Studies of DNA demethylation

Hsin-hao Tsai

Thesis presented for the degree of Doctor of Philosophy

School of Biological Sciences

University of Edinburgh

2004

Declaration

I declare that this thesis was composed by myself and the research presented is my own except where otherwise stated.

Acknowledgements

There are many people I need to thank. Without the help of these very important individuals, I would not have made it this far.

First of all, I am very grateful for Adrian's brave decision of taking me on as a PhD student under his supervision. Thanks to everyone at the Bird Lab for helping me during the course of my PhD and for making my time at the Lab such an enjoyable one. Thanks especially for all the special language lessons, life lessons in drinking and other social aspects. You are always on my mind, even though your names are not mentioned.

Special thanks to Jim Selfridge and Donald Macleod whom I bugged for their "embryos" during the making of the HmC story. Thanks to Aileen Graig for organising my labware and for helping me out when I was in trouble. BIG thanks to Helle Jørgenson for her kindness, friendly company and great understanding to my confusing questions. I would also like to thank Helle, Jim and Ms MacDougal for proofreading my strange "foreigner's English" and their comments.

Thanks too, to all my friends here for dragging me away from my darkness and spreading sunshine over me -- and especially to my friends who are far away for their constant encouragements.

Special thanks go to my brother, Yuan-hao, for his remote technical support, for being my remote entertainment centre and source of moral support. Finally, I'd like to dedicate this thesis to my parents who offered unfailing support at all times.

Abstract

DNA methylation is an epigenetic event which regulates gene expression by transcriptional repression and plays an important role in the biological phenomena of imprinted genes and X inactivation. On the other hand, the reversal of DNA methylation, DNA demethylation, is not as well understood. DNA demethylation has been demonstrated to occur as an active process during embryo development, tumourigenesis and hormone-induced gene activation, or via a replication-dependent passive route. However, the molecular mechanism of active DNA demethylation is yet to be determined, and a precise role of DNA demethylation *in vivo* remains obscure.

Interestingly, bacteriophage T3 has been reported to express S-adenosylmethionine hydrolase (SAMase) during the early stage of phage infection. In order to overcome the restriction-modification (R-M) system, SAMase destroys the modification cofactor *S*-adenosylmethionine (SAM) and hydrolyses SAM into homoserine and methylthioadenosine (MTA). Since SAM is also the major donor of the methyl groups incorporated in DNA methylation, this SAM cleaving activity may be utilized as a demethylating agent, which helps us to understand the mechanism of DNA demethylation. However, the reaction mechanism of T3 SAMase has not been very well studied. To further elucidate this mechanism, experiments were carried out to purify recombinant SAMase and to enable attempts to solve the crystal structure of this enzyme. In addition, we aimed to observe effects of DNA demethylation in mammalian cells by using this hydrolase activity to reduce the cellular level of SAM. SAMase was also chosen to substitute the more commonly used demethylating agent, 5-azaC, in order to avoid the highly toxic impact on drug-receiving cells.

Additionally, RNA interference (RNAi), was utilised to explore the impact of DNA demethylation. Using small interfering RNA (siRNA), we depleted the mRNA

of the enzyme responsible for maintenance methylation, DNA methyltransferase 1 (Dnmt1), and observed a correlation between DNA demethylation and *Xist* expression when the methyl-binding protein MBD2 was removed. Furthermore, the oxidative DNA repair has been suggested as a candidate pathway, which involves demethylation of the 1-methyladenine and 3-methylcytosine via a hydroxmethyl intermediate. To explore this possible mechanism, we investigated the candidate pathway by searching for a putative intermediate of hydroxymethyl cytosine during active DNA demethylation *in vivo*.

Abbreviations

A_x	absorbance at x nm
A	adenine or adenosine
aa	amino acid
ATP	adenosine triphosphate
APS	ammonium persulphate
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine or cytidine
C′	carboxyl
cDNA	complementary DNA
Ci	Curie
cm	centimetre
Ct	threshold cycle
Da	Dalton
DAPI	4′,6-Diamidino-2-phenylindole
DEPC	diethyl pyrocarbonate
ds	double stranded
dH_2O	deionised water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
Dnmt	DNA methyltransferase
dNTP	deoxynucleoside triphosphate
DOC	deoxycholate
DTT	dithiothreitol
E.C.	Enzyme Commission
EDTA	diaminoethanetetraacetic acid
ES	embryonic stem
f	femto (10^{-15})
g	gram
g	relative centrifugal force

G	guanine or guanosine
Gapdh	glyceraldehyde-3-phosphate dehydrogenase
Glu-HmC	glucosylated hydroxymethylcytosine
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HmC; 5-HmC	5-hydroxymethylcytosine
HmU	hydroxymethyluracil
hr	hour
IPTG	isopropul-β-D-thiogalactopyranoside
IVF	*in vitro* ferterlisation
kb	kilobase
kDa	kilodalton
l	litre
LB	Luria-Bertani
m	milli (10^{-3})
M	molar
mA	milli-ampere
MBD	methyl-binding domain
min	minute
mC; 5-mC	5-methylcytosine
MOI	multiplicity of infection
mol	mole
mRNA	messenger RNA
MTA	methyl-thio-adenosine
n	nano (10^{-9})
nm	nanometre
nt	nucleotide
O.D.$_x$	optical density at x nm
oligo	oligonucleotide
ori	origin of replication
p	pico (10^{-12})
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfg	plaque forming unit
pH	$-\log_{10}[H^+]$
PI	propidium iodide
RNA	ribonucleic acid
RNase	ribonuclease

R-M	restriction-modification
rpm	revolutions per minute
RT-PCR	reverse-transcription PCR
ss	single stranded
SAH	*s*-adenosylhomocysteine
SAM	*s*-adenosylmethionine
SAMase	S-adenosylmethionine hydrolase
SDS	sodium dodecyl sulphate
sec	second
T	thymine or thymidine
TAE	Tris-acetate-EDTA (buffer)
TE	Tris-EDTA
TEMED	N,N,N′,N′-tetramethylethyledediamine
TLC	thin layer chromatography
Tris	2-amino-2(hydroxymethyl)-1,2,-propandiol
U	unit; uracil or uridine
UV	ultra violet
V	volts
v/v	volume per volume
w/v	weight per volume
Xist	X inactivate specific transcript
Δ	deletion
μ	micro (10^{-6})
°C	degree Celsius

Amino Acids

Amino acid	3-letter code	1-letter code
Arginine	Arg	R
Asparagine	Asn	N
Glutamine	Gln	Q
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Alanine	Ala	A
Cysteine	Cys	C
Glycine	Gly	G
Aspartte	Asp	D
Glutamate	Glu	E
Phenylalanine	Phe	F

Genetic code

		Second base				
		T	C	A	G	
First base	T	Phe	Ser	Tyr	Cys	T
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	Stop	Stop	A
		Leu	Ser	Stop	Trp	G
	C	Leu	Pro	His	Arg	T
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
	A	Ile	Thr	Asn	Ser	T
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
	G	Val	Ala	Asp	Gly	T
		Val	Ala	Asp	Gly	C
		Val	Ala	Glu	Gly	A
		Val	Ala	Glu	Gly	G

Third base

Table of Contents

List of Figures

List of Tables

Introduction

Chapter 1
Introduction

1.1 Overview

The work presented in this thesis focuses on DNA demethylation. In one line of experiments, attempts were made to create a demethylating environment in cells, by methods other than the commonly used cytosine analogs such as 5-azacytidine, a typical demethylating drug which is known for its toxic effects on cells. Also, we looked into DNA demethylation events *in vivo* with the hope of resolving the long-puzzled mystery concerning the mechanism of DNA demethylation.

1.2 Occurrence of DNA methylation

DNA methylation is a very well-described modification, which involves transfer of a methyl group from S-adenosylmethionine (SAM), the major methyl donor, to the 5' position of cytosine by methyltransferases (Fig. 1.1). In eukaryotes, this modification usually resides symmetrically in the context of CpG dinucleotides. However, methylation in prokaryotes can occur in a variety of sequence contexts, and is also found at the N-6 position of adenine. Additionally, the methyl group can be added to unusual sites, such as the O-6 position of guanine, by DNA damage from alkylating agents such as methyl methanesulfonate.
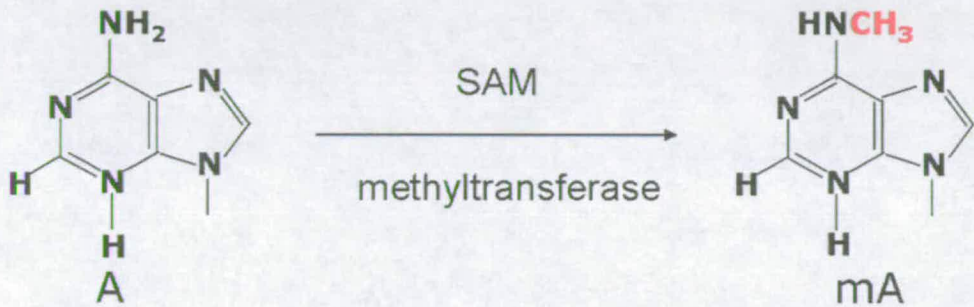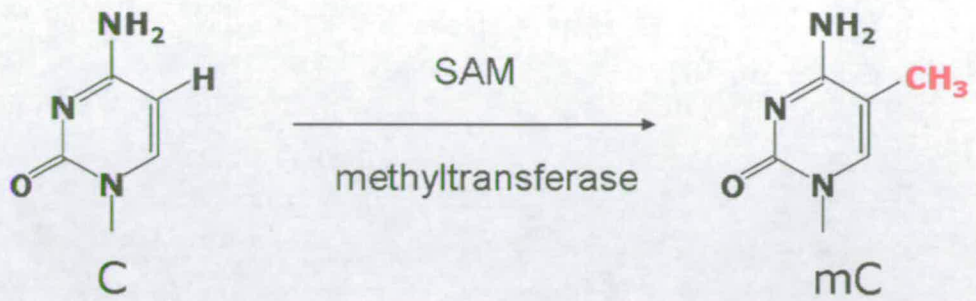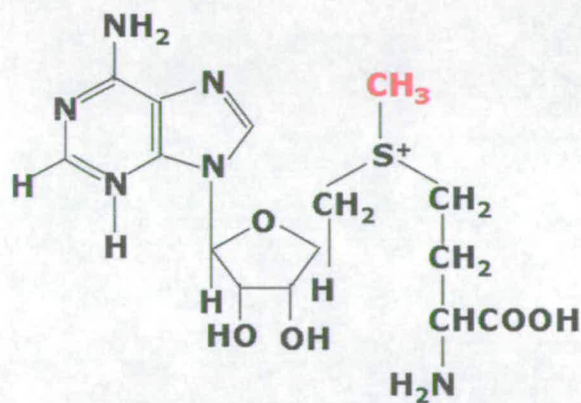
Figure 1.1: Occurrence of DNA methylation. (A) DNA in eukaryotes is normally methylated on cytosine at the C-5 position. Additionally, methylation occurs at the N-6 position of adenine in prokaryotes. The reaction consumes SAM and is carried out by methyltransferases. (B) Structure of SAM. The methyl group involved in methylation is labeled in red.

Although methylation at the 5′ position of cytosine is a widespread phenomenon ranging from mammals to bacteria, the level and pattern of methylation varies along the evolution tree. In prokaryotes, the methylation patterns are very strain- and species-specific. In eukaryotes, the lowest level occurs in *Caenorhabditis elegans*, which has no detectable cytosine methylation (Simpson et al., 1986). *Drosophila melanogaster* has a low level of 5-methylcytosine detected in the context of CpT, CpA and CpC, instead of the common CpG target in animals (Kunert et al., 2003; Lyko et al., 2000). Most invertebrate genomes are predominantly nonmethylated, but some are modified at an intermediate level with less than 50% methylation (Bird and Taggart, 1980). In the sea urchin, 10-40% of the genome is heavily methylated, while the remainder of the genome contains long stretches of nonmethylated DNA (Bird et al., 1979). At the other end of the scale, vertebrate genomes are globally methylated with 60-90% of CpG dinucleotides containing 5′-methylcytosine (5-mC; Singer et al., 1979) and the nonmethylated CpG fraction constitutes less than 2% of the genome (Cooper et al., 1983).

1.3 DNA methylation in eukaryotes

1.3.1 DNA (cytosine-5)-methyltransferases

Genomic methylation patterns are established and maintained by DNA (cytosine-5)-methyltransferases. When a CpG site that was previously unmethylated becomes methylated, this site is said to be methylated *de novo*. On the other hand, when methylation occurs at a hemimethylated site arising from events such as DNA replication, the methylation is called maintenance methylation, because it restores the symmetrically methylated state of the parental CpG. The reaction mechanism of
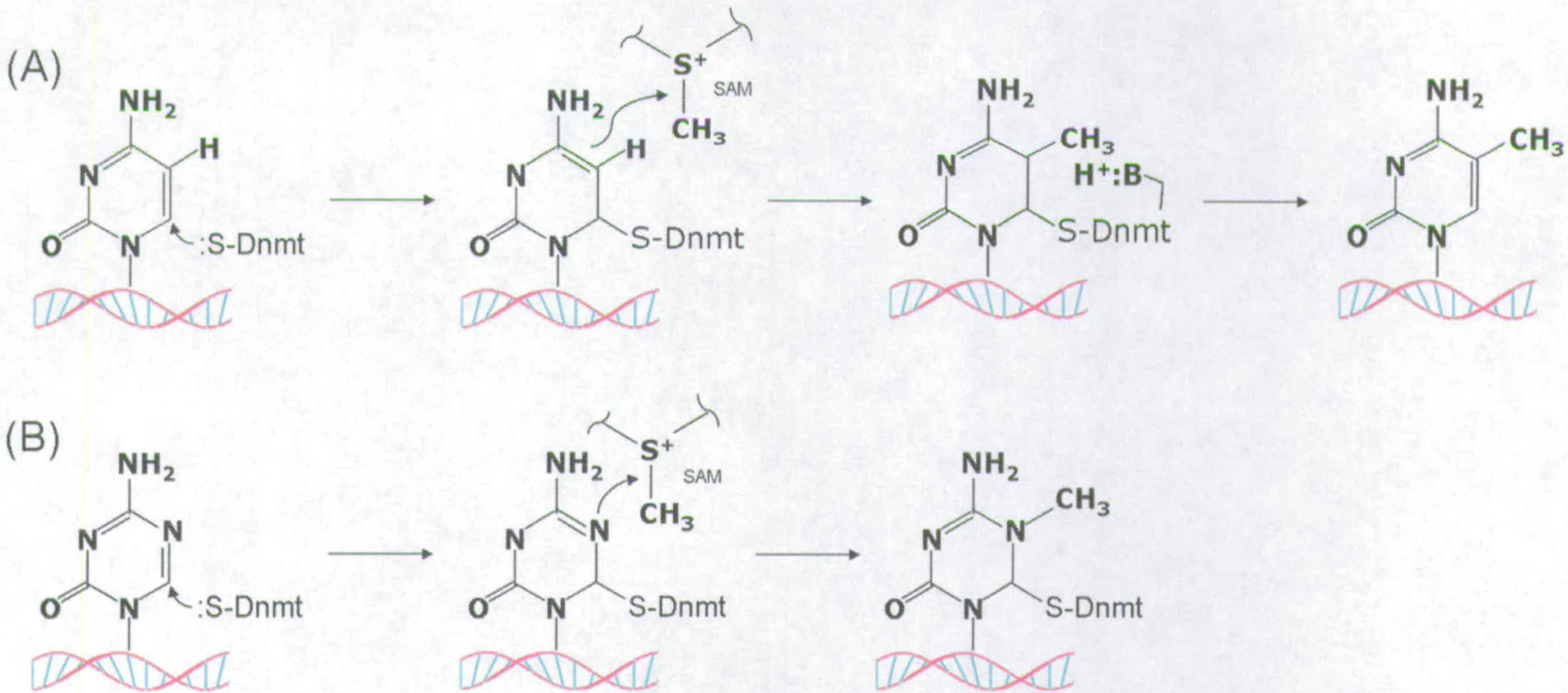
Figure 1. 2: Mechanisms of DNA methylation (A) catalysed by DNA methyltransferase and (B) inhibited by 5-azacytidine. (A) A thiolate residue of DNA methyltransferase attacks the cytosine ring at the C-6 position and opens up the double bond which attracts a methyl group from SAM. The proton then transfers from cytosine to Dnmt to allow double bond reformation, which also releases the enzyme. (B) When methylation inhibitors, such as 5-azacytidine, are incorporated into DNA, the replacement of C-5 by a nitrogen still allows the transfer of methyl residue from SAM to Dnmt but the consequent complex is a very stable complex, which is not in favour to reform the double bond. This engages the enzyme on DNA all the time and, therefore, inhibits DNA methylation.

the methyltransferase activity is depicted in Figure 1.2A. First, the enzyme forms an

cysteinyl thiolate with SAM at the catalytic site. This thiolate adds covalently to the

C-6 position of cytosine and follows by the protonation of the N-3 position creating

the 4,5 enamine, which in turn attacks the methyl group of SAM. Finally, abstraction

of the C-5 proton allows reformation of the 5,6 double bond and β-elimination to

release the enzyme, resulting in methylated cytosine (reviewed in Bestor, 2000b).

In mammals, the DNA methyltransferase (Dnmt) proteins are divided into three

families: Dnmt1, Dnmt2 and Dnmt3 (Figure 1.3). All the Dnmt proteins, except

Dnmt3L, are identified via a C-terminal catalytic domain, comprising six highly

conserved motifs (reviewed in Bestor, 2000a). Dnmt1 was the first mammalian Dnmt
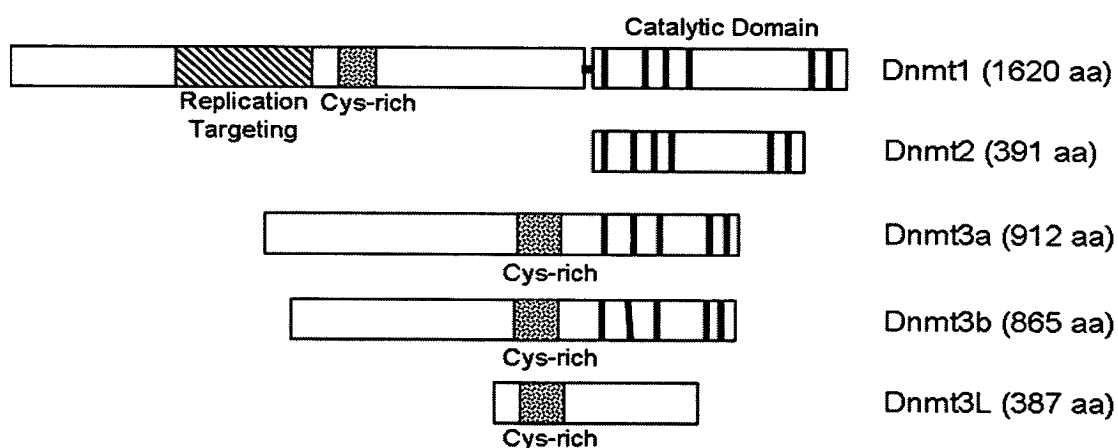
Figure 1.3: The general organisation of the five mammalian DNA methyltransferase family members. The catalytic domain resides in the C-terminal of all Dnmts, except Dnmt3L. The vertical bars inside the C-terminus correspond to the highly conserved motifs found in most DNA methyltransferases. The linker region between the N- and C-terminus of Dnmt1 is represented by a short thick horizontal line. The Cys–rich region of Dnmt1 includes three CXXC domains that are related to that of methyl-CpG binding protein 1 (MBD1), whereas the Cys-rich region of Dnmt3 proteins is an ATRX-like PHD domain. The N-terminus of Dnmt1 also includes a replication targeting sequence that is involved in the targeting of Dnmt1 to replication foci at S phase. The protein sizes given on the right are the sizes of mouse proteins.

to be identified, and is the major methyltransferase in somatic tissues (Bestor *et al.*, 1988). Dnmt1 has 5-to-30 fold preference for hemimethylated DNA when compared to non-methylated DNA (Bestor, 1992; Li et al., 1992; Yoder et al., 1997a). Dnmt1 was found to localise in the nucleus to DNA replication foci during S phase, through the replication targeting site in the N-terminus, and methylates hemimethylated daughter DNA to regenerate the symmetrical methylation status resembling parental DNA (Leonhardt *et al.*, 1992). This ability to maintain methylation status thus classified Dnmt1 as a maintenance methyltransferase.

Dnmt1 also showed little *de novo* activity *in vitro* (Bestor, 1992; Yoder et al., 1997a), though the fact that active Dnmt1 resides only in the cytoplasm of preimplantation embryos and specifically enters the nucleus shortly at the 8-cell stage suggests that Dnmt1 may play a role during the early developmental stage. Interestingly, the early embryonic Dnmt1 is actually a truncated variant of Dnmt1, named Dnmt1o, which expresses specifically in the oocytes and is soon replaced by full-length somatic Dnmt1 after implantation (Cardoso and Leonhardt, 1999; Mertineit et al., 1998; Ratnam et al., 2002). However, the biological significance of this variant trafficking is currently unknown.

In contrast to Dnmt1, little is known about Dnmt2. The Dnmt2 protein lacks the entire N-terminal regulatory region, retaining only the catalytic domain. Although Dnmt2 was initially shown to perform no methyltransferase activity in mammals (Okano et al., 1998b; Tweedie et al., 1997), a very weak activity (~5% of the activity of Dnmt1) of Dnmt2 was recently observed in mouse and human cell lines (Liu et al., 2003). Also, a Dnmt2-like protein was found in *D. melanogaster* and has been shown to mediate a weak but significant methyltransferase activity targeting CpT or CpA instead of the usual CpG context (Kunert et al., 2003; Lyko et al., 2000). However,

Dnmt2 appears dispensable for both *de novo* and maintenance methylation in mammals, because Dnmt2-deficient ES cells are normal in growth and morphology (Okano *et al.*, 1998b).

Dnmt3a and 3b were first identified by a sequence search of the dbEST database using bacterial cytosine methyltransferase as the enquiry input (Okano et al., 1998a; Yoder and Bestor, 1998). Although conserved in the methylase catalytic domain, neither Dnmt3a nor Dnmt3b contains the N-terminal regulatory region found in Dnmt1. Instead, the N-termini of Dnmt3 family members all have a cysteine-rich ATRX-like PHD domain, which is present in many chromatin-associated proteins. On the basis of homology to this ATRX-like domain, Dnmt3L (Dnmt3-like) was identified and included in this family. However, Dnmt3L lacks most of the C-terminal catalytic domain, which accounts for its lack of methyltransferase activity.

Dnmt3a and 3b are thought to be responsible for *de novo* DNA methylation, which is mainly observed in the postimplantation embryos and the primordial germ cells in the midgestation embryos. Both Dnmt3a and Dnmt3b have been shown to perform *de novo* methylation *in vivo*, but the *de novo* methylation activity of Dnmt3b seemed weaker than that of Dnmt3a (Hsieh, 1999b). Moreover, Dnmt3a- or Dnmt3b-deficient mouse ES cells and embryos showed little or no change of the *de novo* methylation pattern in retroviral DNA and major satellite DNA. A significantly reduced effect on *de novo* methylation activity only appeared when *Dnmt3a* and *Dnmt3b* were both deleted (Okano *et al.*, 1999), suggesting an overlapping function of the two *de novo* methyltransferases. Despite the overlapping function, Dnmt3b has a specific role in the methylation of centromeric minor satellite repeats. The *Dnmt3b* locus was mapped to chromosome 20q11.2, a locus associated with ICF

(immunodeficiency, centromeric instability, facial anomalies) syndrome, and several *Dnmt3b* mutations were found in ICF patients (Hansen et al., 1999; Okano et al., 1999). Thus DNMT3B becomes the first and only DNA methyltransferase identified in a human autosomal recessive disorder.

Furthermore, DNA methyltransferases are involved in transcriptional repression. Dnmt1 has been shown to interact with histone deacetylase 1 (HDAC1), HDAC2 and other co-repressors (Fuks et al., 2000; Robertson et al., 2000; Rountree et al., 2000). Dnmt3a and 3b co-localise with HP1-positive hetereochromatin, and both repress transcription via interaction with HDAC activity (Bachman et al., 2001). Dnmt3L has also been to act as a transcriptional repressor associated with HDAC1, and required HDAC activity for its repression activity (Deplus et al., 2002). This shows that the transcriptional repression activity of the Dnmt proteins can be independent of the Dnmt enzymatic activity since Dnmt3L lacks a complete catalytic domain.

Interestingly, there are accumulating data revealing interplay between the Dnmt family members. First, the N-terminus of Dnmt1 was found to interact with its C-terminus and can allosterically facilitate its methyltransferase activity (Fatemi *et al.*, 2001). Moreover, gene targeting of human DNMT1 alone only reduced global methylation to 80%, whereas a double knockout of DNMT1 and DNMT3B showed a marked decrease of overall genomic methylation to less than 1% in the human colorectal carcinoma cancer cell line HCT116 (Rhee et al., 2000; Rhee et al., 2002). These observations imply cooperation between Dnmt1 and Dnmt3b. Also, Dnmt3a was shown to have higher *de novo* methylation activity following incubation with Dnmt1, and therefore provides a link between Dnmt1 and Dnmt3a (Fatemi *et al.*, 2002). Another gene knockout study demonstrated that both Dnmt1 and Dnmt3a

and/or Dnmt3b were required for methylation of repeat sequences, such as LINE-1, in mouse ES cells, although Dnmt1 alone can perfectly maintain methylation of most CpG-poor sequences (Liang *et al.*, 2002). Finally, the DNA methyltransferase-like protein, Dnmt3L, has been shown in physical contact with other Dnmt3 family members (Hata et al., 2002). However, in an *in vitro* study, DNMT3L only stimulates *de novo* methylation carried out by DNMT3A, not by DNMT3B (Chedin et al., 2002). To date, Dnmt2 has not been shown to interact with any other Dnmt family member (Margot et al., 2003).

1.3.2 The methyl-CpG binding domain (MBD) proteins

In a search for proteins that can specifically bind to methylated DNA, MeCP1 and MeCP2 were first identified, but MeCP1 was shown to be a multisubunit complex (Meehan et al., 1989). The MBD domain of MeCP2 was mapped and used to identify the MBD protein family, which consists of MBD1-4 and MeCP2 (Figure 1.4; Nan et al., 1993).

