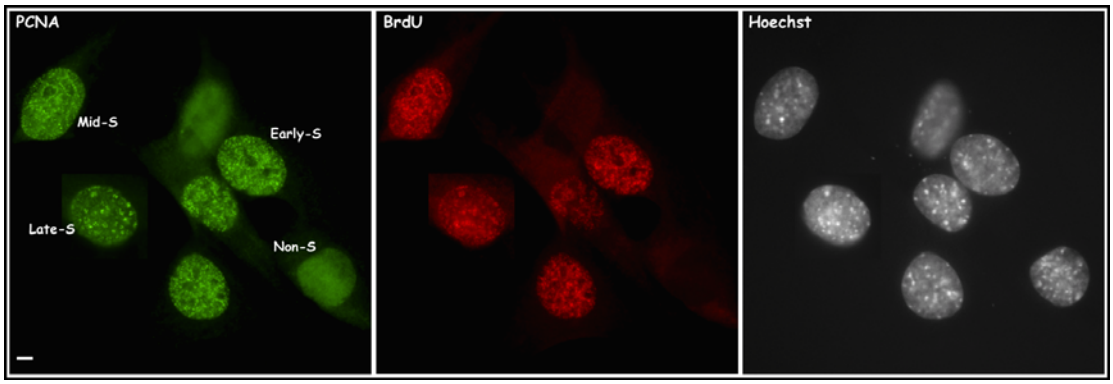


Fig. 3.13. Different stages during S phase can be identified in an asynchronously growing population of mammalian cells by the pattern of RF in the nucleus. *Mouse C2C12 cells were pulse labeled with 10μM BrdU for 10 min and fixed with ice cold methanol and immunostained with FITC conjugated anti-PCNA antibody (Pharmingen). The cells were fixed again with 3.7% formaldehyde and immunostained with anti-BrdU (Harlan Seralab) antibody followed by detection with Cy5 conjugated secondary antibody. Cells showing different patterns of RF typical of early-, mid- and late-S phases are shown. Non-S phase cells are not labeled with BrdU and show diffused distribution of PCNA. Scale Bar = 2 μm*

DNA replication is organized in such a way that the heterochromatin (densely methylated) replicates during late-S phase while euchromatin (sparsely methylated) replicates earlier (Goldman et al., 1984) (Hatton et al., 1988). Enhanced association of DNMT1 with late-RF might be required for efficient maintenance of the highly methylated DNA in the late replicating heterochromatic regions. In order to determine whether the multiple RFTS in DNMT1 have any selective preference to target to early or late replication foci, various deletions of the N-terminal domain of DNMT1 fused to GFP/YFP were made as summarized in Fig. 3.14A. Western blot analysis of COS7 cells transfected with the various plasmid constructs gave bands at expected sizes indicating that all the fusion proteins are correctly expressed (Fig. 3.14B).

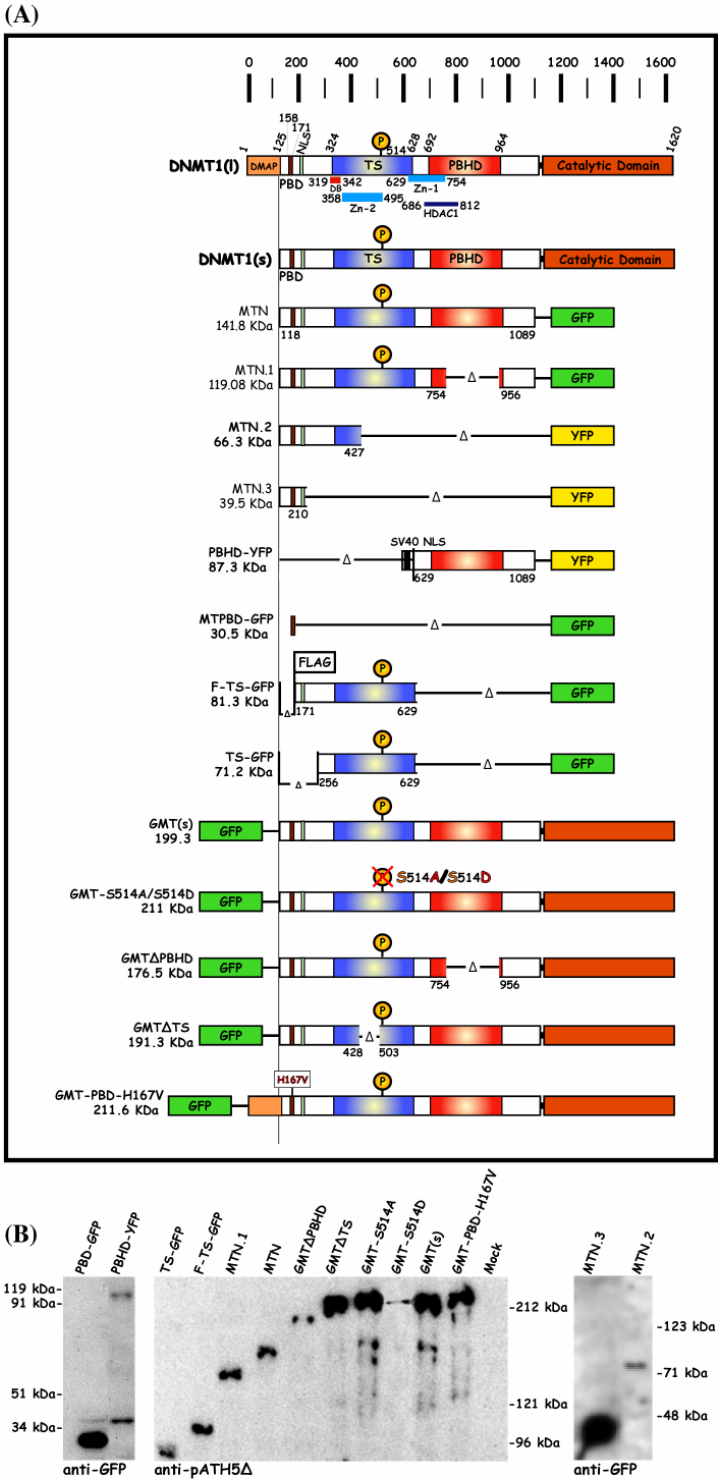

(A)
DNMT1(l)
DNMT1(s)
MTN
141.8 KDa
MTN.1
119.08 KDa
MTN.2
66.3 KDa
MTN.3
39.5 KDa
PBHD-YFP
87.3 KDa
MTPBD-GFP
30.5 KDa
F-TS-GFP
81.3 KDa
TS-GFP
71.2 KDa
GMT(s)
199.3
GMT-S514A/S514D
211 KDa
GMTΔPBHD
176.5 KDa
GMTΔTS
191.3 KDa
GMT-PBD-H167V
211.6 KDa
DMAP
PBD
TS
PBHD
Catalytic Domain
Zn-1
Zn-2
HDAC1
SV40 NLS
FLAG
GFP
YFP
S514A/S514D
H167V
(B)
PBD-GFP
PBHD-YFP
TS-GFP
F-TS-GFP
MTN.1
MTN
GMTΔPBHD
GMTΔTS
GMT-S514A
GMT-S514D
GMT(s)
GMT-PBD-H167V
Mock
MTN.3
MTN.2
119 kDa-
91 kDa-
51 kDa-
34 kDa-
-212 kDa
-121 kDa
-96 kDa
-123 kDa
-71 kDa
-48 kDa
anti-GFP
anti-pATH5Δ
anti-GFP

Fig. 3.14. Deletions of DNMT1 used for studying the preference of PBD, TS and PBHD to associate with RF in mammalian cells. (A) *The two isoforms of DNMT1 are shown at the top. DNMT1(l) is the larger somatic isoform and DNMT1(s) is the short isoform expressed in some cells, e.g. oocytes, that differ only in the first 118 amino acids encoding the DMAP binding region. Various deletions of DNMT1(s) were fused to GFP/YFP as listed. The predicted molecular weights of the fusion proteins are indicated. The boundaries of the different domains and deletions are indicated by the amino acid numbers corresponding to the DNMT(l) isoform. The DNMT1(s) isoform was used in making the deletions because the DMAP had no influence on targeting of the protein to RF. See Fig. 1.7 for description of the domains. An SV40 NLS was added in PBHD-YFP for efficient nuclear uptake. F-TS-GFP contains an N-terminal FLAG epitope tag.* ***(B)*** *Western blot analysis of the chimeric proteins showing they are correctly expressed. COS7 or 293T cells were transfected with the fusion constructs shown in (A) and after 24 hrs of incubation the cells were lysed by boiling in Laemmli's loading buffer and analyzed by immunoblotting with the mentioned antibodies.*

Firstly, we analyzed whether endogenous DNMT1 is present at RF throughout S phase. C2C12 cells were pulse labeled with BrdU for 10 minutes and immunostained with antibodies against DNMT1 (anti-PATH52 Ab) and BrdU. As shown in Fig. 3.15, endogenous DNMT1 colocalizes with BrdU labeled RF typical of the different stages of S phase. In late-S phase cells two populations of RF are observed - a population of large toroidal shaped foci at the pericentric heterochromatin and a population of smaller foci. The nature of the chromatin in the latter is not known and it could be either euchromatin or heterochromatin. A closer look shows that DNMT1 colocalizes with both populations of RF during late-S phase (marked with arrow and arrowhead). Moreover, DNMT1 colocalizes with RF in all S phase cells. Taken together, this indicates that DNMT1 associates with all sites of DNA replication throughout S phase.

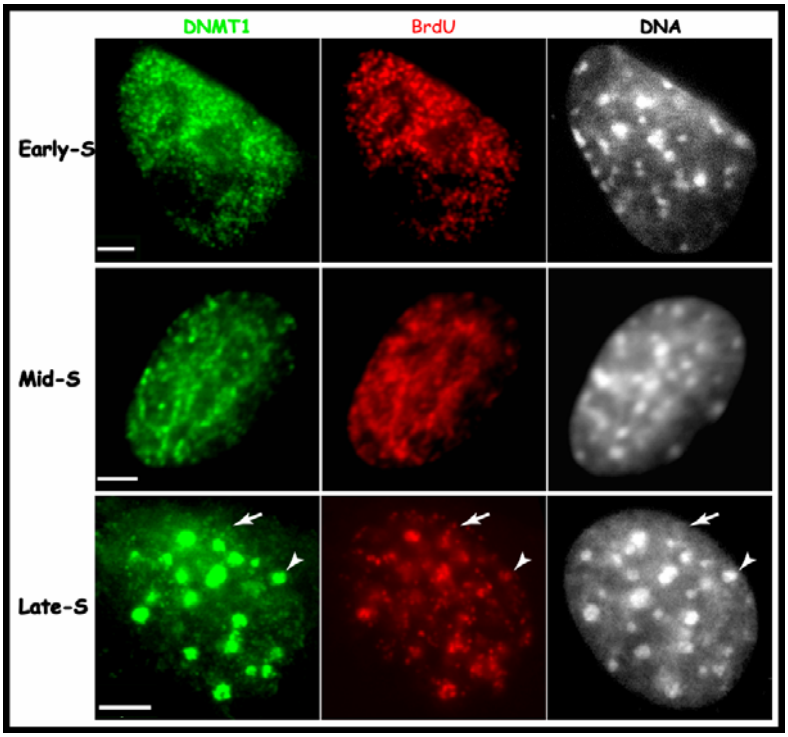


Fig. 3.15. Endogenous DNMT1 associates with RF throughout S phase. *C2C12 cells were pulse labeled with 100 µM BrdU for 10 min and fixed with 3.7% formaldehyde and immunostained for DNMT1 with anti-PATH52 antibody. In the late-S phase cell, arrowhead illustrates an exemplary late replicating centromeric heterochromatin which colocalizes with the Hoechst bright spots, and the arrow shows smaller foci which do not colocalize with Hoechst spots. Scale Bar = 2 µm.*

Next, the preference of the various N-terminal deletion proteins to associate with RF during early-, mid- and late-S phase was evaluated. The N-terminal domain fused to GFP (MTN) associated with RF throughout S phase (Fig. 3.16A). Deletion of the PBHD alone (MTN.1, Fig. 3.16B), or both TS and PBHD (MTN.2 and MTN.3, Fig. 3.16C and D) did not affect the ability of the fusion protein to associate with RF. This indicates that the PBD is sufficient for association with RF throughout S phase. Also, a minimal PBD-GFP construct carrying only the 10 amino acid PBD and the 10 basic residues following the PBD reported to be important for association with RF (Montecucco et al., 1998) efficiently associated with the RF throughout S phase (MTPBD-GFP, Fig. 3.16E). Taken together, these observations indicate that the deletion of PBHD and TS does not affect targeting and that the PBD is sufficient for targeting to RF during early, mid (data not shown) and late-S phase. Contrary to earlier reports (Liu et al., 1998), PBHD-YFP fusion protein did not associate with RF at any stage during S phase (Fig. 3.16F). This discrepancy could be because: (a) in their report replication sites were labeled by staining for endogenous DNMT1 and we observe that DNMT1 foci are not always associated with active replication (shown later); (b) their study shows that many DNMT1 foci are excluded of PBHD fusion protein. We observe that in few cells overexpressing PBHD-YFP, the fusion protein accumulates in some structures and these could partially colocalize with RF just as a matter of chance. From our results we conclude that PBHD has no ability to associate with RF. The F-TS-GFP and TS-GFP fusion constructs (Fig. 3.16G and H) are interesting in that they associate with RF only during late-S phase while in early S phase they do not colocalize with RF. Same results were observed with the betagal epitope tagged version which was initially used in mapping the TS in DNMT1

(Leonhardt et al., 1992) (Fig. 3.16I). Thus, we consistently observe a specific preference of the TS for targeting to RF during late-S phase. The results of the targeting preference are summarized in Fig. 3.17.

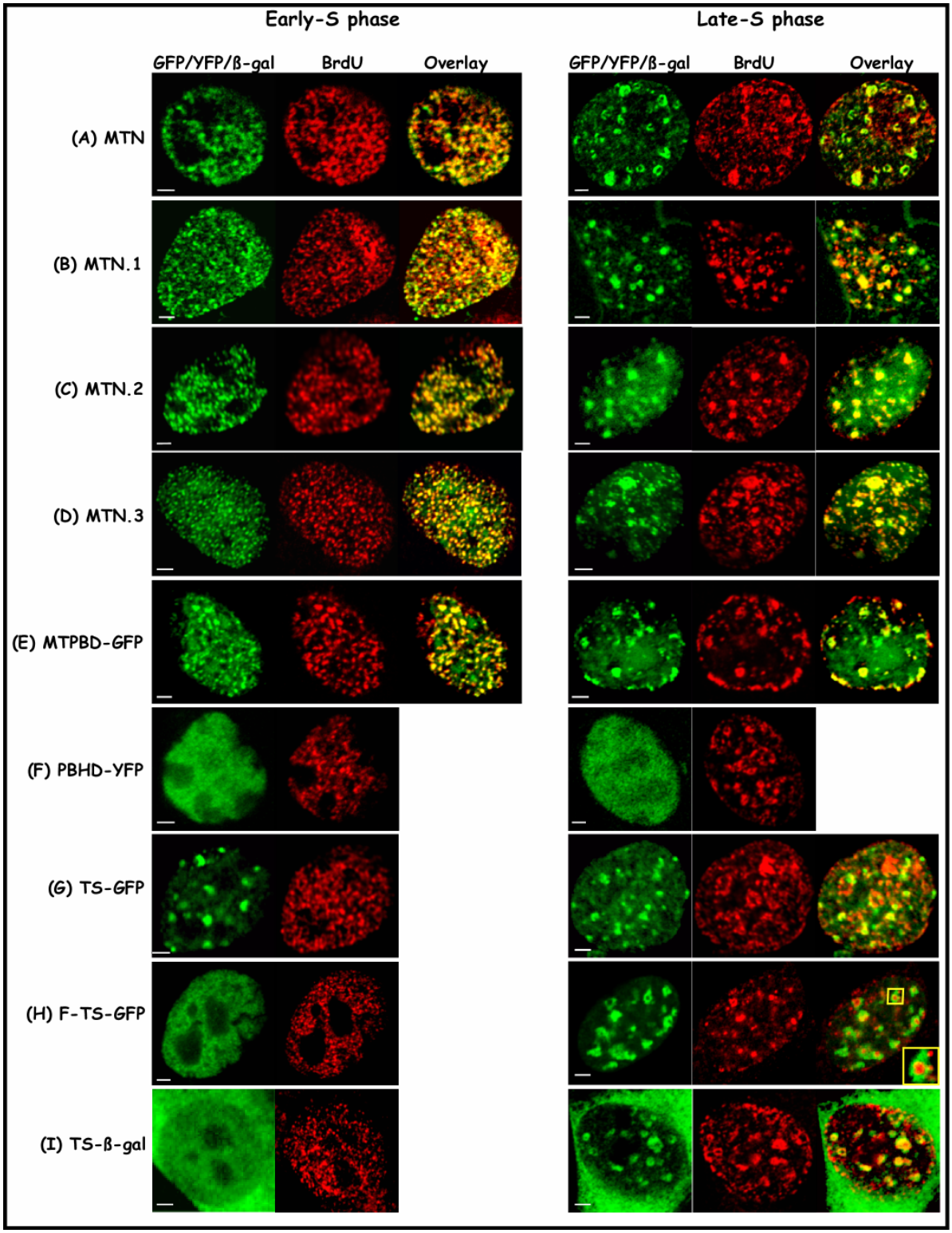


Fig. 3.16. Subnuclear localization of various deletions of DNMT1 in early- and late-S phase nuclei. *Mouse C2C12 cells were transfected with the indicated plasmid constructs and BrdU labeled 24 hrs later followed by indirect immunostaining for BrdU. Cells transfected with TS-ß-gal were immunostained with antibody against ß-gal. The TS associates with RF specifically during late-S phase. Overlay of green and red images are only shown in cases where there is colocalization of the fusion protein and BrdU foci. Scale bar = 2 µm.*

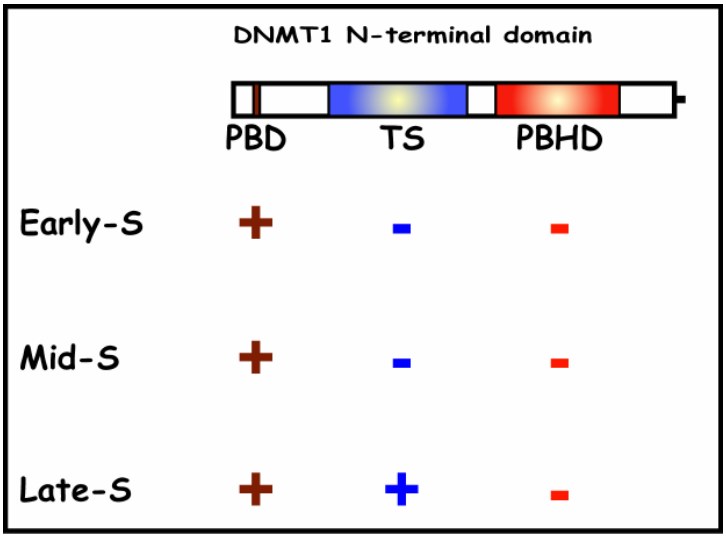


Fig. 3.17. TS associates with replication foci specifically during late-S phase*. Summary of the preference of PBD, TS and PBHD to associate with replication foci is illustrated. '+' indicates association; '-' indicates no association. PBD associates with RF throughout S phase while TS associates with RF only during late-S phase. PBHD does not associate with RF at any stage during S phase.*

3.2.2. TS associates with late replicating pericentric heterochromatin

A closer look at Fig. 3.16H (boxed region) brings out two interesting features of the association of F-TS-GFP with RF: (a) F-TS-GFP associates only with late-replicating pericentric heterochromatin which forms large toroidal foci and is excluded from the small RF; (b) even in the large toroidal shaped RF, TS-GFP is actually just around the site of DNA synthesis (BrdU incorporation site) and does not colocalize with the BrdU foci. This is in contrast to the late-S pattern in Fig. 3.16A-E wherein the GFP/YFP fusion proteins are also present at the small RF and completely colocalizes with the toroidal shaped large RF during late-S phase indicating that they associate with all sites of active DNA replication. The large RF are the late-replicating pericentric heterochromatin (Fig. 3.13 and 3.15). These observations suggest that TS *per se* associates with pericentric regions and not to sites of active DNA synthesis. Moreover, in Fig. 3.16G, it can be observed that during early-S phase, TS-GFP is present in some structures resembling late-S phase structures (such a pattern was also observed in some cells expressing F-TS-GFP). Probably this indicates that TS can associate with pericentric heterochromatin even during early and mid-S phase. This raises two important questions: (1) Is the association with pericentric heterochromatin confined only to the S phase or does it occur during other cell cycle stages? (2) When in S phase does the TS associate with pericentric heterochromatin? The latter question is of significance because during S phase core replication factors (like PCNA, DNA Ligase I) are in a dynamic state of assembly and disassembly which follows a strict spatio-temporal pattern (Leonhardt et al., 2000). The spatio-temporal pattern is a result of a yet undefined higher order chromatin arrangement organizing replicons in clusters (Ma et al., 1998) (Berezney et al., 1995) which fire at specific times during S phase. From the studies on the firing of replicon clusters and the dynamics of PCNA, one can derive that assembly of core replication factors at a site is concomitant with firing of DNA replication while disassembly is accompanied with termination of DNA replication at that site (Sporbert et al., 2002).