Each of the MBD family members except MBD3 can bind to a single symmetrically methylated CpG, and localise to highly methylated DNA regions such as the major satellite DNA (Hendrich and Bird, 1998; Nan et al., 1993). All MBDs, except MBD4, are involved in transcriptional repression. MeCP2, MBD2 and MBD3 function as part of transcriptional repression complexes, which include either HDAC1 and/or HDAC2 (Nan et al., 1997; Ng et al., 1999). MeCP2 has a transcriptional repression domain (TRD) which associates with a corepressor complex containing transcriptional repressor Sin3 and a histone deacetylase (Nan et al., 1998). MBD2 does not contain a TRD domain but can similarly repress
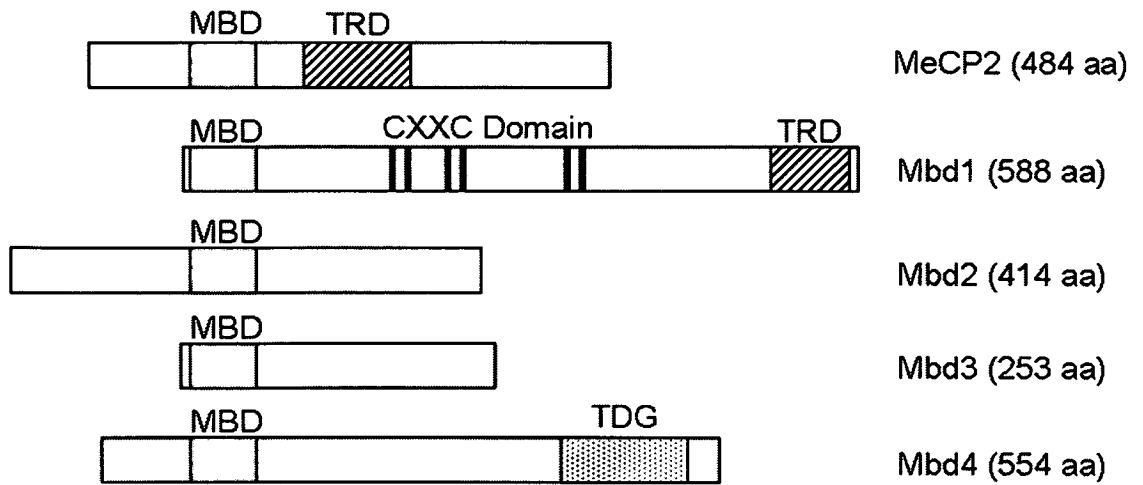
Figure 1.4: Overview of the MBD family members. The main feature of these proteins is the methyl-CpG binding domain (MBD) that characterises this protein family. Both MeCP2 and Mbd1 have a transcription repression domain (TRD) that can mediate transcriptional repression. The Cys-rich region of MBD1 (CXXC domain, horizontal bars) is shown to bind nonmethylated DNA so Mbd1 repression activity can work on not only methylated but also nonmethylated promoters. The thymine DNA glycosylase (TDG) domain of MBD4 is responsible for the T:G mismatch repair. Sizes of the mouse MBD proteins are listed.

transcription in a deacetylase-dependent manner. Mbd2 is also a component of the MeCP1 complex that includes HDAC1, HDAC2, Sin3a and histone binding protein RbAp48/46 (Ng et al., 1999). Although MBD3 does not bind to methylated DNA and contains no TRD motif, MBD3 still performs transcriptional repression as a part of a deacetylase complex, the Mi-2/NuRD complex, which has a nucleosome remodelling activity (Zhang et al., 1999). On the other hand, MBD1 represses transcription through its TRD domain regardless of HDAC activity, and has recently been shown to repress both methylated and non-methylated DNA via its MBD domain and its CXXC-3 domain, respectively (Jorgensen et al., 2004; Ng et al., 2000).

However, the precise roles of these proteins *in vivo* are not yet clear. Mice that lack MBD3 show embryonic lethality, despite the fact that MBD3 is the only MBD protein that does not bind to methylated DNA (Hendrich et al., 2001).

MeCP2-deficient mice are viable until several months after birth and display symptoms similar to human Rett Syndrome, a neurological disorder in which many MeCP2 mutations have been identified (Guy et al., 2001). *Mbd2*-null mice are also viable without inappropriate activation of endogenous imprinted genes or retroviral sequences, but the females exhibit defects in maternal behaviours, such as nurturing (Hendrich et al., 2001). The weak phenotypes of the MBD protein-deficient mice relative to the *Dnmt*-deficient mice suggest that the MBD proteins are functionally redundant.

Although MBD4 can bind specifically to methylated DNA, it prefers binding to a T:G mismatch both *in vivo* and *in vitro* (Hendrich et al., 1999a; Hendrich et al., 1999b). MBD4 has a helix-hairpin-helix domain similar to that of bacterial DNA glycosylase and can initiate base excision repair (BER) by excising thymidine from T:G mismatch in the context of CpG dinucleotides. T:G mismatches arise from the spontaneous deamination of 5-mC to thymidine and give rise to C:G to T:A mutations upon replication. *Mbd4$^{-/-}$* mice have an increased frequency of such transition mutations at CpG sites and show an increase in tumourigenesis on a tumour-prone background (Millar et al., 2002). Thus, while the other MBD family members mediate the biological readout from 5-mC, MBD4 acts to maintain this hypermutable base within the genome.

1.3.3 Roles of DNA methylation

Since the discovery of 5-methylcytosine, scientists have tried to elucidate the biological significance of DNA methylation. Though there is no definitive answer as yet, it is generally considered that DNA methylation acts to regulate gene expression,

X inactivation, genomic imprinting and mobile elements, via transcriptional repression.

In a heavily methylated mammalian genome, the appearance and location of CpG islands provided a base to investigate the role of DNA methylation and revealed its role in gene suppression. CpG islands are DNA domains of approximately 1kb in length, which have a high CG content and generally lack cytosine methylation (Bird, 1987). About 70-80% of total nonmethylated cytosines (about 1% bulk genome) are found in CpG islands, which are located at the 5' end of all housekeeping genes and many tissue-specific genes. Methylation of the 5' CpG islands represses expression of their associated genes, such as imprinted genes and genes on the inactive X chromosome (Monk, 2002; reviewed in Verona et al., 2003). This suggests a positive but not reciprocal link between DNA methylation and transcriptional repression, because CpG island promoters of tissue-specific genes (e.g. human alpha-globin and alpha 2(1) collagen, embryonic histone gene) are normally unmethylated in both expressing and non-expressing tissues (Bird et al., 1979; reviewed in Bird, 2002). Gene repression by CpG promoter methylation is also seen in cancer cells (reviewed in Jones, 2002). Promoter hypermethylation of tumour suppressor genes such as p16 represses their expression and promotes tumour proliferation (Gonzalez-Zulueta et al., 1995). In spite of CpG island hypermethylation, a general hypomethylation of total genomic DNA is also a common feature of tumourigenesis. Methylation changes prior to tumour formation suggest that DNA hypomethylation has a causal role in tumourigenesis (Di Croce et al., 2002; Eden et al., 2003; Gaudet et al., 2003). *In vitro*, methylation of plasmid DNA prior to introduction into tissue culture cells generally inhibits reporter gene expression. Experiments with episomes have shown that methylation of as few as

7% of CpG sites can effectively repression gene expression (Hsieh, 1994). Introduction of an *in vitro*-methylated DNA at two genomic sites by Cre-mediated genomic targeting resulted in the loss of reporter gene expression, whereas the reporter gene was properly expressed when an identical but unmethylated DNA was introduced (Schubeler et al., 2000). This methylated construct also brought about the localisation of hypoacetylated histone H3 and H4, which is consistent with observations that DNA methylation renders active chromatin to nuclease-insensitive structure (Kass et al., 1997; Keshet et al., 1986). It has also been shown that DNMT1 associates with HDAC1 and forms a complex together with tumour-suppressor protein Rb and sequence-specific transcriptional activator E2F1 (Robertson et al., 2000). The resulting complex repressed transcription from E2F-responsive promoters and improperly silenced CpG island-associated tumour suppressor gene. Apart from the mechanism directly involving Dnmt1, methylated DNA may also repress transcription by the impediment of the transcriptional factors binding their DNA recognition elements, such as CTCF binding in Igf2/H19 imprinting control region (Bell and Felsenfeld, 2000; Hark et al., 2000). On the other hand, methyl-CpG binding proteins, as discussed earlier, may act indirectly as mediators that bind methylated DNA and recruit HDAC to achieve transcriptional repression. Collectively, the dominant effect of DNA methylation represses transcription through the modulation of the protein environment, surrounding control elements, and chromatin structure of the gene.

Current knowledge concerning the role of DNA methylation in gene imprinting has largely come from studies involving gene targeting of DNA methyltransferase genes in mice. Imprinting is the phenomenon whereby an allele may have a different effect on the offspring depending on the sex of the contributing parent.

Dnmt1-null embryos suffered disturbance of gene imprinting and X inactivation, as several imprinted genes showed biallelic expression (Lei et al., 1996; Li et al., 1992) and transient inactivation of genes on X chromosomes is observed (Panning and Jaenisch, 1996). Also, these embryos do not survive beyond embryonic day (E) 9.5, which suggests Dnmt1 is essential for early development. Disruption of the oocyte-specific Dnmt1o, on the other hand, does not lead to complete genome-wide demethylation or embryonic lethality. However, mutated homozygous females fail to carry heterozygous offspring and some maternally imprinted genes appear demethylated, giving biallelic expression, which suggests a role of Dnmt1o in the maintenance of gene imprinting (Howell et al., 2001).

Deletion of the catalytic domain of both Dnmt3a and Dnmt3b disrupted *de novo* methylation in ES cells and embryos (Okano et al., 1999). In comparison to *Dnmt3a* and *3b* double knockout, mutation of *Dnmt3a* alone or *Dnmt3b* alone had a much weaker phenotype. However, neither of these disruptions was as severe as observed in *Dnmt1* null animals. *Dnmt3a$^{-/-}$* mice did not die until four weeks after birth, whereas *Dnmt3b$^{-/-}$* mice developed normally before E9.5 but were not viable at birth. Due to abnormal development, *Dnmt3a* and *3b* double knockouts showed abnormal morphology from E8.5 and could not survive longer than E11.5 (Okano *et al.*, 1999). In the ES cells of these double, but not single, knockout mice, the differentially methylated region of *Igf2* and the 5′ region of *Xist* did not undergo *de novo* methylation after differentiation *in vitro* (Okano *et al.*, 1999). In addition, [*Dnmt3a$^{-/-}$*, *Dnmt3b$^{+/-}$*] embryos derived from ovary transplantation lose methylation at the differentially methylated regions of imprinted genes such as *Igf2r*, *Peg1*, *Peg3* and *Snrpn* (Hata et al., 2002). Male mice lacking Dnmt3L, on the other hand, were viable but infertile due to defects in spermatogenesis, while *Dnmt3L$^{-/-}$* females could no

longer establish maternal imprinting nor produce viable offspring, due to aberrant acquisition of genomic methylation during oogenesis (Bourc'his et al., 2001b; Hata et al., 2002). These observations all indicate that DNA methylation plays a role in early development and in the maintenance and establishment of imprinted genes.

During early embryogenesis, one of the two X chromosomes in female mammalian cells such as human and mouse is inactivated to achieve an equivalent dosage from X-linked genes between male and female cells. The association between methylation and X inactivation is supported by several observations. The methylation inhibitor, 5-azadeoxycytidine (5-azadC), has been used to derepress X-linked genes inactivated on inactive X chromosome (Xi; Mohandas et al., 1981), and the methylation of CpG island of hypoxanthine phosphoribosyltransferase (HPRT) gene follows X inactivation by several days (Lock et al., 1987). Also, *Dnmt1*-deficient male embryos are hypomethylated at the 5' end of the *Xist* (X inactive specific transcript) locus, which induces *Xist* expression and the repression of X-linked genes (Beard et al., 1995). CpG islands of X-linked genes are usually heavily methylated on the Xi but barely methylated on the active X, whereas the promoter of the *Xist* gene works in the totally opposite way, being methylated on active X chromosome and hypomethylated on the Xi (Vasques et al., 2002). However, methylation of *Xist* on the future inactive X does not seem to be essential for the initiation of X inactivation since DNA methylation happens several days after X inactivation (Lock et al., 1987). This suggests DNA methylation plays a downstream, maintenance role in the process of X inactivation. Nonetheless, a study of *Dnmt1*-knockout mice indicated that the zygotic function of Dnmt1 may not be essential for the establishment of X inactivation in the extraembryonic tissues while imprinted genes of the same pool showed aberrant expression (Sado et al., 2000). On

the other hand, the embryonic lineages of Dnmt1 mutant embryos do not have stable X inactivation which consistently suggests DNA methylation may have a maintenance role (Sado *et al.*, 2000). Moreover, ICF patients have hypomethylated CpG islands on the inactive X, which could explain the instability of silencing on Xi (Hansen et al., 2000). Since ICF is caused by *DNMT3B* mutation, this suggests another link between methylation and X inactivation. However, a recent report seems to refute this idea because *Dnmt3a-* and *3b*-deficient mice do not show abnormal X inactivation. Expression of *Xist* was not affected in these mice either, even though they do show highly demethylated promoter regions of both *Xist* and *Hprt* (Sado et al., 2004). Still, hypomethylated DNA in Dnmt1 mutant embryos and embryonic cells result in aberrant *Xist* expression and subsequent cell death (Panning and Jaenisch, 1996). Besides, there is synergistic effect of DNA methylation and histone deacetylation on *Xist* reactivation, because *Xist* reactivation is observed in cells simultaneously treated with the deacetylation inihibitor, Trichostatin A (TSA), and the methylation inhibitor, 5-azadC (Csankovszki et al., 2001; Keohane et al., 1996). From these lines of evidence, DNA methylation is also important to correctly inactivate X chromosome.

One of the proposed roles of DNA methylation is the suppression of parasitic sequences, such as transposable elements and proviral DNA (Yoder et al., 1997b). Cytosine methylation inactivates the promoters of most viruses and transposons (Harbers et al., 1981; Schmid, 1996). The repression of LINE and SINE retrotransposon promoters is disrupted in the human genome when DNA methylation is reduced (Liu et al., 1994; Woodcock et al., 1997) and the transcription of intracisternal A particle (IAP) elements is massively induced in mouse embryos lacking Dnmt1 (Walsh et al., 1998). Also, the majority of 5-mC content in the

mammalian genome occurs in these parasitic sequences (Bestor et al., 1984; Yoder et al., 1997b). Intriguingly, the nonmethylated genomes of *D. melanogaster* and *C. elegans* do harbor transposons, such as P elements. Besides, mosaic methylation of the *C. intestinalis* genome showed that multiple copies of transposons and most interspersed elements were free of methylation, whereas three-quarters of the genes were methylated along with some of the short interspersed elements (Simmen et al., 1999). This suggests that methylation in the invertebrate genome is not targeted to silence transposable elements. In mammals, only a small fraction of these elements remain active in the heavily methylated genome, which cannot count as a fair base for the genome defense theory. It has also been reported that some promoters of these elements can be activated without altering DNA methylation (Chu et al., 1998; Liu et al., 1995), and several retrotransposons are reportedly nonmethylated in mammalian germ cells and early embryos, where protection against transposition is thought to be important (reviewed in Bird, 1997). Therefore, the hypothesis that transposon silencing by methylation is a part of the genome defense system remains debatable.

1.4 DNA methylation in prokaryotes

1.4.1 Prokaryotic DNA methylation

DNA methylation is also widespread among prokaryotes and occurs at adenine and at cytosine, giving the possible products of methylation as N6-methyladenine and C5-methylcytosine (Figure 1.1). It has been suggested that DNA methylation in prokaryotes may play a role in DNA mismatch repair, modulation of gene expression and coordinating DNA replication initiation, chromosome partition and

cell division (reviewed in Palmer and Marinus, 1994). In the best studied system of the bacterium *Escherichia coli*, cytosine C-5 methylation is carried out by the *dcm* gene and methylates the internal cytosine in the sequence CC(A/T)GG. On the other hand, the methylase encoded by the *dam* gene is in charge of adenine methylation in the sequence context of GATC. However, both *dcm⁻* and *dam⁻* cells are viable with minor defects such as susceptibility to restriction (Kruger et al., 1989; Urieli-Shoval et al., 1983). Moreover, the laboratory strain *E. coli* B is actually *dcm* free by nature (Marinus, 1984). Since there is a lack of phenotypic evidence, the biological significance of cytosine methylation in *E. coli*, therefore, remains speculative.

One important role of DNA methylation in prokaryotes is to protect DNA against corresponding restriction endonucleases (reviewed in Dryden et al., 2001). Both adenine N6- and cytosine C5-methylation can arise as a result of the restriction-modification (R-M) mechanism, which is achieved by the methylase part of various R-M enzymes, such as EcoKI. R-M enzymes are divided into several classes and each utilises a slightly different pathway to perform host restriction-modification defence. Generally speaking, all the R-M enzymes comprise an endonuclease activity and a methylase activity, sharing the same specificity for its target DNA sequence. The methyltransferase part of an R-M system modifies the host DNA, so it cannot be cleaved by the corresponding endonuclease. Upon phage invasion, foreign DNA entering the host without carrying specific modification markers will be digested by the bacterial endonuclease. In this way, the adenine methylation marks self from non-self DNA, thus protecting the bacteria from bacteriophage infection (reviewed in Dryden et al., 2001).

1.4.2 Phage-bacteria antagonism

To counteract the bacterial immune system, phages have anti-restriction mechanisms that can resist bacterial discrimination, allowing them to successfully invade bacteria using three general strategies. Firstly, particular bases in the specific sequence recognised by the endonuclease may be methylated by a corresponding modification by DNA methyltransferase, rendering phage DNA resistant to restriction. Secondly, when the host nuclease specifically requires a modified sequence, the absence of methylation in phage genome provides protection from bacterial nuclease. Thirdly, they utilise an unusual base to substitute one of the usual four throughout the genome. The third strategy provides virtually unlimited protection against nucleases. T-even phages such as T2 and T4 take up 5′-hydroxymethylcytosine (5-HmC) from bacteria, to avoid nucleases that recognize unmodified DNA and then modify HmC by glucosylation, which subsequently protects phage DNA from nucleases that recognise specific modification.

The closely-related bacteriophage T7 and T3 take an entirely different approach in addition to the first strategy, which collectively allow them to grow interchangeably on a range of *E. coli* strains. By recruiting gene 0.3 to overcome host restriction, they grow unaffected by the DNA restriction system. This early gene is also named overcome host restriction (ocr) for its function to overcome host restriction. The ocr protein binds to restriction endonucleases and consequently block their restriction activity (Kruger et al., 1983). In addition to the ocr function, bacteriophage T3 has developed a unique property to strengthen its anti-restriction mechanism. Another early enzymatic activity of T3 was first detected in infected *E. coli* B (Gefter et al., 1966). This activity catalyses the hydrolysis of SAM, the methyl

donor, into homoserine and methyl-thio-adenosine (MTA; Figure 1.5). The breakdown of SAM facilitates R-M enzyme inhibition, because SAM is a cofactor required to activate type I R-M enzymes in *E.coli* B and K strains. Both ocr and the SAM cleaving activities are early expressed in T3, and it has been shown that the SAM hydrolysis activity and the ocr activity actually reside on the same peptide, later named S-adenosylmethionine hydrolase (SAMase, E.C. 3.3.1.2; Hughes et al., 1987a; Spoerel et al., 1979; Studier and Movva, 1976). However, the ability to hydrolyze SAM appears not to be the primary reason for anti-restriction, as a mutant that has lost SAMase activity is still capable of overcoming host restriction (Studier and Movva, 1976). Also, the impairment of host restriction by SAMase, instead of being catalytic, is a stoichiometric mechanism and the interaction between SAMase and endonuclease is reversible (Spoerel et al., 1979). This beautifully designed dual system has drawn further investigation on SAMase, especially the unique property of SAM destruction.

S-adenosylmethionine   Methyl-thio-adenosine   homoserine
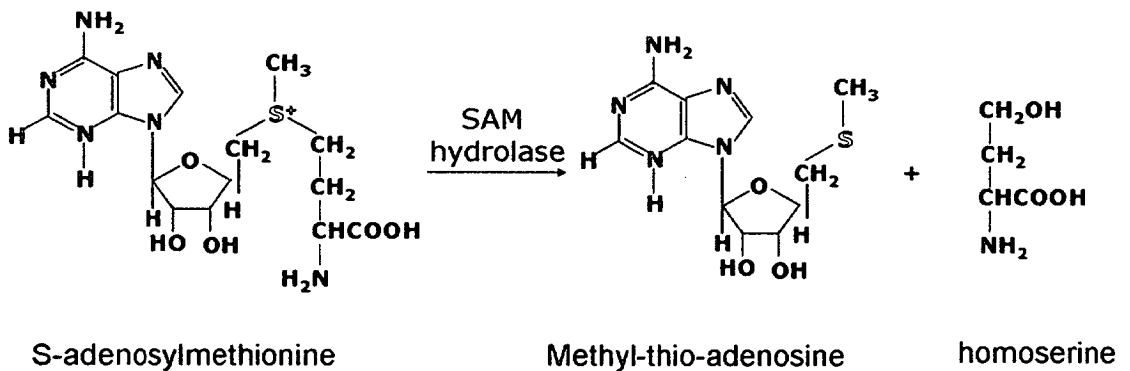
Figure 1.5: SAM hydrolysis by SAMase. S-adenosylemthionine is cleaved by SAMase into homoserine and methyl-thio-adenosine. This is a special reaction only found *in vivo* in bacteriophage T3. Enzyme reaction occurs at the site shown in red.

1.5 DNA demethylation

1.5.1 Demethylation *in vivo*

The reversal of DNA methylation is termed DNA demethylation. Two mechanisms of DNA demethylation have been observed *in vivo*. The more common form is termed passive DNA demethylation and occurs upon DNA replication. If DNA methyltransferases cannot access methylatable sites on a newly synthesised, unmethylated daughter strand, DNA methylation is lost over a number of cell divisions. Binding of specific proteins on nascent DNA, such as Sp1 proteins, in postimplantation development and in germinal tissues may protect DNA from methylation (Chesnokov and Schmid, 1995; Macleod et al., 1994). Protein association specifically required for DNA demethylation of, at least, the first strand during replication has been shown (Hsieh, 1999a). This is likely to hinder subsequent *de novo* methylation by Dnmt3a and result in DNA demethylation (Han et al., 2001; Hsieh, 1999a; Lin et al., 2000). Since unmethylated sites on hemimethylated daughter strands are only possible during replication as a result of the failure in maintenance methylation, this demethylation is therefore replication dependent and termed passive demethylation.

DNA demethylation can also proceed in a replication-independent style, termed active demethylation. Since the removal of the stably-formed C-C bonding is not thermodynamically favourable, this reaction is unlikely to occur without enzyme catalysis. Active demethylation has been reported on several occasions, either upon cell differentiation or under the control of hormones. For examples, the δ-crystallin gene was hypomethylated during development of the chicken lens, where a large fraction of cells in the lens stopped dividing (Sullivan and Grainger,

1987). In the avian vitellogenin gene, three CpG sites underwent demethylation in a two-step fashion after the hormonal activation by estradiol, while the fourth site became demethylated from a hemimethylated strand after glucocorticoid induction. Both sets of CpGs remained demethylated even after cessation of gene transcription and their occurrence was independent of replication (Saluz et al., 1986). In L8 myoblasts, the α-actin gene underwent full demethylation on an unintegrated pre-labelled plasmid by the aid of a cis-acting sequence in the upstream regulatory region (Paroush et al., 1990). Naïve T cells also showed site-specific demethylation in the promoter-enhancer of the interleukin-2 (IL-2) gene. Here, demethylation was observed as early as 20 minutes after stimulation at selective sites. Demethylation preceded DNA replication which occured hours later, and it remained stably demethylated up to nine days after stimulation (Bruniquel and Schwartz, 2003). In addition, mouse erythroleukemia cells underwent genome-wide transient demethylation in response to the treatment with hexamethylene bisacetamide (HMBA), when the DNA was not yet replicating in the cell cycle (Razin et al., 1986).

1.5.2 Demethylation during development

During development, the process of DNA demethylation was first observed when a net loss of methylcytosine was found in early embryos and germ cell lineages (Monk et al., 1987). Using methylation-sensitive digestion and hybridization techniques, it was observed that DNA from oocytes was undermethylated whereas sperm DNA was relatively methylated just before fertilization. However, the sperm genome underwent demethylation during cleavage because DNA in blastocyst stage has very little methylation. From E6.5 onwards, fetal DNA methylation level

increases again (Howlett and Reik, 1991; Rougier et al., 1998). Individual DNA sequences, such as *Pgk2* gene and satellite DNA, were examined and all displayed similar, but slightly different, methylation patterns in gametes (Howlett and Reik, 1991; Kafri et al., 1992). However, most of the above results point to the direction of passive demethylation as a mechanism for these events, which coincided with the replication time of embryos, because the cells under examination are either later than one-cell stage or tested by using whole cell DNA, in which the differential demethylation between two parental genomes was not distinguishable.

The first striking evidence of a *bona fide* active demethylation was published in 2000. Using indirect immunofluorescence staining against 5-methylcytosine, Mayer *et al.* (2000a) demonstrated that a high methylation level was equally maintained 3-6 hours after fertilization, but only the paternal genome underwent rapid loss of 5'-methylcytosine within 6-8 hours after fertilization in the one-cell embryo. However, the first replication event does not occur until the S phase of the two-cell embryo. Besides, the maternal genome only experienced gradual and genome-wide demethylation after the two-cell embryo stage, which was in accordance with cell division (Mayer et al., 2000a; Rougier et al., 1998). The pattern of maternal demethylation is consistent with passive demethylation as a result of lacking maintenance methylation in the genome following DNA replication and cell division. Bisulfite sequencing of sperm, oocyte and early embryo also revealed similar results, showing that highly methylated sperm was rapidly demethylated soon after fertilization and the maternal genome either remains methylated, or became further *de novo* methylated after fertilization (Oswald et al., 2000). However, the exact timing of remethylation is still unclear. Slightly contradictory to the observation made by Mayer *et al.* (2000a), a later publication showed that the

process of paternal demethylation took only 90-120 minutes and was complete within 4-5 hours after *in vitro* fertilization (Santos et al., 2002). In addition, the reduction of genome-wide methylation in the whole embryo, except imprinted genes, recorded in most literature is complete at the blastocyst stage and remethylation begins soon after implantation (Howlett and Reik, 1991; Mayer et al., 2000a; Monk et al., 1987), whereas a demonstration of demethylation completing at morula stage contrasts the above statement (Santos et al., 2002). This publication also showed that the first *de novo* methylation was detected only in the inner cell mass of blastocyst, which suggests the re-establishment of methylation patterns begins in preimplantation embryos and is lineage-specific. Whichever is the correct timing, the active demethylation of the paternal pronucleus in the non-replicating embryo is undoubtedly true.

During mammalian development, another active demethylation event occurs when primordial germ cells (PGCs) enter the genital ridge. PGCs are cells derived after acquiring genome-wide *de novo* methylation from E6.5. These cells later proliferate and migrate towards the genital ridge between 10.5 and 11.5 day postcoitum (dpc) with a high methylation level. From 12.5 dpc, methylation marks in PGCs, including those on repetitive sequences, non-imprinted and imprinted genes, are all removed. Even the X chromosome, which originally undergoes random inactivation, is demethylated while the counterpart in somatic tissues retains its methylation status (Hajkova et al., 2002; Howlett and Reik, 1991; Kafri et al., 1992; Monk et al., 1987). This wave of demethylation peaks around 13.5 dpc. Although some repetitive sequences are not erased completely, the imprinting marks and the majority of the DNA obtain an equal epigenetic state prior to the differentiation of the definitive male and female germ cells. Once this ground zero is

established, *de novo* methylation is again carried out in germ cells, starting from day 15.5, and is completed around 18.5 dpc. Interestingly, embryonic germ cells, which are derived from PGCs, are able to demethylate foreign, somatic DNA when fused with thymic lymphocytes (Beaujean et al., 2004; Hajkova et al., 2002; Lee et al., 2002; Tada et al., 1997). This suggests that the demethylation ability is carried in the cell itself and, unlike the demethylation in early preimplantation embryos, does not require a sense of parental origin.

The precise role of demethylation in early embryos and of some tissue-specific genes is not yet clear. First of all, the genome-wide demethylation of paternal pronuclei does not seem to be a prerequisite for mammalian development, because it appears common only among mouse, rat, pig, and human but not in species like zebrafish, rabbit and sheep, whereas an intermediate level of demethylation is observed in the cow (Beaujean et al., 2004; Dean et al., 2001; Macleod et al., 1999). Furthermore, both successful and unsuccessful rapid demethylation in one-cell embryos has been observed in cloned animals, but the overall reprogramming of methylation pattern is aberrant (Bourc'his et al., 2001a; Dean et al., 2001; Lee et al., 2002). In manipulated embryos, the pronuclei of paternal origin are always subjected to active demethylation while those of maternal origin are always methylated, regardless of the ploidy of the embryos (Barton et al., 2001; Mayer et al., 2000b). Polyspermic embryos containing up to four sperm pronuclei underwent full demethylation of all male pronuclei (Santos et al., 2002). Interestingly, embryos with abnormal demethylation obtained from normal matings actually resemble these manipulated embryos (Dean et al., 2001; Shi and Haaf, 2002). These observations suggest that asymmetric methylation between the maternal and paternal pronuclei of the early embryos is functionally important, perhaps in the epigenetic

reprogramming for future somatic development.

1.5.3 Demethylation driven by external forces

Prior to the application of gene-targeting technology, drug treatment of cells was the only available means of reducing DNA methylation levels in cells. The cytidine analogs such as 5-azacytidine (5-azaC) and 5-aza-2'-deoxycytidine (5-azadC) (Figure. 1.6) are the most widely used and best studied DNA methylation inhibitors. Yet, analogs such as 5-6-dihydro-5-azacytidine and 6-azacytidine have little or no effect at all (Jones and Taylor, 1980). The mechanism by which these demethylation agents function is through their covalent binding to DNA methyltransferases. When the normal deoxycytidine residue in DNA is replaced by azacytidine analogs, the cysteine thiolate of the enzyme binds to the C-6 of 5-azadC and forms a stable covalent bond which no longer allows DNA methyltransferases to break free (Figure. 1.2B). This irreversible binding of azacytidine traps the DNA methyltransferases, decreases the cellular concentration of the enzymes and thus leads to subsequent passive demethylation of genomic DNA. Supportive evidence for this mechanism is that Dnmt1 deficient cells appear more resistant to the toxicity of 5-azadeoxycytidine treatment than wild-type or heterozygous mutant cells (Juttermann et al., 1994).

5-azaC-induced DNA hypomethylation has an established link with changes in gene expression and cellular differentiation. *Hprt* gene on the inactive X chromosome, tyrosine aminotransferase gene in fetal rat liver, adenosine deaminase gene in adenosine deaminase-deficient mouse cells are all examples of genes reactivated by treatments of 5-azaC analogs (reviewed in Haaf, 1995). There are also
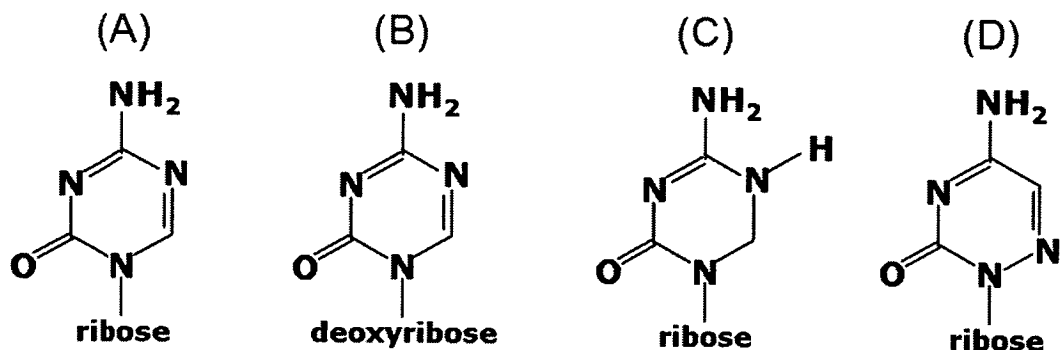
Figure 1.6: DNA methylation inhibitors and non-inhibitor analogs. (A) 5-azacytidine. (B) 5-azadeoxycytidine. (C) 5-6-dihydro-5-azacytidine (D) 6-azacytidine.

observations on the induction of endogenous and exogenous viral gene expression in the host genome. Injection of 5-azaC into mice results in active trancription of previously inactive viral genes (Jaenisch et al., 1985). Besides, 5-azaC treatment of undifferentiated cells can induce blood cell differentiation in human and mouse cell lines, or liver differentiation in fetal and neonatal rats (Haaf, 1995). Such differentiation may result from the activation of several related genes of that particular differentiation pathway. Along the same line, tumor suppressor genes, like *p15* and *p73* of acute myeloid leukemia, have also been reported to undergo gene activation upon treatment with these drugs (Cameron et al., 1999; Kawano et al., 1999). In fact, 5-azaC analogs were originally utilised as potent anticancer agents for certain types of tumors such as acute nonlymphoblastic leukemia (Charache et al., 1983; Ley et al., 1982). However, they cannot act as a general anticancer drug because several carcinogens are actually known to induce hypomethylation. 5-azaC actually promotes tumors in rat, mice and human, and cells exposed to 5-aza-C and analogs have increased tumor aggressiveness (Haaf, 1995). Moreover, a recent publication also shows that *Dnmt1*-deficient mice have a higher rate of tumourigenesis (Eden et al., 2003; Gaudet et al., 2003).

It is difficult to say whether demethylation induced by 5-azaC is responsible for changes in gene activation and cell differentiation, since other direct effects of 5-azaC analogs on chromosome stability, replication timing and tumourigenesis raise doubts about results obtained from 5-azaC-analog treated cells. Treatments with azacytidine analogs are usually at a concentration of 10^{-5} to 10^{-4} M. With such treatment, cells have shown a temporal shift in the timing of DNA replication (reviewed in Haaf, 1995). Constitutive heterochromatin and the inactive X chromosome change their late replication time forward to an earlier time-point after a pulse treatment with 5-azaC (Haaf and Schmid, 1989; Jablonka et al., 1985). Chromosomal instability has also been shown in 5-azaC treated cells. A low dose of pulse treatment (0.1-10 μM for a few hours) is sufficient to induce the formation of both single-strand and double-strand breaks in DNA, decondensation of heterochromatin, dose-dependent increase of sister chromatid exchanges and gene amplification (Haaf, 1995; Jones et al., 1982; Jones and Taylor, 1980). The latter two changes, sister chromatid exchanges and gene amplification, appear to be heritable even in the absence of continuous drug treatment. In addition, strongly increased frequency of endoreplication is observed during metaphase (Poot et al., 1990). It is not clear if chromosomal decondensation promotes subsequent chromosome instability or, conversely, the instability from DNA breaks results in chromosome decondensation and consequently brings about translocations, deletions and micronuclei formation during the following cell cycles.

5-azaC and 5-azadC are also cytotoxic to cells as a result of incorporation into ribosomal RNA (rRNA) and transfer RNA (tRNA), which inhibits the processing and methylation, respectively (Haaf, 1995). They also interfere with synthesis and methylation of nuclear and cytoplasmic rRNA and receptor activity, which

consequently induces disassembly of polyribosomes. 5-azaC modification of rRNA and tRNA also blocks protein synthesis. In contrast, 5-azadC exerts its cytotoxicity mostly on DNA. Spontaneous degradation of the incorporated analog may result in DNA damage. 5-azadC is also a more potent DNA strand breaker and inhibitor of chromosome condensation, although it is approximately 5- to 10-fold more effective in gene induction (reviewed in Haaf, 1995).

Other than pharmaceutical methods and gene knockout technology, an additional tool for reducing DNA methylation is derived from the phenomenon of post-transcriptional gene silencing. When double-stranded RNA (dsRNA) is produced after transcription, RNA interference (RNAi) is initiated to knock down the expression of specific genes. This mechanism occurs in various organisms, ranging from plant, fungi, yeast, Drosophila to mammals and the underlying idea is that dsRNA leads to the destruction of nascent mRNA (reviewed in Dorsett and Tuschl, 2004a). Long double-stranded RNAs and microRNAs (miRNA) are produced endogenously in cells and processed by Dicer, a RNA nuclease that cleaves long dsRNA into short fragments of 21-28 nucleotides. These short double-stranded RNAs are called small interfering RNA (siRNA). siRNA then binds to the RNA-induced silencing complex (RISC) and is guided to a complementary RNA target. This type of gene silencing is highly specific, even one nucleotide difference within the siRNA sequence is able to prevent siRNA targeting. However, not all positions of a siRNA contribute equally to the efficiency of target recognition (Elbashir et al., 2001a). Mismatches in the center of the siRNA duplex are most critical and essentially abolish target RNA cleavage. Therefore, selection of the target sequences is a trial-and-error process (reviewed in Elbashir et al., 2002).

There has been some publication using siRNA targeting Dnmt proteins (Leu et al., 2003; Robert et al., 2003). Down-regulation of both Dnmt1 and Dnmt3b using siRNAs in the ovarian cancer cell line CP70 caused a greater reduction of DNA methylation than single siRNA treatment. These siRNA-treated cells also partially removed DNA methylation from three inactive promoter CpG islands, *TWIST*, *RASSF1A*, and *HIN-1*, and restored the expression of these genes (Leu et al., 2003). This finding supports the cooperative relationships between these two enzymes and sets an example of siRNA disruption in DNA methylation. Further investigation using siRNA certainly is worth the effort.

1.6 Proposed mechanisms of active demethylation

Other than the obvious cause of demethylation by inhibition of the maintenance and *de novo* DNA methyltransferases during replication, a major mechanism which can explain active demethylation and/or fill in details of passive demethylation is still unavailable. The loss of methylation has been suggested to proceed by either site-specific DNA binding proteins, by triggering site-specific repair activity, by site-specific nucleotide replacement, or by direct removal of the methyl group on methylcytosine.

1.6.1 Direct demethylation by a DNA demethylase

The existence of an enzyme capable of recognising methylcytosine residues and removing the methyl group catalytically has been postulated for some time. MBD2b, a truncated form of MBD2, has been proposed to be a candidate demethylase carrying out DNA demethylation (Bhattacharya et al., 1999). This protein was

identified by TBLASTN search for sequences homologous to the MBD domain of MeCP2 (Hendrich and Bird, 1998). The resulting cDNA, which encodes a protein previously identified as MBD2b (Hendrich and Bird, 1998), was cloned from HeLa cells and *in vitro* translated to test for its putative demethylase activity (Bhattacharya et al., 1999). In this paper, the authors found this 40kDa protein was able to perform such a task with selectivity for binding to DNA containing symmetrically methylated CpG dinucleotides. The same enzyme activity was also detected and partially purified from human lung cancer cell line A594, and the reaction did not release single nucleotide, phosphorylated base, or phosphate as reaction products, which suggested this enzyme was a direct demethylase that transforms methylcytosine to cytosine without disrupting the integrity of the DNA substrate (Bhattacharya et al., 1999). Subsequent reports claimed this demethylase was a processive enzyme that demethylates continuously along one DNA strand (Bhattacharya et al., 1999). Also, this demethylase activity was reported to be directed by histone acetylation, especially at the promoter region, and was responsible for demethylation induced by valproic acid (Cervoni et al., 1999).

However, these demethylase properties appear paradoxical. First, the removal of a methyl group from methylcytosine involves the breaking of a C-C bond. Thermodynamically, this reaction requires high activation energy and is thus chemically unfavourable. To explain this, the authors proposed that participation of water molecules might allow the reaction to become thermodynamically competent by forming the reaction byproduct of methanol, which involves the breaking of O-H and C-C bonds and the subsequent formation of C-O and C-H bonds. Although a "quick calculation" was made to show such possibility, no further evidence has yet been provided to support such property in this demethylase (Detich et al., 2003). In

addition, MBD2 in HeLa cells is demonstrated to be associated with histone deacetylase (HDAC) in the MeCP1 repressor complex and represses transcription, which is relieved by the treatment of deacetylase inhibitor, TSA (Feng and Zhang, 2001; Ng et al., 1999). This suggests that, totally opposite to the demethylase activity, MBD2 acts as a transcriptional repressor after binding to the methylated DNA, instead of removing DNA methylation and resulting in transcription activation. Besides, attempts to reproduce such activity have not been successful in other systems (Boeke et al., 2000; Ng et al., 1999; Wade et al., 1999). Above all, MBD2-deficient mice are viable, and the paternal demethylation, which is an important active demethylation event in early mouse embryos, progresses normally in one-cell *Mbd2*[-/-] mouse embryos (Hendrich et al., 2001; Santos et al., 2002). Therefore, MBD2b is not a likely participant in the demethylation reaction.

1.6.2 Indirect demethylation by base excision repair

Another possible mechanism of active DNA demethylation is the removal of part or the entire methylated nucleotide. This has been reported in Friend mouse erythroleukemia cells, where genome-wide demethylation occurs rapidly within 24 hours after induction by hexamethylene bisacetamide (Razin et al., 1986). Replacement of 5-mC by radioactively labelled cytosine is transiently detected by radioactively-labelled two-dimension thin layer chromatography (2D-TLC), in which the [H^3]-cytosine signal correlates with the peak of demethylation. As nonreplicating DNA incorporates only [H^3]-cytosine, not [H^3]-adenosine, it is likely that this replacement occurs at cytosine residues only, instead of a patch of DNA (Razin et al., 1986). Later, this excision demethylation activity was reported both in

HeLa cell extract and in chicken embryo extract (Jost, 1993; Vairapandi and Duker, 1993). In HeLa nuclear extract, the decrease of the 5-mC content correlates with the stoichiometrical generation of apyrimidinic sites (Vairapandi and Duker, 1993). The recovery of thymines as released pyrimidines indicates a base excision pathway involving glycosylase, endonuclase and deaminase activities followed by repair. In chicken, 12-day embryo extract was used to demonstrate an excision-repair activity removing methylated base by nicking strictly at methyl CpG sites, regardless of the sequence context (Jost, 1993). Although no methylcytosine glycosylase activity was detected at that time, such activity and enzyme were later observed and purified (Jost et al., 1995). This 5-methylcytosine DNA glycosylase (5-MCDG) activity has a preference toward hemi- and non-methylated DNA, in contrast to the substrate in HeLa cells, which prefer fully methylated DNA (Jost, 1993; Vairapandi and Duker, 1993). Also, the chicken 5-MCDG performs its enzyme activity in combination with AP nuclease, and copurifies with mismatch-specific thymine DNA glycosylase activity. Furthermore, this activity needs additional RNA complementary to the sequence subject for demethylation (Fremont et al., 1997; Jost et al., 1997). RNA has been suggested to play an active role as an acceptor of the entire 5-mC nucleotide in an *in vitro* demethylation assay, where demethylase activity in rat L8 myoblsts was sensitive to RNase but insensitive to protease K (Weiss et al., 1996). Later attempts to repeat this experiment with more rigorous purification and storage conditions reevaluated the role of RNA, since RNA sensitivity was no longer detectable when DNA degradation was removed by adding unmethylated DNA, storing cells in liquid nitrogen and centrifuging at higher speed (Swisher et al., 1998). These moves retained substrate DNA in a better state and subsequent digestion with or without protease K made no difference in resulted cleavage pattern. The addition of RNase

in substrates containing unlabelled unmethylated DNA did not inhibit demethylase activity any more, which suggests previously observed ribonuclease sensitivity came possibly from the tight binding of the nucleases to the methylated DNA substrate (Swisher et al., 1998). However, it does not rule out the possibility of RNA involvement. Besides, another paper again argued for RNA participation by showing restoration of 5-MCDG activity requires an RNA specifically complementary to the methylated strand (Jost et al., 1999). Furthermore, after analysis by tandem mass spectrometry, it was found that 5-methylcytosine DNA glycosylase activity is present in the chicken homolog of the mammalian G/T mismatch DNA glycosylase (Zhu et al., 2000b). Both G/T mismatch DNA glycosylase and 5-MCDG activities localised in the same region at the C- and N-terminus of the enzyme, respectively. Addition of RNA that was complementary to the methylated strand greatly reduced G/T mismatch DNA glycosylase activity but stimulated 5-MCDG activity. On the other hand, the RNA complementary to non-methylated strand did not affect G/T mismatch DNA glycosylase activity, whereas the presence of RNA helicase and ATP enhanced 5-MCDG activity (Schwarz et al., 2000; Zhu et al., 2000b). Since chicken 5-MCDG is also a G/T mismatch DNA glycosylase, another G/T mismatch DNA glycosylase, human MBD4, was tested for its demethylation activity and revealed that 5-MCDG activity was present in both chicken and human MBD4, although the chicken MBD4 homolog does not contain a MBD domain (Zhu et al., 2000a). In this publication, this human MBD4 has a G/T mismatch DNA glycosylase activity about 30-40 times higher than its 5-MCDG activity. Also, the G/T mismatch DNA glycosylase activity resided in the N-terminus of the MBD4 while 5-MCDG activity localised in the last 48 aa of its C terminus. This protein binds to fully, hemi- and non-methylated DNA,

but the best substrate for both activities is hemi-methylated DNA. However, unlike 5-MCDG from chicken embryo that released a whole nucleotide, the activity of human MBD4 cleaved only partially on the abasic sugar at the 3' position. Complete release of the abasic sugar was only detectable after alkaline hydrolysis. One sharp contrast between 5-MCDG and MBD4 is that RNA, which stimulates 5-MCDG, inhibited strongly both glycosylase activities in MBD4 (Zhu et al., 2000a). Just like the putative demethylase Mbd2b, the results stated above are also controversial, because a different laboratory showed mammalian MBD4 does not have any methylcytosine DNA glycosylase or endonuclease activity (Hendrich et al., 1999b). Instead, mammalian MBD4 efficiently removes thymidine or uracil from mismatched CpG sites. This MBD4 also prefers G/T mismatch DNA substrate over fully, hemi- or non-methylated DNA, especially hemi-methylated DNA is a rather weak binding target compared to fully methylated DNA. In terms of functional domain, the G/T mismatch DNA glycosylase activity is found to reside in the C-terminus of MBD4, instead of the N-terminus as shown in the Jost lab (Hendrich et al., 1999b). In addition, there is also a discrepancy in the enzyme substrate, as the Jost lab claimed MBD4 preferred hemi-methylated DNA in an AT-rich context whereas the Bird lab demonstrated MBD4 had a preference for CG-rich DNA substrate (Hendrich et al., 1999b; Zhu et al., 2000a). Above all, the substrate preference of 5-MCDG in hemimethylated DNA does not explain active demethylation since at least one round of replication is needed for the production of hemimethylated DNA. 5-MCDG has been reported to participate in active genome-wide DNA demethylation during mouse myoblast differentiation. Antisense morpholino oligonucleotides targeting 5-MCDG reduced about 80% of demethylation, which means residual demethylation was processed directly from

symmetrically methylated DNA, maybe by 5-MCDG or different proteins in the same complex (Jost et al., 2001). However, such property is not yet demonstrated and further evidence is required to verify the role of 5-MCDG in active demethylation. Still, we cannot yet discount a possible role of glycosylase in the DNA demethylation mechanism, especially given that 5-mC is indeed a hotspot for mC-to-T mutations, which may serve as a start point for the excision repair pathway.

1.6.3 AlkB oxidation

A new candidate mechanism for DNA demethylation is via an oxidative DNA repair pathway. Originally identified from *E. coli* mutants defective in processing methylation damage of single-stranded DNA (ssDNA), the AlkB protein was later found to catalyse oxidative demethylation of 1-methyladenine and 3-methylcytosine (Dinglay et al., 2000; Kondo et al., 1986). From theoretical protein fold prediction, AlkB is suggested to be a member of the 2-oxoglutarate-dependent and iron-dependent dioxygenase (2-OG-Fe(II)-dioxygenase) family and is highly conserved from bacteria to mammals (Aravind and Koonin, 2001). With no detectable nuclease, DNA glycosylase or methyltransferase activity, AlkB was later demonstrated to repair DNA alkylation damage by coupling oxidative decarboxylation of α–ketoglutarate (α–KG) to hydroxylation of the methylated bases in DNA (Figure 1.7). This reaction consumes oxygen and releases carbon dioxide and succinate to generate hydroxymethyl intermediates. The hydroxymethyl group on the nitrogen site of the ring is then spontaneously released as formaldehyde, and consequently result in unmodified cytosine
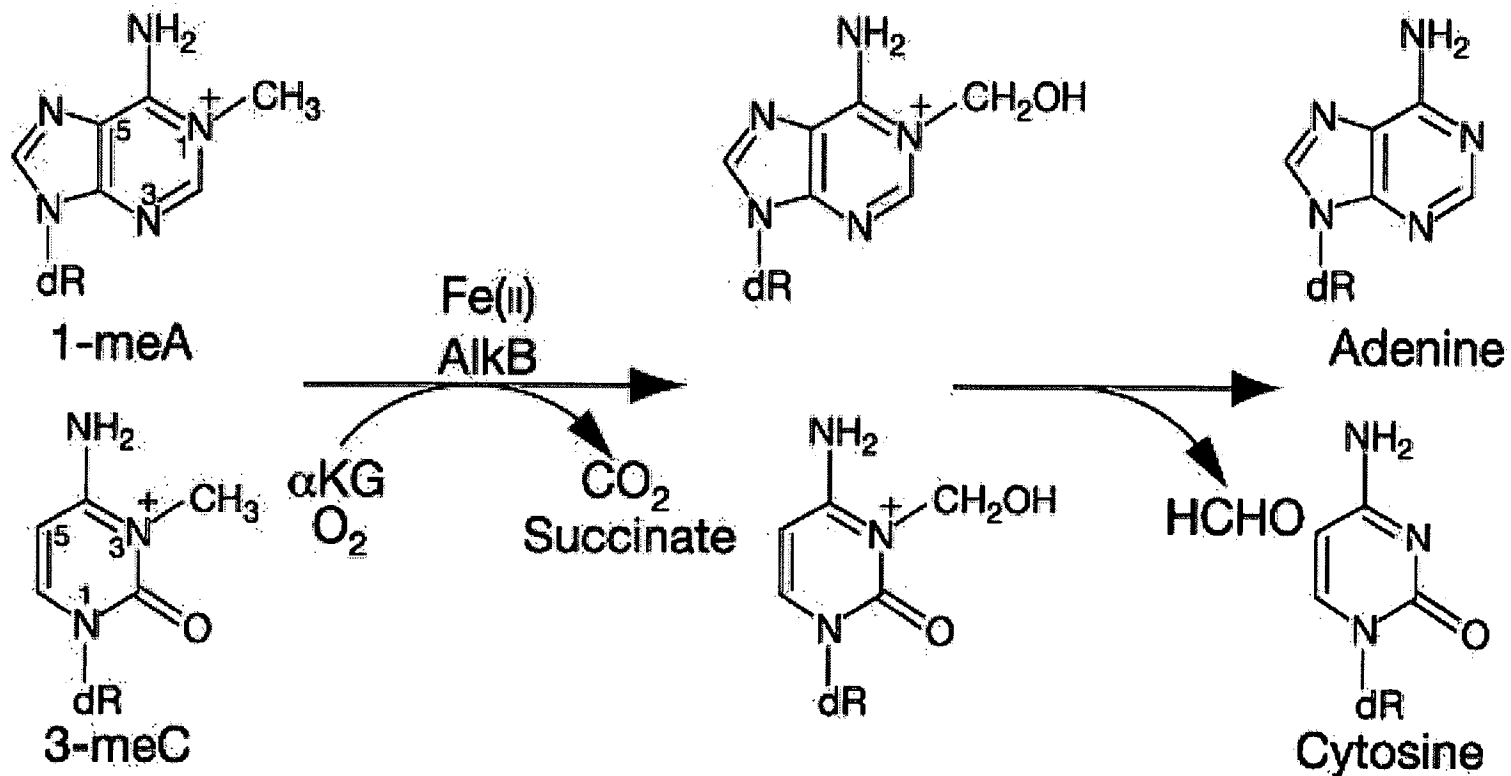
Figure 1. 7: Oxidative demethylation by AlkB. AlkB consumes oxygen and uses Fe (II) and α-ketoglutarate (αKG) as reaction cofactors to hydroxylate the methyl group on 1-methyladenine or 3-methylcytosine. Carbon dioxide and succinate are reaction byproducts resulted from AlkB catalysis, whereas formaldehyde is spontaneously released from the repulsive bonding between the hydroxymethyl group and the nitrogen on the ring (Trewick et al., 2002).

(Falnes et al., 2002; Trewick et al., 2002). This oxidative demethylation generally uses ssDNA as substrate, but in some cases, dsDNA and RNA can be a target for the reaction, although this is not a general feature among all AlkB homologs. AlkB and its human homologs, hABH2 and hABH3, all process oxidative DNA demethylation with a preference of hABH2 for dsDNA substrates, and that of AlkB and hABH3 for ssDNA as their substrate (Aas et al., 2003; Duncan et al., 2002). Moreover, AlkB and hABH3, but not hABH2, are involved in RNA damage repair. On the other hand, hABH1 does not exhibit any damage repair ability and does not bind to DNA probes in cross-linking studies, regardless of its highly conserved sequence compared to other AlkB homologs (Aas et al., 2003; Mishina et al., 2004). Although this DNA damage repair pathway only applies to 1-methyladenine and 3-methylcytosine so far, it does provide another candidate mechanism for 5-methylcytosine demethylation, if 5-methylcytosine can undergo oxidative demethylation through a hydroxymethyl intermediate and release this intermediate spontaneously or by another enzymatic reaction. Moreover, it is plausible to take advantage of such an enzymatic mechanism to perform a rapid, replication-independent demethylation. However, there is yet no evidence showing direct demethylation of 5-mC by AlkB and homologs. Further investigation is required to test such possibility.