Hence recruitment to the RF should follow some temporal cue emanating from the RF. Whereas PCNA/DNA ligase I are recruited at all RF throughout S phase, some proteins like pol ε (Fuss and Linn, 2002), HDAC2 (Rountree et al., 2000) and WSTF-ISWI (Bozhenok et al., 2002) are recruited to only late-RF like the TS. Thus, recruitment of proteins to RF is a temporally controlled process and there must be specific temporal cues which discriminate the recruitment of different proteins to RF. By seeking to determine the timing of association of the TS to the late replicating pericentric heterochromatin, we were interested in finding out the temporal cue that determines this association with respect to the one which determines recruitment of PCNA/ DNA ligase I.

To answer these two questions, we monitored the dynamics of F-TS-GFP throughout the cell cycle, and its association with RF during S phase by live cell imaging. A cell cycle marker for live cell analysis was developed by fusing HsDNA Lig I to the DsRed red fluorescent protein (RFP-Ligase) allowing us to follow progression of a cell through different cell cycle stages as well as through the different stages of S-phase. By western blotting we determined that the RFP-Ligase fusion protein is correctly expressed (Fig. 3.18A). Comparison of the subnuclear localization pattern of RFP-Ligase and endogenous DNA Ligase I in C2C12 cells at different cell cycle stages shows that RFP-Ligase behaves like the endogenous DNA Ligase I (Fig. 3.18B). Thus RFP-Ligase can be used to mark actively replicating sites in live cells. To determine the dynamic localization of TS throughout the cell cycle, C2C12 cells were cotransfected with plasmids expressing F-TS-GFP and RFP-Ligase and 24 hrs after transfection the cells which were expressing low levels of the fusion proteins were imaged at about 1 hr intervals for a total of 21 hrs in a chamber maintained at 37ºC. Fig. 3.19 shows the main images from this time series (for complete time lapse, see video 1). At the time when imaging of this cell was started (0 hr), the cell was in early to mid-S phase as can be deciphered from the pattern of RFP-Ligase. At this time, F-TS-GFP was present throughout the nucleus and in some large structures (arrows). Notably, F-TS-GFP did not show any specific association with the early- S phase RF labeled by RFP-Ligase. After about an hour, the cell had entered mid-S phase (image not shown). At 3 hrs, the cell shows a typical mid-S phase pattern and F-TS-GFP still showed the same pattern as at the 0 hr time point. The boxed region shows a newly forming RF which partially colocalizes with the pre-existing F-TS-GFP structure (compare with the 0 hr time point). At about 3.5 hrs from the start, the cell has entered late-S phase where the typical late RF pattern can be observed. Complete colocalization of newly formed large RF with pre-existing F-TS-GFP structures is evident (see boxed region). More importantly, the small RFP-Ligase foci (see boxed region) do not contain F-TS-GFP which is similar to that observed in the fixed cell studies (Fig. 3.16h). At about 8.1 hrs, RFP-Ligase has completely dis-assembled from the RF and is diffused in the nucleus indicating that the cell has entered G2. F-TS-GFP has progressively accumulated at the same sites where it was already visible at 0 hrs. Since RFP-Ligase associated with these sites forming large foci only during late-S phase (3.5 hrs), it can be concluded that these sites correspond to late replicating pericentric heterochromatin. Interestingly, at 8.1 hrs, F-TS-GFP had concentrated more at the large foci and the diffused population had decreased. Probably this indicates that F-TS-GFP gradually accumulates at the pericentric heterochromatin. At about 15.5 hrs, when the cell is in mitosis, F-TS-GFP is retained at the chromatin and in the following G1 (21 hrs), F-TS-GFP is still associated with large structures resembling the pericentric heterochromatin. Thus, F-TS-GFP associates with pericentric heterochromatin throughout the S and G2 phase, and is

retained at these sites during mitosis and in the following G1. From these observations it can be concluded that: (1) TS has a strong affinity for pericentric heterochromatin. (2) TS does not function as a replication foci targeting sequence and instead the observation that it has a preference for late RF is due to its affinity for pericentric heterochromatin. (3) There is no temporal cue that determines the association of TS with pericentric heterochromatin and this is independent of the stage of S phase. (4) The association with pericentric heterochromatin is cell cycle independent.

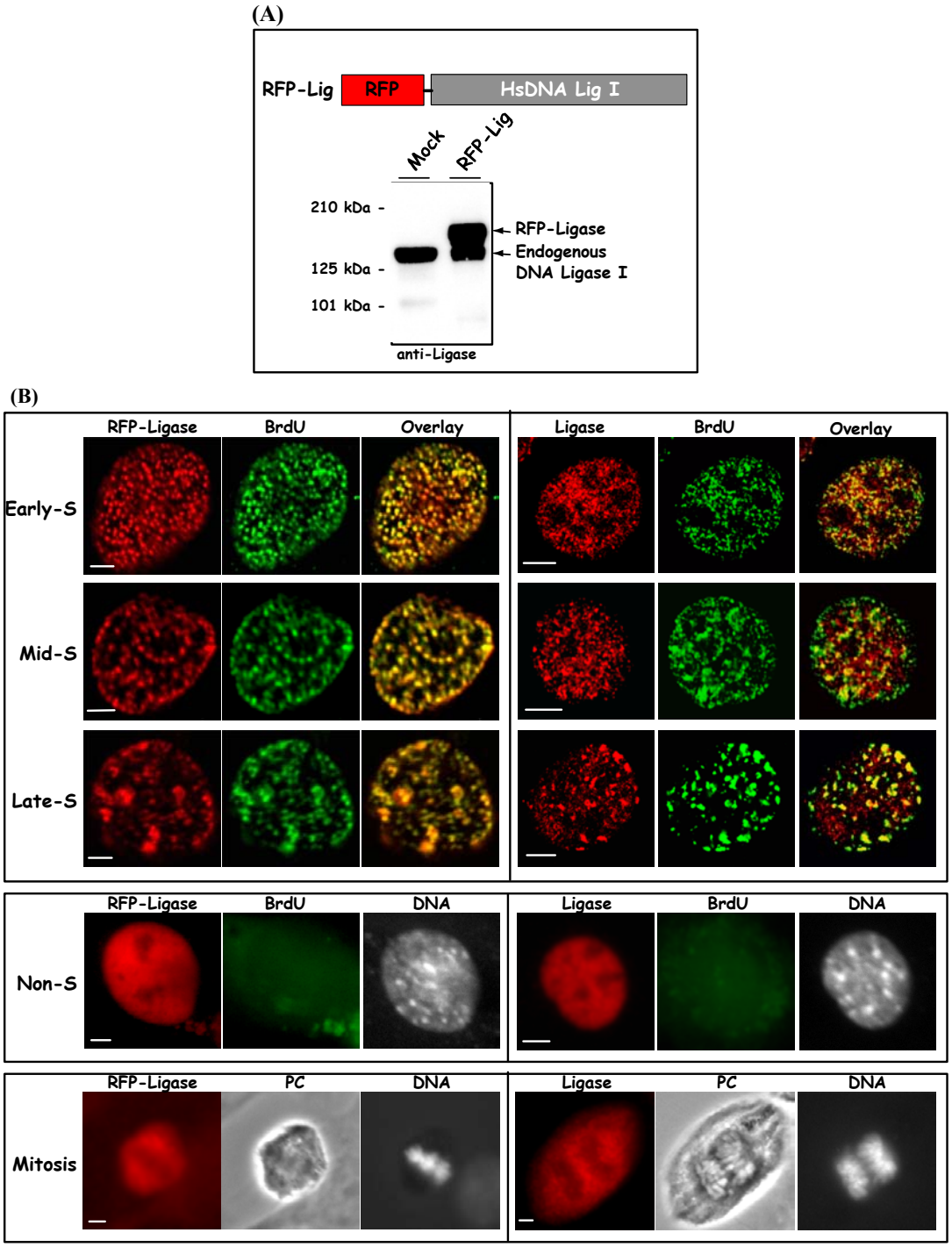


Fig. 3.18. Development of a cell cycle marker for live cell analysis. *Characterization of RFP-Ligase fusion protein expression and subnuclear localization.* ***(A)*** *COS7 cells were transfected with the RFP-Ligase fusion plasmid (shown at top) and after 24 hrs the cells were harvested and extracted by boiling in Laemmli's sample buffer and analyzed by immunoblotting with an antibody against DNA Ligase I (Cardoso et al., 1997). The fusion protein is correctly expressed. Mock represents transfection of cells in which the plasmid DNA is excluded. (B) RFP-ligase behaves like the endogenous DNA Ligase I throughout the cell cycle in mouse cells. On the left, C2C12 cells were transfected with RFP-Ligase fusion plasmid and pulse labeled with BrdU 24 hrs later followed by indirect immunostaining for BrdU. On the right, C2C12 cells were pulse labeled with BrdU and immunostained with antibodies against DNA Ligase I and BrdU.*

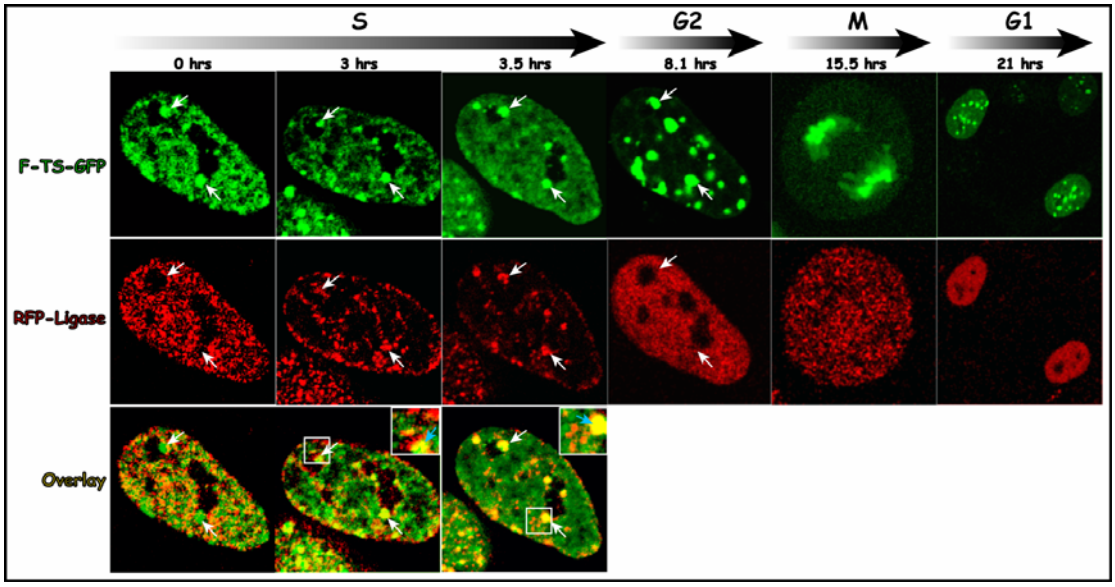


Fig. 3.19. Time lapse analysis of mouse cells expressing F-TS-GFP and RFP-Ligase. *C2C12 cells were cotransfected with plasmids encoding F-TS-GFP and RFP-Lig. After 24 hrs the cells were imaged at 1 hr intervals in a Zeiss LSM 510. Three Z sections of 1µm each were taken at each time point. Movement in the Z plane was corrected by comparing the Z sections at different time points and arranging them sequentially so as to include the same structures (RF and F-TS-GFP structures) in each sequential image. Only indicated time points are shown here. Overlay shows the merged image of green and red channels. White arrow indicates two exemplary F-TS-GFP structures that existed at 0 hrs and persist throughout the cell cycle. Box shows zoomed regions. Blue arrows in boxed region indicate colocalization of pre-existing F-TS-GFP structure with newly formed RFP-Lig foci. The full time lapse is in Video-1.*

3.2.3. Subnuclear localization of DNMT1 throughout the cell cycle

The strong affinity of TS for pericentric heterochromatin is striking because once associated with these regions, it is visualized there through following cell cycles. This raises the question whether this observation is an artifact or DNMT1 uses the TS to associate with pericentric heterochromatin at any stage during the cell cycle. In order to determine this, we established the localization of endogenous DNMT1 as well as DNMT1(s) fused to GFP (GMT) at G1, S, G2 and M. Fig. 3.20A describes the strategy we used for identifying cells in G1 and G2. To get cells in G1, loosely attached mitotic cells were harvested by mechanical shake off, concentrated by centrifuging and were then laid onto a coverslip. The cells were incubated at 37˚C for about 2 hrs and allowed to become adherent and enter G1. Subsequently the cells were fixed and immunostained. To identify cells in G2, C2C12 cells were pulse labeled with 10 µM BrdU for 10 min and chased for 2-3 hrs in medium containing 100µM thymidine and subsequently fixed and stained for BrdU, DNMT1 and PCNA. It is known that PCNA is localized at RF during S phase and is diffused in the nucleus of non-S phase cells. Cells showing positive BrdU staining and diffused PCNA are cells which were in the late-S phase during the BrdU pulse and have entered G2 during the 2-3 hrs chase. We rule out these cells to be in G1 because the G2 phase in C2C12 cells takes about 4-5 hrs and with our strategy of chasing for 2-3 hrs after BrdU labeling, the BrdU positive cells should be in G2. In G2 phase, DNMT1 is present in large structures reminiscent of the late-RF while PCNA is diffused (Fig. 3.20B). The DNMT1 foci colocalize with BrdU foci indicating that DNMT1 is retained at these late replicating sites even after the cell has exited S phase. This is

similar to the behaviour of the TS. In mitotic cells, DNMT1 is associated with the chromatin like the TS (Fig. 3.20C). G1 cells showed a diffused distribution of DNMT1 in the nucleus like PCNA (Fig. 3.20E). Thus, it seems that DNMT1 associates with pericentric heterochromatin during late-S phase and is retained there through G2 and M, and is released from these sites in the following G1. In early-G1 cells, DNMT1 is still associated with the pericentric heterochromatin (Fig. 3.20E). From these studies on fixed cells it is not clear whether DNMT1 is retained at late replicating pericentric heterochromatin in G2 or it is released at the end of S phase and again associates with the pericentric heterochromatin sometime in G2. To sort this out, C2C12 cells were cotransfected with plasmids expressing GFP-DNMT1(s) and RFP-Ligase and the dynamics of GFP-DNMT1(s) was monitored in live cells. The results are shown in Fig. 3.21, Video 2 and Video-3. At the time imaging of this cell was started (0 hrs) the cell was in mid-S phase when DNMT1 completely colocalizes with RFP-Ligase. At 3 hrs, the cell has entered late-S phase when both GFP-DNMT1(s) and RFP-Ligase are present in large structures. The boxed region shows that GFP-DNMT1(s) is present even at the small RF during late-S phase indicating that, like RFP-Ligase, GFP-DNMT1(s) associates with all RF (unlike the TS, see Fig. 3.19, 3.5 hrs). At 5 hrs, the cell is in the very late-S phase when some of the large RFP-Ligase foci present at 3 hrs have disassembled (arrow heads) while GFP-DNMT1(s) is retained at these sites. At 7 hrs, all the RFP-Ligase foci have disassembled and the cell has entered G2. At this point, GFP-DNMT1 is retained at the late-replicated regions (arrowheads). During mitosis (8.6 hrs), GFP-DNMT1 is associated with the chromatin and in the following G1 (9.6 hrs) it is diffused. These results match the observations on fixed cell studies and confirm that DNMT1 is retained at late-replicating pericentric heterochromatin through G2 and M and released from these sites only in the next G1.

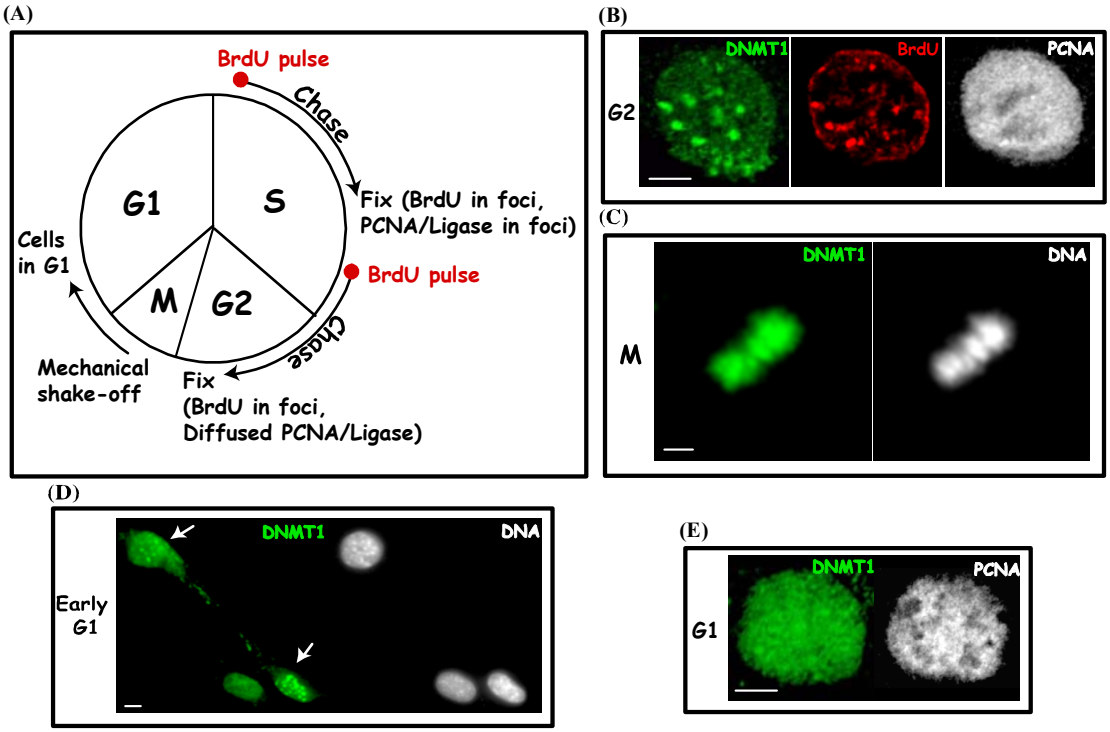


Fig. 3.20. Elucidation of the subnuclear localization of endogenous DNMT1 throughout the cell cycle. ***(A)*** *Strategy for identifying cells in G1, S and G2.* ***(B)*** *DNMT1 is retained at late replicating centromeric heterochromatin during G2. C2C12 cells were pulse-labeled with 10μM BrdU and chased for 2-3 hrs followed by fixation. Cells were immunostained with anti-BrdU, anti-DNMT1 and anti-PCNA antibodies. G2 cells were identified by their positive staining for BrdU and diffused PCNA.* ***(C)*** *DNMT1 is associated with the chromatin during mitosis. Mitotic cells were identified by their morphology and condensed chromosomes.* ***(D)*** *In early G1 cells, when the daughter cells are still connected (cells marked with arrow), DNMT1 is present at the centromeric heterochromatin.* ***(E)*** *DNMT1 shows a diffused subnuclear localization in G1 cells. G1 cells were obtained by seeding mitotic cells collected by mechanical shake off onto cover slips and allowing them to adhere for 2 hrs. The association of DNMT1 with RF during S phase is shown in Fig 3.15. Scale Bar = 5 μm.*

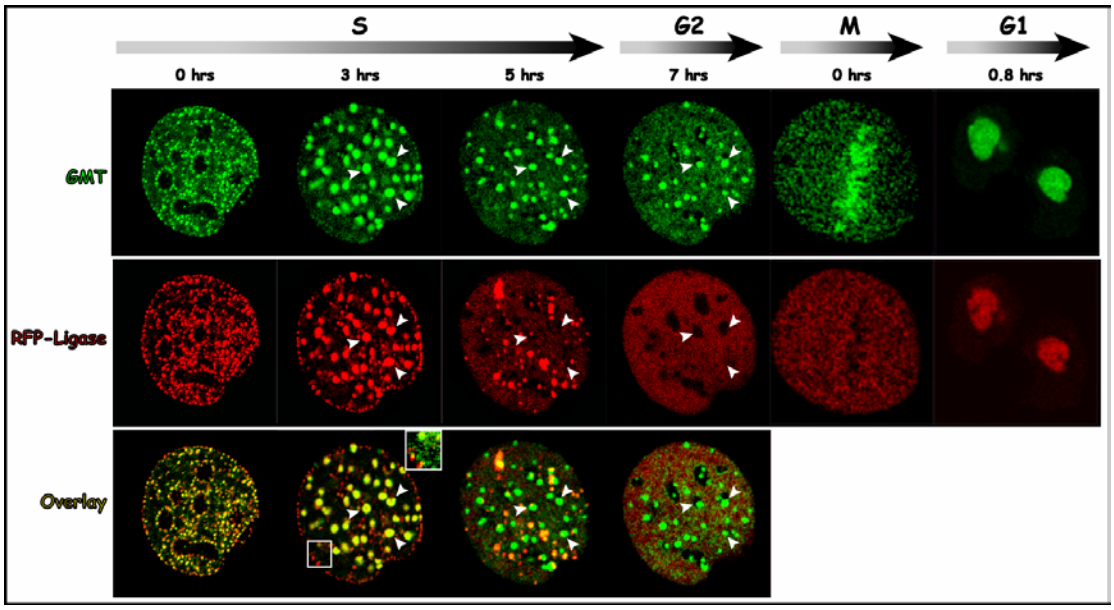


Fig. 3.21. Time lapse analysis of mouse cells expressing GMT and RFP-Ligase. *C2C12 cells were cotransfected with plasmids encoding GMT and RFP-Lig. After 24 hrs the cells were imaged at 1 hr intervals as mentioned in Fig. 3.19. Only indicated time points are shown here. Overlay shows the merged image of green and red channels. White arrowheads point to three exemplary GMT foci which assemble concomitantly with RFP-Lig during late-S phase and are retained through the rest of S phase and in G2. Box in the overlay at 3 hrs shows zoomed region where GMT also associates with smaller foci. Green levels are increased to show this. The images of the M phase to G1 phase are from another cell, hence it starts again from 0 hrs. During M, GMT is associated with the chromatin. The daughter cells in G1 show diffused nuclear distribution of GMT. Full time lapse in Video-2.*

To test whether retention of DNMT1 at late replicating pericentric heterochromatin in G2 and its association to mitotic chromatin during M is dependent on the TS, a region in the TS was deleted. This deletion (GMTΔTS, Fig. 3.14A) associated with RF throughout S phase (Fig. 3.22) but was not retained at pericentric heterochromatin in G2 and rather showed a diffused distribution (Fig. 3.22). GMTΔTS was not bound to chromatin during mitosis (Fig. 3.22) and in G1 it was again diffused in the nucleus (Fig. 3.22). These results strongly indicate that retention and binding to pericentric heterochromatin during G2 and M is dependent on the TS. Binding of DNMT1 to pericentric regions during their replication might also involve the TS in addition to the PBD, however this cannot be conclusively discerned here. The fact that the TS alone can bind to pericentric regions independent of the cell cycle stage indicates that this interaction could occur also during S phase in the context of the full length DNMT1.

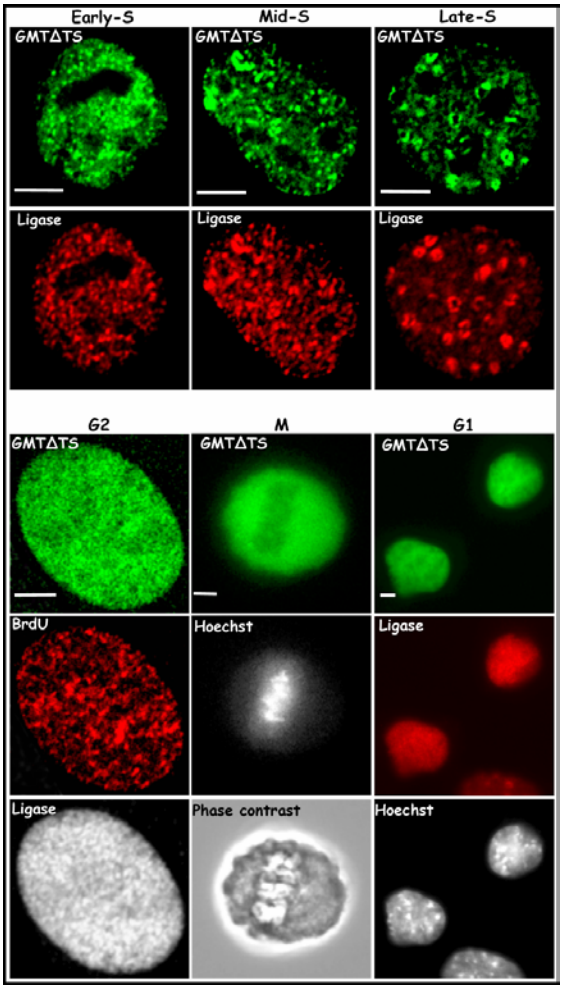


Fig. 3.22. TS is required for retention of DNMT1 at pericentric regions during G2 and M phases. *C2C12 cells were transfected with plasmid encoding GMTΔTS. G1 and G2 cells were identified as described in Fig. 3.20. After 24 hrs of transfection, cells were immunostained with antibody against DNA Ligase I to label RF. GMTΔTS colocalizes with sites of DNA replication throughout S phase. During G2, GMTΔTS shows diffused nuclear distribution, and is excluded from mitotic chromatin. G1 cells show diffused nuclear distribution of GMTΔTS. Scale bar = 5 μm*

3.2.4. Comparison of dynamics of TS with DNMT1: what determines release of DNMT1 from pericentric heterochromatin during G1?

Comparison of the localization of TS-GFP (Fig. 3.19) and GFP-DNMT1 (Fig. 3.21) at different cell cycle stages shows that TS associates with pericentric heterochromatin during all cell cycle stages while GFP-DNMT1 is present at these sites only during G2 and M. Two features of GFP-DNMT1 are striking: (1) During S phase, GFP-DNMT1 is targeted to sites of replication and its association with pericentric heterochromatin is delayed until late-S phase when these regions start replicating whereas TS can bind to these sites anytime in S phase. (2) During G1, GFP-DNMT1 becomes diffused while TS stays bound to the pericentric heterochromatin. These observations indicate that: (a) some feature(s) of DNMT1 prevents the association of TS with pericentric heterochromatin until late-S phase and (b) mediates its release in G1 thereby making it diffused.

In order to determine the region responsible for this control, we analyzed the

localization of a series of deletion and point mutants of DNMT1 fused to GFP (Fig. 3.14) throughout the cell cycle in fixed or live cells. Previous studies have identified phosphorylation of serine at position 514, which is within the TS, and to which no function has yet been attributed (Glickman et al., 1997). We tested whether phosphorylation of this site controls association of DNMT1 with pericentric heterochromatin by mutating it to a phosphorylation-off (S514A) and phosphorylation-on (S514D) state. All mutants except MTN.1 and GMT-PBD-H167V behaved like DNMT1 in that they associated with RF throughout S phase and assembled at the pericentric heterochromatin only during late S phase, were retained at the pericentric regions during G2 and M phases, and were diffused in the nucleus during G1 (Table 3.2). Deletion of PBHD in MTN.1 did not affect association of the fusion protein to RF, and its retention at pericentric heterochromatin in G2 and M (Fig. 3.23A and B). But during G1, MTN.1 was not released from the pericentric heterochromatin indicating that PBHD might be important for the release of DNMT1 from pericentric heterochromatin. Full length GMT carrying a deletion of the PBHD, GMTΔPBHD, was not efficiently localized to the nucleus and so its subnuclear distribution could not be evaluated. Previous work has shown that there is an NLS in this region (Cardoso and Leonhardt, 1999).

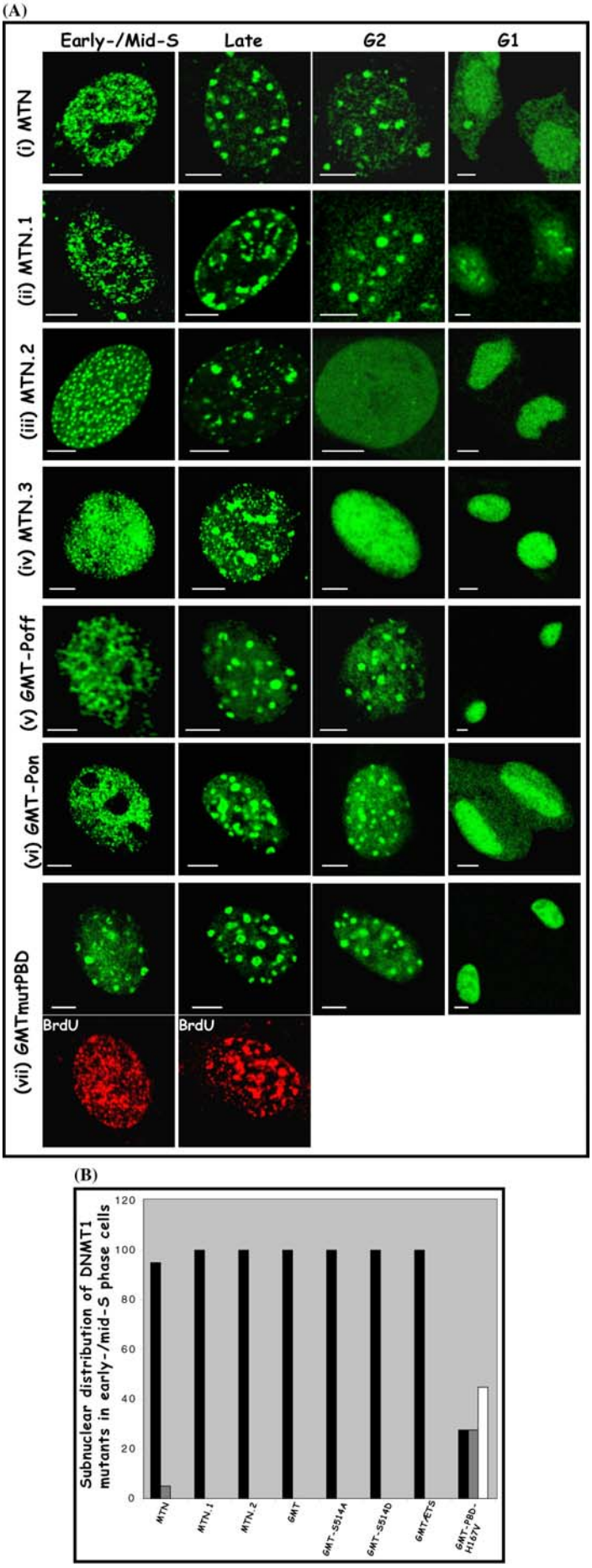

(A)
Early-/Mid-S
Late
G2
G1
(i) MTN
(ii) MTN.1
(iii) MTN.2
(iv) MTN.3
(v) GMT-Poff
(vi) GMT-Pon
(vii) GMTmutPBD
BrdU
BrdU
(B)
Subnuclear distribution of DNMT1 mutants in early-/mid-S phase cells
120
100
80
60
40
20
0
MTN
MTN.1
MTN.2
GMT
GMT-S514A
GMT-S514D
GMT/ETS
GMT-PBD-H15Y

Fig. 3.23. PBD delays association of TS in DNMT1 with pericentric heterochromatin until late-S phase. (A) *C2C12 cells were transfected with plasmids encoding the various mutants of DNMT1 fused to GFP. 24 hrs after transfection the subnuclear localization of the GFP-fusion proteins at S, G2, M and G1 phases were analyzed as described in Fig. 3.20. MTN.1 (ii), which lacks the PBHD, is retained in pericentric heterochromatin during G1. GMT-PBD-H167V (vii) is associated with pericentric heterochromatin right during early- and mid-S phase like the TS-GFP fusion protein. BrdU staining is shown only in the case of GMT-PBD-H167V. Scale Bar = 5 µm.* ***(B)*** *C2C12 cells transfected with various plasmids were pulse labeled with BrdU 24 hrs later and incorporation of BrdU was detected by immunostaining. The percentage of transfected cells in early-/mid-S phase exhibiting colocalization of the GFP-fusion protein with either BrdU labeled RF or pericentric heterochromatin, or showing diffused nuclear distribution was scored. Black is colocalization with RF, gray is colocalization with pericentric heterochromatin, white is diffused nuclear distribution. Only GMT-PBDmut associates with pericentric heterochromatin during early-/mid-S phase.*

Table 3.2: Summary of association (+) or no association (-) of various proteins with RF during different stages of S phase or with pericentric heterochromatin during different stages of the cell cycle. 'Late-S' and 'Throughout' mean association with pericentric heterochromatin occurs only during late-S phase or throughout S phase, respectively.

Fusion protein	**Association with RF during:**			**Association with pericentric heterochromatin during:**			
	Early-S	**Mid-S**	**Late-S**	**G2**	**M**	**G1**	**S**
DNMT1 (Endogenous)	+	+	+	+	+	-	Late-S
MTN	+	+	+	+	+	-	Late-S
MTN.1	+	+	+	+	+	+	Late-S
MTN.2	+	+	+	-	-	-	Late-S
MTN.3	+	+	+	-	-	-	Late-S
TS-GFP	-	-	-	+	+	+	Throughout
F-TS-GFP	-	-	-	+	+	+	Throughout
GMT	+	+	+	+	+	-	Late-S
GMT-S514A	+	+	+	+	+	-	Late-S
GMT-S514D	+	+	+	+	+	-	Late-S
GMTΔTS	+	+	+	-	-	-	Late-S
GMT-PBD-H167V	-/+	-/+	-/+	+	+	-	Throughout

Mutating the PBD (GMT-PBD-H167V) to a form that cannot bind to PCNA (Chuang et al., 1997) reduced targeting to RF accompanied by association with pericentric heterochromatin right in early-S reminiscent of F-TS-GFP. Like DNMT1, GMT-PBD-H167V was retained at pericentric heterochromatin during G2 and mitosis and showed diffused nuclear distribution in G1. Thus, PBD drives the orchestrated assembly of DNMT1 to RF throughout S phase by delaying TS-mediated association of DNMT1 with the pericentric heterochromatin until late-S phase when these regions replicate. It can be envisaged that the PBD brings DNMT1 to the pericentric heterochromatin during late-S which would then allow the TS to bind to some component of the pericentric heterochromatin and get retained here until the end of mitosis. These results indicate that PBD prevents binding of the TS in DNMT1 with pericentric heterochromatin until late-S phase while PBHD influences its release during G1.

Since in transfected cells there is a high variability in expression of the fusion proteins within the same population of cells, the observed association of GMT-PBD-H167V with pericentric heterochromatin during early-/mid-S phase cells could be a result of excess of the fusion protein available after saturation at the RF. To correct for any mis-representation arising from singular observations, C2C12 cells expressing moderate levels of the transfected plasmids encoding the various deletions, which were in early-/mid-S phase were randomly selected and the percentage cells exhibiting colocalization of the deletion protein with RF or pericentric heterochromatin was evaluated. Only GMT-PBD-H167V was associated with pericentric heterochromatin in a significant number (27.6%) of early-/mid-S phase cells (Fig. 3.23D). Notably, GMT was targeted to RF in all cells and in none of the early-/mid-S phase cells it associated with pericentric heterochromatin. In few cells GMT-PBD-H167V was targeted to RF indicating that the H to V mutation in PBD did not completely abolish binding to PCNA, and therefore association with RF. The N-terminal deletion proteins were associated with the RF in all early-/mid-S phase cells scored. Thus the association of GMT-PBD-H167V with pericentric heterochromatin during early-/mid-S phase cells is not due to overexpression and rather is due to a lack of ability to target to RF.

3.3. Biological effects of TS overexpression

The results in the previous section brings out a novel role for the TS in the retention of DNMT1at pericentric heterochromatin in G2 and M. This leads us to the question as to what role DNMT1 has at pericentric regions during G2 and M. We approached this question by asking what is the effect of expression of TS-GFP on cells, arguing that excess TS would prevent association of endogenous DNMT1 with pericentric heterochromatin. Since dense methylation of pericentric regions is important for chromatin stability and cell viability (reviewed in Robertson and Wolffe, 2000), the effect of expression of TS-GFP on cell growth and nuclear morphology was monitored.

3.3.1. Effect on cell viability

From the observations during live cell microscopy, it seems that short-term expression of F-TS-GFP/TS-GFP did not affect cell viability because the cells traversed through the different cell cycle stages and divided normally. To determine the long term effect of expressing TS on cell viability, C2C12 cells were transfected with the different DNMT1 deletion constructs (which have a neomycin resistance marker) and the cells were grown in the presence of neomycin (G418) to select for stably transfected cells. If the fusion protein expressed from the plasmid has a negative influence on cell growth, these cells would not survive and would be lost (Fig. 3.24A). Transfected cells would multiply to form colonies if the protein expressed from the plasmid is not toxic to the cell. Thus, transfected cells would be positively selected for by *neo*^{r} and negatively selected by how detrimental the expression of the recombinant transgene would be. The cells were cultured in the presence of G418 until all the cells in the mock transfection (no *neo*^{r} encoding plasmid) were dead. Subsequently, the cells were fixed with methanol, stained with Giemsa stain and colonies were counted. The result shown in Fig. 3.24B is normalized against the GFP control transfection. As can be seen, there is more than 40% reduction in colony formation when cells are expressing the TS fusion protein indicating that the TS has a deleterious effect on cell growth in the long term. A similar reduction was observed in the case of MTPBD-GFP. This would be expected due to the crucial role of PCNA in various cellular processes and binding of MTPBD-GFP would prevent the authentic partners of PCNA from associating with it thereby hampering cell growth. Similar results for the PBD have been reported previously (Mattock et al., 2001). The observation that expression of GMT does not affect cell viability indicates that the TS affects cell growth only when expressed/presented outside the context of the full length protein. This emphasizes the previous observations on the role of other domains in DNMT1, viz the PBD and PBHD, in controlling the activity of the TS. Also, it indicates that fusion of GFP does not have any affect on the normal activity of these proteins. Although we see a negative effect of the TS on cell viability, this assay does not yet pinpoint the process which is affected.