1.7 Project summary

DNA methylation is an epigenetic event which regulates gene expression by transcriptional repression. However, the reversal of DNA methylation, DNA demethylation, is not as well understood. As discussed in section 1.6, DNA

demethylation has been demonstrated to occur as an active process during embryo development, tumourigenesis and hormone-induced gene activation, or via a replication-dependent passive route. Nevertheless, the molecular mechanism of active DNA demethylation is yet to be determined, and the precise role of DNA demethylation *in vivo* remains obscure. The work presented in the following chapters is thus focused to bring about artificial DNA demethylation via various approaches and to investigate potential mechanism of DNA demethylation. Initially, S-adenosylmethionine hydrolase (SAMase) from bacteriophage T3 was purified in both non-recombinant and recombinant forms and protein crystallisation was attempted. This part of study was aimed to facilitate the application of SAMase as a demethylating agent and may additionally expand our knowledge of the SAMase-related phage R-M system. Expression of SAMase and transfection of small interference RNA (siRNA) in mammalian cells were utilised to bring about artificial DNA demethylation. Cells treated with these applications were used to analyse the effects of reduced methylation levels. To exam a candidate mechanism of DNA demethylation *in vivo*, a monoclonal antibody was raised and characterised to target 5-HmC, a putative reaction intermediate of oxidative demethylation. Further analysis in preimplantation embryos was carried out to verify the presence of such an intermediate.

Materials and Methods

Chapter 2
Materials and Methods

2.1 Standard solutions and reagents

1M Na$_x$H$_x$PO$_4$ pH 7.2 (NaPi): mix 1 M NaH$_2$PO$_4$ and 1 M Na$_2$HPO$_4$ in a 1:4 ratio.

Agarose gel loading dye (5x): 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol in dH$_2$O, stored at 4 ℃.

Alkaline wash buffer for southern blotting: 1.5 M NaCl, 5 M NaOH.

Blocking buffer: 5% (w/v) non-fat milk powder in TBS-T.

Buffer A: 10 mM Tris·HCl, pH 7.5, 22 mM NH$_4$Cl, 1mM dithiothreitol (DTT), 10% (v/v) glycerol

Coomassie staining solution: 0.1% (w/v) Coomassie Brilliant Blue R250, 50% (v/v) methanol, 10% (v/v) glacial acetic acid.

CTAB: 5% (w/v) CTAB in 0.5 M NaCl.

DEPC-dH2O: Add 0.1% (v/v) diethylpyrocarbonate (DEPC) to water, mix and incubate at 37 ℃ overnight. Autoclave before use.

ECL solution I: 25 mM luminol, 396 μM p-coumaric acid, 100 mM Tris·HCl, pH8.5.

ECL solution II: 5.6 mM H$_2$O$_2$, 100 mM Tris·HCl, pH 8.5.

HBS (2x): 280 mM NaCl, 10 mM KCl, 1.5 mM Na2HPO4•2H2O, 12 mM glucose, 50 mM HEPES. pH was adjusted using NaOH to pH 7.05. Sterile filter and store in aliquots at -20 ℃.

Hybridisation buffer: 0.5 M NaPi pH 7.2, 7% (v/v) SDS, 12.5 mM EDTA.

Laemmli buffer: 60 mM Tris·HCl, pH 6.8, 100 mM DTT, 10% (v/v) glycerol.

Lysis buffer for DNA extraction: 0.05% (w/v) protease K, 1% SDS, 95% (v/v) TE-8

Neutralisation buffer for Southern blotting: 0.5 M Tris·HCl pH 7.5, 1.5 M NaCl

Nickel agarose binding buffer: 300 mM NaCl, 10 mM imidazole, 50 mM $NaH_2PO_4·H_2O$, pH8.0.

Nickel agarose wash buffer: 300 mM NaCl, 20 mM imidazole, 50 mM $NaH_2PO_4·H_2O$, pH8.0.

Nickel agarose elution buffer: 300 mM NaCl, 250 mM imidazole, 50 mM $NaH_2PO_4·H_2O$, pH8.0.

Glutathione elution buffer: 10 mM glutathionine in Tris 9.5 buffer or 1x PBS.

PBS (10x): 10 Tablets of phosphate buffered saline, consisting of 0.2mg KH_2PO_4, 1.15mg Na_2HPO_4, 8mg NaCl, 0.2 mg KCl, were dissolved in 1 litre of dH_2O to give a solution of pH 7.3. The solution was autoclaved and stored at room temperature.

PBS-T: 0.05% (v/v) Tween-20 in PBS.

Random priming buffer (10x): 100 mM $MgCl_2$, 900 mM HEPES, pH 6.6.

SAMase activity assay buffer (10x): 50 mM DTT, 1 M Tris·HCl pH7.5, 100 mM EDTA

SDS-PAGE loading buffer (2x): 6% β-mercaptoethanol, 6% SDS, 0.6% bromophenol blue, 20% glycerol, stored at -20 ℃.

SDS-PAGE resolving gel solution: 8-15% (v/v) 40% 29:1 Acrylamide:bis-acrylamid, 0.1% (v/v) SDS, 375 mM Tris·HCl, pH 8.8, 0.1% (v/v) TEMED, 1% (v/v) APS.

SDS-PAGE stacking gel solution: 3.9% (v/v) 40% 29:1 Acrylamide:bis-acrylamide, 0.1% (v/v) SDS, 125 mM Tris·HCl, pH 6.8, 0.1% (v/v) TEMED, 2% (v/v) APS.

SSC (20x): 3 M NaCl, 0.3 M Na Citrate, pH 7.0.

TAE buffer (10x): 0.4M Tris-acetate, 100 mM EDTA.

TBS-T: 20 mM Tris, pH 8.0, 100 mM NaCl, 0.05% (v/v) Tween-20.

TE buffer (1x): 10 mM Tris, 1 mM EDTA, pH 8.0.

TE-8: 500 mM Tris-HCl, pH 8.0, 20 mM EDTA, pH 8.0, 10 mM NaCl.

Transfer buffer (10x): 50 mM Tris, 385 mM glycine, 0.1% (v/v) SDS.

Tris 7.5 buffer: 100 mM Tris pH7.5, 0.5 mM DTT, 0.05% deoxycholate.

Tris 9.5 buffer: 100mM Tris pH9.5, 0.5 mM DTT, 0.05% deoxycholate.

Tris-glycine electrophoresis buffer (5x): 125 mM Tris, 1.25M glycine, 1%SDS.

Wash solution I for southern hybridisation: 0.04 M NaPi pH 7.2, 1% SDS, 2 mM EDTA.

Wash solution II for southern hybridisation: 0.04 M NaPi pH 7.2, 0.1% SDS, 2 mM EDTA.

Wash solution III for southern hybridisation 0.2 M NaPi pH 7.2, 4 mM EDTA.

2.2 Enzymes, antibodies, isotopes and chemicals

All restriction enzymes, calf intestinal phosphatase (CIP) were purchased from New England Biolabs Ltd. T4 DNA ligase was purchased from Roche Molecular Biochemicals Ltd. RedHot DNA polymerase and *Pfu* DNA polymerase were from ABgene Inc. and Promega Co., respectively. C^{14}-SAM and α-P^{32}-ATP were purchased from Amersham Bioscience plc. Antibodies used are summarised in Table 2.1. Anti-HmC antibody was raised by us, in collaboration with Lawrence Sowers Lab of Lomo Linda University, and the antibody unit of Babraham Institute. The anti-Dnmt1 PATH52 antibody was a kind gift from Dr. Timothy Bestor. Mounting agents containing DAPI (4',6' diamidino-2-phenylindole) or propedium iodine

were purchased from Vector Laboratories Inc.

Table 2.1: Antibodies used in this thesis.

Antibody	Reference/source	Working condition
mouse anti-FLAG M2	Sigma-Aldrich Co.	1:1,000
mouse anti-ssDNA	Alexis Co.	1:100
mouse anti-5mC	Eurogentec Ltd.	1:50
mouse anti-5mC	Oncogene Inc.	1:1,000
rabbit anti-Dnmt1 (PATH52)	(Bestor, 1992)	1:10,000
rabbit anti-MBD1 (M245)	Santa Cruz Inc.	1:1,000
sheep anti-MBD2 (S923)	(Ng et al., 1999)	1:2,500
rat anti-HmC	This thesis	1:1-1:10,000
AMCA conjugated anti-rat IgG secondary antibody	Vector Labs Inc	1:200
Texas Red conjugated anti-mouse IgG secondary antibody	Vector Labs Inc	1:200
anti-mouse IgG 594 secondary antibody	Molecular probes Inc.	1:1,000
HRP conjugated anti-mouse IgG secondary antibody	Amersham Bioscience plc.	1:1,000
HRP conjugated anti-rat IgG secondary antibody	Amersham Bioscience plc.	1:1,000
HRP conjugated anti-sheep IgG secondary antibody	Amersham Bioscience plc.	1:1,000
FITC conjugated anti-mouse IgM secondary antibody	Oxford Biotechnology Ltd.	1:100

2.3 Microbiological media and antibiotics

Ampicillin stock: 10 mg/ml ampicillin in sterile dH_2O, filter sterilised and stored at -20 °C.

Bacterial freezing medium: 0.1M $CaCl_2$, 20% glycerol.

BBL top agar: 10 g/l BBL trypticase, 5 g/l NaCl, 6.5 g/l DIFCO agar. Add dH_2O up to 1 litre. Autoclave.

Kanamycin stock: 10 mg/ml kanamycin in sterile dH_2O, filter sterilised and stored at -20 °C.

LB agar: LB broth plus 15 g/l OXOID agar.

LB broth: 10 g/l NaCl, 10 g/l Difco bacto-tryptone, 5 g/l Difco yeast extract. Adjust to pH 7.2 with 10M NaOH. Autoclave.

phage buffer: 10 g/l Na_2HPO_4 (anhydrous), 3 g/l KH_2PO_4 (anhydrous), 5 g/l NaCl, 10 ml 0.1M $MgSO_4$, 10 ml 0.01M $CaCl_2$, 1ml 1% gelatin. Add dH_2O up to 1 litre. Autoclave.

Tetracycline stock: 10 mg/ml tetracycline in 50% ethanol, stored at -20 °C.

2.4 Plasmids and bacterial strains

The plasmids and the bacterial strains used are presented in Table 2.2.and Table 2.3, respectively.

Table 2.2: Plasmids used in this thesis.

Plasmid	Reference/ source	Selectable markers	Description
pBAD24	(Newman and Fuqua, 1999)	Amp^r	Bacterial expression vector for native SAMase, with P_{BAD} promoter controlled by arabinose operon.
pJF118HE	(Furste et al., 1986)	Amp^r	Bacterial expression vector for native SAMase with IPTG-inducible *tac* promoter.
pGEX-4T-2	Amersham	Amp^r	Bacterial expression vector for fusion protein GST-SAMase with IPTG-inducible *tac* promoter.

Table 2.2: *Continued.*

Plasmid	Reference/ source	SelecTable markers	Description
pET30b(+)	Novagen	Kan^r	Bacterial expression vector for recombinant His-SAMase fusion with IPTG-inducible T7 promoter.
pTet-OFF	Clontech	Ampr	Regulatory vector expressing tTA protein.
pBI-EGFP	Clontech	Ampr	Bidirectional mammalian expression vector with P$_{CMV}$ promoters regulated by the TRE, reciprocal to the dose of deoxycycline.

Table 2.3: Bacterial strains used in this thesis.

Strain	Genotype	Description
DH5α	*supE44 ΔlacU169 ($\emptyset$80 lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	for recombinant DNA manipulation
BL21(DE3)	F^- *opmT hsdS$_B$(R$_B^-$M$_B^-$) gal dcm met*(DE3) (Cm^R)	for high-level protein expression
NM494	genotype pop101, Δ(hsdS, mcrB)	for phage infection, T4 permissive (Raleigh et al., 1988)
NM654	genotype C600, ΔhsdRM	for phage infection, T4 sensitive (Loenen et al., 1987)
W 4975	genotype W3110, *galU$^-$*	for phage infection, producing unglucosylated DNA (Hattman and Fukusawa, 1963b)

2.5 Manipulation of DNA and cloning procedures

2.5.1 Synthetic oligonucleotides

Custom oligos from Sigma-Genosys Ltd. were resuspened in TE buffer to make stock solutions of 1 µg/µl and stored at -20 ℃. Dilutions of the stock solutions for polymerase chain reaction and DNA sequencing were made in TE. The sequences and applications of all oligonucleotides used are presented in Table 2.4.

Table 2.4: Oligonucleotides used in this thesis

Oligo-nucleotide	5'->3' Sequence	Description
SAMase F1	TAAGGATCCATGATTTTCAC	PCR and sequencing primer for SAMase gene
SAMase R1	TTGAATTCTTATTGTACTTGCC	PCR and sequencing primer for SAMase gene
ET6H-F	ATCCATGGAGATGATTTTCAC	PCR and sequencing primer for SAMase gene
ET6H-R	AGCCGGATCCTTATTGTACTTGCC	PCR and sequencing primer for SAMase gene
FN-SAM	CCGATGGACTACAAGGACGACG ATGACAAGGATCCAAAAAAGAA GAGAAAGGTAATGATTTTCAC	PCR primer for the construction of recombinant SAMase cDNA with FLAG and NLS DNA sequences at the N-terminus of SAMase gene
SH-6His	ATTGCTAGCTTAATGGTGATGGTG ATGGTGTTGTACTTGC	PCR primer for the insertion of a 6xHis-tag DNA sequence at the C-terminus of SAMase gene
k-FLAG	AATACGCGTCCACCATGGACTAC AAGG	PCR primer for the insertion of Kozak consensus
GFP-F	ATGGTGAGCAAGG	RT-PCR primers for EGFP in BI-EGFP vector
GFP-R	ATGTGATCGCGCT	
Gapdh p3	TACCCCCAATGTGTCCGTCG	Real-time PCR primers for *Gapdh*
Gapdh p4	CCTGCTTCACCACCTTCTTG	
XIST 7F	CAGCAAGCCCACAATTCTGG	Real-time PCR primers for *Xist*
XIST 11R	GGACTGCCAGCAGCCTATAC	
Dnmt1-3F	AAG TGC CCC GAG TGTG	Real-time PCR primers for *Dnmt1*
Dnmt1-3R	AGGTGGAGTCGTAGATGG	
sat-for	CTGTAGGACGTGGAATATGGC	PCR primers for the production of major satellite probes (Jorgensen et al., 2004)
sat-rev	CCGTGATTTTCAGTTTTCTCGC	

2.5.2 The polymerase chain reaction

The polymerase chain reaction (PCR) was used to amplify DNA sequence for cloning or analysis. Between 50 ng (plasmid) and 200 ng (genomic DNA) of template DNA was used for amplification with 0.3 μ g of each primer. The reactions were performed in reaction buffer for proofreading polymerase *Pfu*, with 0.2 mM

dNTPs, and 1.5 units of polymerase. The cycling conditions used are shown below:

Cycle	Temperature (℃)	Time (min/sec)
1	94	1min
30	94	30sec
	55-65	30sec
	72	15-30sec
1	72	10min

When PCR was used to screen bacterial colonies, RedHot *Taq* DNA polymerase was used in place of *Pfu*. In this case, 3 mM $MgCl_2$ was added to the reaction buffer, and 1 unit of polymerase was used per reaction. The annealing temperature for primer-template combination in each reaction was defined empirically.

2.5.3 Restriction digests

Restriction digestion was carried out in 10-50 μl volume with appropriate reaction buffer for the chosen enzyme and, if required, 100 μg/ml BSA. At least 1 unit of enzyme was added per μg of DNA to be digested and the reaction was mixed thoroughly before incubation at 37 ℃ for at least 1 hour. The reaction was stopped by heat inactivation, where appropriate.

2.5.4 Purification of DNA from agarose gels

DNA bands resolved by agarose gel electrophoresis were visualised using long wave UV light. DNA exposure to UV light was kept to a minimum. The required DNA band(s) was excised from agarose gel using a clean scalpel. DNA was then purified from the agarose slice using the Qiagen Gel Extraction Kit according to the

manufacturer's instructions.

2.5.5 Ligation of DNA fragments

DNA ligation was performed with T4 DNA ligase (Roche) in supplemented reaction buffer. Molar ratios of 1:1, 1:3 and 1:5 (vector:insert) were used to give optimum ligation conditions. 50-200 ng of total DNA with 1 unit of ligase was used in each reaction. The reaction was incubated at 16 ℃ overnight. Approximately 1/10th of each reaction was used to transform competent *E. coli* cells.

2.5.6 Preparation of plasmid DNA

2.5.6.1 Small scale

Plasmid DNA was prepared from cultures of various bacterial strains. Cells were pelleted from 2 ml of an overnight culture by centrifugation, and then treated with the QIAGEN Miniprep Kit according to manufacturer's instructions. Eluted DNA in TE buffer were used immediately or stored at -20 ℃.

2.5.6.2 Large scale

The QIAGEN Maxiprep Kit was used to prepare larger amounts of plasmid, according to manufacturer's instructions. DNA was prepared from an appropriate volume of culture depending on the copy number of the plasmid. Precipitated DNA was resuspened in TE and used immediately or stored at -20 ℃.

2.5.7 DNA sequencing

The nucleotide sequences of DNA templates were identified by automated

sequencing using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit and an ABI 3100 DNA sequencer. Purified templates, 5-20 ng of PCR product or 150-300 ng of plasmid DNA, were mixed with 3.2 pmol of the appropriate primer, 4 μl of BigDye Terminator v3.1 and dH$_2$O to a final volume of 10 μl before cycle sequencing using the following conditions:

Cycle	Temperature (℃)	Time (min/sec)
1	96	1min
25	96	30sec
	50	20sec
	60	4min

Sample purification, electrophoresis and data collection were carried out by the ICMB sequencing service. Sequencing data was analysed using Lasergene software.

2.5.8 Radioactive labelling of oligonucleotides

DNA probes were random prime-labelled with α-^{32}P-dCTP using Klenow enzyme and hexanucleotide mix (Roche). 200ng of total DNA was incubated at 37 ℃ with 450 ng random hexanucleotides, 3 μl 10x random priming buffer, 3 μl 20 mM DTT, 3 μl 10mM dATP+dTTP+dGTP mix, 3 μl α-^{32}P-dCTP, 2 μl Klenow enzyme and dH$_2$O to a final volume of 30 μl. Radioactive-labelled DNA probes were purified using PCR purification kit according to manufacturer's instructions. The resulting probes were eluted in TE buffer and used immediately.

2.6 Bacterial and phage methods

2.6.1 Bacterial growth condition

2.6.1.1 Plate culture

Bacterial cells were streaked onto nutrient agar plates with appropriate antibiotics and incubated for at least 16 hours at 37 ℃ to allow colony formation.

2.6.1.2 Liquid culture

Single colonies from freshly streaked agar plates were used to inoculate the cultures, which were grown in LB (2-3 ml) containing appropriate antibiotics at 37 ℃ with shaking (220rpm) for 12-16 hours. Large cultures were initiated by inoculation of an appropriate volume of LB with cells from a freshly grown overnight culture, split into a suitable density and grown till a desired O.D. was reached.

2.6.2 Preparation of competent cells

200 ml of LB was inoculated with 1 ml of an overnight culture and grown at 37 ℃ until the culture reached an $O.D._{600}$ of 0.5. The cells were pelleted by centrifugation and resuspended in ice-cold 0.1 M $MgCl_2$. Resuspended cells were again pelleted by centrifugation and resuspended in ice-cold 0.1 M $CaCl_2$. After a final centrifugation step, the pelleted cells were resuspended in freezing medium, flash frozen in liquid nitrogen and stored at -80 ℃.

2.6.3 Bacterial transformation by heat shock

10 ng of plasmid DNA in a volume of 1-5 μl was mixed with 100 μl competent

bacteria on ice for 10-15 minutes, to allow binding of the DNA to the cells. The cells were then heat shocked at 42 ℃ for 45 seconds to transform and put back on ice for 2 minutes. 900 μl of LB medium was added to the cells, which were incubated at 37 ℃ for 1 hour before plating on LB plates containing the appropriate antibiotics. Plates were inverted and incubated overnight at 37 ℃.

2.6.4 Preparation of T4 phage stocks

2.6.4.1 Preparation of phage lysates

Wild type bacteriophage T4 was cultured by infection of *E. coli* strain NM494 in LB. When the bacterial culture reached 0.5 O.D.$_{600}$, an appropriate amount of wild type T4 was added to result in a multiplicity of infection (MOI) of 2. Ten minutes after the first infection, the cells were superinfected with the same MOI to induce lysis-inhibition. Superinfected bacteria were cultured for another 2.5 hours and further lysed with chloroform at room temperature overnight. Pancreatic DNase I, at a final concentration of 1 μl/ml, was added to the lysed cells and incubated at room temperature for 30 minutes. Treated cultures were then collected by centrifugation at 4,000 x g for 20 minutes and the phage-containing lysates were transferred into a clean bottle and stored at 4 ℃. T4 phage with unglucosylated hydroxymethylcytidine DNA was grown in the same way, but the infected bacteria strain was replaced by *E. coli* W4975.

2.6.4.2 Purification of T4 phage by PEG Precipitation

The volume of the clear phage lysate was measured and poured into a centrifuge bottle with a magnetic stirrer. NaCl was slowly added to the lysate to a final

concentration of 0.5 M with constant vigorous stirring. PEG 8000 was then added to a final concentration of 10% and the phage solution was stirred for at least an hour. After the stirring bar was removed, phage was collected by centrifuging at 4,000 x g for 30 minutes. The supernatant was then discarded and the pellet was resuspended in phage buffer. Purified phage was stored at 4 ℃.

2.7 DNA methods

2.7.1 Isolation of phage DNA

For 1 ml of purified phage stock, 200 μl 1 M Tris·HCl pH 8.0, 200 μl 0.5 M EDTA, 380 μl dH$_2$O, and 10 μl 10 mg/ml protease K were added. After an incubation of 30 minutes at 37 ℃, 100 μl of hexadecyl-trimethyl-ammonium bromide (CTAB) was added into the phage solution and incubated for another 5 minutes at 65 ℃ before a centrifugation at maximum speed for 10 minutes. The CTAB-DNA pellet was jarred loose in 2 ml 1.2 M NaCl, after the supernatant was discarded. The salt solution containing CTAB-DNA pellet was then transferred to a new tube and another 2 ml of 1.2 M NaCl was added to resuspend the pellet thoroughly, followed by an addition of 4 ml 95% ethanol. The precipitating DNA was spooled out with a fused Pasteur pipette and rinsed with 70% ethanol before air dry. Extracted DNA was dissolved in appropriate amount of dH$_2$O and stored at 4 ℃ or -20 ℃.

2.7.2 Genomic DNA extraction from cultured cells

Cells were cultured in flasks or plates until 100% confluence and resuspended in 2 ml lysis buffer per 75 cm^2 culture area after a rinse with PBS. The suspension was transferred to centrifuge tubes. After an overnight incubation at 37 ℃, 1/4th volume

of saturated NaCl was added to the suspension, mixed and centrifuged at 900 x g for 30 minutes. The supernatant was then transferred to a clean tube and the genomic DNA was recovered using ethanol precipitation. Precipitated DNA was spooled out with a Pasteur pipette and resuspended in TE. DNA was treated with 100 μg / ml RNase A at 37 ℃ for at least 3 hours. Genomic DNA was again purified using ethanol precipitation and resuspended in TE.

2.7.3 Southern blotting and hybridisation

DNA for analysis was separated by 1.2% agarose gel electrophoresis. The gel was stained with 4 μg/ml ethidium bromide for 30 minutes, destained in dH_2O for 10 minutes, and photographed alongside a fluorescent ruler. The gel was then depurinated in 0.25 M HCl for 15 minutes with gentle shaking, washed twice in alkaline wash buffer for 15 minutes and washed twice in neutralisation buffer for 10 minutes. Southern transfer was set up as follows: A platform was first prepared on a tray filled with 20x SSC transfer buffer, covered with a wick made from 2 sheet of Whatman 3MM paper saturated in 20x SSC. The treated gel is then placed on the wick platform followed by a sheet of Hybond N+ (Amersham), 3 sheet of Whatman 3MM paper, a stack of paper towels and finally a glass plate and a weight about 750g to weigh down the assembled transfer stack. Air bubbles trapped between each layers were avoided. The assembly was left overnight. The following morning, the membrane was rinsed in 2x SSC and the transferred DNA was crosslinked to the membrane by UV crosslinking in a Stratalinker UV crosslinker 1800 (Stratagene) using the autocrosslink function at an energy of 120,000 μjoule.

Place the membrane in a hybridisation bottle with 15 ml hybridisation buffer to

prehybridise at 65 °C for at least 15 minutes. DNA probes were added to the hybridisation buffer after prehybridisation and allowed to incubate overnight at 65 °C. Membranes were then washed once with solution I for 1 hour and solution II for another hour. The membrane was briefly dried and wrapped in a transparent wrap before exposure in a PhosphorImage cassette and analysed on a Storm 840 apparatus.

2.8 RNA methods

2.8.1 RNA extraction

RNA was isolated using TRI reagent from Sigma-Aldrich Co., following manufacture's instructions. RNA pellets were resuspended in an appropriate amount of DEPC-treated, RNase-free dH_2O. RNA samples were stored at -20 °C.

2.8.2 RNase-free DNase treatment

Removal of genomic DNA from RNA samples was carried out using DNA-free kit from Ambion Inc., according to manufacturer's instructions. For each 2.5 μg total RNA subjected to DNase treatment, 1 μl DNase was used in a total reaction volume of 25 μl. After DNase treatment, the supernatant containing RNA was transferred to a new tube and used immediately.

2.8.3 Reverse transcription

Reverse transcription was carried out using M-MLV reverse transcriptase from Promega Co. DEPC-dH_2O was added to the DNase treated RNA samples to make

up a volume of 34 μl and incubated at 65 ℃ for 5 minutes. Samples were chilled on ice for 1 minute. 6 μl of 5 μM random hexamer primers (Roche), 6 μl 10mM dNTP mix, 12 μl 5x reaction buffer, 1 μl RNAsin (Promega) and 1 μl M-MLV reverse transcriptase were added to the samples and the reaction was completed in a PCR cycler using the following program:

Cycle	Temperature (℃)	Time (min/sec)
3	20	8 min
	25	8 min
	37	30 min
1	70	15 min

The resulting samples were stored at -20 ℃ or used immediately for PCR.