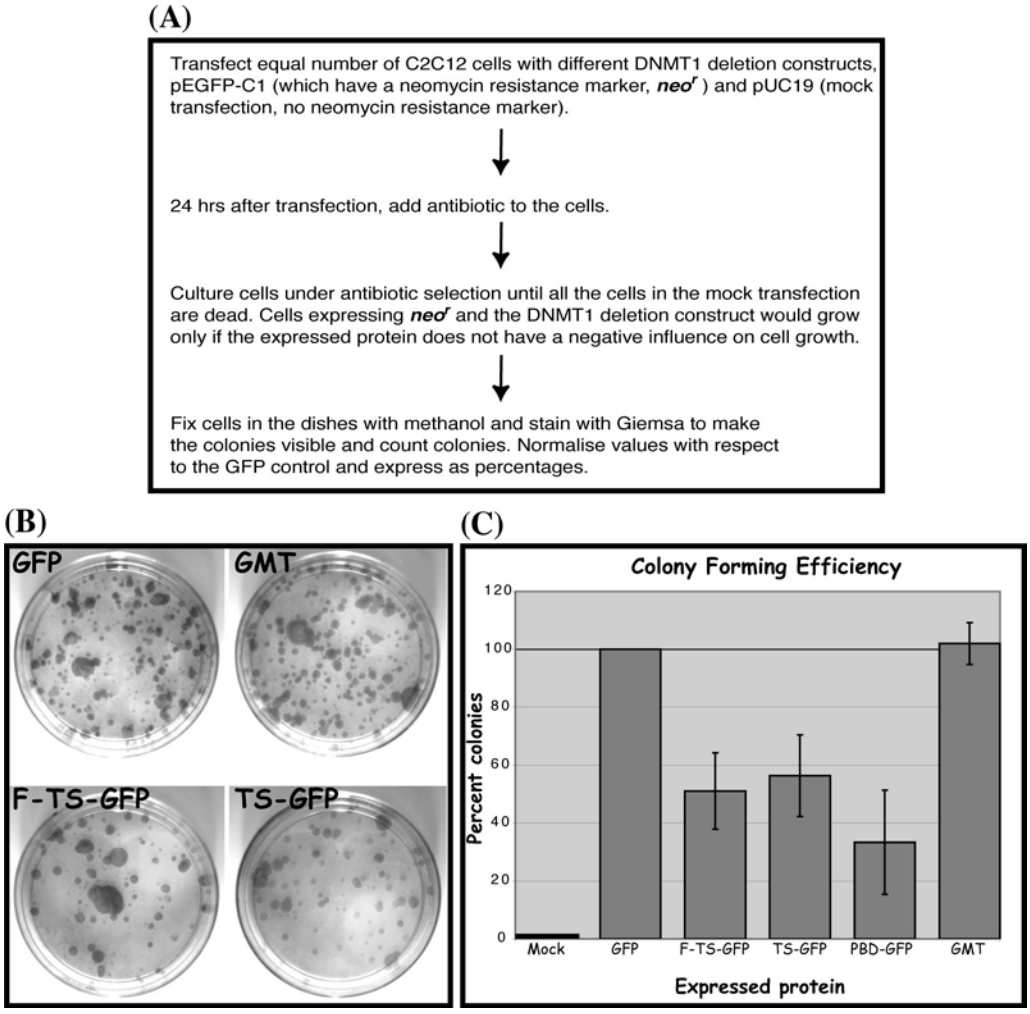


Fig. 3.24. Expression of TS reduces viability of cells. (A) *Methodology for colony forming assay.* ***(B)*** *Images of Giemsa stained colonies.* ***(C)*** *Colony forming efficiency of C2C12 cells transfected with plasmids encoding the mentioned proteins. The number of colonies in each experiment was normalized to the GFP control of the corresponding experiment and expressed as percentage. Results from three independent experiments were averaged; error bars show standard deviation.*

3.3.2. Effect on nuclear organization and morphology

To get an insight into the effect of TS in normal chromatin organization, fixed cells expressing the F-TS-GFP/TS-GFP fusion proteins and stained for DNA with Hoechst 33258/TOPRO-3 were screened under the microscope for unusual nuclear structures. Such an analysis brought forth two interesting defects in these cells:

Micronuclei Formation:

Cells expressing F-TS-GFP/TS-GFP and MTPBD-GFP showed almost 2-3 fold increase in micronuclei formation compared to cells expressing GFP or GMT (Fig. 3.25A). Since micronuclei formation reflects chromatin instability (Bonassi and Au, 2002) (Tucker and Preston, 1996), expression of TS-GFP seems to interfere with the integrity of chromatin. The affinity of TS for pericentric heterochromatin indicates that it might influence the maintenance of pericentric heterochromatin. Since many DNA synthesis and DNA repair factors have a PBD, it is quite likely that

overexpression of MTPBD-GFP interferes with these basic processes leading to micronuclei formation.

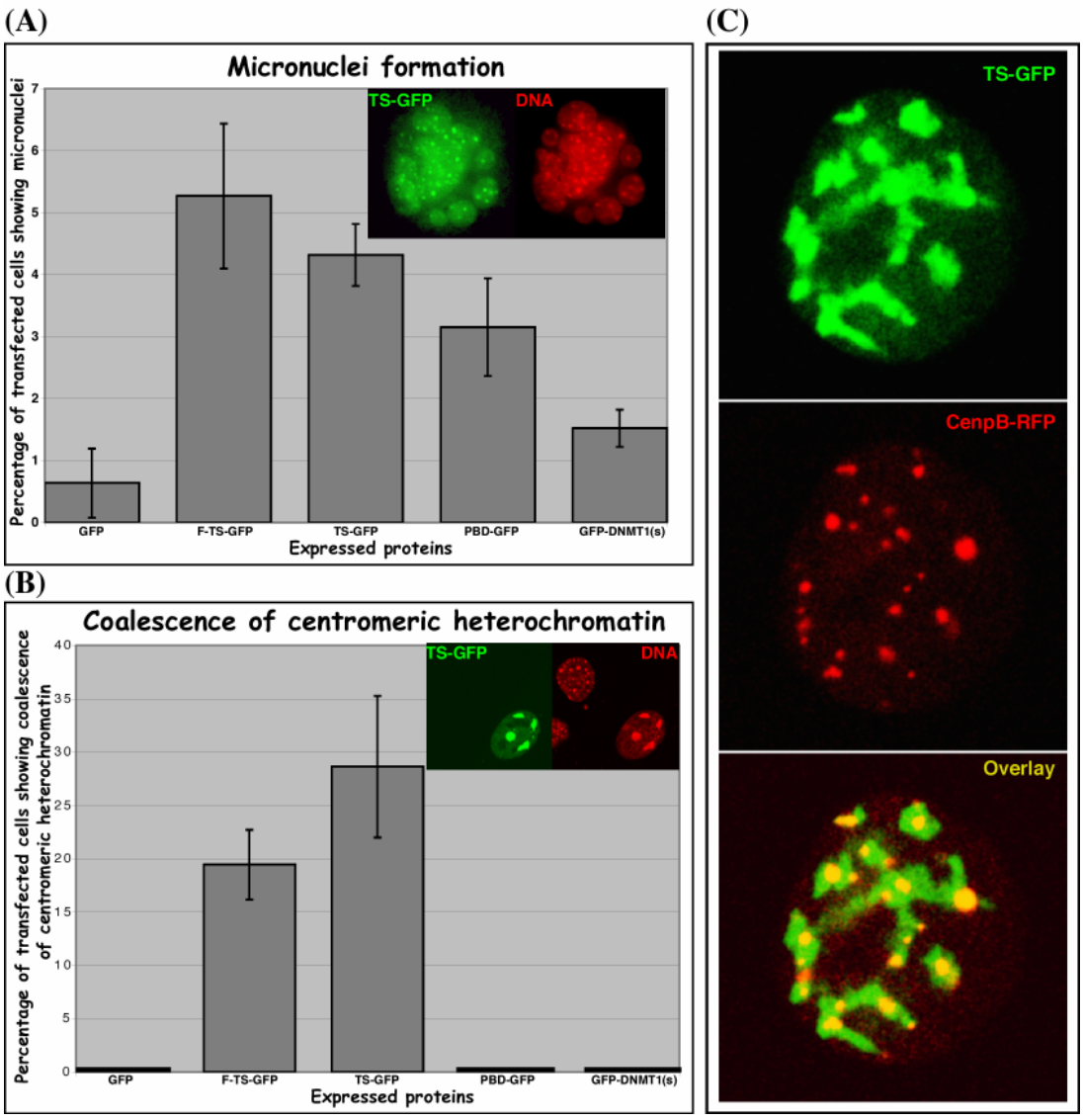


Fig. 3.25. Micronuclei formation and coalescence of centromeric heterochromatin. *Transfected cells expressing the mentioned proteins were fixed and stained for DNA with Hoechst 33258/TOPRO-3 and cells exhibiting micronuclei or coalescence of centromeric heterochromatin were scored. About 200-300 transfected cells were counted in each case and the percent values were calculated. The values obtained from three different transfections were averaged and plotted. Bars indicate standard deviation.* ***(A)*** *Percent transfected cells exhibiting micronuclei formation. Inset shows an image of a cell transfected with plasmid expressing TS-GFP exhibiting several micronuclei.* ***(B)*** *Percent transfected cells in which the centromeric heterochromatin is coalesced. Inset shows an image of a cell transfected with plasmid expressing TS-GFP exhibiting coalesced centromeric heterochromatin and a non-transfected cell showing normal centromeric heterochromatin.* ***(C)*** *C2C12 cells were cotransfected with plasmids encoding TS-GFP and CENPB-RFP to label the centromeres. Coalesced heterochromatin consists of relocated and clustered centromeric regions.*

Coalescence of Centromeric Heterochromatin:

Some cells overexpressing the F-TS-GFP and TS-GFP exhibit coalescence of centromeric heterochromatin into large chromocenters. Fig. 3.25B shows the percentage of transfected cells showing coalescent centromeric heterochromatin. In untransfected cells, centromeric heterochromatin is visible as discrete structures that are densely stained with DNA dyes like Hoechst 33258 or TOPRO-3. In cells overexpressing TS-GFP, the centromeric regions coalesce together to form large structures. Cells expressing GMT or MTPBD-GFP never show such structures (Fig. 3.25B). To test whether centromeres had relocated in these coalesced heterochromatin, plasmids encoding TS-GFP and CENP-B-RFP (which labels centromeres) were co-transfected in C2C12 cells. As shown in Fig. 3.25C, these coalesced heterochromatin centers have CENP-B-RFP indicating that indeed these regions are formed by fusion of the centromeric heterochromatin.

These results suggest a role of the TS in the large scale organization of centromeric heterochromatin.

3.4. Evolutionary conservation of the regulatory domain of DNMT1

C-5-methyltransferases are classified into five families based on phylogenetic analysis of their catalytic domains (Colot and Rossignol, 1999). One of this is the DNMT1 subfamily constituted by mammalian DNMT1 and counterparts from other animals, plants and fungi. DNMT1 is attributed a maintenance of methylation function primarily due to its strong preference for hemimethylated DNA (Bestor, 1992) (Yoder et al., 1997a) and its redistribution to replication foci during S phase (Leonhardt et al., 1992). However, these studies have been done only with mammalian DNMT1 and it is not known whether the DNMT1 in other organisms has a similar role. Our studies on the ability of the PBD, TS and PBHD of mouse DNMT1 to associate with RF in *Drosophila* and mammalian cells have shown that only the PBD is capable of mediating association of DNMT1 with RF. Recruitment of several proteins involved in DNA replication/repair to RF seems to be mediated by interaction with PCNA via the PBD suggesting that this is a general mechanism for targeting proteins to RF. In order to check whether the other members of the DNMT1 family can potentially be targeted to replication structures and thereby function as maintenance MTase, we performed sequence analysis on the DNMT1 family and also all known C-5-methyltransferases in eukaryotes to detect the presence of the PBD. Moreover, the role of the TS as a pericentric heterochromatin binding domain is interesting and we sought to identify other proteins harbouring a TS. The results from these sequence analyses are described below.

3.4.1. PBD is present only in metazoan DNMT1 family

Alignment of the PBD from various proteins (Fig. 3.2) shows that conservation of the PBD is confined to just 3-4 residues spanning a region of about 10 amino acid forming a consensus motif QxxI/LxxFF. This is the minimal sequence required for binding to PCNA (Zheleva et al., 2000). A region of about 10 amino acids rich in basic residues immediately following the consensus motif is required for efficient association with RF (Montecucco et al., 1998). To determine whether any of the C-5-methyltransferases from eukaryotes other than the reported mammalian DNMT1 have a PBD that could associate with RF, a profile of the PBD from the known PBDs (including the region rich in basic residues) from different proteins was made and used to search for significant matches in all the known C-5-methyltransferases. Fig. 3.26 shows the alignment of the sequences obtained from such a search. Significant hits for the PBD were obtained in all DNMT1 proteins from metazoan origin. No significant hits were observed in proteins from the metazoan DNMT2 and DNMT3 families. This is consistent with the observation that DNMT3 proteins are not associated with RF (Margot et al., 2001) and DNMT2 is an inactive protein which is not required for *de novo* or maintenance methylation in embryonic stem cells (Okano et al., 1998b). Some of the plant and fungal MTases show similarity to the consensus PBD motif but lack the adjacent region rich in basic residues, which is important for association with RF. Moreover, analysis of the location of PBD shows that the PBD is present at similar locations in homologous proteins. For example, the PBD in all DNA Ligase I homologues is present in the extreme N-terminus and in the case of metazoan DNMT1, it is located 1/5-th into the protein from the N-terminus. This is not the case for any of the PBD-like sequences

identified in the MTases in plants and fungus. The PBD motif being a small region of about 10 aa, many unrelated proteins contain sequences that match the consensus (Dalrymple et al., 2001). A search of the Swissprot database with the PBD profile returns many hits of proteins from bacteria to human (some of which are mitochondrial proteins) whose annotation does not attribute them any role in cell cycle or DNA metabolism indicating that these are false hits. The PBDs identified in the plant and fungal C-5-methyltransferases most probably are false hits considering their unusual locations in the protein and the absence of an adjacent region rich in basic residues. Thus, the PBD-like sequences identified in plants and fungus most likely do not function as authentic PBDs.

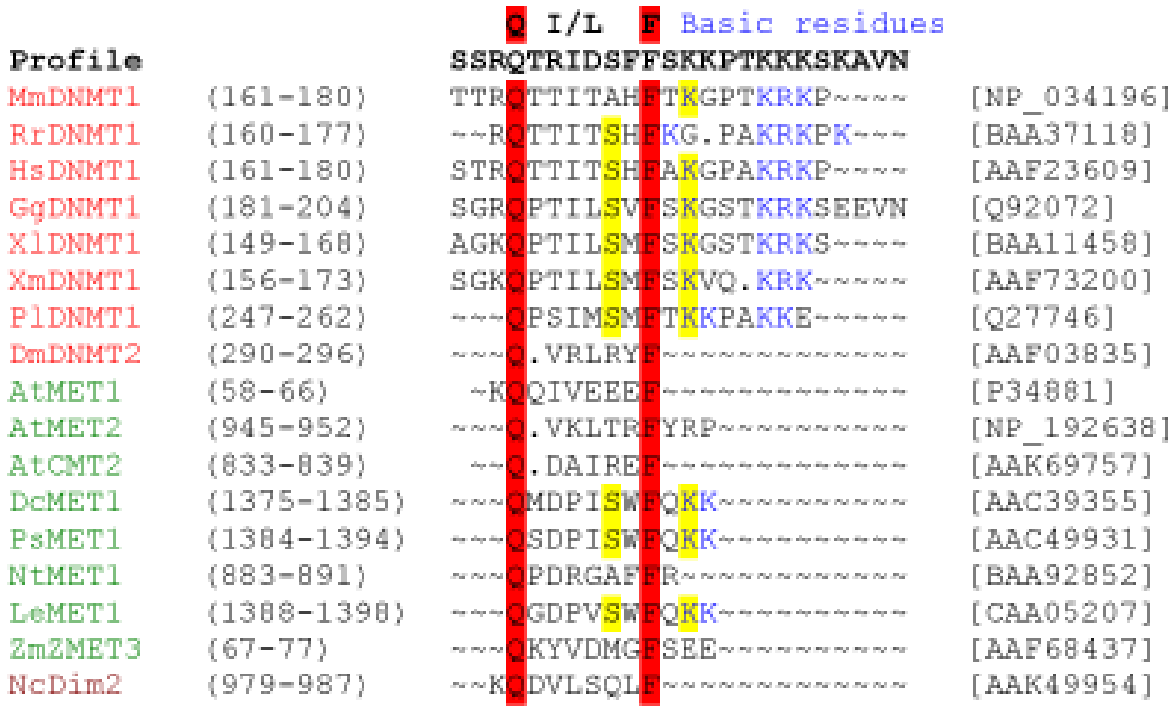


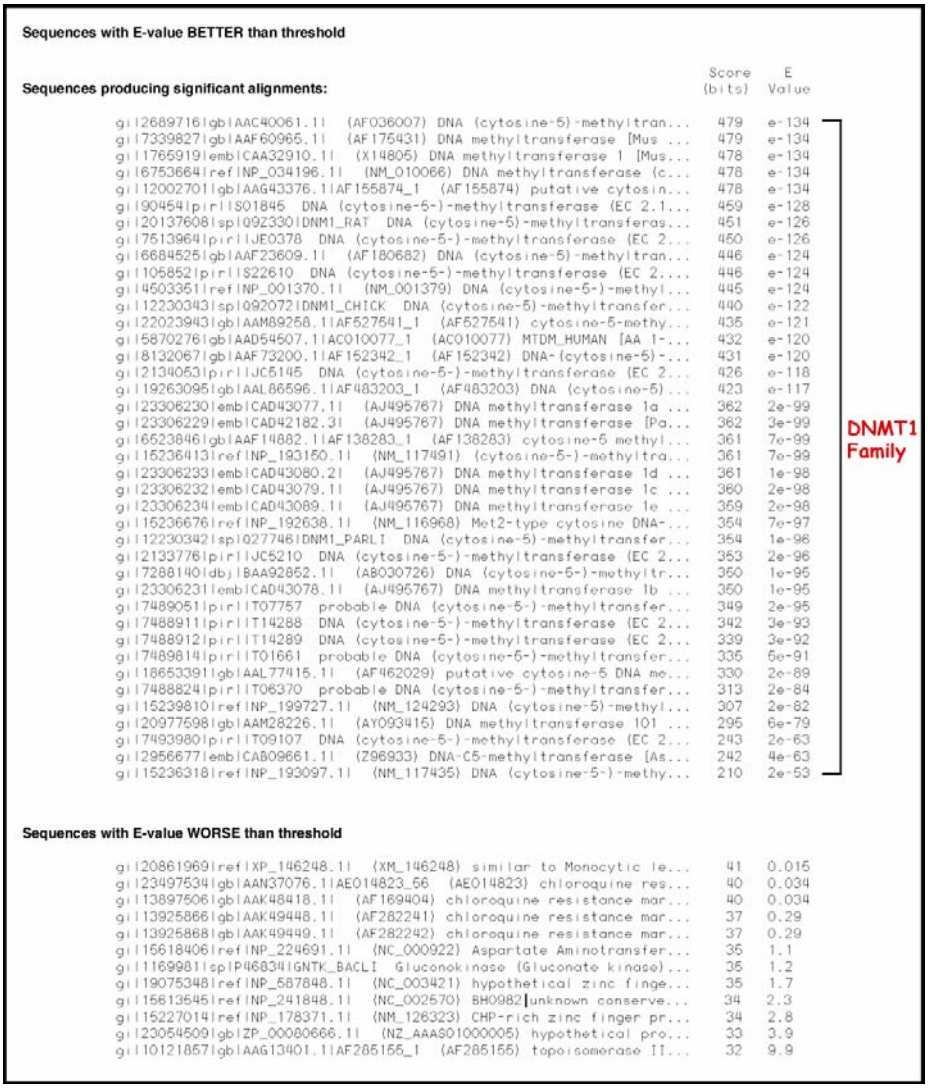
Fig. 3.26. Search for a PBD in all C-5-methyltransferases. *PBDs characterized in different proteins from metazoa were aligned by PileUp and a profile was generated using ProfileMake. This profile, shown at the top, was used to search all the C-5-methyltransferase families in metazoans, plants and fungi using the program ProfileSearch. The PBDs identified were aligned by Clustal and realigned by visual inspection to give maximal alignment. 100% identity is highlighted in red colour, more than 50% identity is highlighted in yellow, residues in blue form the region rich in basic residues. Sequence names in red are from metazoans, in green are from plants and brown is from fungus. Sequence positions are shown in numbers in parentheses. Square brackets at the right indicate protein accession numbers. The sequence of the most conserved residues in the PBD required for binding to PCNA and association with RF is shown at the top. Most of the PBDs identified in the plant and fungal methyltransferases are either in the C-terminal catalytic domain or in the PBHD or chromodomain and they lack the stretch of basic residues. Mm: Mus musculus; Rr: Rattus rattus; Hs: Homo sapiens; Gg: Gallus gallus; Xl: Xenopus laevis; Xm: Xiphophorus maculatus x helleri; Pl: Paracentrotus lividus; Dm: Drosophila melanogaster; At: Arabidopsis thaliana; Dc: Daucus carota; Ps: Pisum sativum; Nt: Nicotiana tabacum; Le: Lycopersicon esculentum; Zm: Zea mays; Nc: Neurospora crassa.*