2.9 Protein methods

2.9.1 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gel electrophoresis under denaturing conditions was used to separate polypeptides based on their molecular size. The Bio-Rad vertical electrophoresis system (Mini PROTEAN) was used. Resolving gel solution was made to the required acrylamide concentration and the mixture was poured into the gel plates, leaving room at the top for the stacking gel. dH$_2$O was layered over the solution and it was allowed to polymerise for 30 minutes. Once set, the water was removed from the gel surface and stacking gel solution was added and allowed to polymerise for 30 minutes. Protein samples were heated to 99℃ for 3-5 minutes in SDS-PAGE loading buffer before loading. Samples were electrophoresed with 1x

Tris-Glycine electrophoresis buffer at 10-20 V/cm for an appropriate period of time.

2.9.2 Electrophoretic transfer of protein blotting

SDS-PAGE gels were rinsed briefly in transfer buffer to remove SDS before transfer. The resolved proteins on SDS-PAGE were transferred onto 0.2 μm Protran nitrocellulose membranes (Schleicher & Schuell) with a Mini Trans Blot cell (Bio-Rad) at 200mA for 2 hours at room temperature or with a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) at 200 mA for 1 hour.

2.9.3 Western detection of proteins

After transfer to nitrocellulose membranes, non-specific protein binding sites were blocked by incubation in blocking buffer for at least 1 hour at room temperature or 4 ℃ overnight. Primary antibody incubation was carried out for 2 hours at room temperature in blocking buffer with appropriate dilution of antibody. Membranes were washed 3 times in blocking buffer for 10-15 minutes, before being incubated for 1 hour with an appropriate concentration of secondary antibody (conjugated with horseradish peroxidase) in blocking buffer. Following the secondary antibody incubation, two more washes in blocking buffer and one wash in TBS-T were carried out before ECL detection of bound antibody. An equal volume of ECL solution I and II were mixed and then incubated with membranes for 1 minute. Membranes were wrapped in Saran wrap and exposed to ECL Hyperfilm (Amersham) for an appropriate length of time.

2.9.4 Protein extraction from bacteria

Induced bacterial cultures were harvested by centrifugation down bacteria at 5,000 rpm, 4 ℃, for 15 minutes. Wash bacteria pellets twice by resuspending bacteria in 100mM Tris·HCl pH 7.5 and centrifuge at 5,000 rpm at 4 ℃, for 15 minutes. Bacteria were resuspended in appropriate volume of Tris 7.5 buffer containing 1 mg/ml lysozme. After an incubation of 30 minutes on ice, bacterial cells were sonicated in Sonifier 250 (Branson) with appropriate output power and time depending on the volume of resuspended bacteria.

2.9.5 Whole cell extract from tissue culture cells

Tissue culture cells were harvested from the surface of the culture plate and transferred to an Eppendorf tube. Cells were pelleted by centrifugation at 1300 rpm for 5 minutes, washed once in PBS, and then resuspended in Laemmli buffer with 1/10th volume of 20% SDS. After boiling for 5 minutes on a heat block, cell extracts were cooled on ice for 1 minute and briefly centrifuged before use or kept at -20 ℃.

2.9.6 Measurement of protein concentration

2.9.6.1 Protein concentration measured by Bradford assay

The Bio-Rad Protein Assay reagent was used to measure protein concentrations. The reagent was diluted 1:10 and filtered before use. BSA was used to prepare a standard curve each time a protein concentration measurement was made. Protein samples were incubated with diluted assay reagent for 5 minutes before the $O.D._{595}$ measurement of each sample. Protein concentrations were calculated by comparison to the BSA standard curve.

2.9.6.2 Protein concentration measured by UV detector

Concentrations of purified protein samples were determined by a UV detector scanning through wavelengths from 340 to 240 nm. The buffer solution used to store sample proteins was used as a baseline standard in comparison to the diluted protein samples. Protein concentrations were calculated by the absorbance of A_{280} with calibration from the absorbance curve.

The protein concentrations of whole cell extracts were determined by a UV detector measuring $O.D._{280}$ and $O.D._{260}$. The readings of these two wavelengths were then applied to the following equation, to give a concentration of mg/ml. C (mg/ml) $= 1.55 \times A_{280} - 0.76 \times A_{260}$

2.10 Tissue culture

Tissue culture cells were grown in the appropriate growth medium, supplemented with 10% (v/v) bovine calf serum and 10 units/ml penicillin-streptomycin solution (Gibco-BRL), at 37 °C in 5% CO_2. 1x Trypsin-EDTA solution (Gibco-BRL) was used to detach cells from tissue culture flasks for passaging. Cells were frozen in growth medium containing 10% (v/v) DMSO and kept in liquid nitrogen for long-term storage. Where requested, deacetylation inhibitor, TSA, was added to cells at a final concentration of 1 μ g/ml for 18 hours before cell were harvested.

2.11 Expression of proteins in E. coli

2.11.1 Construction of plasmids expressing SAMase

The T3 SAMase gene (NCBI accession No. X04791; Appendix I) was amplified with

primer set SAMase F1/SAMase R1 and primer set pET6H-F/pET6H-R to produce cDNA, which includes either BamHI/EcoRI or NcoI/BamHI cloning sites, and subsequently cloned into appropriate vectors to give plasmid construct producing proteins with correct reading frames. A summary of the cloning strategy and recombinant plasmids is in Table 2.5.

Table 2.5: Plasmid constructions for bacterial expressions.

plasmid name	original vector	primers used for cDNA amplification (Forward/Reverse)	insertion sites (5′/3′)
pJF-SH	pJF118HE	SAMase F1/SAMase R1	BamHI/EcoRI
pBAD-SH	pBAD24	SAMase F1/SAMase R1	BamHI/EcoRI
pGEX-SH	pGEX-4T-2	SAMase F1/SAMase R1	BamHI/EcoRI
pET-SH	pET30b (+)	ET6H-F/ET6H-R	NcoI/BamHI

2.11.2 Induction of protein expression in bacteria

Bacteria of appropriate strains were transformed with desired plasmids. An overnight culture was diluted 1:100 into 500 ml LB containing the appropriate antibiotics in a large flask and grown to an O.D.$_{600}$ of 0.5 before induction with IPTG of various concentrations for various lengths of time according to the conditions tested. However, induction of plasmids containing araBAD promoter (pBAD24 based constructs) was carried out by adding 0.2% arabinose rather than IPTG and induced for four hours.

2.12 Expression of protein in mammalian cells

2.12.1 Construction of plasmids expressing SAMase

The Tet-Off Expression System (Clontech) in combination with the bidirectional

expression vector pBI-EGFP (Clontech) was used to express tagged SAMase in human or mouse cells under the control of deoxycycline. The phage SAMase cDNA was amplified with primer set, FN-SAM and SH-6His, to produce cDNA flanked by an N-terminal FLAG-NLS DNA sequence and a C-terminal 6xHis tag sequence. The recombinant cDNA was then modified by k-FLAG primer to insert a Kozak consensus upstream of the start codon ATG included the FLAG sequence. The resulting cDNA was cloned into the bidirectional vector pBI-EGFP at the MluI/NheI sites. A schematic illustration of this construct is shown in Figure 4.2.

2.12.2 Transfection of mammalian cells

Cells for transfection were freshly plated and allowed to grow to 70% confluence over 16-24 hours. HeLa cells were transfected using Lipofectamine (Invitrogen) according to the manufacturer's instructions. A total of 1-2 μg plasmid DNA was used to transfect a 35mm well. Human embryonic kidney cells, HEK293, were transfected using a calcium phosphate-mediated transfection method according to the Maniatis manual (Sambrook et al., 1989). For each well, 102.67 μl of 40 μg/ml DNA in 0.1x TE pH 8.0 was mixed with 14.47 μl of 2M $CaCl_2$ and 116.67 μl 2x HBS. The mixture was incubated at room temperature for 30 minutes before adding dropwise into the medium over cells. Cells were washed in PBS, 20 hours after transfection and harvested 48 hours after transfection.

2.13 Purification of native and recombinant protein

2.13.1 Native SAMase purification

This method is adapted from the purification of SAMase A, published in Spoerel

and Herrlich (1979). For every 100 ml of crude cell supernatant, 10 ml of saturating polymin P (prepared by Mr. Laurie Cooper) was added and incubated for 20 min at 4 ℃ to allow the formation of polymin-binding precipitate. The precipitate was removed by centrifugation at 12,000 x g for 20 min. Ammonium sulfate at 80% saturation was then added to the non-absorbing supernatant containing SAMase, in order to remove any residual soluble polymin P. After one wash of the pellet with 80% ammonium sulfate, the precipitate was isolated and dissolved in 10 ml of buffer A containing 35% glycerol. The dissolved protein solution was dialysed against buffer A and applied to a 40 ml-DEAE cellulose column (Whatman DE-52). The flow-through (about 10 ml) from the column was collected and subjected to an S-adenosylhomocysteine (SAH) affinity column. The DEAE column was washed with 50 ml buffer A and eluted with 0.15-0.2 M NH$_4$Cl in a total volume of 10 ml. The flow-through was applied to a 10-ml SAH affinity column (prepared by Mr. Laurie Cooper), washed with 16 ml of 2 M NH$_4$Cl in buffer A. SAMase was eluted in 10 mM adenosine, 2 M NH$_4$Cl in buffer A with a total volume of 12 ml collected in 12 fractions (1 ml per fraction).

2.13.2 His-tagged protein purification

Proteins containing a 6xHis tag were produced in *E. coli* strain BL21 (DE3) and purified by nickel agarose chromatography. Bacterial cultures expressing the protein were pelleted by centrifugation at 5000 rpm for 15 minutes, washed once in TE buffer, and lysed in 5 ml nickel agarose binding buffer by incubation with lysozyme (1mg/ml) for 30 minutes on ice. After sonication for 5 minutes on ice, cell debris was pelleted by centrifugation at 12,500 rpm for 30min at 4℃. Ni-NTA superflow

beads (Qiagen) were washed three times in PBS and then added to the clear supernatant from the lysed cells. Bacterial cell lysates were incubated with the nickel agarose beads on a roller at 4 °C for at least 1 hour to allow binding of His-tagged proteins to the resin. After binding, the lysate was transferred to a Bio-Rad Econo-Pac Chromatography Column and the flow-through was collected. The column was washed four times with wash buffer of 10-time bed volume. His-tagged protein was eluted in four elution buffer of 1 bed volume.

2.13.3 GST affinity purification

2.13.3.1 Equilibration of Glutathione Sepharose 4B beads

Glutathione Sepharose 4B (Amersham) was used to purify GST fusion protein. For each ml bed volume of Sepharose gel needed, 1.33 ml of 75% Sepharose slurry was transferred to a Falcon tube, after gently mixing the slurry. The gel was then centrifuged at 500 x g for 5 minutes before the storage buffer was discarded. Following a wash in 10 bed volume of cold 1x PBS or Tris 9.5 buffer, the gel was equilibrated with 1 bed volume of the corresponding wash buffer. The equilibrated column was stored at 4 °C.

2.13.3.2 Preparation of GST fusion protein from cell lysate

GST fusion proteins were produced in *E. coli* strain BL21 (DE3). Bacterial cultures expressing the proteins of interest were pelleted by centrifugation at 5000 rpm for 15 minutes, washed twice, and lysed by sonification in 50mM Tris•Cl pH 7.5. After a centrifugation at 12500 rpm, at 4 °C for 30 minutes, the supernatant was discarded, whereas the cell debris containing inclusion bodies was subjected to protein

refolding protocols and was later used in the glutathione sepharose purification procedure.

2.13.3.3 Preparation of GST fusion protein from inclusion bodies

Inclusion bodies containing GST fusion proteins were pelleted by centrifugation of the cell lysate at 12500 rpm, 4 ℃ for 30 minutes. The resultant pellet was dissolved in an equal volume of 8M urea at 4 ℃ overnight and placed on a roller to denature proteins for refolding. After urea denaturation, the solution was centrifuged at 12500 rpm, at4 ℃, for 15 minutes and the clear supernatant was transferred to a beaker containing 4x or 9x supernatant volume of Tris 9.5 buffer with a stir bar stirring constantly. The refolded protein solution was then centrifuged at maximum speed at 4 ℃ for 15 minutes to remove any remaining insoluble fraction. The soluble protein was then transferred to a tube containing equilibrated glutathione sepharose.

2.13.3.4 Thrombin cleavage of GST fusion protein

GST fusion protein was cleaved from the GST tag by Thrombin digestion. Protease Thrombin (Amersham) was dissolved to obtain a concentration of 1 U/µl and added to the fusion protein sample for a treatment of 10 U/mg fusion protein (approximate). The incubation was performed at 4 ℃ or room temperature for four hours before analysis on SDS-PAGE gels or applied to purification procedure.

2.13.3.5 Purification of GST fusion proteins

The soluble protein applied to the equilibrated sepharose was incubated in batch at 4 ℃ for an hour. In batch purification, the mixture was centrifuged at 500 x g for 5

minutes followed by the removal of supernatant. The gel was washed thrice with 10 bed volume of wash buffer, each centrifuged at 500 x *g* for 5 minutes and transferred to a new tube for further analysis. The protein was eluted thrice with 1 bed volume equilibration buffer containing 100 mM reduced glutathione (Amersham). For column purification, the protein-sepharose mixture was transferred to a Bio-Rad Econo-Pac Chromatography Column after incubation. The purification was performed in the same protocol, but each fraction was collected by draining the column instead of centrifugation.

2.14 Protein concentration

Eluates containing purified fusion proteins were combined and transferred to Centriprep or Centricon centrifugal filter devices (Millipore), depending on the total volume of purified protein solution. The centrifugal filter devices were then cetrifugated at 3000 x *g* for Centriprep or 7500 x *g* for Centricon until the solution volume was reduced to minimum. When the protein solution reached a final concentration of at least 5 mg/ml, the concentrated protein was then transferred to a dialysis tube with clips sealing each end. Dialysis was performed in a total of three litres of Tris 7.5 buffer overnight with changes of buffer every 4-8 hours.

2.15 Protein Crystallisation

The Hanging Drop Vapour Diffusion method was used to perform crystallisation. Cover slips of 22mm in diameter (BDH; VWR International) were first siliconised by placing cover slips in a vacuum device with dichlorodimethylsilane (Sigma) for 2 hours at room temperature. Purified protein samples with a concentration higher

than 5mg/ml were used in crystallisation. Precipitants were prepared from 15-40% $(NH_4)_2SO_4$ and 5-30% PEG 8000 at pH 4.6, 6.5, 7.5, 9.0, or purchased from Hampton Research. 24-well plates (VDX) were used to set up hanging drop crystallisation by adding 1 ml of the precipitant to a well and carefully pipetting 1.5 μl of protein sample next to an equal volume of the precipitant reagent on the cover slip. The cover slip was carefully inverted and placed over the well with grease around the well rim to seal the cover slip onto the well. After assembly, the crystallisation apparatus was kept at 4 °C or 17 °C in a stable environment to avoid vibration. Careful examination of the drops under a microscope was carried out immediately after setting up the screen and once everyday for the first week, then once a week there after.

2.16 SAMase activity assay

A method adapted from Hughes et al (Hughes et al., 1987a) was used to determine the enzyme activity according to the production of the radioactive cleavage product, MTA, in an enzymatic reaction. Appropriate volumes of bacterial protein extracts was added to a total reaction volume of 20 μl containing 10x assay buffer and 3.2 μl of 60 mCi/mmol S-adenosyl-$_L$-[methyl-^{14}C]methionine (Amersham). After an incubation of 30 minutes at 37 °C, the assay tubes were placed in a bath of dry ice and acetone for 10 seconds to stop the reaction. The reaction solution was quickly thawed and 5 μl of the reaction was spotted onto a cellulose sheet (Merck). The reaction products were then separated by thin layer chromatography in a close chamber with a separation solvent comprising 1-butanol:acetone:HOAc:H_2O = 7:7:2:4 (v/v/v/v). When the solvent front was about 5 cm from the top edge of the

sheet, the sheet was taken from the chamber and dried briefly before exposure to a radiosensitive film overnight and detection on a PhosphorImager, Storm 840 apparatus.

2.17 Real-time PCR

20 μl of the cDNA product resulted from reverse transcription was diluted in dH$_2$O at a 1-in-5 ratio to give a final volume of 100 μl. For each real-time PCR reaction, 7.5 μl of the diluted cDNA sample was mixed with 2 μl of appropriate primers (3.2 pmol/μl), 12.5 μl of iQ SYBR Green Supermix (Bio-Rad) and 3 μl dH$_2$O. Reactions were set up in 96 well PCR plates and sealed with Optical Adhesive Covers (Applied Biosystems). Covers were carefully handled without touching the surface, to avoid interference of fluorescent reading in a real-time PCR machine. The sealed plate was placed in the iCycler iQ Multicolor Real-Time PCR Detection System (Bio-Rad) programmed as follow:

Cycle	Temperature (℃)	Time (min/sec)
1	95	1 min
45	95	30 sec
	60	30 sec
	72	15 sec
1	35	2 min
120	35	10 min

The fluorescent reading was monitored at the annealing stage (60℃) when SYBR green fluorescent dye was incorporated into the DNA product. Measurements of mRNA levels were recorded by iCycler program (Bio-Rad) as threshold cycles (Ct). Ct was the cycle number recorded when the system began to detect the increase in

the fluorescent signal resulting from the exponential growth of PCR product. Ct values were analysed by the comparative Ct method, which first normalises the expression of sample input with an amount of internal control and then compares the differences between normalised samples amounts (ABI-7700 User Bulletin #2; reviewed in Livak and Schmittgen, 2001). When normalised to an endogenous reference gene and relative to a calibrator, the expression rate of a target was given by $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct$ was derived as follow:

From the amplification of target gene X and reference gene I in target sample a and control sample b, Ct values of each PCR product were recorded as $Ct_{X,a}$, $Ct_{I,a}$, $Ct_{X,b}$ and $Ct_{I,b}$, respectively.

In sample a, the normalised Ct of gene X,

$$\text{in reference to gene } I, \text{ is } \Delta Ct_a = Ct_{X,a} - Ct_{I,a};$$

in sample b, the normalised Ct of gene X,

$$\text{in reference to gene } I, \text{ is } \Delta Ct_b = Ct_{X,b} - Ct_{I,b}.$$

Using sample b as a control Ct of gene X,

$$\text{the relative Ct of gene } X \text{ in sample } a \text{ is } \Delta\Delta Ct = \Delta Ct_a - \Delta Ct_b.$$

For each sample, the average, standard deviation, and standard error of mean for relative expression levels were calculated from quadruplicated PCR reactions and the result was plotted using Microsoft Excel software.

2.18 Detection of base modification in DNA

DNA for analysis was denatured by boiling for 5 minutes and blotted on Optitran nitrocellulose membrane (Schleicher & Schuell) using a Bio-Dot SF Microfiltration

Apparatus (Bio-rad) under vacuum. The apparatus were first cleaned with a large amount of water and ethanol wash to eliminate DNA contamination. Three layers of Bio-Dot SF filter paper was placed on the gasket support plate followed by a membrane of appropriate size, before screwing the gasket and sample template tightly in position. Appropriate amounts of samples were applied in each slot and bound to the membrane by vacuum. When all liquid had been filtered through the paper, the apparatus was dissembled and the DNA was crosslinked to the membrane by UV crosslinking in a Stratalinker UV crosslinker (Stratagene) using autocrosslink function at an energy of 120,000 μjoule.

The detection of bases was detected in the same manner as western detection of proteins but using appropriate anti-base antibodies as primary antibodies. The membrane crosslinked with DNA was blocked by incubation in blocking buffer for at least 1 hour at room temperature or at 4 ℃ overnight. Primary antibody incubation was carried out in blocking buffer with an appropriate dilution at room temperature for 2 hours. The membranes was washed three times in blocking buffer for 10-15 minutes, before being incubated for 1 hour with an appropriate concentration of secondary antibody (conjugated with horseradish peroxidase) in blocking buffer. Following the secondary antibody incubation, two washes in blocking buffer and one wash in TBS-T were carried out before ECL detection of bound antibody. An equal volume of ECL solution I and II were mixed and then incubated with membranes for 1 minute. Membranes were wrapped in Saran wrap and exposed to ECL Hyperfilm (Amersham) for an appropriate length of time.

2.19 Immunocytochemistry

2.19.1 Immunocytochemistry of cultured cells

Cells for immunofluorescence analysis were grown on 22 x 22 mm cover slips in 6-well plates. After a wash in PBS, cells were fixed in 4% paraformaldehyde for 20 minutes at room temperature. Cells were washed twice in PBS and permeablised using 0.2% Triton-X in PBS for 10 minutes at room temperature. Prior to antibody staining, cells were washed twice in PBS and incubated in 3% BSA in PBS for 45 minutes at room temperature to block non-specific antibody binding. Antibodies were diluted in 3%BSA in PBS. Primary antibody incubations were carried out on parafilm by placing cover slips upside-down on antibody solutions for 1 hour at room temperature. Cells were washed 3 times in PBS for 2-3 minutes each. Cover slips were then placed upside-down and incubated with appropriate secondary antibodies on parafilm for one hour at room temperature but covered in dark to avoid light emission. After the second incubation, cells were washed 3 times in PBS for 2-3 minutes each and dried briefly before the cover slips were mounted onto slides using Vector Shield Mounting Medium with DAPI (Vector Labs). Slides were sealed with nail varnish and stored at 4 ℃.

2.19.2 Immunocytochemistry of embryos

Animal protocols leading to the harvesting of embryos were performed by Dr. Jim Selfridge. After washing in M2 medium, mouse embryos were incubated in M16 medium at 37 ℃, CO_2 incubator for 4 hours before fixing in 2% paraformaldehyde at room temperature for 20 minutes. Fixed embryos were permeabilised with 0.5% Triton-X in PBS-T followed by DNA denaturation in 4 N HCl at 37 ℃ for 30 minutes.

After washing in PBS-T for 30 minutes at room temperature, embryos were then blocked in 1% BSA in PBS-T for 1 hour at room temperature. Antibodies were diluted in 1% PBS-T and spun for 10 minutes at 14,000 rpm before use. Incubations of primary antibodies were carried out at 37 ℃ for 1 hour. Before secondary antibody incubations, the embryos were washed in PBS-T for 30 minutes at room temperature. Secondary incubations were carried out at room temperature for 1 hour in the dark. Before fixing embryos on slides, another wash in PBS-T was performed for 30 minutes at room temperature. Embryos were mounted onto slides using Vectashield Mounting Medium with propedium iodine (Vector Labs). Slides were sealed with nail varnish and kept at 4 ℃.

2.19.3 Microscopy

Slides were examined under appropriate magnifications using Zeiss Axioskop 2 microscope coupling with IPLab imaging program or Olympus IX70 coupling with DeltaVision Spectris Restoration Microscope system.

2.20 siRNA transfection

Double-stranded siRNAs targeting Dnmt1 were designed following the online siRNA user guide (http://www.rockefeller.edu/labheads/tuschl/sirna.html; Elbashir et al., 2001b) and synthesised by the QIAGEN custom siRNA service. Sequences chosen for targeting are listed in Table 2.6. Male mouse tail fibroblasts subjected to siRNA treatment were seeded at 50% confluence and grown for 16-24 hours in alpha MEM medium supplemented with 10% calf bovine serum and 10 units/ml penicillin-streptomycin solution. The next day, cells were transfected with

siRNA(s) at an appropriate concentration, using Oligofectamine Reagent (Invitrogen) following manufacturer's instructions. Cells were passaged 48 hours afterwards and further transfected once before being harvested for analysis.

2.21 Codon usage analysis

The codon usage table for human and bacteriophage T3 (Appendix II) was acquired from the Codon Usage Database (www.kazusa.or.jp/codon/; Nakamura et al., 2000). Codon usage analysis was carried out using European Molecular Biology Open Software Suite (EMBOSS; Rice et al., 2000) on an internet platform maintained by Centre for Computational Research, the State University of New York (http://bioinformatics.ccr.buffalo.edu/biotool/EMBOSS/Nucleic/codon.html). For both analyses (cai for CAI and chips for Nc), the input sequence was the coding region of the T3 SAMase gene (Appendix I) and the codon usage Table was selected as "Ehuman.cut."

Table 2.6: Oligonucleotide sequences for siRNAs used in this thesis.

siRNA	Targeting sequence (5'-3')	Description
D1-4	AAGTCGGACAGTGACACCCTT	targeting exon 4
D1-18	AATCAGTGGTGGCTCAGTGGCTT	targeting exon 18
D1-22	AAGTGCAAGGCGTGCAAAGAT	targeting exon 22
D1-33	AACTTCGTGTCCTACAGACGC	targeting exon 33

Results

Chapter 3
Studies of a bacteriophage protein: SAMase expression in bacteria

3.1 Introduction

The unique SAM-cleaving property of SAMase has inspired various attempts to reduce DNA methylation levels in cells. In order to facilitate the later application of SAMase as a demethylating agent, it may be helpful to gain a good understanding of the enzymatic activity and mechanism of SAMase. However, little is known about SAMase, especially in the context of the phage antirestriction system. Therefore, in addition to the application of SAMase in mammalian cells, purification of SAMase is first carried out to facilitate the potential use of SAMase.

After the discovery of the SAM cleaving activity, two forms of SAMase have been purified (Spoerel and Herrlich, 1979). One is recorded as SAMase A, with a size of 17 kDa; the other is called SAMase B, which is a complex of SAMase A plus a bacterial host protein of 49 kDa. In addition to the major 17 kDa peptide, SAMase also has a minor form of 20 kDa. However, both the 20-kDa SAMase peptide and the SAMase B complex appear to be suboptimal products resulting from phage infection at the stationary phase. As the biological function of interest is actually carried out by SAMase A, hereafter SAMase is termed to represent SAMase A. Hughes *et al.*

(1987b) used the 0.3 gene of bacteriophage T7 as a template to map the gene encoding the SAMase activity from T3. They isolated a gene fragment 600-bp long, including the T3SH gene of 459 bp and adjacent sequences containing RNA polymerase binding sites. Using protein and DNA databases including GenBank, dbEST and SwissProt to perform BLAST searches, no sequence currently shows any degree of similarity to T3 SAMase or its encoding gene.

Biochemical analysis has shown that the enzyme SAMase uniquely hydrolyses SAM to methyl-thio-adenosine (MTA) and homoserine with a K_M for SAM being 0.2 mM and an optimal reaction pH at 7.0 (Gefter et al., 1966). SAMase tolerates ionic conditions up to 2.5 M salt but is inactivated at temperatures above 40℃. SAMase functions at a rate proportional to the substrate concentration when the SAM concentration is below 5×10^{-4} M. The enzyme activity is inhibited by Mn^{++}, Cu^{++}, p-chloromercuribenzoate, but is unaffected by Mg^{++} or EDTA. Reagents that can modify cysteine residues in proteins were also found to inactivate SAMase activity. In contrast, the presence of sulfhydryl groups enhances SAMase activity (Gefter et al., 1966; Spoerel and Herrlich, 1979). Apart from the information presented above, little is known about the mechanism of SAMase activity, nor have there been significant studies of SAMase under physiological conditions, even though it has been more than 30 years since the first demonstration of this function.