These sequence analyses show that only metazoan C-5-methyltransferases of the DNMT1 family have a PBD while their plant and fungal counterparts lack a PBD and therefore might not associate with RF. Or, the association of the latter with RF is mediated independent of the PBD.

3.4.2. TS is a unique domain present only in DNMT1 family from animals, plant and fungi

The mouse DNMT1 TS mapped earlier (Leonhardt et al., 1992) was used as query to search the non-redundant (nr) database using PSI-BLAST (Altschul et al., 1997), which is a sensitive method to detect weak similarities in protein and nucleotide sequences (Callebaut et al., 1999). Following convergence after four iterations, the sequences with an E-value better than the threshold comprised only C5-methyltransferases from metazoans, plants and fungi belonging to the DNMT1 family (Fig. 3.27). None of the sequences with an E-value worse than threshold, except monocyte leukemia zinc finger protein (mouse), hypothetical zinc finger protein (*S. pombe*) and CHP rich zinc finger protein, seemed to have any link with chromatin function. The regions of these proteins showed very little similarity to mouse TS and their expected values are very high (> 0.015) to be homologous to the TS. The fact that sequences with a significant E-value all belonged to the C5-methyltransferases of the DNMT1 family indicates that the TS is unique to the DNMT1 family.

The results of the PSI-BLAST showed that in all the plant C-5-methyltransferases there are two regions in the N-terminus that show similarity to mouse TS, designated here as TS-1 and TS-2 (Fig. 3.28). Previous studies have reported two TS domains in MET1 from *Arabidopsis* (Colot and Rossignol, 1999). Such duplicated TS domains is not observed in fungal and metazoan DNMT1. Though both TS show some similarity to mammalian TS, there is very little similarity within them. TS-2 is detected only through iterative BLAST searches using PSI-BLAST. The expected values obtained from performing a BLAST2 comparison of the mouse TS and *Arabidopsis* TS-1, mouse TS and *Arabidopsis* TS-2, and *Arabidopsis* TS-1 and TS-2 are $2e^{-12}$, 0.93, and 3418, respectively. This high E-value between TS-1 and TS-2 is true for all the plant sequences indicating that these duplicated TS might have originated very early in evolution in a common ancestor and subsequently diverged by accumulating many mutations. However, closer inspection of the alignment of the two plant TS and the mouse TS obtained from PSI-BLAST shows that a region is conserved in both (Fig. 3.28, boxes). Analysis of the alignment of all the TS identified shows that this region forms a conserved motif (Fig. 3.29). In fungi, only *A. immersus* Masc2 has a TS which also has this conserved motif.



Sequences with E-value BETTER than threshold

Sequences producing significant alignments:

	Score (bits)	E Value
gi|26897161|gb|AAC40061.1| (AF036007) DNA (cytosine-5)-methyltran...	479	e-134
gi|7339827|gb|AAF60965.1| (AF175431) DNA methyltransferase [Mus ...	479	e-134
gi|1765919|emb|CAA32910.1| (X14805) DNA methyltransferase 1 [Mus...	478	e-134
gi|6753664|ref|NP_034196.1| (NM_010066) DNA methyltransferase (c...	478	e-134
gi|12002701|gb|AAG43376.1|AF155874_1 (AF155874) putative cytosin...	478	e-134
gi|904541|pir||S01845 DNA (cytosine-5-)-methyltransferase (EC 2.1....	459	e-128
gi|20137608|sp|Q9Z330|DNM1_RAT DNA (cytosine-5)-methyltransferas...	451	e-126
gi|7513964|pir||JE0378 DNA (cytosine-5-)-methyltransferase (EC 2...	450	e-126
gi|6684525|gb|AAF23609.1| (AF180682) DNA (cytosine-5)-methyltran...	446	e-124
gi|105852|pir||S22610 DNA (cytosine-5-)-methyltransferase (EC 2....	446	e-124
gi|4503361|ref|NP_001370.1| (NM_001379) DNA (cytosine-5-)-methyl...	445	e-124
gi|12230343|sp|Q92072|DNM1_CHICK DNA (cytosine-5)-methyltransfer...	440	e-122
gi|22023943|gb|AAM89258.1|AF527541_1 (AF527541) cytosine-5-methy...	435	e-121
gi|5870276|gb|AAD54507.1|AC010077_1 (AC010077) MTDM_HUMAN [AA 1-...	432	e-120
gi|8132067|gb|AAF73200.1|AF152342_1 (AF152342) DNA-(cytosine-5)-...	431	e-120
gi|2134053|pir||JC5145 DNA (cytosine-5-)-methyltransferase (EC 2....	426	e-118
gi|19263095|gb|AAL86596.1|AF483203_1 (AF483203) DNA (cytosine-5)...	423	e-117
gi|23306230|emb|CAD42077.1| (AJ495767) DNA methyltransferase 1a ...	362	2e-99
gi|23306229|emb|CAD42182.3| (AJ495767) DNA methyltransferase [Pa...	362	3e-99
gi|6523846|gb|AAF14882.1|AF138283_1 (AF138283) cytosine-5 methyl...	361	7e-99
gi|15236413|ref|NP_193150.1| (NM_117491) (cytosine-5-)-methyltra...	361	7e-99
gi|23306233|emb|CAD43080.2| (AJ495767) DNA methyltransferase 1d ...	361	1e-98
gi|23306232|emb|CAD43079.1| (AJ495767) DNA methyltransferase 1c ...	360	2e-98
gi|23306234|emb|CAD43089.1| (AJ495767) DNA methyltransferase 1e ...	359	2e-98
gi|15236676|ref|NP_192638.1| (NM_116968) Met2-type cytosine DNA-...	354	7e-97
gi|12230342|sp|Q27746|DNM1_PARLI DNA (cytosine-5)-methyltransfer...	354	1e-96
gi|2133776|pir||JC5210 DNA (cytosine-5-)-methyltransferase (EC 2....	353	2e-96
gi|7288140|dbj|BAA92852.1| (AB030726) DNA (cytosine-5-)-methyltr...	350	1e-95
gi|23306231|emb|CAD43078.1| (AJ495767) DNA methyltransferase 1b ...	350	1e-95
gi|7489051|pir||T07757 probable DNA (cytosine-5-)-methyltransfer...	349	2e-95
gi|7488911|pir||T14288 DNA (cytosine-5-)-methyltransferase (EC 2...	342	3e-93
gi|7488912|pir||T14289 DNA (cytosine-5-)-methyltransferase (EC 2...	339	3e-92
gi|7489814|pir||T01661 probable DNA (cytosine-5-)-methyltransfer...	335	5e-91
gi|18653391|gb|AAL77415.1| (AF462029) putative cytosine-5 DNA me...	330	2e-89
gi|7488824|pir||T06370 probable DNA (cytosine-5-)-methyltransfer...	313	2e-84
gi|15239810|ref|NP_199727.1| (NM_124293) DNA (cytosine-5)-methyl...	307	2e-82
gi|20977698|gb|AAM28226.1| (AY093415) DNA methyltransferase 101 ...	295	6e-79
gi|7493980|pir||T09107 DNA (cytosine-5-)-methyltransferase (EC 2...	243	2e-63
gi|2956677|emb|CAB09661.1| (Z96933) DNA-C5-methyltransferase [As...	242	4e-63
gi|15236318|ref|NP_193097.1| (NM_117435) DNA (cytosine-5-)-methy...	210	2e-53

DNMT1 Family

Sequences with E-value WORSE than threshold

gi|20861969|ref|XP_146248.1| (XM_146248) similar to Monocytic le...	41	0.015
gi|23497534|gb|AAN37076.1|AE014823_56 (AE014823) chloroquine res...	40	0.034
gi|13897506|gb|AAK48418.1| (AF169404) chloroquine resistance mar...	40	0.034
gi|13925866|gb|AAK49448.1| (AF282241) chloroquine resistance mar...	37	0.29
gi|13925868|gb|AAK49449.1| (AF282242) chloroquine resistance mar...	37	0.29
gi|15618406|ref|NP_224691.1| (NC_000922) Aspartate Aminotransfer...	35	1.1
gi|1169981|sp|P46834|GNTK_BACLI Gluconokinase (Gluconate kinase)...	35	1.2
gi|19075348|ref|NP_587848.1| (NC_003421) hypothetical zinc finge...	35	1.7
gi|15613645|ref|NP_241848.1| (NC_002570) BH0982 unknown conserve...	34	2.3
gi|15227014|ref|NP_178371.1| (NM_126323) CHP-rich zinc finger pr...	34	2.8
gi|23054509|gb|ZP_00080666.1| (NZ_AAAS01000005) hypothetical pro...	33	3.9
gi|10121857|gb|AAG13401.1|AF285155_1 (AF285155) topoisomerase II...	32	9.9

Fig. 3.27. TS is an unique domain present only in the DNMT1 family. *The nr database was searched using PSI-BLAST with the mouse TS as query. Subsequent iterations were performed by including the sequences (which were all C-5 methyltransferases) which showed an E value better than threshold. Following four iterations convergence was attained and no new sequences were retrieved. All the sequences showing significant E value belonged to the DNMT1 family.*

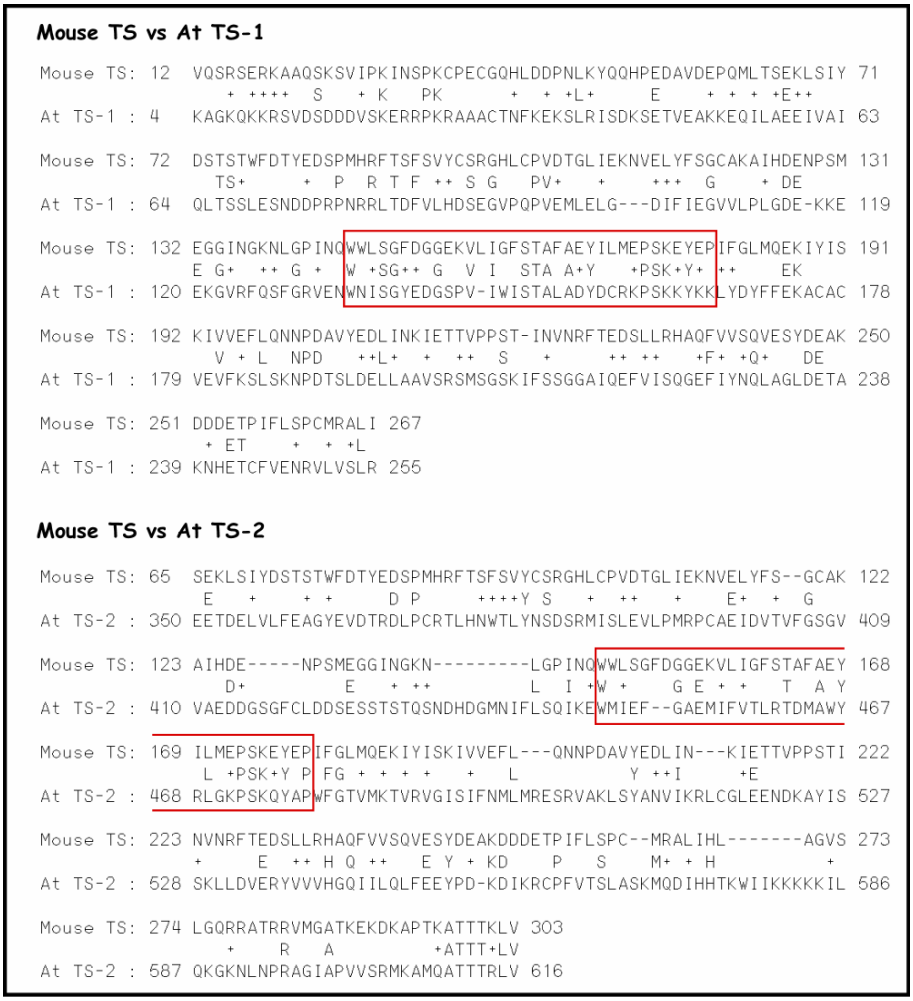


Fig. 3.28. TS domain is duplicated in plants and shows a conserved motif*. A typical example of the result of PSI-BLAST showing duplicated TS in plants, in this case A. thaliana* (accession no. AAF14882)*. Box shows the region that has a high degree of conservation in both TS.*

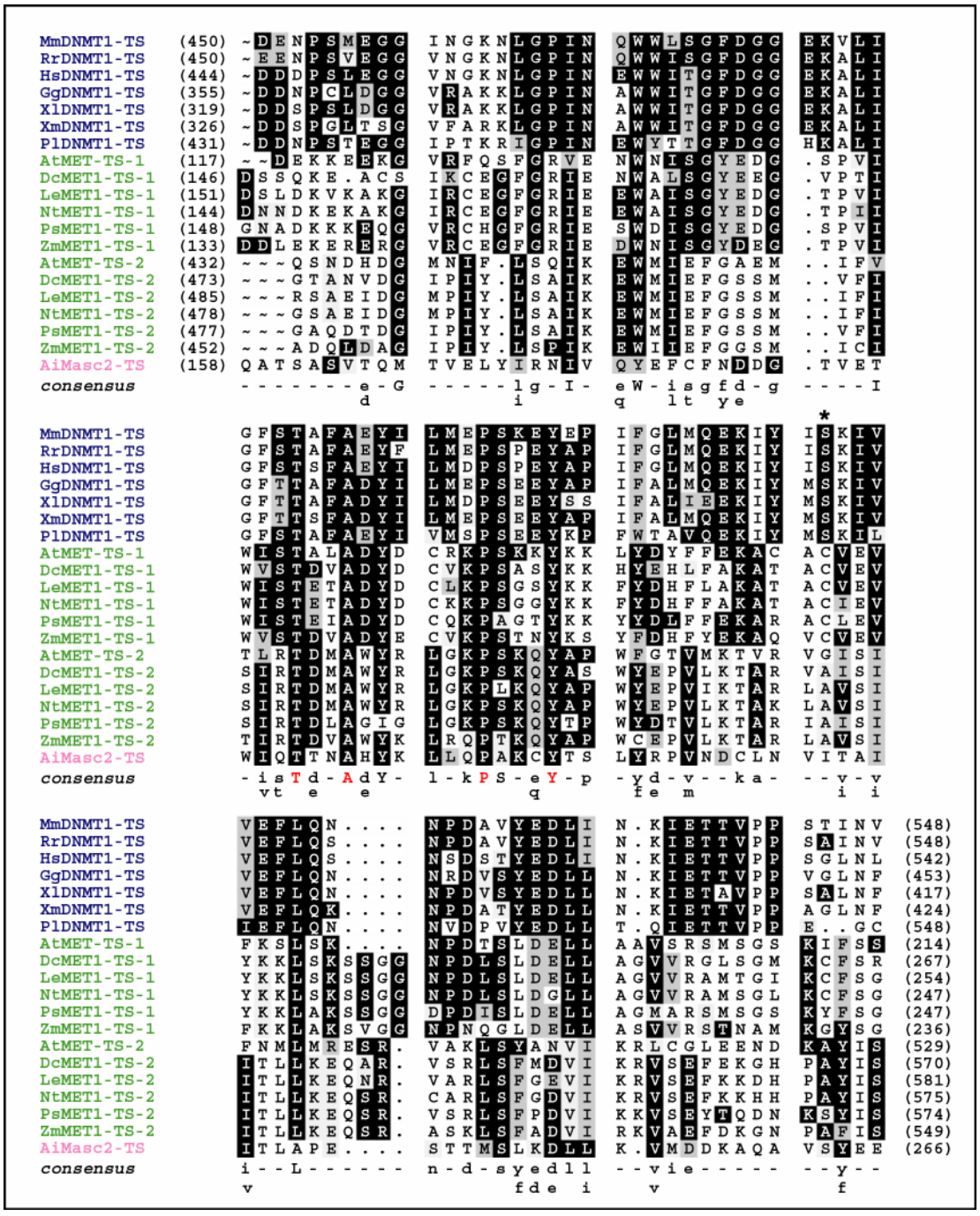


Fig. 3.29. A conserved motif in the TS from metazoans, plants and fungus. *TS from mouse (Leonhardt et al., 1992) was used as a query to search the nr database for similar sequences using PSI-BLAST. All significant hits obtained belonged to the DNMT1 family from metazoans, plants and fungus (see also Fig 3.27). These were aligned using Clustal. Here a region of the alignment with maximum conservation is shown. Residues that are identical to the consensus are shaded black, highly similar residues are shaded grey. Asterisk indicates the phosphorylation site (S at position 514) that is conserved in metazoans. In the consensus sequence, letters in red are 100% identical, uppercase is greater than 80% identity and lowercase is similar residues that are conserved at least between two of the three groups of TS: metazoan TS, TS-1 and TS-2. Conserved changes are shown below the consensus. In sequence names Ai is Ascobolus immersus. Other sequence names are as in Fig 3.26.* (AtMET is AAF14882).