In bacteriophage T7, the ocr protein has shown to exhibit similar antirestriction function and gene location as T3 SAMase. Comparisons between T3 SAMase and T7 ocr proteins have revealed that there is no sequence homology in their nucleotide and amino acid sequences (Hughes et al., 1987b), neither is there any report of a similar enzyme inhibitory mechanism. The determined structure of T7 ocr has shown that the highly acidic ocr protein mimics B-form DNA (Atanasiu et al., 2001;

Walkinshaw et al., 2002). With the disguise as DNA, T7 ocr forms a strong affinity between ocr and type I R-M enzymes and consequently out-competes DNA binding of the R-M enzymes (Atanasiu et al., 2002). Moreover, the interaction between ocr and type I R-M enzymes involves the entire surface of ocr and leads to the R-M enzyme being fully wrapped around ocr (Atanasiu et al., 2002). This explains why ocr can defend against all known type I R-M enzymes, regardless of the particular recognition sequence by the R-M enzyme. Conversely, it has been reported that SAMase does not bind to DNA to inhibit endonuclease activity. Instead, SAMase inhibits EcoK consumption of ATP after the endonuclease binding to DNA (Spoerel et al., 1979). Furthermore, the putative pI of SAMase is about 10, which is very basic compared to that of the ocr protein.

It is thus very interesting to know how SAMase overcomes restriction activity and how the SAM cleaving activity works, when a similar function is carried out by the T7 ocr protein which has a very different isoelectric point. It is also very useful for the further application of SAM depleting activity in regulating methylation levels. To explore the mechanism and properties of SAMase activities, we first aimed to acquire homogeneous protein for further structural and enzymatic analyses.

3.2 Expression of SAMase in bacteria

In order to find an optimal expression system for SAMase, the SAMase gene was cloned into various vectors. pJF118HE and pBAD24 vectors both allow native protein expression (Furste et al., 1986; Guzman et al., 1995), whereas pGEX-4T-2 and pET-30b(+) produce recombinant proteins with a fusion tag of GST or 6xHis,

respectively. All four vectors are under control of inducible promoters which only express the protein of interest when the appropriate inducer is present. pJF118HE, pGEX-4T-2 and pET-30b(+) are induced by IPTG, while pBAD24 is induced by the addition of arabinose. Detailed properties of each vector in use are listed in Table 2.2. The T3 SAMase cDNA was first amplified from phage genome and engineered to obtain recombinant constructs expressing native SAMase or fusion SAMase with either an N-terminal GST tag or an N-terminal 6xHis tag, as described in Section 2.11.1 and Table 2.5. The resulting constructs were accordingly named pJF-SH, pBAD-SH, pGEX-SH and pET-SH. Isolated plasmids were individually transformed into the bacterial strain BL21(DE3). Protein expression in BL21(DE3) was induced by either 1 mM IPTG, for pJF-SH, pGEX-SH and pET-SH expression, or 0.02% arabinose, for pBAD-SH expression. After induction, harvested cells were lysed by sonication and fractioned as supernatant and pellet by centrifugation.

The presence of the target protein was examined on SDS-PAGE gels of appropriate percentage, as shown in Figure 3.1. Expression of SAMase in the native form was not clear from either cell supernatant or cell pellet by Coomassie blue staining of the gels (Fig. 3.1A). Expression of His-tagged SAMase from pET-SH expression was also unclear as shown in Figure 3.1B. A protein size of about 17 kDa was expected for both native SAMase and His-tagged SAMase, but no recognizable band was observed at this size. The uninduced samples were indistinguishable from the induced fractions. There was no visible difference between fractions from SAMase constructs and empty vectors either. On the other hand, the protein expression pattern was altered following the expression of GST-SAMase from pGEX-SH plasmid. Unmodified blank vector clearly expressed GST protein in the induced cell supernatant, with a band at the expected size of approximately 26

Figure 3.1: SAMase protein expression was analysed by SDS-PAGE. Native protein expression using pBAD-SH and pJF-SH vectors (A) were examined on 15% SDS-PAGE gels. Fusion proteins expressed from pET-SH (B) and pGEX-SH (C) vectors were examined on 15% and 12% SDS-PAGE gels, respectively. An equal volume of protein samples was loaded and gels were stained with Coomassie brilliant blue for visualization. M represents protein marker with sizes labelled on the left-hand side. + and - indicate cell culture grown with and without expression inducer, respectively. Unmodified vectors, pBAD, pJF, pET30b, and pGEX were also tested to compare protein expression patterns of SAMase expressing and non-expressing cells. Expected sizes of native SAMase or fusion proteins are indicated by the arrows.

kDa (Figure 3.1C, lane 1 and 2). However, recombinant plasmid subjected to IPTG induction did not result in a GST fusion protein in the cell supernatant at the expected size of about 43 kDa. Instead, a GST fusion protein was seen in the cell pellet. This analysis showed that GST-SAMase was overexpressed and present in the cell pellet in the form of inclusion bodies (Figure 3.1C, lane 9).

Inability to observe protein bands on SDS-PAGE gels does not prove absence of protein expression. To analyse if SAMase protein was expressed at a low level on SDS-PAGE gels, enzyme activity assays were carried out, since an antibody against SAMase is not available for sensitive methods such as western blot. Moreover, it is important to express functionally active enzymes. To analyse SAMase activity, which causes hydrolysis of SAM into homoserine and MTA, cell extracts potentially containing SAMase were incubated with a radioactively-labelled substrate of ^{14}C-SAM. The enzyme activity was verified by the presence of radioactive MTA, after its separation from the rest of the reaction system by thin-layer chromatography (TLC). According to previous data, MTA had a relative mobility (R_f) of ~0.68 in the separation solvent, while the uncleaved substrate, SAM, hardly migrated and had an R_f of 0.08 in the same system (Hughes et al., 1987a). All cell supernatants used in the SDS-PAGE analyses were examined along with the cell pellet obtained from pGEX-SH induction, which contained highly-expressed protein in the inclusion body. The assay result is shown in Figure 3.2, where negative controls (Ctrl; lane 1 in Fig. 3.2A, 2B and 2C), induced (+) and uninduced (-) samples containing empty vectors (Fig. 3.2A, lanes 2, 3, 6 and 7; Fig. 3.2B, lanes 4 and 5; Fig. 3.2C, lanes 2, 3, 6 and 7) gave no product spot of ^{14}C-MTA. To the contrary, bacteria transformed with recombinant plasmids all exhibited SAMase activity in the induced cell fractions (Fig. 3.2A, lanes 4 and 8; Fig. 3.2B, lane 2; Fig. 3.2C, lanes 4
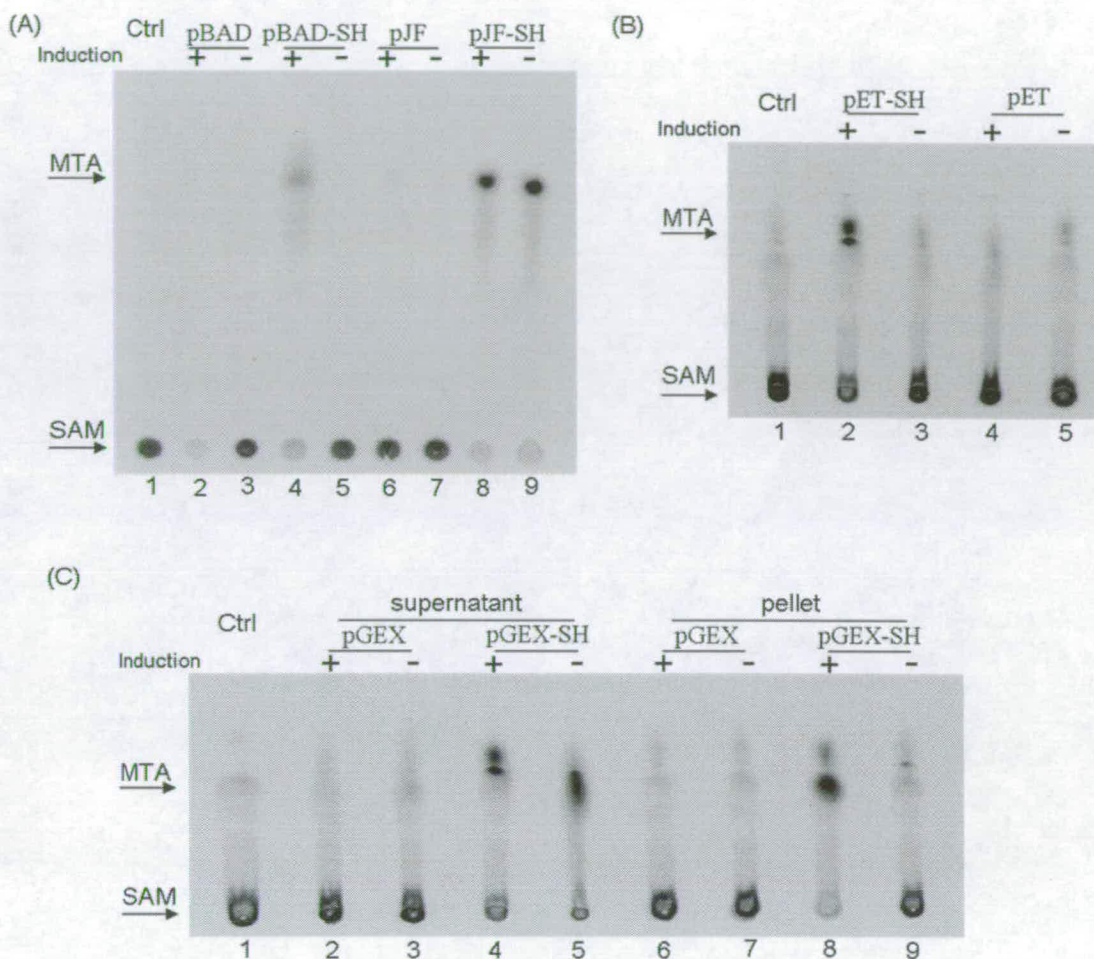
Figure 3.2: Presence of SAMase was determined by enzyme activity assays. Cells transformed with empty vectors and SAMase-expressing constructs pJF-SH and pBAD-SH (A), pET-SH (B), and pGEX-SH (C) were grown with (+) and without (-) the induction of SAMase expression. The supernatants from each expression construct and the resuspended pellet of pGEX-transformed cells were tested for the SAM cleaving activity using radioactively-labelled ^{14}C-SAM incubating with 2 µl protein sample. Each reaction (5 µl) was spotted onto a cellulose sheet and analysed by TLC. The migration of reaction substrate (SAM) and product (MTA) is indicated by arrows. Control reactions, Ctrl, without addition of protein are shown in lane 1.

and 8). Although the total radioactivity observed in Figure 3.2A, lanes 2 and 4 were lower than loadings in other lanes, it is sufficient to see the difference between active and non-active lanes. On the other hand, radioactive MTA was also detected in the incubations containing uninduced pJF-SH and pGEX-SH supernatants (Fig. 3.2A, lane 9 and Fig. 3.2C, lane 5). This may result from a leaky *tac* promoter, as SAMase expression was not observed in the tightly-controlled pBAD-SH under uninduced condition (Fig. 3.2A, lane 5). The SAM cleaving activity of SAMase was also demonstrated in the cell pellet of inclusion bodies, which contained the GST-SAMase fusion protein (Fig. 3.2C, lane 9), suggesting the overexpression of GST-SAMase was achieved in an active form. Although this protein was assayed in the insoluble form, the composition of enzyme activity buffer may have affected the protein conformation and assisted the performance of the enzyme activity. This suggests that it is possible to use this abundant protein as a source for the acquisition of homogeneous active protein.

3.3 Purification of soluble SAMase

Although SAMase was not expressed in visible amounts from both pJF-SH and pBAD-SH constructs, the presence of SAMase in the supernatants of induced cells was confirmed by the enzyme activity assays. Hence, it may be possible to obtain homogeneous SAMase from a large volume of the soluble fractions using appropriate purification procedures. With soluble native SAMase, a previously-published method could be employed to purify SAMase via its affinity for S-adenosylhomocysteine (SAH), a competitive inhibitor of SAMase (Spoerel and Herrlich, 1979). For soluble His-tag SAMase, commercial nickel agarose affinity

chromatography (Ni-NTA) can be used to separate the target protein from the rest of cell extract using the affinitive binding of 6xHis tags to nickle ligands. The same principle applies to the purification of GST-SAMase, which can be purified using glutathione sepharose beads.

3.3.1 Purification of native SAMase protein

A protocol for purification of SAMase has been described in Spoerel and Herrlich (1979). In their work, two slightly different approaches were applied to purify two forms of SAMase, which arose from different infection conditions. Since direct transfection of bacteriophage T3 was bypassed to produce only the 17 kDa SAMase, the method utilised in this study focused on the purification of the so-called SAMase A, which was obtained from bacterial transformation of plasmid pBAD-SH. A summary of the purification procedure is depicted in Figure 3.3A. According to Spoerel and Herrlich, cell lysates were subjected to polymin P anion-exchange resin. This cleaning-up stage uses the cationic polymer to remove DNA from cell extracts while the basic protein of SAMase remains in solution. Proteins bound to polymin P formed precipitates and were removed from the solution by centrifugation. To remove residual polymin P, soluble non-absorbing proteins were precipitated by 80% ammonium sulfate, and the precipitate was isolated and dissolved in buffer A containing 35% glycerol, in order to prevent protein precipitation. After dialysis against buffer A, the dissolved protein was applied to a DEAE-cellulose column. SAMase in the flow-through from the DEAE column was subsequently purified by affinity chromatography to immobilised SAH. After the non-specific binding protein was washed off, homogeneous SAMase was eluted using a buffer containing

adenosine by competition between SAH and adenosine for SAMase binding.

Because SAMase is not highly expressed in the bacterial culture, as demonstrated in Figure 3.1A, a rather large culture will be required to purify a large quantity of protein. To test the efficiency and feasibility of the purification approach using a SAH affinity column, a culture of 1 litre bacteria, transformed with plasmid pBAD-SH, was grown for a pilot experiment and suggested that this method was not efficient, as shown in Figure 3.3B and 3.3C. The resulting protein in the fractions after DEAE cellulose and the SAH affinity chromatography was analysed on 15% SDS-PAGE and showed that the majority of the ammonium sulfate-dissolved protein was recovered in the flow-through fraction (Fig. 3.3B, lane 1). The result of the enzyme activity assay shown in Figure 3.3C shows that SAMase activity was present in the flow-through (Fig. 3.3C, lane 1), which is consistent with the result described in Spoerel and Herrlich (1979). A low activity of SAMase was also found in the eluate (Fig. 3.3C, lane 2). This is contradictory to the finding by Spoerel and Herrlich that no further SAMase activity was eluted from the column. Nonetheless, active SAMase was recovered mainly in the flow-through from the DEAE column. When the flow-through was applied to the SAH affinity column, most of the protein either did not bind to the column or appeared in the wash fraction (Fig. 3.3B, lane 4), but surprisingly this wash fraction contained most of the enzyme activity (Fig. 3.3C, lane 4). Although some SAMase activity was detected in the eluates from SAH affinity chromatography (Fig. 3.3C, lanes 5-9), no protein was detectable by SDS-PAGE analysis (Fig. 3.3B, lanes 5-9).

This result shows that SAH affinity chromatography is not an ideal method for the purification of large quantities of SAMase.
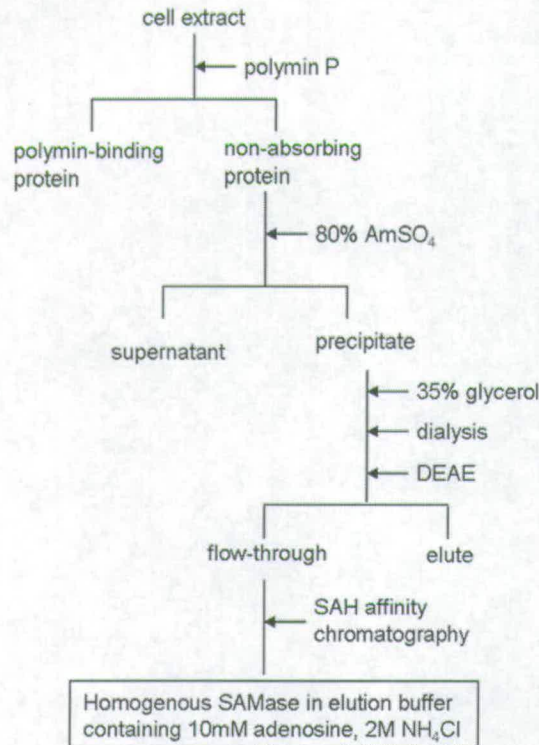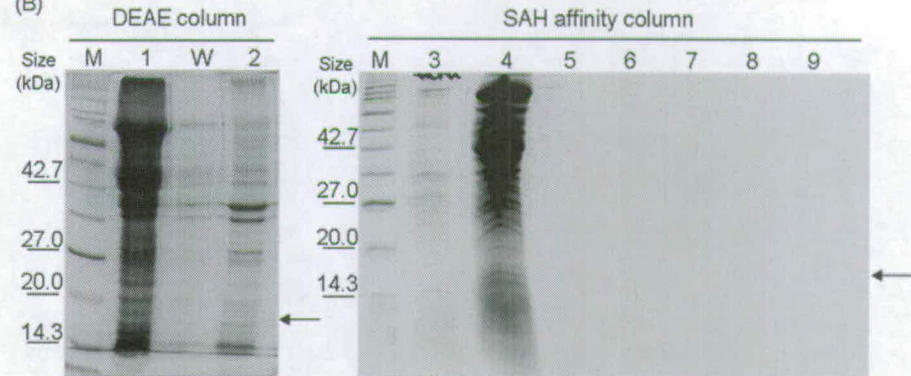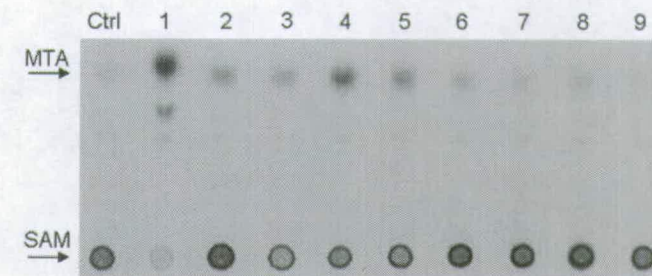
Figure 3.3: Purification of SAMase using SAH affinity column. (A) Purification flowchart of native SAMase purification. Details of the methodology are described in the materials and methods and in the text. (B) An equal volume of protein samples from the flow-throw (1), wash (W), and eluate (2) from the DEAE column and samples from the flow-through (3), wash (4), and eluates (5-9) from the SAH affinity column was analysed on 15% SDS-PAGE gels. Marker (M) sizes are labelled on the left and SAMase is indicated by the arrows. (C) Samples analysed in (B) were tested for the presence of SAMase in the activity assay. Migration of reaction substrate (SAM) and product (MTA) are as indicated. Control reaction without addition of protein is in lane Ctrl. The volume of protein analysed was in proportion to the volume of each fraction.

3.3.2 Purification of His-tag SAMase

In order to purify His-tag SAMase, a one-litre culture of bacteria expressing this fusion protein was inoculated, induced, and harvested. The cell supernatant was obtained from the lysed bacteria and the protein was purified by nickel agarose affinity chromatography, making use of the 6xHis tag at the N-terminus of SAMase. The SDS-PAGE analysis in Figure 3.4 shows a typical purification experiment of His-SAMase on nickel agarose. After a one-hour incubation which allowed protein binding to the nickel-agarose beads, the majority of the protein from the cell supernatant was found in the flow-through fraction (Fig. 3.4A, lane F). As shown in Figure 3.4A, lanes w1-w4, additional non-binding protein was washed off from the agarose beads in the wash buffer. However, the purification is not satisfactory as proteins of various sizes were seen in the four elution fractions but no band was observed at the position of 17 kDa, where His-SAMase would appear (Fig. 3.4A, lanes e1-e4). The presence of His-SAMase was further examined using the enzyme activity assay. Figure 3.4B shows that His-SAMase was present and active in cell supernatant and the flow-throw fraction after affinity binding (Fig. 3.4B, lane S and F). Although the radioactive spots had migrated differently, which may be due to the increasing concentration of imidazole in the buffer, the proteins obtained in the wash and eluation fractions did not include active His-SAMase. Combining these two results, we draw the conclusion that His-tag SAMase is not suitable for purification using nickel-affinity chromatography. The fusion tag of SAMase did not bind to the nickel ligand as expected, and the active fusion SAMase was found in the flow-throw fraction. This may due to an unexpected protein conformation which wrapped the 6xHis tag within SAMase structure. It is also possible that the 6xHis
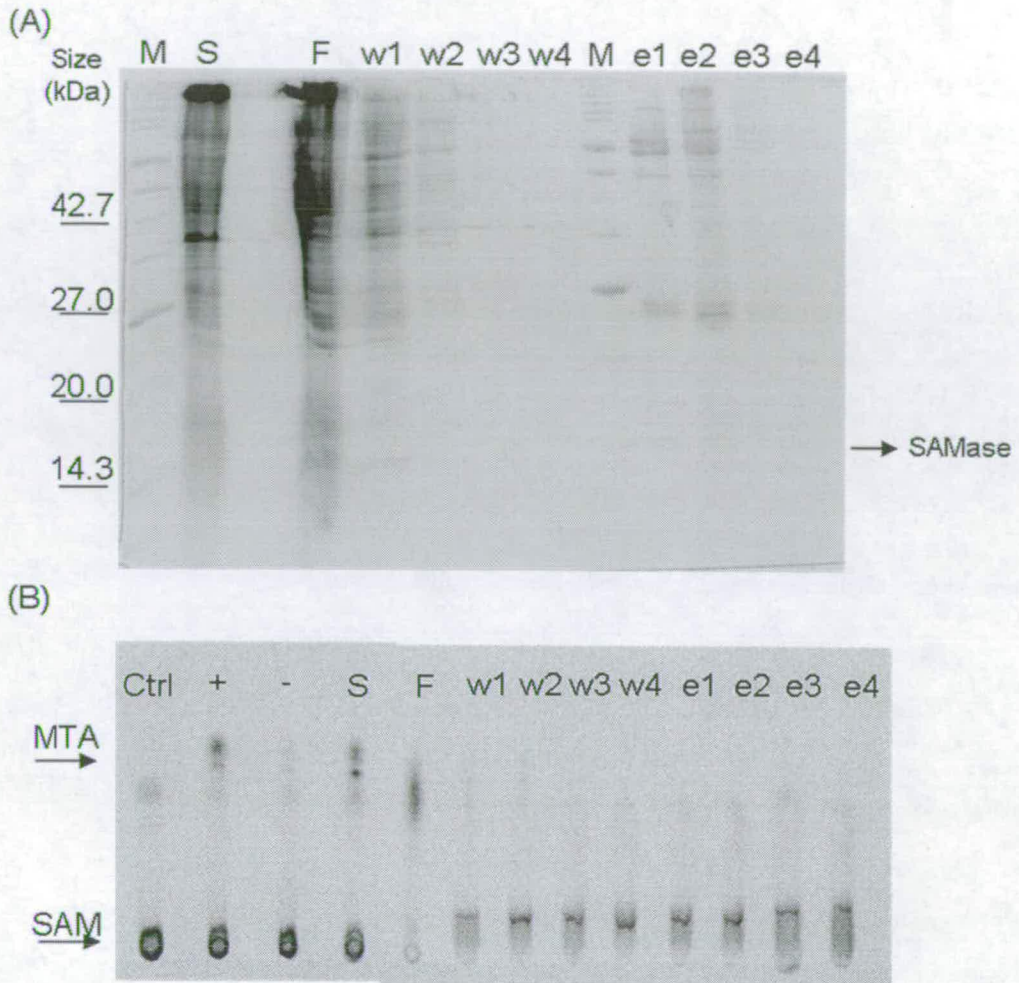
Figure 3.4: Purification of His-tag SAMase using nickel affinity chromatography. (A) Protein fractions from cell supernatant (S), flow-through (F), washes (w1-w4) and eluates (e1-e4) were analyzed on a 15% SDS-PAGE gel. Marker (M) sizes are labelled on the left and GST-SAMase is indicated by the arrow. (B) Protein fractions shown in (A) were tested for the enzyme activity. The change in the appearance of SAM spots in washes and eluates may result from the increasing use of imidazole. Reactions without (Ctrl) additional protein and with cell supernatant expressing (+) and not expressing (-) active SAMase were used as controls. Reaction substrate (SAM) and product (MTA) are as indicated. The volume of protein analysed was in proportion to the volume of each fraction, except that only half of the volume in proportion to wash fractions were used.

tag was no longer covalently attached to SAMase during fusion protein production, although further investigation is required to decide the reasons.

3.3.3 Purification of GST-SAMase

As the purifications of native SAMase and His-tagged SAMase were unsuccessful, purification of GST-SAMase using glutathione affinity chromatography was then tested. However, the majority of GST-SAMase was expressed in inclusion bodies, as shown in Figure 3.1C. This insoluble protein in the cell pellet therefore needed to be recovered in the soluble fraction before this rich source of protein could be utilised in any further analysis.

3.3.3.1 Optimisation of GST-SAMase Expression

To achieve a maximum yield of soluble protein, recovery of the insoluble protein from inclusion bodies is essential, but optimisation of the expression conditions can increase the protein availability. As this protein is expressed only after IPTG induction, the most important factors affecting protein expression are the induction period and the inducer concentration.