In order to determine whether this conserved motif is present in other proteins, a profile of the TS from metazoans, plants and fungus was made and used to search for similar sequences in the Swissprot database. No significant hits were obtained outside the members of the DNMT1 family indicating that this domain is specific to the DNMT1 family. Many of the low scoring hits showed similarity to the TS motif WIS**T**DF**A**D**Y**DLMK**P**SKE**Y** (residues that are more than 90% conserved in the various TS from DNMT1 family are in bold) but further analysis of their annotation in the database or their homologues did not indicate any function related to chromatin. Two of them, telomerase reverse transcriptase catalytic subunit (mouse accession no. 070372) and RNA-directed RNA polymerase (yeast, accession no. P25328), are chromatin related proteins and the presence of a TS-like motif could be significant but the scores were very low.

Thus, these results indicate that the TS is a unique domain present only in the DNMT1 family. A conserved motif in the TS has been identified and mutational analysis of these residues should provide further insight into their roles.

4. Discussion

In mammalian cells replication of the genome takes place at distinct subnuclear sites called replication foci (RF). Various replication factors are localized to these sites via specific protein sequences called RFTS. Also DNMT1, an enzyme that maintains the DNA methylation pattern after DNA replication, localizes at RF. Association of DNMT1 with RF is reported to be mediated independently by three regulatory sequences. To understand the evolutionary conservation in the organization of these subnuclear structures and the mechanisms for the assembly of the replication machinery, we have analyzed the ability of mammalian replication factors and DNMT1 to associate with RF in *Drosophila* and mammalian cells. To dissect the function of the three regulatory sequences of DNMT1 and their role in controlling the subnuclear distribution of DNMT1 throughout the cell cycle, detailed analyses of various DNMT1 mutants to associate with RF in mammalian cells were carried out.

4.1. The mammalian DNMT1 does not associate with RF in Drosophila cells

The observation that immunostaining for PCNA and BrdU incorporation sites gives a punctate pattern indicates that in *Drosophila* S2 cells, like mammalian cells, DNA replication is organized in discrete structures or RF (Fig. 3.3). This is consistent with previous observation that replication in *Drosophila* Kc cells occurs at subnuclear foci (Ahmad and Henikoff, 2001). PCNA is highly conserved between human and *Drosophila* (Fig. 3.1), and HsPCNA and DmPCNA can efficiently associate with RF across the two organisms (Fig. 3.3). This indicates that PCNA, a conserved central core at the replication fork, is structurally and functionally exchangeable at the RF between mammals and *Drosophila*. However, in *Drosophila* cells neither the N-terminal domain of DNMT1 nor the three individual RFTS of DNMT1 showed any association with RF (Fig. 3.5B and Fig. 3.7). The full length DNMT1(s) showed partial association with RF in about 10% of *Drosophila* cells in S phase (Fig. 3.5D and Table 3.1). One possible explanation for this partial association is that DNMT1 associates with particular features of the genome, which in an unsynchronized cell population would yield a partial association with RF (see also Section 4.3 and Outlook for further discussion). It is surprising that DNMT1 having a functional PBD, which is reported in homologous proteins from a wide range of eucaryotes including *Drosophila* (Warbrick et al., 1998), is unable to associate with RF in *Drosophila* cells. Due to the general inability of PBD-GFP fusions from different origins (DNMT1, HsDNA Ligase I, DmDNA Ligase I) to associate with RF in *Drosophila* cells, we argued that in *Drosophila* another domain might function as an RFTS. To identify a domain that functions as an RFTS in *Drosophila*, a series of deletions of DmDNA Ligase I fused to GFP were analyzed for their ability to associate with RF in *Drosophila* cells. Surprisingly neither the full length DmDNA Ligase I nor any of the deletions associated with RF in *Drosophila* cells (Fig. 3.11A). However, DmDNA Ligase I associated with RF in mammalian cells, which is mediated by the PBD as deletion of the PBD abrogated this association (Fig. 3.11B). This also indicates that DmDNA Ligase I fused to GFP is expressed and folded properly and that the inability

to associate with RF in *Drosophila* cells cannot be attributed to improper expression or misfolding. The fact that no localization at *Drosophila* RF was observed does not rule out transient and short lived associations but clearly no distinguishable RF structures assemble as in mammalian cells. These results suggest a fundamental difference between *Drosophila* and mammalian cells. While replication in *Drosophila* cells seems to rely on transient interaction with highly mobile factors, mammalian cells seem to have evolved stable higher order structures possibly to cope with the challenges of replicating a genome 22 times as big as the fly genome.

4.2. Targeting of DNMT1 to RF in mammalian cells is driven solely by the PBD

Analysis of the association of various deletions of DNMT1 fused to GFP/YFP with RF showed that the PBD is sufficient for targeting of DNMT1 to RF (Fig. 3.16 A-E). Contrary to earlier reports (Liu et al, 1998), PBHD was unable to associate with RF in any of the cells in S phase (Fig. 3.16F). However, in their work they observe only „partial“ targeting of PBHD to RF in „some“ cells. We observe that the PBHD-YFP fusion forms some aggregates in the nucleus in overexpressing cells (judged by YFP fluorescence intensity) and as a matter of chance these aggregates could colocalize with RF. Moreover, PBHD (also called BAH domain) is present in many proteins involved in chromatin remodeling and gene silencing (Callebaut et al., 1999). The only other example of a protein involved in DNA replication that contains the PBHD/BAH is *S. cerivisiae* OrcI where the BAH domain is dispensable for DNA replication while it is required for transcriptional silencing (Bell et al., 1995). Thus similar to these proteins, PBHD/BAH domain in DNMT1 might have a role pertaining to chromatin alterations rather than having a role in targeting the protein to replication structures. The crystal structure solution of the BAH domain from *S. cerevisiae* Orc1 suggests it is primarily a protein interaction domain whose specificity is determined by a highly divergent region which juts out of a main scaffold formed by the conserved region (Zhang et al., 2002). It is not yet known what binds to the PBHD/BAH domain of DNMT1. However, it has been shown that HDAC1 interacts with DNMT1 and a part of this interaction domain lies within the conserved region of PBHD/BAH of DNMT1 (Fuks et al., 2000). Thus, based on these reports the PBHD domain in DNMT1 might be involved in some chromatin remodeling activities.

Here we observe that the TS associates only with late-RF (Fig. 3.16G, H), which is consistent with the previous observation that the TS associates mostly with late-RF (Leonhardt et al., 1992). Further analysis of the temporal course of TS association with late-RF showed that this association is attributable to the affinity of the TS for pericentric heterochromatin (Fig. 3.19). Thus, the TS does not have any inherent ability to associate with RF *per se*, and the observed association with late-RF is coincidental with its association with pericentric heterochromatin.

Taking together these observations on the ability of the three reported RFTS from DNMT1 to target to RF, it is clear that PBD is the only domain which mediates targeting of DNMT1 to RF. A point mutation in the PBD (H at position 167 was mutated to V) to a form that abrogates PCNA binding ability (Chuang et al., 1997) resulted in reduced association of the mutant GMT-PBD-H167V with replication sites (Fig. 3.23A and Table 3.2). About 30% of cells in S phase showed association of

GMT-PBD-H167V with RF. We attribute this to some weak binding of the mutated PBD with PCNA. It was suggested that the PBD from DNMT1 might play a role in targeting to RF during early-S while the other domains would target DNMT1 to RF during other S phase stages (Chuang et al., 1997). In opposition to this view, our results show that PBD solely mediates association of DNMT1 with RF throughout S phase.

4.3. The TS mediates cell cycle dependent binding of DNMT1 to pericentric heterochromatin

Previous studies have shown that DNMT1 forms a punctate pattern colocalizing with RF during S phase (Leonhardt et al., 1992), and is diffused in the nucleus in G2 (Rountree et al., 2000). In this study we observe that the TS domain alone showed a strong affinity for pericentric heterochromatin and remains associated with these sites in later cell cycle stages (Fig. 3.16 and Fig. 3.19). The subnuclear localization of DNMT1 and deletion mutants throughout the cell cycle was analyzed by two independent approaches: by immunostaining of endogenous DNMT1 in fixed cells at different cell cycle stages and by following the dynamics of a GFP-DNMT1(s) fusion protein (GMT) in live cells throughout the cell cycle. Both approaches unequivocally showed that from late-S through G2 and M until early-G1 DNMT1 is present at pericentric heterochromatin. It is diffused in the nucleus during G1 phase and associates again with the RF in S phase (Fig. 3.20 and Fig. 3.21). During M phase, endogenous DNMT1 (Fig. 3.20E) and GMT (Fig. 3.21) apparently are bound to the whole chromatin and is not confined only to the pericentric heterochromatin unlike F-TS-GFP, which is concentrated at pericentric heterochromatin (Fig. 3.19). This could be due to the other functional domains in DNMT1. Deletion of the TS from DNMT1 resulted in loss of association with pericentric heterochromatin during G2 and M phases (Fig. 3.22) indicating that the TS is solely responsible for retention of DNMT1 at these sites following S phase in G2 and M phases. Importantly, this deletion mutant efficiently associated with the RF at pericentric heterochromatin during late S phase, which should be mediated by the PBD. Thus, the TS is responsible for the specific association of DNMT1 with late-replicating pericentric heterochromatin during G2 and M.

The TS domain alone and the full length DNMT1 display differing subnuclear localization throughout the cell cycle. Whereas TS-GFP has a strong affinity for pericentric heterochromatin throughout the cell cycle (Fig. 3.19), DNMT1 is present at these sites only during late S phase and is retained here through G2 and M, to be released in G1 (Fig. 3.20 and Fig. 3.21). Thus, the activity of the TS in DNMT1 is controlled in such a way that it can associate with its target sites only during late-S, G2 and M. By analyzing the subnuclear localization of a series of point and deletion mutations in the DNMT1 fused to GFP throughout the cell cycle, we observe that a point mutation in PBD, abrogating PCNA binding ability (Chuang et al., 1997), resulted in reduced association of the mutant with replication sites and led to association with pericentric heterochromatin right during early- and mid-S phase (Table 3.2, Fig. 3.23). This pattern resembles the localization of the TS-GFP fusion protein. Thus, during S phase, PBD controls the localization of DNMT1 and overrides the strong affinity of the TS for pericentric heterochromatin. Three possibilities can be

envisaged to mediate this overriding effect of the PBD over the TS: (*i*) The PBD-PCNA interaction at the replication fork has a higher affinity and this might be dominant over the TS binding to pericentric heterochromatin. However, it is observed that overexpressed GFP-DNMT1 fusion does not associate with pericentric heterochromatin during early-/mid-S phase indicating that the overriding effect of the PBD over the TS cannot be explained as a simple competition between PCNA for the PBD and the pericentric heterochromatin for the TS. (*ii*) The TS is masked by other domains in DNMT1 which prevents its binding to pericentric heterochromatin. Binding of PCNA would result in a conformational change resulting in unmasking of the TS that makes it available for interaction with pericentric heterochromatin. (*iii*) Interaction of DNMT1 with PCNA allosterically activates the TS allowing binding to pericentric heterochromatin. In the latter two possibilities, the TS would get unmasked or allosterically activated only at the RF upon binding to PCNA. Thus the TS would get an opportunity to bind pericentric heterochromatin only during replication of these regions. Further experiments are required to sort out these three possibilities.

A fusion protein lacking the PBHD behaved similar to GFP-DNMT1 throughout S, G2 and M but was retained at pericentric regions during G1 (Table 3.2, Fig. 3.23 (*ii*)). Thus, the PBHD domain via its yet unidentified interacting partner seems to play a role in the release of DNMT1 from pericentric heterochromatin during G1.

4.4. Alternative mechanism of inheritance of DNA methylation in plants and fungi

It is interesting to note that maintenance of the two forms of epigenetic information, nucleosome state in *S. cerevisiae* (Zhang et al., 2000) and DNA methylation (this study), both are linked to PCNA. This makes sense considering the central place PCNA occupies in the replication machinery. Since DNA replication involves opening of the chromatin structure and formation of nascent unmethylated DNA, the proteins involved in epigenetic maintenance might have evolved in such a way that their activity is coupled to the DNA replication process itself. Sequence analysis aimed at identifying a PBD in the DNMT1 homologues in plants and fungi showed that they lack such a motif (Fig. 3.26) while the N-terminal domain of these proteins have the TS (duplicated in the case of plants; illustrated in Fig. 4.1) and the PBHD/BAH domains (Fig. 3.28) (Callebaut et al., 1999). Considering the conservation of PBD from archaebacteria to higher eucaryotes (Warbrick et al., 1998) (Dalrymple et al., 2001), the absence of a PBD in the DNMT1 homologues in plants and fungi is intriguing and suggests that maintenance of DNA methylation is uncoupled from the very process of DNA replication in these organisms. We propose that this could reflect the role DNA methylation has in these organisms and the distribution of methylated residues in the genome. In mammals, DNA methylation occurs in repetitive elements (transposons and other repeats) and also in coding regions except for CpG islands of active promoters (Yoder et al., 1997b). Methylation of coding regions is proposed to play a role in transcriptional regulation and there exists evidence correlating tissue specific methylation of promoters with transcriptional silencing in mammals (Futscher et al., 2002). The distribution of DNA

methylation in repeats as well as coding regions would have demanded from DNMT1 a mechanism that scans the whole genome indiscriminately for methylation patterns, which have to be inherited in the newly synthesized DNA. This could be best accomplished by targeting DNMT1 to sites of DNA replication, thereby creating a selective advantage confered by the PBD in DNMT1. In plants and fungi, genomic methylation is restricted mostly to transposons and other repeat elements (Rabinowicz et al., 1999) (Goyon et al., 1996), which are typically excluded from gene rich regions. Here, it is believed that DNA methylation mainly protects host genome from insults by invading foreign DNA/transposons (Martienssen and Colot, 2001). We hypothesize that, in these organisms, the *modus operandi* of the DNA methylation machinery might be by specifically targeting to repeats. This could be mediated by the TS sequence in the DNMT1 homologues in plants and fungi, which is supported by our observation that the TS has an innate ability to target to pericentric repeats. Thus, maintenance of DNA methylation in plants and fungi might be operating by an alternative, and probably evolutionarily older mechanism that targets DNMT1 to repeat elements. Consistent with this hypothesis is the observation that an N-terminal region in Dim2, required for methylation of duplicated genes inactivated by RIP (**r**epeat **i**nduced **p**oint mutation) (Kouzminova and Selker, 2001), has a TS-like motif identified in this study. Dim2 is a DNMT1 homologue in *N. crassa* that is required for all known DNA methylation in this fungus. Sequence analyses using a profile of the conserved motif in TS detected the TS motif in the N-terminal domain of Dim2 at an analogous position as that of TS in Masc2 (illustrated in Fig. 4.1), which is the DNMT1 homologue in the fungus *A. immersus*. In this context, it is worth pointing out that DNMT1 might also function as a *de novo* methyltransferase by acting as the primary step in methylating foreign DNA/repetitive elements by virtue of being targeted to DNA repeats. This is supported by deletion of dim2 in *N. crassa* which causes complete loss of all known DNA methylation indicating that it is a *de novo* methyltransferase (Foss et al., 1993). Thus, TS mediated targeting to repetitive elements might be a conserved mechanism for silencing foreign DNA by both *de novo* and maintenance methylation. It would be very interesting to know how this targeting occurs. In the case of human DNMT1, the TS harbours a Zn binding domain and a bipartite DNA binding domain (Zn-2 and DB in Fig. 3.14) identified *in vitro* (Chuang et al., 1996), which might be involved in targeting to repetitive elements. However, the corresponding region from the closely related murine DNMT1 is reported to have no DNA binding activity. Moreover, extensive deletion analysis of the TS (Leonhardt et al., 1992) has shown that presence of the DNA binding domain and the Zn binding region is not sufficient for association with pericentric heterochromatin. Also, deletion of one of the motifs of the bipartite DNA binding domain does not abrogate association with pericentric heterochromatin (TS-ß-gal, Fig. 3.16i). Alternatively, targeting of TS to repetitive DNA elements might occur via protein-protein interactions with other not yet identified proteins bound to the TS. Since the TS associates with pericentric heterochromatin at any cell cycle stage, these factors should be constitutive components of these heterochromatin regions.

Fig. 4.1. Domain organization in DNMT1 from metazoan, plants and fungi. *The TS is duplicated in plants. A TS-like motif identified in N. crassa Dim2 is also illustrated. BAH corresponds to PBHD. Names of organisms are abbreviated as in Fig. 3.26.*

Our hypothesis that plant and fungal DNMT1 are not targeted to RF is based only on sequence analyses and it could still be possible that the DNMT1 homologues in plants and fungi are targeted to RF by a mechanism independent of the PBD. It will be interesting to test this experimentally by studying the subnuclear distribution of DNMT1 during S phase and other cell cycle stages in plants and fungi. It would be also interesting to analyze the organization of functional domains of the DNA methyltransferases in the invertebrate chordate *Ciona intestinalis* in which the DNA repeats in the genome are not methylated while genes are methylated (Simmen et al., 1999).

4.5. Mechanism of maintenance of epigenetic information by DNMT1 in mammals

Genetic information is modulated by two epigenetic marks, DNA methylation and histone modifications, which have to be maintained at every round of DNA replication for controlled gene expression and genomic integrity. It has been shown that DNMT1 interacts with HDAC1 and 2 which has led to the proposal that DNMT1 plays a role in maintenance of deacetylated histones in late-replicating heterochromatin (Robertson et al., 2000) (Rountree et al., 2000). Based on our results, we propose the following model describing the mechanism of epigenetic maintenance by DNMT1 (Fig. 4.2). As the cell enters S phase, PCNA assembles throughout the nucleus at sites where DNA replication initiates (Leonhardt et al., 2000) (Somanathan et al., 2001) and DNMT1 associates with these sites via its PBD (Step-1, Fig. 4.2). DNMT1, with its preference for hemi-methylated DNA (Gruenbaum et al., 1982) (Bestor and Ingram, 1983) (Yoder et al., 1997a), then methylates the newly replicated DNA. This dynamic distribution of DNMT1 follows that of PCNA throughout S phase. At late RF, where pericentric alpha- and gamma-satellite DNA also replicate (Ten Hagen et al., 1990) (O'Keefe et al., 1992) (Leonhardt et al., 1992), DNMT1 is primarily recruited via interaction with PCNA followed by binding to a yet unknown but constitutive component of pericentric heterochromatin via its TS (Step-2, Fig. 4.2). At the cessation of DNA replication, PCNA disassembles from these sites while DNMT1 is retained there (Step-3, Fig. 4.2). We propose that the TS mediated retention of DNMT1 at late replicating pericentric heterochromatin throughout G2 has the following roles in the maturation of centromeric regions: (*i*) By being retained at pericentric heterochromatin throughout G2, DNMT1 would complete methylation of

newly synthesized DNA (Step-4, Fig. 4.2). This is important because DNMT1 being catalytically slow (Pradhan et al., 1999) would require more time to finish methylation of the densely methylated pericentric heterochromatin. This is supported by observations that a significant amount of micrococcal nuclease resistant DNA is methylated several hours following DNA synthesis (Geraci et al., 1974) (Woodcock et al., 1982) (Davis et al., 1985). (*ii*) Retention of DNMT1 at pericentric heterochromatin during G2 might play a role in the formation of condensed chromatin by recruiting HDACs which would deacetylate the newly deposited histones. It has been proposed that DNMT1 mediates chromatin maturation by recruiting HDAC2 to late RF (Rountree et al., 2000). Our observation that TS mediates retention of DNMT1 at pericentric regions during G2 supports this proposal and extends this proposed role of DNMT1 in chromatin maturation to G2.

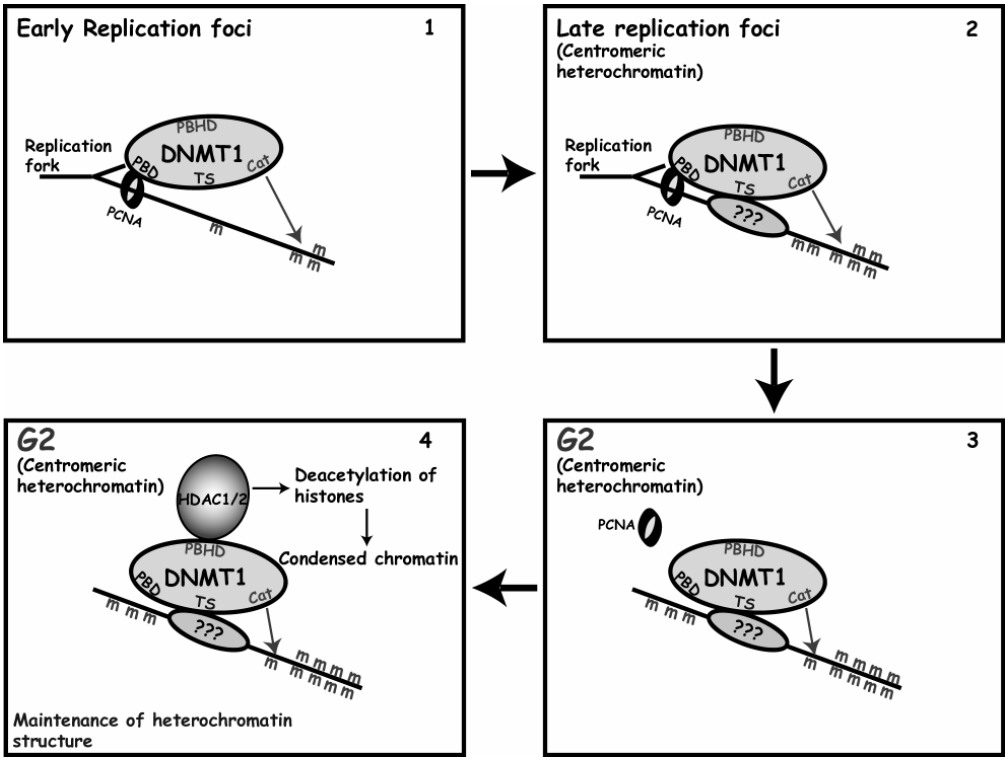


Fig. 4.2. Hypothetical model explaining the mechanism of inheritance of DNA methylation pattern by DNMT1. *The events at a replication fork during early-S and late-S are shown. (1) DNMT1 is recruited at early replication foci interacting with PCNA via the PBD. The catalytic domain (cat) in DNMT1 methylates the newly synthesized DNA strand. (2) At late replication foci, DNMT1 is again recruited by PCNA and methylates the newly synthesized strand. (3) At the cessation of DNA replication, PCNA disassembles from the DNA strand but DNMT1 is retained at these sites via the TS. DNMT1 continues methylating the pericentric heterochromatin. (4) During G2, DNMT1 is retained at the pericentric regions where it completes DNA methylation and acts as a docking site for histone modifying enzymes like HDAC1/2 and thereby aids in maturation of newly synthesized heterochromatin.*

The *in vivo* role of TS in maintaining and/or establishing DNA methylation patterns by DNMT1 is supported by rescue experiments with ES cells deficient in DNMT1 ($Dnmt1^{-/-}$). Wild type ES cells are capable of differentiation while $Dnmt1^{-/-}$ ES cells are not. This inability to differentiate can be rescued by expression of Dnmt1 but not mutant forms that lack the TS (Gaudet, F. and Leonhardt, H., unpublished results). Deletion of the TS in DNMT1, however, does not abolish catalytic activity

(Margot et al., 2000). Thus, the role of the TS in mediating association of DNMT1 with pericentric heterochromatin is essential for maintenance of DNA methylation *in vivo*. Furthermore, overexpression of TS in mammalian cells, likely preventing the access of endogenous DNMT1 to pericentric heterochromatin, causes large scale chromatin reorganization and increases micronuclei formation suggestive of genomic instability (Fig. 3.24 and Fig. 3.25). Thus the association of DNMT1 with heterochromatin mediated by the TS is vital for the maintenance of epigenetic information and the organization and stability of the genome.

5. Outlook

During this work a few new perspectives and ideas for future research developed that could not be pursued due to time constraints. Some of these were mentioned in the discussion and are briefly outlined below:

- Based on the hitherto unknown association of DNMT1 with centromeric heterochromatin in mammalian cells, the results obtained from studies on *Drosophila* cells should be revisited. DNMT1 associates with some subnuclear structures in about 10% *Drosophila* cells, which could reflect association with heterochromatin via the TS. This has to be further analyzed by costaining with markers for heterochromatin, like HP1. The TS might bind to some component of heterochromatin that is conserved in *Drosophila* and mammals.

- The protein component in heterochromatin to which the TS binds has to be identified by 2 hybrid screens or biochemical approaches like affinity purification. It cannot be ruled out that the TS directly binds to DNA at heterochromatin and this has to be tested by studying the ability of the TS to bind to specific DNA sequences from heterochromatic regions.

- The dependence of TS binding to pericentric heterochromatin on other epigenetic modifications such as acetylation and methylation states of histones should be investigated.

- The dynamics of the interaction of DNMT1 with various subnuclear structures at different stages of the cell cycle should be analyzed by fluorescence photobleaching techniques.

- The role of TS mediated retention of DNMT1 in G2 and M phases has to be established.

- The subcellular localization of the DNMT proteins from other organisms like plant and fungi should be analyzed, and the role of the TS in the DNMT1 from these organisms should be elucidated.

6. Postface

The subnuclear localization of DNMT1 throughout the cell cycle was determined by following the distribution of GFP-DNMT1 fusion protein by live cell microscopy. The specific cell cycle stages in live cells were identified by co-expressing DNA Ligase I fused to DsRed fluorescent protein (RFP-Ligase). Normally only mitosis and interphase can be identifed in live cells (Fig. I-A). Expression of RFP-Ligase allows identification of all the cell cycle stages in live cells. During S phase RFP-Ligase associates with RF and, thus, by following a cell through S phase the transition to G2 can be identified followed by mitosis and the transition to G1 (Fig. I-B). We observe that DNMT1 shows a diffused nuclear distribution in G1 phase, associates with RF during S phase and is bound to pericentric heterochromatin in G2 cells. Thus, the combination of both, the subnuclear distribution of GFP-DNMT1 and RFP-Ligase can be used as a marker to identify the cell cycle stage in fixed cells, or in a snapshot of live cells without the need for following the transition through different cell cycle stages (Fig. I-C). In other words, at any given time the cell cycle phase can clearly be deduced from the distribution of these two markers. Therefore, these markers should be valuable for the study of cell cycle dependent processes.

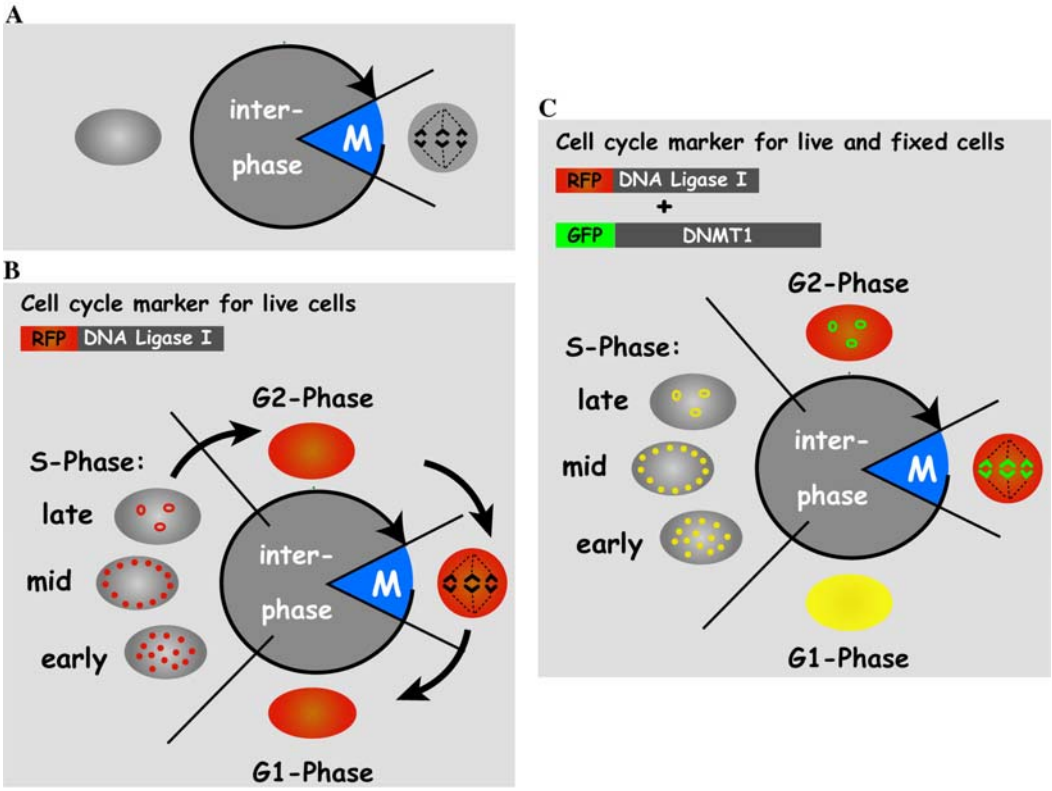


Fig. I. Cell cycle marker for live and fixed cells. ***(A)*** *Only mitosis (M) and interphase can be identified in live cells based on the cellular and nuclear morphology.* ***(B)*** *Expression of RFP-Ligase allows identification of the different cell cycle stages by live cell microscopy. In live cells expressing RFP-Ligase, G2 phase can be identified by following the transition of a cell from S phase (punctate RFP-Ligase pattern corresponding to replication sites) into G2 when RFP-Ligase is diffused in the nucleus. The progression through S phase can be followed by the different patterns of RF. During M phase, the nuclear membrane is broken down and RFP-Ligase is excluded from the chromatin. Cells in G1 phase can be identified by following a cell through mitosis into G1 when RFP-Ligase is diffused in the nucleus.* ***(C)*** *Expression of both GFP-DNMT1 and RFP-Ligase allows identification of the cell cycle stage in fixed as well as live cells. Colocalization of RFP-Ligase and GFP-DNMT1 at RF during S phase is shown in yellow. During G2, only GFP-DNMT1 is bound to pericentric heterochromatin (shown as green "donut" shaped structures) while RFP-Ligase is diffused. During mitosis GFP-DNMT1 is at the chromatin. In G1, both RFP-Ligase and GFP-DNMT1 are diffused (shown as yellow).*

7. Videos

Video-1: Time lapse series of a mouse cell expressing F-TS-GFP and RFP-Ligase. See Fig 3.19 for details. In the panel showing the RFP-Ligase image, the colour of the white arrows changes to yellow on cessation of DNA replication when RFP-Ligase disassembles from these sites.

Video-2: Time lapse series of a mouse cell expressing GMT and RFP-Ligase. See Fig 3.21 for details. The arrow marks indicate the chromatin during mitosis. Since the daughter cells were in different Z-planes they are shown in separate frames (Daughter cell-1 and Daughter cell-2).

Video-3: Time lapse series of a mouse cell in mitosis expressing GMT. See Fig 3.21 for details. DIC images of each time point are also shown. At 26 min, green fluorescence from only one of the daughter nuclei is visible (Daughter cell-1) as the other nucleus is outside the plane of focus.

8. Bibliography

Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M., and Falaschi, A. (2000). Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 287, 2023-2026.

Achwal, C. W., Ganguly, P., and Chandra, H. S. (1984). Estimation of the amount of 5-methylcytosine in Drosophila melanogaster DNA by amplified ELISA and photoacoustic spectroscopy. EMBO J 3, 263-266.

Ahmad, K., and Henikoff, S. (2001). Centromeres are specialized replication domains in heterochromatin. J Cell Biol 153, 101-110.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol 215, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Bachman, K. E., Park, B. H., Rhee, I., Rajagopalan, H., Herman, J. G., Baylin, S. B., Kinzler, K. W., and Vogelstein, B. (2003). Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene. Cancer Cell 3, 89-95.

Bachman, K. E., Rountree, M. R., and Baylin, S. B. (2001). Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. J Biol Chem 276, 32282-32287.

Baker, T. A., and Bell, S. P. (1998). Polymerases and the replisome: machines within machines. Cell 92, 295-305.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120-124.

Becker, P. B., Ruppert, S., and Schutz, G. (1987). Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. Cell 51, 435-443.

Bell, S. P., Mitchell, J., Leber, J., Kobayashi, R., and Stillman, B. (1995). The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. Cell 83, 563-568.

Berezney, R., Dubey, D. D., and Huberman, J. A. (2000). Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma 108, 471-484.

Berezney, R., Mortillaro, M. J., Ma, H., Wei, X., and Samarabandu, J. (1995). The nuclear matrix: a structural milieu for genomic function. Int Rev Cytol, 1-65.

Berger, S. L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev 12, 142-148.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 203, 971-983.

Bestor, T. H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. EMBO J 11, 2611-2617.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. Hum Mol Genet 9, 2395-2402.

Bestor, T. H., and Ingram, V. M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A 80, 5559-5563.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev 16, 6-21.

Bird, A. P., Taggart, M. H., Nicholls, R. D., and Higgs, D. R. (1987). Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. EMBO J 6, 999-1004.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7, 1513-1523.

Blow, J. J. (2001). Control of chromosomal DNA replication in the early Xenopus embryo. EMBO J 20, 3293-3297.

Bonassi, S., and Au, W. W. (2002). Biomarkers in molecular epidemiology studies for health risk prediction. Mutat Res 511, 73-86.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., and Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science 294, 2536-2539.

Bozhenok, L., Wade, P. A., and Varga-Weisz, P. (2002). WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. EMBO J 21, 2231-2241.

Bravo, R., Frank, R., Blundell, P. A., and Macdonald-Bravo, H. (1987). Cyclin/PCNA is the auxiliary protein of DNA polymerase-delta. Nature 326, 515-517.

Bravo, R., and Macdonald-Bravo, H. (1987). Existence of two populations of cyclin/proliferating cell nuclear antigen during the cell cycle: association with DNA replication sites. J Cell Biol 105, 1549-1554.

Callebaut, I., Courvalin, J. C., and Mornon, J. P. (1999a). The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. FEBS Lett 446, 189-193.

Callebaut, I., Courvalin, J. C., and Mornon, J. P. (1999b). The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. FEBS Lett 446, 189-193.

Cardoso, M. C., Joseph, C., Rahn, H. P., Reusch, R., Nadal-Ginard, B., and Leonhardt, H. (1997). Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication in vivo. J Cell Biol 139, 579-587.

Cardoso, M. C., and Leonhardt, H. (1999a). „Differentiation, development and programmed cell death.“ in The Molecular Basis of Cell Cycle and Growth Control, G. Stein, R. Baserga, D. Denhardt, and A. Giordano, eds. (New York, John Wiley & Sons), pp. 305-347.

Cardoso, M. C., and Leonhardt, H. (1999b). DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol 147, 25-32.

Cardoso, M. C., Leonhardt, H., and Nadal-Ginard, B. (1993). Reversal of terminal differentiation and control of DNA replication: cyclin A and Cdk2 specifically localize at subnuclear sites of DNA replication. Cell 74, 979-992.

Cardoso, M. C., Sporbert, A., and Leonhardt, H. (1999). Structure and function in the nucleus: subnuclear trafficking of DNA replication factors. J Cell Biochem Suppl, 15-23.

Carlson, L. L., Page, A. W., and Bestor, T. H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev 6, 2536-2541.

Celis, J. E., and Celis, A. (1985). Cell cycle-dependent variations in the distribution of the nuclear protein cyclin proliferating cell nuclear antigen in cultured cells: subdivision of S phase. Proc Natl Acad Sci U S A 82, 3262-3266.

Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem 277, 38746-38754.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., and Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277, 1996-2000.

Chuang, L. S., Ng, H. H., Chia, J. N., and Li, B. F. (1996). Characterisation of independent DNA and multiple Zn-binding domains at the N terminus of human DNA-(cytosine-5) methyltransferase: modulating the property of a DNA-binding domain by contiguous Zn-binding motifs. J Mol Biol 257, 935-948.

Colot, V., and Rossignol, J. L. (1999). Eukaryotic DNA methylation as an evolutionary device. Bioessays 21, 402-411.

Cox, L. S., and Laskey, R. A. (1991). DNA replication occurs at discrete sites in pseudonuclei assembled from purified DNA in vitro. Cell 66, 271-275.

Czank, A., Hauselmann, R., Page, A. W., Leonhardt, H., Bestor, T. H., Schaffner, W., and Hergersberg, M. (1991). Expression in mammalian cells of a cloned gene encoding murine DNA methyltransferase. Gene 109, 259-263.

Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E., and Jennings, P. A. (2001). A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. Proc Natl Acad Sci U S A 98, 11627-11632.

Davis, T., Kirk, D., Rinaldi, A., Burdon, R. H., and Adams, R. L. (1985). Delayed methylation and the matrix bound DNA methylase. Biochem Biophys Res Commun 126, 678-684.

DePamphilis, M. L. (1999). Replication origins in metazoan chromosomes: fact or fiction? Bioessays 21, 5-16.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. Oncogene 21, 5400-5413.

Ehrlich, M., and Wang, R. Y. (1981). 5-Methylcytosine in eukaryotic DNA. Science 212, 1350-1357.

Foss, H. M., Roberts, C. J., Claeys, K. M., and Selker, E. U. (1993). Abnormal chromosome behavior in Neurospora mutants defective in DNA methylation. Science 262, 1737-1741.

Fox, M. H., Arndt-Jovin, D. J., Jovin, T. M., Baumann, P. H., and Robert-Nicoud, M. (1991). Spatial and temporal distribution of DNA replication sites localized by immunofluorescence and confocal microscopy in mouse fibroblasts. J Cell Sci 99, 247-253.

Franceschini, P., Martino, S., Ciocchini, M., Ciuti, E., Vardeu, M. P., Guala, A., Signorile, F., Camerano, P., Franceschini, D., and Tovo, P. A. (1995). Variability of clinical and immunological phenotype in immunodeficiency-centromeric instability-facial anomalies syndrome. Report of two new patients and review of the literature. Eur J Pediatr 154, 840-846.

Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 24, 88-91.

Fuss, J., and Linn, S. (2002). Human DNA polymerase epsilon colocalizes with proliferating cell nuclear antigen and DNA replication late, but not early, in S phase. J Biol Chem 277, 8658-8666.

Futscher, B. W., Oshiro, M. M., Wozniak, R. J., Holtan, N., Hanigan, C. L., Duan, H., and Domann, F. E. (2002). Role for DNA methylation in the control of cell type specific maspin expression. Nat Genet 31, 175-179.

Gaudet, F., Talbot, D., Leonhardt, H., and Jaenisch, R. (1998). A short DNA methyltransferase isoform restores methylation in vivo. J Biol Chem 273, 32725-32729.

Geraci, D., Eremenko, T., Cocchiara, R., Granieri, A., Scarano, E., and Volpe, P. (1974). Correlation between synthesis and methylation of DNA in HeLa cells. Biochem Biophys Res Commun 57, 353-358.