As shown in Figure 3.5, changes in these two factors greatly varied the expression of GST-SAMase. Although it was still not possible to identify GST-SAMase expression in cell supernatants, the amount of total protein increased with time at each inducer concentration (Fig. 3.5, lanes 2-10). The same applied to cell pellets, where three hours of induction produced the largest amount of GST-SAMase (Fig. 3.5, lanes 11-19). Changes in inducer concentration did not make a difference in cell supernatants, but it clearly changed the yield within the inclusion bodies. When
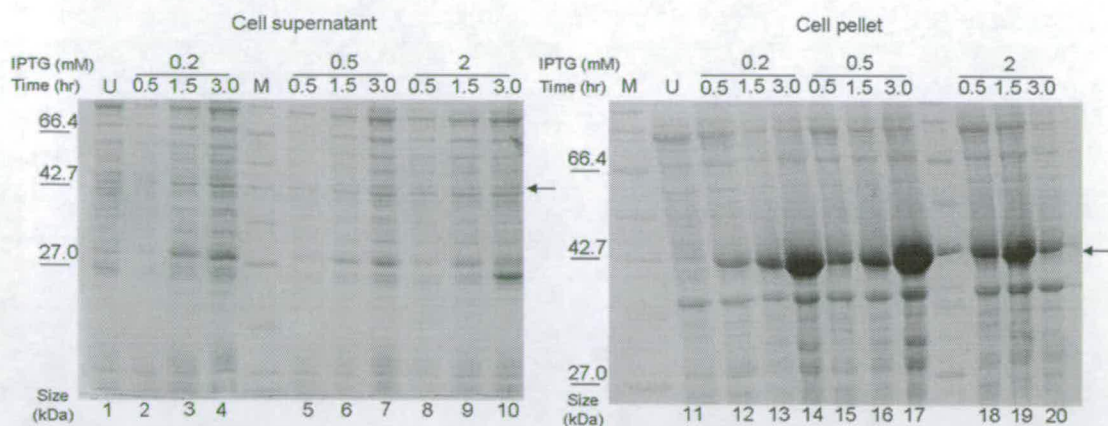
Figure 3.5: Optimisation of GST-SAMase expression. GST-SAMase expression was optimised by changing the length of induction time (0.5-3 hours) and the concentration of inducer (0.2-2 mM). Cell supernatant and pellet collected from each condition were separated and analysed in an equal volume on 12% SDS-PAGE gels. Marker (M) sizes are labeled on the left and arrows indicate where GST-SAMase appears. U represents cell fractions without induction.

expression was induced for 0.5 hour and 1.5 hours, 2 mM of inducer IPTG gave the highest amount of the protein, compared to expression induced by 0.2 and 0.5 mM IPTG (Fig. 3.5 lanes 11, 12, 15, 16 compared to lane 18 and 19). However, 2 mM of IPTG no longer gave the highest protein production when the incubation extended to 3 hours (Fig. 3.5 lane 20). Instead, pellets from 0.2 and 0.5 mM IPTG contained more GST-SAMase (Fig. 3.5 lane 14 and 17). This may be due to the toxicity from prolonged high expression at 2mM IPTG. As a whole, the milder and longer induction, 0.5mM IPTG for 3 hours (Fig 3.5 lane 17), produced the most GST-SAMase and was used for the subsequent GST-SAMase expression.

3.3.3.2 Recovery of soluble GST-SAMase

Production of insoluble protein in the form of inclusion bodies may result from the overexpression of a toxic protein. Host cells form incorrectly folded, insoluble protein and deposit the protein in inclusion bodies to alleviate the toxicity of an

exogenous protein. Although insoluble protein is not useful in the practice of enzymology, it is advantageous for purification because most host proteins are already separated from the target protein. To recover protein in its soluble form, the protein needs to be dissolved in a denaturant, such as 8M urea or 6M HCl-Guanidine. Protein can gradually refold in the correct conformation during removal of the denaturant, which provides an appropriate environment consisting of reducing agents such as dithiothreitol (DTT). Furthermore, altering the ionic strength of the solution, using salt, detergent and a suitable pH can optimise protein solubilisation. However, this simple solubilisation of the protein may be meaningless, if the protein recovered is in an inactive state.

To obtain active protein, the refolding conditions for denatured GST-SAMase, obtained by dissolving insoluble GST-SAMase in 8M urea, was optimised. The denatured protein was subjected to refolding by dilution. A dilution of urea will quickly change the denaturing condition to a composition suitable for refolding. Although refolding generally prefers gentle changes to allow formation of correct disulfide bonds and tertiary conformation, a direct dilution is favourable for its time-efficiency and its avoidance of less favourable structure formation. Since detergent, salt, reducing agents and pH are all possible factors that facilitate protein refolding, refolding systems were chosen to test the following elements in a DTT-based solution: 0.1 % (v/v) deoxycholate, 0.1 % (v/v) Triton X-100, 5mM NaCl, 5 % (v/v) glycerol at four pH values ranging from pH 6.5-9.5. In combination, this gives a total of 48 different refolding conditions to test at two different dilutions, 5 fold and 10 fold, which both reduce urea concentration to an ineffective level. All 96 combinations of dilutions were performed and each reaction was centrifuged to separate soluble and the remaining insoluble fractions. The resulting soluble and

insoluble fractions were analysed on SDS-PAGE gels, and successful solubilisation was detected by transfer of GST-SAMase from the insoluble to the soluble fraction. Two examples of dilution experiments showing the highest level of refolded soluble fractions are displayed in Figure 3.6A and 3.6B. From these two PAGE gels, it is evident that the amount of recombinant protein gradually increased in the soluble fractions (Fig. 3.6A and 3.6B, lanes 1-4) as the pH increased, whereas the GST-SAMase band in the corresponding insoluble fractions gradually disappeared (Fig.3.6A and 3.6B, lanes 5-8). The addition of glycerol did not affect the protein solubility as the amount of soluble protein was approximately the same, when comparing the two fractions in lane 4 of Figure 3.6A and lane 4 of Figure 3.6B, which contained most soluble proteins. Neither salt nor the detergent Triton X-100 had any effect on refolding efficiency (data not shown). After examination of all dilution conditions, the system for futher refinement was chosen to include DTT, deoxycholate and high pH.

A more detailed optimisation of deoxycholate concentration and pH was carried out. Deoxycholate concentrations of 0.01, 0.05 and 0.1% (v/v) were tested at pH 8.5 or pH 9.5. At pH 9.5, all three deoxycholate concentrations gave very high level of protein refolding with slightly less protein in 0.01% deoxycholate solution (Fig. 3.6C, lane 4-6), but at pH 8.5 only 0.1% deoxycholate showed a high amount of the soluble protein (Fig. 3.6C, lane 3) and most of the protein remained in the insoluble form at 0.01% deoxycholate (Fig. 3.6C, lane 7). When comparing pH, a high pH of 9.5 retained more protein in the soluble form at the same level of deoxycholate (Fig. 3.6C, lanes 1 and 4, lanes 2 and 5, lanes 3 and 6). This analysis suggested an optimal refolding condition at a high pH and a high deoxycholate concentration.
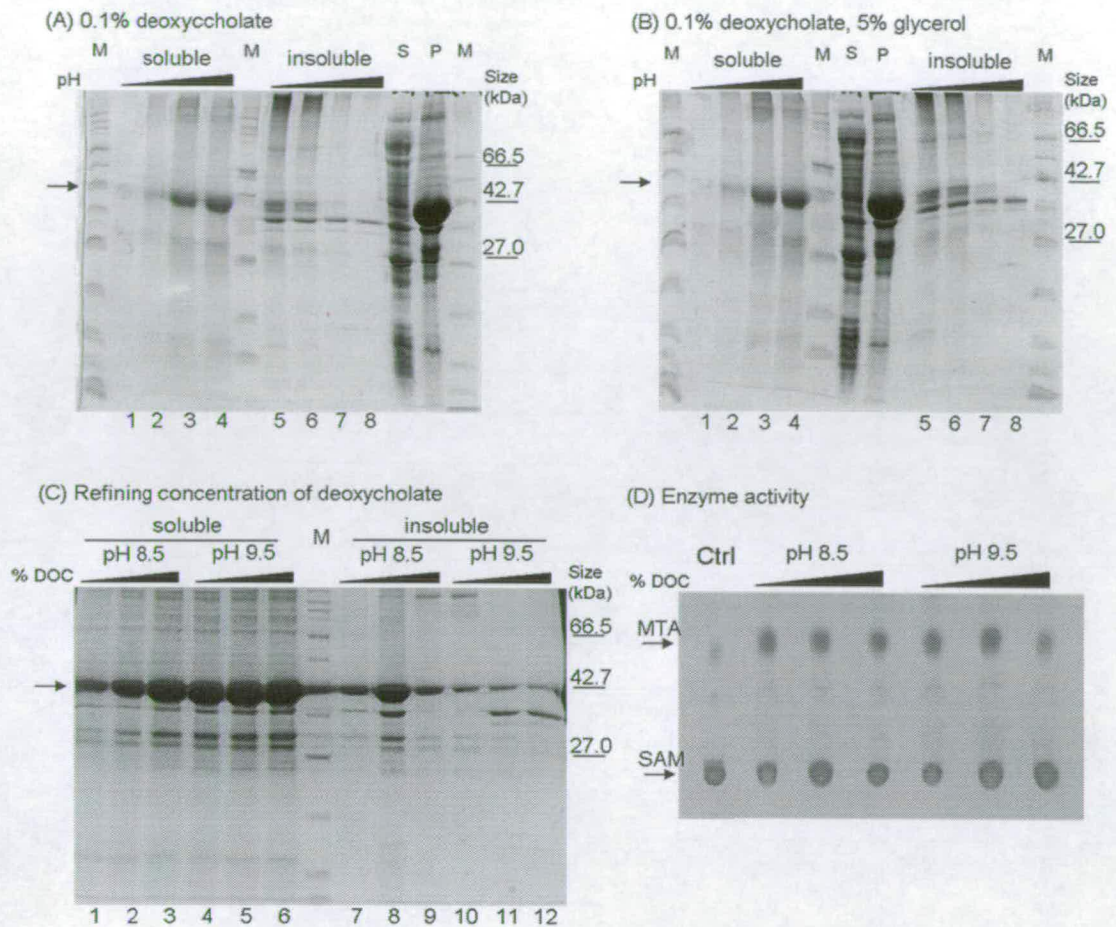
Figure 3.6: Optimisation of conditions for the recovery of insoluble protein. Refolded protein solutions were divided into soluble and insoluble fractions by centrifugation. The starting materials of cell supernatant (S) and pellet (P) were also examined. An equal volume of samples was analyzed on 12% SDS-PAGE gels. Marker (M) sizes are labelled on the right and arrow indicates GST-SAMase. (A) Protein refolding was performed in 0.1% deoxycholate at pH 6.5 (lanes 1 and 5), 7.5 (lanes 2 and 6), 8.5 (lanes 3 and 7), or 9.5 (lanes 4 and 8). (B) Protein refolding was performed in 0.1% deoxycholate with 5% glycerol at pH 6.5 (lanes 1 and 5), 7.5 (lanes 2 and 6), 8.5 (lanes 3 and 7), or 9.5 (lanes 4 and 8). (C) Refolding conditions were also tested by changing the percentage of deoxycholate (%DOC). Concentration of 0.01% (lanes 1, 4, 7, 10), 0.05% (lanes 2, 5, 8, 11) and 0.1% (lanes 3, 6, 9, 12) deoxycholate were tested under pH 8.5 (lanes 1-3 and 7-9) or pH 9.5 (lanes 3-6 and 10-12). (D) Refolded protein from the soluble fraction of (C) was tested for its enzyme activity. A reaction without additional protein was used as a control (Ctrl). Reaction substrate (SAM) and product (MTA) are as indicated.

Following the recovery of soluble GST-SAMase, the enzyme activity of each soluble fraction was examined. The protein in the soluble fractions showed good activity at all refolding conditions, except for the condition using 0.1% deoxycholate at pH 9.5 which showed slightly less activity (Fig. 3.6D). This minor reduction of activity may result from the strong detergent and high pH which may interfere with the enzyme activity by changing composition of the reaction buffer. From the results of the refolding experiments and the enzyme activity assay, a solution system containing 0.05% deoxycholate and 5 mM DTT at pH 9.5 was utilised to refold denatured protein. It is also worth noticing that protein recovered from the inclusion bodies had reached more than 80% purity, as seen in lanes 4-6 of Figure 3.6C. This confirms the merit of inclusion bodies in protein purification.

The presence and activity of the fusion protein was shown in enzymatic assays, and we further used mass spectrometry to provide confirmation of the target protein by virtue of its molecular weight. The predicted weight of GST-SAMase is 43327.07 Da. As shown in Figure 3.7, a peak at the size of 43,623.89 Da representing the fusion GST-SAMase, while the sizes of other peaks matched the proteins seen by SDS-PAGE of the analysed fraction (Figure 3.6C, lane 5). From this analysis, the identity of GST-SAMase was confirmed and further purification to isolate GST-SAMase from the contaminating proteins was carried out. It is worth noting that the difference between the predicted and the empirical weight of SAMase may result from post-translational modifications or detergent adducts in the protein solution. To define the modifications on SAMase, proteolytic treatments such as trypsin digestions may be carried out to produce smaller peptides, which allow a refined resolution of mass spectrometry and provide a precise identification of the molecular weight of SAMase.
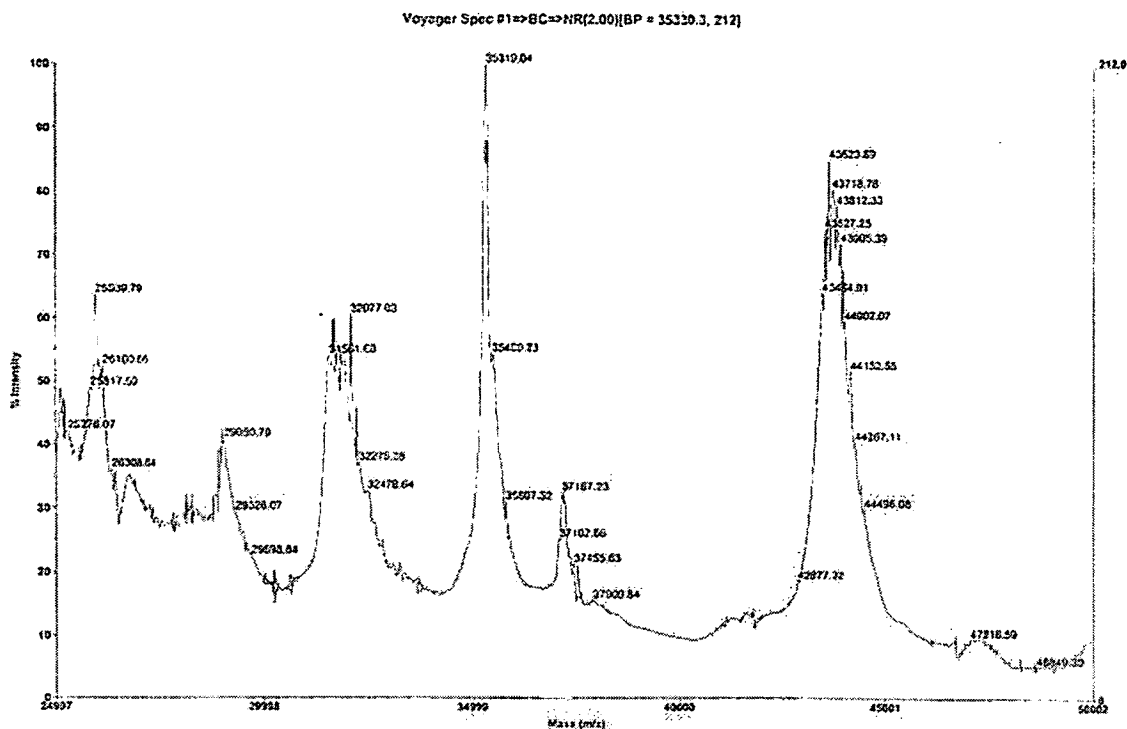
Figure 3.7: Soluble protein recovered from inclusion bodies was analysed by mass spectrometry. The peak of GST-SAMase has a size of 43,623 Da, which is very close to the predicted size of 43,327 Da. Other peaks represent proteins that match the pattern shown in Figure 3.6C lane 5.

3.3.3.3 Purification of soluble GST-SAMase

Purification of proteins containing a GST tag was achieved by application of glutathione sepharose. Glutathione conjugated-sepharose beads allow proteins containing the GST motif to bind glutathione residues and separate protein from non-GST proteins. After removing non-specific proteins, GST fusion protein can be released from the solid glutathione support by the reduced glutathione which competes for GST binding. A test of the purification method given by the manufacturer of the sepharose was carried out in a PBS-based neutral (pH 7.3)

system. As seen in Figure 3.8 lane 4, GST-SAMase was successfully purified using elution buffer of 100 mM reduced glutathione in PBS. However, this protein is more stable in a high pH environment; therefore, purification using Tris 9.5 buffer which contains 0.5 mM DTT, 100 mM EDTA at pH 9.5 was also tested by replacing all the PBS during the process. Although the manufacturer suggests carrying out the procedure below pH 8.4, the purified product (Fig. 3.8, lane 7) demonstrates that GST-SAMase can be purified at a high pH 9.5 which gives higher yield than using buffers based on PBS. In order to obtain a maximal amount of GST-SAMase, the following purification was thus performed at pH 9.5.
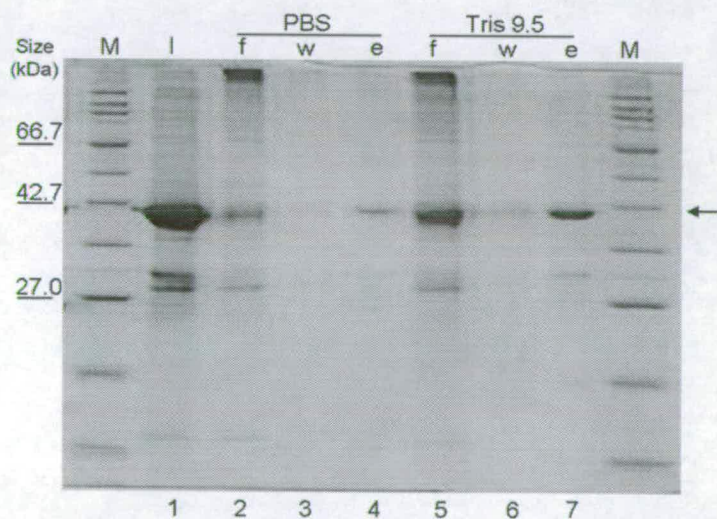
Figure 3.8: Soluble GST-SAMase after refolding was purified by affinity chromatography under two different conditions, either PBS or Tris 9.5 buffers. Samples (10 µl) of input (I; lane 1), flow-through (f; lanes 2 and 5), wash (w; lanes 3 and 6) and elution (e; lanes 4 and 7) were analysed on a 12% SDS-PAGE gel. Marker (M) sizes are labelled on the left and the arrow indicates GST-SAMase.

3.4 Optimisation

3.4.1 GST tag Cleavage

When a protein is tagged, it is preferable to remove the tag in order to analyse the target protein in its native form. The built-in thrombin cleavage site in the pGEX-SH construct allows cleavage between GST and SAMase by incubating the fusion protein in the presence of the protease, thrombin. Although an overnight digestion is recommended by the manufacturer, a four-hour incubation of thrombin at either room temperature or 4°C was carried out. The result of cleavage reaction was analysed on a 12% SDS-PAGE gel. After four hours at room temperature, free GST protein was detectable but no fusion protein or native SAMase was detected (Fig. 3.9, lane RT), whereas at 4°C free GST tag and native SAMase were both detected after four hours of incubation (Fig. 3.9, lane 4°). Although residual GST-SAMase fusion remained after cleavage, suggesting that the digestion was not complete, the presence of the SAMase band shows that the low reaction temperature favoured the stability of SAMase. The cleaved products were used to isolate SAMase from other components. However, SAMase which no longer bound to the column did not appear in the flow-through (Fig. 3.9, F) or wash (Fig. 3.9, W), whereas GST and GST-SAMase fusion protein were both eluted in the elution buffer as expected (Fig. 3.9, E). A very small amount of SAMase was actually present in the eluate too. However, the majority of native SAMase seemed to be degraded during the final purification process, which may be due to a higher instability following the removal of GST. Besides, the purification is not efficient since SAMase was found in the eluate.

The 26kDa GST tag is larger than the 17kDa native SAMase, which makes
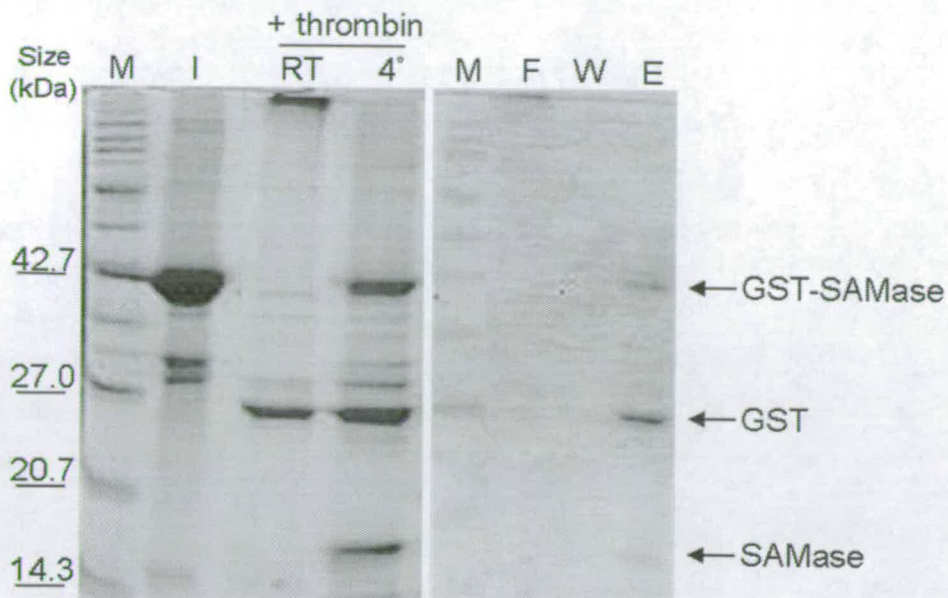
Figure 3.9: Thrombin cleavage of GST-SAMase fusion. Refolded GST-SAMase (I) was incubate with thrombin protease at either room temperature (RT) or 4 ℃ (4°) and analysed on a 12% gel. The reaction performed at 4 ℃ was applied to the glutathione affinity sepharose. The flow-through (F), wash (W), and elution (E) fractions was analyzed on a 12% SDS-PAGE gel. Marker (M) sizes are labeled on the left and the reaction substrate, GST-SAMase, and the products, GST and SAMase, are indicated on the right.

SAMase a minor component of the GST fusion. Nonetheless, the GST tag does not seem to interfere with SAMase activity, as shown by the activity assays. Therefore, GST-SAMase fusion remained uncleaved in the subsequent analysis in order to maintain the maximal stability of the enzyme.

3.4.2 Purification for crystallisation

The purification shown in Figure 3.8 was carried out in batch of 0.1 ml bed volume, which purified the fusion protein to 90% homogeneity. However, a large proportion of GST-SAMase was not bound to the glutathione beads. This may result from an insufficient binding capacity of the gel since a very large amount of input GST-SAMase was applied. Another problem of the pilot purification was that a

small amount of impurity was still detectable on the PAGE gel. This may be solved by increased washes. For the optimised purification experiment, the gel volume was increased to 5 ml and the soluble protein was incubated with the gel at 4 °C for two hours in batch, followed by an extensive washing on a column before elution of the protein. An analysis of the protein fractions is shown in Figure 3.10A. This time, very little GST-SAMase flowed through the column (Fig. 3.10A, lane 2). However, a large amount of GST-SAMase was washed off the column (Fig. 3.10A, lane 6), which suggests that the input protein quantity is too large even for a 5-ml bed volume of glutathione sepharose. However, the GST-SAMase protein in the elution fraction 1 and 2 was very pure (Fig. 3.10A, lanes 3-5). The protein was in a large volume with a low concentration of 0.5 mg/ml, measured by Bradford assay. In order to obtain a highly concentrated pure protein sample, all three fractions of eluates were pooled (Figure 3.10A, lane 8) and concentrated using an ultracentrifugation device. In this instance, a low cut-off size of 3 kDa was used to prevent the loss of the 43 kDa protein while the buffer components could be filtered through. The resulting filtrate (Fig. 3.10A, lane 9) and the concentrated fractions after 30 minutes and two hours (Fig. 3.10A, lanes 10 and 11) of centrifugation were analysed on a 12% SDS-PAGE gel and tested for the enzyme activity (Fig. 3.10B), which demonstrated that the fusion protein retained good activity even after concentration. A small amount of protein contaminants were present in the final concentrated fraction, but GST-SAMase was estimated reaching more than 95% homogeneity (Fig. 3.10A, lane 11). Before an accurate measurement of protein concentration, the concentrated fraction was dialysed against Tris 7.5 buffer which contains 0.05 mM DTT at pH 7.5. The resulting concentrated pure protein was ready for crystallisation.
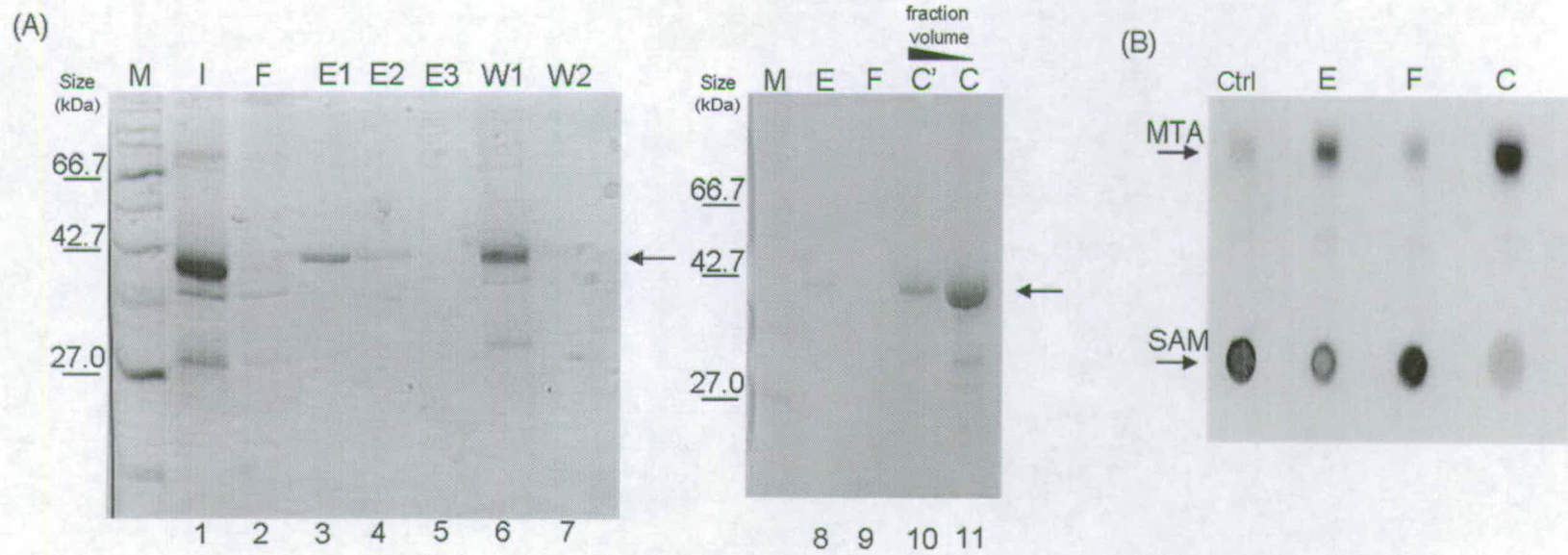
Figure 3.10: (A) The procedure for GST-SAMase purification was optimised by intensive washing (lane 1-7) and the resulting eluate of GST-SAMase was concentrated by ultracentrifugation (lane 8-11). After the soluble protein input (I, lane 1) was applied to the affinity column and the flow-through (F, lane 2) was collected, the beads were washed in 3 bed volumes of wash buffer (W1 and W2, lane 6 and 7) followed by three elution in one bed volume of elution buffer (E1, E2 and E3, lane 3-5). All three eluates (E, lane 8) were pooled and transferred to an ultracentrifugation device with a cut-off of 3kDa. During centrifugation, filtered buffer was collected (F', lane 9) and the concentrated fraction was tested when the volume was reduced to 1ml (C', lane 10) and 300μl (C, lane 11), respectively. All protein fractions were analysed on 12% SDS-PAGE gel with protein size marker (M) labelled on the left. Arrows indicate the position of GST-SAMase on the gel. (B) Filtrate (F) and protein fractions before (E), during (C') and after (C) concentration were tested for the activity of SAMase. Reaction substrate (SAM) and product (MTA) are indicated. Incubation without enzyme is used as a negative control (Ctrl) for the reaction. The samples were analysed in volumes proportional to the fraction volumes.