Giacca, M., Zentilin, L., Norio, P., Diviacco, S., Dimitrova, D., Contreas, G., Biamonti, G., Perini, G., Weighardt, F., Riva, S., and et al. (1994). Fine mapping of a replication origin of human DNA. Proc Natl Acad Sci U S A 91, 7119-7123.

Gilbert, D. M. (1998). Replication origins in yeast versus metazoa: separation of the haves and the have nots. Curr Opin Genet Dev 8, 194-199.

Glickman, J. F., Pavlovich, J. G., and Reich, N. O. (1997). Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. J Biol Chem 272, 17851-17857.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23, 175-182.

Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A., and Nag, A. (1984). Replication timing of genes and middle repetitive sequences. Science 224, 686-692.

Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of Drosophila melanogaster contains 5-methylcytosine. EMBO J 19, 6918-6923.

Goyon, C., Rossignol, J. L., and Faugeron, G. (1996). Native DNA repeats and methylation in Ascobolus. Nucleic Acids Res 24, 3348-3356.

Graham, F. L., and van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52, 456-467.

Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci U S A 84, 4355-4358.

Gruenbaum, Y., Cedar, H., and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. Nature 295, 620-622.

Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M., and Kuriyan, J. (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. Cell 87, 297-306.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development 129, 1983-1993.

Hatton, K. S., Dhar, V., Brown, E. H., Iqbal, M. A., Stuart, S., Didamo, V. T., and Schildkraut, C. L. (1988). Replication program of active and inactive multigene families in mammalian cells. Mol Cell Biol 8, 2149-2158.

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D., and Spector, D. L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. Cell 107, 727-738.

Hein, J. (1990). Unified approach to alignment and phylogenies. Methods Enzymol 183, 626-645.

Holliday, R., and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. Science 187, 226-232.

Hozak, P., Hassan, A. B., Jackson, D. A., and Cook, P. R. (1993). Visualization of replication factories attached to nucleoskeleton. Cell 73, 361-373.

Issa, J. P., Ottaviano, Y. L., Celano, P., Hamilton, S. R., Davidson, N. E., and Baylin, S. B. (1994). Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet 7, 536-540.

Jackson, D. A., and Cook, P. R. (1986). Replication occurs at a nucleoskeleton. EMBO J 5, 1403-1410.

Jackson, D. A., and Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. J Cell Biol 140, 1285-1295.

Jacob, F., Brenner, S., and Cuzin, F. (1964). On the regulation of DNA replication in bacteria. Cold Spring Harbor Sump Quant Biol 28, 329-345.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum Mol Genet 2, 731-735.

Jones, P. A., and Taylor, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85-93.

Jost, J. P., and Jost, Y. C. (1994). Transient DNA demethylation in differentiating mouse myoblasts correlates with higher activity of 5-methyldeoxycytidine excision repair. J Biol Chem 269, 10040-10043.

Kannouche, P., Broughton, B. C., Volker, M., Hanaoka, F., Mullenders, L. H., and Lehmann, A. R. (2001). Domain structure, localization, and function of DNA polymerase eta, defective in xeroderma pigmentosum variant cells. Genes Dev 15, 158-172.

Kouzminova, E., and Selker, E. U. (2001). dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in Neurospora. EMBO J 20, 4309-4323.

Kress, C., Thomassin, H., and Grange, T. (2001). Local DNA demethylation in vertebrates: how could it be performed and targeted? FEBS Lett 494, 135-140.

Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J. (1994). Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell 79, 1233-1243.

Lachner, M., and Jenuwein, T. (2002). The many faces of histone lysine methylation. Curr Opin Cell Biol 14, 286-298.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410, 116-120.

Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122, 3195-3205.

Leipe, D. D., Aravind, L., and Koonin, E. V. (1999). Did DNA replication evolve twice independently? Nucleic Acids Res 27, 3389-3401.

Leonhardt, H., and Cardoso, M. C. (1995). Targeting and association of proteins with functional domains in the nucleus: the insoluble solution. Int Rev Cytol 162B, 303-335.

Leonhardt, H., and Cardoso, M. C. (2000). DNA methylation, nuclear structure, gene expression and cancer. J Cell Biochem Suppl, 78-83.

Leonhardt, H., Page, A. W., Weier, H. U., and Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71, 865-873.

Leonhardt, H., Rahn, H. P., and Cardoso, M. C. (1998). Intranuclear targeting of DNA replication factors. J Cell Biochem Suppl 31, 243-249.

Leonhardt, H., Rahn, H. P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D., and Cardoso, M. C. (2000a). Dynamics of DNA replication factories in living cells. J Cell Biol 149, 271-280.

Leonhardt, H., Sporbert, A., and Cardoso, M. C. (2000b). Targeting regulatory factors to intranuclear replication sites. Crit Rev Eukaryot Gene Expr 10, 127-133.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature 366, 362-365.

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69, 915-926.

Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D., and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science 293, 2453-2455.

Liu, Y., Oakeley, E. J., Sun, L., and Jost, J. P. (1998). Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. Nucleic Acids Res 26, 1038-1045.

Lock, L. F., Takagi, N., and Martin, G. R. (1987). Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell 48, 39-46.

Lyko, F., Whittaker, A. J., Orr-Weaver, T. L., and Jaenisch, R. (2000). The putative Drosophila methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries. Mech Dev 95, 215-217.

Ma, H., Samarabandu, J., Devdhar, R. S., Acharya, R., Cheng, P. C., Meng, C., and Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. J Cell Biol 143, 1415-1425.

Margot, J. B., Aguirre-Arteta, A. M., Di Giacco, B. V., Pradhan, S., Roberts, R. J., Cardoso, M. C., and Leonhardt, H. (2000). Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. J Mol Biol 297, 293-300.

Margot, J. B., Cardoso, M. C., and Leonhardt, H. (2001). Mammalian DNA methyltransferases show different subnuclear distributions. J Cell Biochem 83, 373-379.

Martienssen, R. A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science 293, 1070-1074.

Matthias, P., Müller, M. M., Schreiber, E., Rusconi, S., and Schaffner, W. (1989). Eukaryotic expression vectors for the analysis of mutant proteins. Nucl Acids Res 17, 6418.

Mattock, H., Lane, D. P., and Warbrick, E. (2001). Inhibition of cell proliferation by the PCNA-binding region of p21 expressed as a GFP miniprotein. Exp Cell Res 265, 234-241.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature 403, 501-502.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development 99, 371-382.

Montecucco, A., Rossi, R., Levin, D. S., Gary, R., Park, M. S., Motycka, T. A., Ciarrocchi, G., Villa, A., Biamonti, G., and Tomkinson, A. E. (1998). DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. EMBO J 17, 3786-3795.

Montecucco, A., Savini, E., Weighardt, F., Rossi, R., Ciarrocchi, G., Villa, A., and Biamonti, G. (1995). The N-terminal domain of human DNA ligase I contains the nuclear localization signal and directs the enzyme to sites of DNA replication. EMBO J 14, 5379-5386.

Nakamura, H., Morita, T., and Sato, C. (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. Exp Cell Res 165, 291-297.

Nakayasu, H., and Berezney, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. J Cell Biol 108, 1-11.

Newlon, C. S., and Theis, J. F. (1993). The structure and function of yeast ARS elements. Curr Opin Genet Dev 3, 752-758.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P., and Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev 16, 479-489.

Noma, K., Allis, C. D., and Grewal, S. I. (2001). Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science 293, 1150-1155.

O'Keefe, R. T., Henderson, S. C., and Spector, D. L. (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. J Cell Biol 116, 1095-1110.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247-257.

Okano, M., Xie, S., and Li, E. (1998a). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet 19, 219-220.

Okano, M., Xie, S., and Li, E. (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res 26, 2536-2540.

Ono, T., Takahashi, N., and Okada, S. (1989). Age-associated changes in DNA methylation and mRNA level of the c-myc gene in spleen and liver of mice. Mutat Res 219, 39-50.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol 10, 475-478.

Otterlei, M., Warbrick, E., Nagelhus, T. A., Haug, T., Slupphaug, G., Akbari, M., Aas, P. A., Steinsbekk, K., Bakke, O., and Krokan, H. E. (1999). Post-replicative base excision repair in replication foci. EMBO J 18, 3834-3844.

Parker, B. A., and Stark, G. R. (1979). Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present early in infections by virus or viral DNA. J Virol 31, 360-369.

Pinarbasi, E., Elliott, J., and Hornby, D. P. (1996). Activation of a yeast pseudo DNA methyltransferase by deletion of a single amino acid. J Mol Biol 257, 804-813.

Pradhan, S., Bacolla, A., Wells, R. D., and Roberts, R. J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem 274, 33002-33010.

Prelich, G., Tan, C. K., Kostura, M., Mathews, M. B., So, A. G., Downey, K. M., and Stillman, B. (1987). Functional identity of proliferating cell nuclear antigen and a DNA polymerase-delta auxiliary protein. Nature 326, 517-520.

Rabinowicz, P. D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L. D., Stein, L., McCombie, W. R., and Martienssen, R. A. (1999). Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet 23, 305-308.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D., and Cantoni, G. L. (1986). Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA demethylation during differentiation. Proc Natl Acad Sci U S A 83, 2827-2831.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science 293, 1089-1093.

Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14, 9-25.

Roberts, D. B. (1998). Drosophila, A Practical Approach (Second Edition). Oxford University Press.

Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L., and Wolffe, A. P. (2000). DNMT1 forms a complex with rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet 25, 338-342.

Robertson, K. D., and Wolffe, A. P. (2000). DNA methylation in health and disease. Nat Rev Genet 1, 11-19.

Rougier, N., Bourc'his, D., Gomes, D. M., Niveleau, A., Plachot, M., Paldi, A., and Viegas-Pequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes Dev 12, 2108-2113.

Rountree, M. R., Bachman, K. E., and Baylin, S. B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet 25, 269-277.

Sadoni, N., Langer, S., Fauth, C., Bernardi, G., Cremer, T., Turner, B. M., and Zink, D. (1999). Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. J Cell Biol 146, 1211-1226.

Samaniego, R., de la Torre, C., and Moreno Diaz de la Espina, S. (2002). Dynamics of replication foci and nuclear matrix during S phase in Allium cepa L. cells. Planta 215, 195-204.

Sambrook, J., and Russell, D. W. (2001). Molecular Cloning, A Laboratory Manual (Third Edition). Cold Spring Harbor Laboratory Press.

Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J Embryol Exp Morphol 27, 353-365.

Simmen, M. W., Leitgeb, S., Charlton, J., Jones, S. J., Harris, B. R., Clark, V. H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. Science 283, 1164-1167.

Smith, J. G., and Calos, M. P. (1995). Autonomous replication in Drosophila melanogaster tissue culture cells. Chromosoma 103, 597-605.

Somanathan, S., Suchyna, T. M., Siegel, A. J., and Berezney, R. (2001). Targeting of PCNA to sites of DNA replication in the mammalian cell nucleus. J Cell Biochem 81, 56-67.

Sparvoli, E., Levi, M., and Rossi, E. (1994). Replicon clusters may form structurally stable complexes of chromatin and chromosomes. J Cell Sci 107, 3097-3103.

Sporbert, A., Gahl, A., Ankerhold, R., Leonhardt, H., and Cardoso, M. C. (2002). DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. Mol Cell 10, 1355-1365.

Stewart, C. L., Stuhlmann, H., Jahner, D., and Jaenisch, R. (1982). De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc Natl Acad Sci U S A 79, 4098-4102.

Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979). Isolation and characterisation of a yeast chromosomal replicator. Nature 282, 39-43.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature 403, 41-45.

Suetake, L., Shi, L., Watanabe, D., Nakamura, M., and Tajima, S. (2001). Proliferation stage-dependent expression of DNA methyltransferase (Dnmt1) in mouse small intestine. Cell Struct Funct 26, 79-86.

Tate, P. H., and Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev 3, 226-231.

Ten Hagen, K. G., Gilbert, D. M., Willard, H. F., and Cohen, S. N. (1990). Replication timing of DNA sequences associated with human centromeres and telomeres. Mol Cell Biol 10, 6348-6355.

Toth, M., Muller, U., and Doerfler, W. (1990). Establishment of de novo DNA methylation patterns. Transcription factor binding and deoxycytidine methylation at CpG and non-CpG sequences in an integrated adenovirus promoter. J Mol Biol 214, 673-683.

Tucker, J. D., and Preston, R. J. (1996). Chromosome aberrations, micronuclei, aneuploidy, sister chromatid exchanges, and cancer risk assessment. Mutat Res 365, 147-159.

Tweedie, S., Ng, H. H., Barlow, A. L., Turner, B. M., Hendrich, B., and Bird, A. (1999). Vestiges of a DNA methylation system in Drosophila melanogaster? Nat Genet 23, 389-390.

Varga-Weisz, P. D., and Dalgaard, J. Z. (2002). A mark in the core: silence no more! Mol Cell 9, 1154-1156.

Warbrick, E., Heatherington, W., Lane, D. P., and Glover, D. M. (1998). PCNA binding proteins in Drosophila melanogaster : the analysis of a conserved PCNA binding domain. Nucleic Acids Res 26, 3925-3932.

Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. Cell 24, 33-40.

Wilkinson, C. R., Bartlett, R., Nurse, P., and Bird, A. P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Res 23, 203-210.

Wilks, A., Seldran, M., and Jost, J. P. (1984). An estrogen-dependent demethylation at the 5' end of the chicken vitellogenin gene is independent of DNA synthesis. Nucleic Acids Res 12, 1163-1177.

Woodcock, D. M., Adams, J. K., and Cooper, I. A. (1982). Characteristics of enzymatic DNA methylation in cultured cells of human and hamster origin, and the effect of DNA replication inhibition. Biochim Biophys Acta 696, 15-22.

Wu, C., and Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. Science 293, 1103-1105.

Wu, J. C., and Santi, D. V. (1987). Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem 262, 4778-4786.

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402, 187-191.

Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature 270, 725-727.

Yoder, J. A., Soman, N. S., Verdine, G. L., and Bestor, T. H. (1997a). DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol 270, 385-395.

Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997b). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13, 335-340.

Zhang, Z., Hayashi, M. K., Merkel, O., Stillman, B., and Xu, R. M. (2002). Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. EMBO J 21, 4600-4611.

Zhang, Z., Shibahara, K., and Stillman, B. (2000). PCNA connects DNA replication to epigenetic inheritance in yeast. Nature 408, 221-225.

Zheleva, D. I., Zhelev, N. Z., Fischer, P. M., Duff, S. V., Warbrick, E., Blake, D. G., and Lane, D. P. (2000). A quantitative study of the in vitro binding of the C-terminal domain of p21 to PCNA: affinity, stoichiometry, and thermodynamics. Biochemistry 39, 7388-7397.

Appendix

A) Acknowledgements

I would like to express my gratitude to people who have directly or indirectly helped me in this work.

I am grateful to Drs. Cristina Cardoso and Heinrich Leonhardt for their supervision of this work. Their support, ideas and helpful discussions have immensely contributed to this work, and to my personal intellectual development. I greatly appreciate Heinrich's calm and Cristina's understanding that have helped us sail through rough times. I appreciate their critical and patient reading of this manuscript. Their approachability, and especially Cristina's infectious excitement, have made working a pleasure. Thanks, Cristina and Heinrich, for all the help and support.

I greatly appreciate Ana Aguirre Arteta, Hans Peter Rahn and Wolfgang Derer for their congenial presence that helped me a lot in adjusting to a foreign country. Their help within and outside the lab have been very useful. Special thanks to Ana for her company in the lab during late working hours and of course for the homemade dinners delivered to the lab. Jean Margot has been very helpful with his critical eye on my work and interesting discussions. Thanks, Jean, for your pleasant company and much needed advice. I thank Anje Sportbert for helping me with the microscopes and useful comments on this work. Also, thanks for those images that saved me a lot of time peering through the microscope. I thank Ingrid Grunewald, Petra Domaing, Danny Nowak, Anja Gahl and Marion Bengs for their valuable help and creating a cheerful and pleasant atmosphere. Robert Martin and Maik Grohman, thanks for the advice and tips on optimizing the print quality. I thank Jochen Kramer and Christina Quensel for introducing me to some of the lab protocols and techniques.

I would like to thank Dr. Harald Saumweber for his help with the registration in the university, critical comments on my work and encouragement. I thank him and Dr. Peter Becker for agreeing to review this work. I thank Dr. Manfred Gossen for his comments and allowing me to use the fly facility. Thanks also to the lab members of Dr. Saumweber and Dr. Gossen for helping me.

I am grateful to Tina Baldinger for helping with the German translation in a very short time.

Aditya, thanks a lot for your help and useful discussions. Also, thanks for the help with fly work, most importantly a Sunday morning crash course over the phone on identifying virgins (well, in *Drosophila*!). Trupti, Geetu and Deepnil, although far, thanks for being there whenever I wanted. Thanks a lot Swaroop, Sumi, Zhyldyz, Kamal, Dheeraj, Anand, Deb and Rashid for the good company. Thanks to Matthias for being just a phone call away and take me to the wonderful world outside science.

Last, but most importantly, thanks to my parents and brothers for their love and encouragement, and keeping alight the torch of hope that brings me here.

B) CURRICULUM VITÆ

Name: Hariharan P. Easwaran
Address: Max Delbrück Centre for Molecular Medicine, FVK, Wiltbergstrasse 50, D-13125 Berlin, Germany
Date of Birth: 16 Sep. 1974
Place of Birth: Mumbai, India

EDUCATION

1991-1994	B.Sc. (Microbiology); University of Bombay, India
1994-1996	M.Sc. (Microbiology); M.S.University, Baroda, India
1996-1998	Research student, Indian Institute of Science, Bangalore, India
1998-2003	Ph.D., Max Delbrück Centre for Molecular Medicine, Berlin, Germany

AWARDS/SCHOLARSCHIPS

1993	Collegiate seminar presentation award during Bachelor's studies
1996	GATE (Graduate Aptitude Test in Engineering) fellowship for pursuing research
1996	NET (National Eligibility Test) fellowship for pursuing research
2002	MDC-FMP Ph.D. Symposium Presentation Award, 2002, Max Delbrück Centre for Molecular Medicine, Berlin, Germany

PRESENTATIONS AT CONFERENCES

2002	SFB/Transregio5 meeting on "Chromatin Assembly and inheritance of functional states"

RESEARCH EXPERIENCE

1993-1994	Research project titled "Physico-Chemical and Microbiological Examination of Thane Creek Water" during B.Sc. (Bombay, India)
1994-1996	M.Sc thesis project titled " Study of Amylase from *Aspergillus repens*" (Baroda, India)
1996-1998	Research project on " Isolation of protein factors that bind to AU rich elements present at the 3'-UTR of mRNAs in higher eucaryotes" (Bangalore, India)
1998-1999	Research project on " Protein transduction into terminally differentiated cells and tissue regeneration" (Berlin, Germany)
1999-2003	Doctoral thesis project (Berlin, Germany)

LIST OF PUBLICATIONS

Derer, W., Easwaran, H. P., Knopf, C. W., Leonhardt, H., and Cardoso, M. C. (1999). Direct protein transfer to terminally differentiated muscle cells. J. Mol. Med. 77, 609-613.

Derer, W., Easwaran, H. P., Leonhardt, H., and Cardoso, M. C. (2002). A novel approach to induce cell cycle reentry in terminally differentiated muscle cells. The FASEB J. 16: 132-133.

C) Statement

Hiermit erkläre ich, die vorliegende Arbeit selbständig angefertig zu haben. Ich habe keine unerlaubten sowie unerwähnten Hilfen benutzt.

Berlin, 23. April 2003 Hariharan P. Easwaran