3.5 SAMase Crystallisation

Proteins are induced to crystallise by creating solvent conditions that result in a supersaturated protein solution which leads to protein aggregation. A protein concentration higher than 5 mg/ml favours this process. Although protein concentrations in the range of 2-30 mg/ml have been used, a pure protein solution of more than 5 mg/ml was aimed at in this study and obtained by the purification and concentration procedure described above. Highly concentrated homogeneous protein is exposed to environments which maximise the degree of supersaturation. Precipitants, such as ammonium sulfate and polyethylene glycol (PEG), are used to achieve the supersaturated state by changing the ionic strength of the solution, drawing the solution from the protein, and reducing the protein solubility while increasing supersaturation near the protein. In addition, commonly used methods allow a gradual decrease in the solution volume, so the supersaturation state still persists when the protein is removed from solution to form crystals. The "hanging drop" method is one such application (Figure 3.11), in which a small drop of the protein sample is "hung" from a glass cover slip over a pool of precipitant. The drop slowly evaporates to equilibrate with the pool below. In this case, the chamber must be sealed and the sample must contain some of the precipitant.

For aggregation of protein molecules to occur in an orderly repetitive fashion, the protein crystals should grow slowly. Trials to screen for the best crystal-forming condition are usually carried out before the time-consuming process of a large crystal formation. The "hanging drop" method is not ideal for growing large crystals due to the low protein amount in a "drop." However, it is ideal for
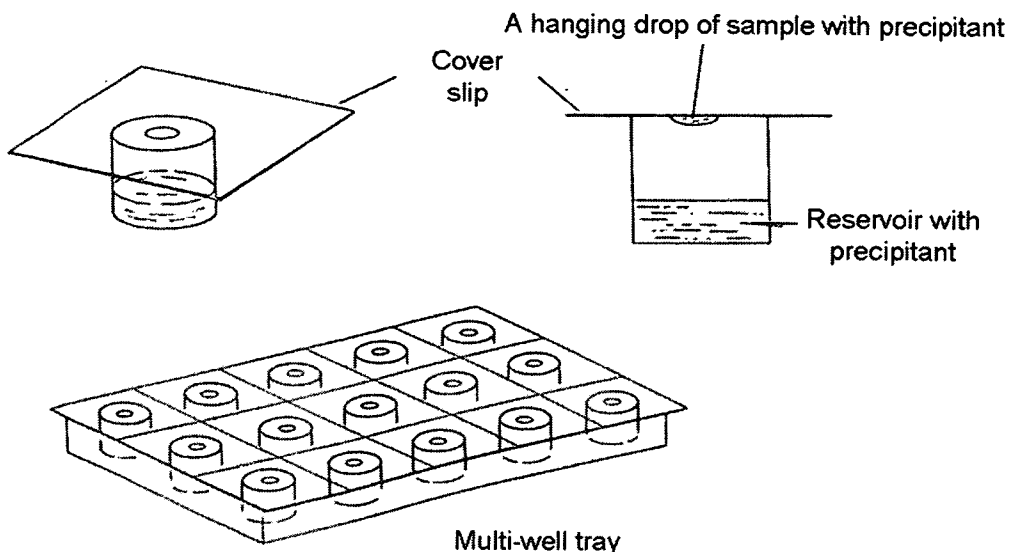
Figure 3.11: "Hanging drop" method for crystallization screen (Modified from Scopes, 1993).

screening since it is very easy to set up a multi-well tray where each well contains one type of test precipitant solution (Figure 3.11). For these reasons, trials with the concentrated homogeneous GST-SAMase was first set up in various precipitant compositions, including 10-60% of saturated ammonium sulfate and 2.5-15% of PEG 4000 (diluted from 50% solution) at pH 4.6-9, in addition to using a Hampton crystal screening kit that provides 50 precipitant solutions. To include precipitant in the sample drop, 1.5 μl of the protein sample was added to 1.5 μl of the precipitant solution with as little mixing as possible. Once the sample was dropped on the cover slip, the cover slip was placed up-side-down over a well containing 1 ml of the corresponding precipitant and sealed with grease. Each trial was observed under microscope immediately after assembly. Amorphous random precipitates of the protein sample were seen right at this point, which was a strong indication of an unsuccessful crystallization. Observations were carried out daily for the first week

after trial set-up, continued weekly for the following four weeks, and once every month afterwards for another 3 months. Protein samples of various concentrations ranging from 3 mg/ml to 7.5mg/ml were tested in screening trials. Unfortunately, no evidence of crystal growth was found, which was most likely due to the rapid precipitation of protein.

3.6 Discussion

In this work, we followed the path of protein purification towards protein crystallisation. Successful expression of SAMase in both native form and as fusion proteins allowed us to purify SAMase for further analysis. It has been reported that depletion of SAM disturbs various biological functions, such as cell division defects (Newman et al., 1998) and a moderate effect on bacterial cell growth and morphology has been observed (Posnick and Samson, 1999). Under the conditions used in our study, we also observed that the presence of SAMase has a mild toxic effect on bacterial cell growth. Uninduced cells and cells expressing only the fusion tags grew about twice as fast as cells expressing SAMase and fusion SAMase (data not shown). This toxicity does not completely hamper cell growth or protein production.

Following SAMase expression, attempts to purify native SAMase were not successful. Both native SAMase and His-SAMase were not successfully purified using published protocols. As most protein was recovered in the flow-through or wash fractions, it is likely that the affinity binding of SAMase is impaired, rather than the stability of SAMase itself. To solve this problem, further analysis needs to be carried out using different purification conditions, i.e. denaturing His-tag protein

to verify the tag conformation in relation to SAMase. In the case of native SAMase purification, although there were minor activities observed in the eluates, a vast amount of input protein yielded only a minimal enzyme activity. The activity observed in eluates (in a loading of 10 μl for SDS-PAGE) came solely from SAMase and the amount of protein in each fraction is just below the 0.1 μg limit for visualisation by eye. A starting volume of 100 litres is required to obtain a protein concentration of 1 mg/ml, which is clearly an inefficient and impractical process. We do not have a control protein sample that can be used to find out whether the failure of purification comes from defects of the SAH affinity column or is due to another aspect of the overall procedure, especially as this method was published more than 20 years ago and materials may have changed since. Another factor that may cause changes in purification efficiency is that we used a plasmid-based expression system rather than a virus infection. This may result in different modifications of the expressed protein and hence alter the biochemical properties for purification. However, further work is required to determine the precise reasons.

The alternative choice of GST-fused SAMase allowed us to obtain purified protein after the recovery of GST-SAMase from inclusion bodies. Insoluble proteins were refolded in a simple dilution system containing deoxycholate, DTT and Tris buffer at a high pH. The refolded soluble protein retained SAMase activity which showed that the protein was refolded into a proper conformation. Nonetheless, after the cleavage of GST from GST-SAMase, degradation of native SAMase was observed during the purification of SAMase from the GST tags. This may be due to the instability of the SAMase protein itself or to unfavourable conditions required for thrombin cleavage. Although native SAMase has been expressed and purified, the actual turnover of SAMase is not yet determined and further analysis is required

to design the optimal purification system for native protein conformation and maintenance of an active protein source.

Protein crystallisation was carried out using the GST fusion protein due to its increased protein stability, but the tag is likely to be the cause of the premature precipitation of the pure GST-SAMase. The structure of GST from blood fluke, *Schistosoma japonicum*, has been solved (McTigue 1995) and preliminary X-ray diffraction results from crystals of GST fusion proteins has been reported (Kuge et al., 1997). Therefore, it would be possible to solve the crystal structure using the GST structure and molecular replacement methods. However, the crystallographic analysis of GST fusions has only been demonstrated in fusions with small peptides of 5-42 amino acids (reviewed in Zhan 2001). In fact, it has long been thought that crystallisation of fusion proteins with large affinity tags is hindered by conformational heterogeneity induced by the fusion tag, making it less conducive to forming a well-ordered, diffracting structure (reviewed in Smyth et al., 2003). Moreover, fusion proteins containing multiple domains are usually too large for NMR studies. Despite this, efforts to improve large fusion protein crystallisation are constantly made. A recent breakthrough has been reported using maltose binding protein (MBP) fused with a 50-residue fragment of the MATa1 protein from *S. cerevisiae* (Ke and Wolberger, 2003). Comparison between this result and that of previous fusion proteins suggests that the linker between the tag and the target protein may be important. The precise nature of the optimal linker remains to be determined (Smyth et al., 2003). Despite the concerns about using the fusion protein for structural analysis, further biophysical analysis using concentrated homogeneous fusion protein can provide information regarding the enzyme activity and mechanism. For example, using purified protein in isothermal titration

calorimeter (ITC), the accurate binding constants, reaction stoichiometry, Gibbs free energy change and enthalpy change can be determined. These results in combination with kinetic studies of the enzyme activity can facilitate the application of SAMase *in vitro* and *in vivo*.

Chapter 4
Artificial DNA demethylation

4.1 Foreword

The conventional laboratory method to reduce DNA methylation is to apply 5-azacytidine (5-azaC) analogs to cultured cells, but 5-azaC analogs have a history of cytotoxicity that results in side effects which includes DNA breakage and chromosomal instability, as described in section 1.5.3. An alternative demethylating agent which avoids the 5-azaC associated cytotoxicity would be potentially useful in elucidating the effects of DNA demethylation, especially if the alternate can bring in long-term modulation of DNA demethylation. With this aim, we intended to take advantage of the SAM depleting property of SAMase in order to reduce SAM level in mammalian cells. Moreover, the application of siRNA may also provide a means of DNA demethylation via reductions in Dnmt1, the major enzyme for the maintenance of DNA methylation.

4.2 DNA demethylation by SAM depletion

4.2.1 Introduction

So far, SAMase is the only known enzyme that can carry out enzymatic SAM

hydrolysis to reduce the level of methylation substrate, SAM, and hence DNA methylation. Some applications of SAMase have been reported utilising this property of the bacteriophage T3 protein. An early report of SAMase expression in the methylation-positive *E. coli* strain JM103 showed that a high level of SAMase was tolerated in the cells (Hughes et al., 1987a). The only noticable phenotype was a tendency of cells to be elongated and occasionally filamentous, which might arise from defects in cell devision. In these SAMase-expressing cells, DNA methylation was down-regulated and led to about 77% inhibition of polyamine synthesis, which demonstrated the feasibility of disrupting SAM-mediated metabolism in living organisms. Later, a few inducible systems were produced to create a SAM-deprived cellular environment. For example, induced expression of SAMase resulted in the reduction of SAM level and consequently the level of DNA methylation in *dam* cells, where potential negative effects would be tolerated (Collier et al., 1994a). SAMase has also been used in the investigation of oxidative DNA damage, because SAM has been suggested to act as a weak alkylating agent *in vivo*. However, there is no obvious contribution of SAM to spontaneous C-to-T mutation rates, when the SAM level was reduced in SAMase expressing cells (Collier et al., 1994b; Posnick and Samson, 1999). In addition, SAMase was taken into practice in the delayed ripening of tomatoes (Good et al., 1994). In the methionine recycling pathway, SAM is converted to aminocyclopropane-1-carboxylic acid and subsequently oxidized to ethylene. Transgenic tomato plants expressing SAMase showed a reduction of SAM and, thereby, ethylene, which in turn delayed the ripening of the tomato fruit. As a whole, these applications gave a good indication that SAMase can mediate alteration of DNA methylation. No application of SAMase in mammalian systems *in vivo* has been reported yet. However, using SAMase to reduce the level of SAM may

provide a useful means in the understanding of effects on DNA demethylation in mammalian cells.

4.2.2 Protein expression strategy

When a bacteriophage gene is intended for expression in mammalian cells, the first question to consider is the synonymous codon usage for the host and recipient organisms. Within the synonymous codons encoding one particular amino acid, each organism has developed usage preferences for a restricted set of codons, the reasons for which are not yet fully understood. Nevertheless, divergence of codon usage bias correlates with evolutionary distance: the "lower" the organism is, the higher the codon usage bias may be (reviewed in Akashi, 2001). This suggests that the codon usage in human is likely to be less biased than the codon usage in bacteriophage T3. Therefore, human codons can translate more required codons for the expression of the bacteriophage T3 gene in human cells.

To confirm whether mammalian codon usage supports SAMase expression, the human codon usage table was first examined to verify if any codon is not at all used. The frequencies of each codon used in the human genome (Appendix II) were acquired from the Codon Usage Database (Nakamura et al., 2000), based on the DNA sequence information from GenBank. No codon of zero frequency was observed in human genome, suggesting that all codons used in the bacteriophage T3 gene can be translated in the human genome.

Next, statistical measurements were carried out to evaluate the likelihood of SAMase expression in human cells. Several methods have been proposed to evaluate the use of the different synonymous codons (reviewed in Comeron and

Aguade, 1998). The two methods utilised in this study were codon adaptation index (CAI; Sharp and Li, 1987) and the effective number of codon usage (Nc; Wright, 1990). CAI estimates the degree of adaptation by the synonymous codons of the gene of interest to the optimal usage in the recipient genome, where values of 1.0 indicate a maximum fit and lower values indicate that the gene contains less preferred codons. Nc represents the effective number of codons used in a gene or genome. An Nc value of 20 indicates the use of only one synonymous codon for each amino acid and a Nc of 61 means a completely uniform use of the different synonymous codons. When referring to a heterologous experimental system, both methods may also provide an approximate indication for the likely success of the expression of a foreign gene. Using bioinformatical programs from the European Molecular Biology Open Software Suite (EMBOSS; Rice et al., 2000), a resulting Nc number of 47 and a CAI number of 0.6 was obtained from the input of T3 SAMase gene sequence in the human codon usage table. This Nc number suggested that the T3 SAMase gene utilised a set of codons that may not be highly biased in the context of human genome. Similarly, the CAI number indicated that the T3 SAMase gene is adequate, but not optimal, for the expression in human cells. These two numbers consistently suggest a likelihood of SAMase expression. Furthermore, successful *in vitro* translation from T3 SAMase mRNA has been reported in a cell-free mammalian system using rabbit ribosomes (Anderson et al., 1976), suggesting the T3 mRNA without eukaryotic stabilising signals, such as 5′ capping, could be translated by mammalian ribosomes. Having considered this information, T3 SAMase expression was carried out using the native T3 DNA sequence.

Due to the potential toxicity of SAMase, an inducible system was chosen to regulate protein expression. For this, the bidirectional expression vector pBI-EGFP

was used in combination with a regulatory plasmid, pTet-OFF. In cells transfected with pTet-OFF plasmid, the regulatory protein tTA is expressed to bind to the Tetracycline Response Element (TRE), when a TRE containing plasmid, such as pBI-EGFP, is cotransfected. The TRE responds to the tTA regulatory protein and allows protein expression to persist unless the regulatory system is differentially turned off according to the dose of the system inhibitor, such as tetracycline analogs. Additionally, the pBI-EGFP expresses bidirectionally from two human cytomegalovirus (CMV) promoters flanking the central TRE in reverse orientations. Simultaneous regulation of the two promoters by the central TRE allows synchronised expression of the downstream genes. In pBI-EGFP vector, one of the CMV promoters has a downstream EGFP (enhanced green fluorescent protein) gene, whereas the other promoter has downstream sites for the insertion of the gene of interest. Thus, one can monitor the transfection efficiency and the expression levels of the target protein in relation to the expression of EGFP.

Additional features were introduced to facilitate protein expression and to enable direct detection of SAMase expression. The resulting construct, named kFNSH, is depicted in Figure 4.1. Firstly, a sequence upstream of the coding region was converted to a Kozak consensus in order to increase translation initiation efficiency. A part of Kozak consensus (5'-CCACC-3') was inserted to locate the adenine at the -3 position from the translation start site (+1). The presence of adenine has been shown to be the most important mark for translation initiation (Kozak, 1999). Following the initiation methionine, an octapeptide FLAG epitope was introduced to permit detection of the FLAG-SAMase fusion protein by immunoblotting or immunocytochemistry. A nuclear localisation signal (NLS) was
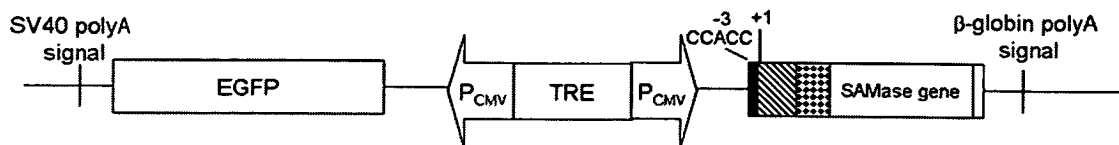
Figure 4.1: Schematic description of the expression construct, kFNSH. The plasmid was designed to include sequences that facilitate protein expression and identification of the fusion SAMase in mammalian cells. Purposes and details of these insertions are stated in the text or in section 2.5.1 and 2.12.1. Vertical lines indicate polyA signals. The regulatory element TRE, CMV promoters, the SAMase gene, and the EGFP gene are shown as indicated. The Kozak sequence is shown in black; the FLAG tag is hatched; the NLS is shown in dots and the 6xHis tag is shown shaded in grey. The diagram is not drawn to scale.

then engineered into the construct at the N-terminal of the SAMase gene. The NLS serves to direct the SAMase fusion protein into the nucleus, in order to reduce the pool of SAM used by DNA methyltransferases. In addition, a 6xHis motif was attached to the C-terminus of SAMase in order to carry out protein purification, if desired. DNA sequences of all features and the construction of the recombinant plasmid are described in Table 2.4 and Section 2.12.1.

4.2.3 SAMase expression in mammalian cells

To examine SAMase expression in mammalian cells, the Tet-Off system was first utilised without drug addition. The expression plasmid kFNSH and the regulatory plasmid pTet-Off were co-transfected into a human embryonic kidney cell line, HEK293, using the calcium phosphate precipitation method. Transfected HEK293 cells were grown for 48 hours before harvest. A whole-cell extract was prepared and subjected to immunoblotting with mouse anti-FLAG antibody. As shown in Figure

Figure 4.2: Analysis of FLAG-SAMase expression in human HEK293 cells. (A) Western blot analysis of the FLAG-SAMase expression. Protein extracts were prepared from cells with (+) or without (-) transfection. Plasmids used for transfections were pBI-EGFP (Vector), kFNSH, or FLAG-Mbd1 (FLAG-M1). Total protein (40 µg) was loaded in each lane and the protein was separated on a 15% SDS-PAGE gel. Arrows indicate positions of the expected proteins on the blot. The primary antibody was anti-FLAG M2 and the secondary was HRP-conjugated anti mouse IgG. FLAG-Mbd1 transfection was used as a positive control. (B) Expression of SAMase was examined by immunofluorescence. SAMase expression in cells transfected with empty vector or kFNSH was detected with anti-FLAG (red) antibody. DNA (blue) was counterstained with DAPI, whereas EGFP (green) automatically emits for detection. All images were taken under 100x magnifications. *(Continued on next page.)*

4.2A, SAMase expression was not detected in cells transfected with the kFNSH plasmid, which looked the same as negative controls using protein extracts from non-transfected cells and cells transfected with an empty pBI-EGFP vector. However, cells transfected with a FLAG-Mbd1 expression construct (Jorgensen et al., 2004) showed positive expression of the FLAG-Mbd1 protein which has a size of 83 kDa. This provided a positive control for the transfection method and western blotting.

To confirm that the negative result was not due to the absence of kFNSH plasmid in the transfected cells, an immunofluorescence experiment was carried out using HEK293 cells transfected with the empty pBI-EGFP vector (Fig. 4.2B, top panel) or kFNSH (Fig. 4.2B, bottom panel). Successful transfection of both plasmids was readily detectable by the green signals from EGFP in cells. The presence of EGFP stained with DAPI suggested the transfection of target plasmids was successful and the regulatory system functioned to express EGFP in cells. However, cells positively transfected with kFNSH expressed exclusively EGFP. Signals from FLAG-SAMase were not observed in the nucleus of EGFP expressing cells, suggesting that the protein of interest was not expressed at a detectable level.

The EGFP expression showed that the plasmid was transfected and functional, the level of SAMase mRNA was thus examined to verify if transcription was

Figure 4.2: *(Continued).* (C) RT-PCR analysis of SAMase transcription. Cells with (+) or without (-) transfection of pBI-EGFP (Vector) or kFNSH plasmids were harvested for RNA extraction. After reverse transcription, the resulting cDNA was subjected to PCR using primer sets for SAMase gene, EGFP gene and *Gapdh.* Positive PCR control for the SAMase gene and the EGFP gene is plasmid kFNSH. Negative control for PCR used H_2O to replace sample cDNA. Negative control for RT reaction (-RT) is shown in the additional panel of kFNSH transfection, which is in comparison to the PCR product of normal reverse transcpription (+RT). Lane M represents marker. Arrows indicate PCR products for the SAMase gene, the EGFP gene and *Gapdh.*

performed. Using RT-PCR, the transcript of the SAMase gene was detected (Fig. 4.2C). A weak band of amplified SAMase cDNA was seen only in cells transfected with kFNSH, but not in cells transfected with pBI-EGFP or non-transfected cells. Meanwhile, EGFP cDNA was evident in cells transfected with pBI-EGFP vector and kFNSH, whereas the internal control of Gapdh was amplified in all cells. Since the RT-PCR does not amplify across any intronic region, a negative RT reaction was performed to distinguish RT-PCR product from DNA contamination. The blank lane (-RT) shown in the right panel of Figure 4.2C confirmed that the PCR product from cells transfected with kFNSH was not amplified from plasmid or integrated SAMase DNA. These experiments suggested that the SAMase gene was transcribed but not at a high level, despite its strong CMV promoter.

From the result of immunoblotting, immunocytochemistry, and RT-PCR, the expression of SAMase protein using the native T3 DNA sequence was not readily observed, although the SAMase gene was transcribed in cells. These results partially suggested that the protein is not readily produced in an in vivo system using a human cell line. However, more experiments are required to reach a comprehensive solution for the expression of SAMase in mammalian cells. Further investigation will be discussed in section 4.4.1.

4.3 DNA demethylation by Dnmt1 mRNA depletion

4.3.1 Introduction

Another approach to investigate the effect of DNA demethylation is to tackle Dnmt1, the major enzyme performing maintenance DNA methylation. The mechanism of

RNA interference drives short double-stranded RNA (dsRNA) to the target transcript which results in its destruction. Using small interfering RNA (siRNA) targeting the *Dnmt1* mRNA, short dsRNA may serve to bring about DNA demethylation by decreasing the level of *Dnmt1* mRNA. As discussed in section 1.5.3, siRNA targeting of DNMT1 has been described and a profound down-regulation of DNMT1 by at least 80% was observed in human cell lines (Leu et al., 2003; Robert et al., 2003). The application of siRNA mediated down-regulation of Dnmt1 has not yet been reported in mouse cells.

To study the effect of DNA demethylation using *Dnmt1* siRNA, the expression of the *Xist* gene was chosen as a readout following such treatment. As discussed in section 1.3.3, DNA methylation has an important role in the regulation of X chromosome dosage compensation. Although the mechanism of X inactivation is not yet fully understood, it has been shown to be initiated by the product of the *Xist* gene (reviewed in Cohen and Lee, 2002). Furthermore, DNA methylation in the promoter of *Xist* is required for the silencing of *Xist* expression (Beard et al., 1995; Norris et al., 1994; Panning and Jaenisch, 1996). The *Xist* gene was shown to be heavily methylated at the CpG island in the promoter region of the non-expressing allele in female cells and in non-expressing male cells, but unmethylated at the expressed allele of the inactive X in female cells (Allaman-Pillet et al., 1998; McDonald et al., 1998). Thus, the level of DNA demethylation may be reflected by the level of *Xist* expression. In addition, the binding of a methylated DNA-dependent protein to the methylated *Xist* promoter has been implicated in the regulation of *Xist* transcription (Huntriss et al., 1997). Recently, the methyl CpG binding protein, Mbd2, has been shown to influence *Xist* expression (Barr and Bird, unpublished observations). In male *Mbd2*[-/-] mouse fibroblasts, the expression level of

the *Xist* gene is elevated by 3-fold in comparison to wildtype male cells, whereas the absence of Mbd1, MeCP2 or Kaiso does not affect *Xist* expression. The additional treatment of cells with TSA, a histone deacetylase inhibitor, strongly enhanced the effect of *Xist* derepression in *Mbd2*$^{-/-}$ cells, in comparison to the wild type. These observations further extend the role of DNA methylation in X inactivation, via Mbd2 mediated transcriptional repression. However, it is not clear whether any other methylation-dependent factor is also involved in the repression of the *Xist* gene. Therefore, *Xist* expression was additionally examined in male *Mbd2*$^{-/-}$ mouse cells in order to see an effect of DNA demethylation and from which we may reveal if any other protein candidates is involved in the maintenance of *Xist* repression via DNA methylation.

4.3.2 *Dnmt1* siRNA down-regulates *Dnmt1* expression

Three siRNAs, D1-18, D1-22, and D1-33, were designed to target *Dnmt1* mRNA at exon 18, 22, and 33, respectively (Fig. 4.3, *upper panel*). In order to evaluate the effects of these siRNAs on the depletion of *Dnmt1* mRNA, quantitative RT-PCR was carried out. The relative expression level was calculated following the comparative threshold cycle (Ct) method described in section 2.16. The cells were exposed to siRNA for 96 hours. This treatment period was chosen to allow turnover of existing Dnmt1 protein and sufficient removal of the established DNA methylation via cell replication. After total RNA extraction, successful production of reverse-transcribed cDNA was confirmed by PCR, using primers which amplify across intronic regions of genomic DNA in order to distinguish amplicons of mRNA from that of genomic DNA. As shown in the middle panel of Figure 4.3, bands of 124 bp and 99 bp were
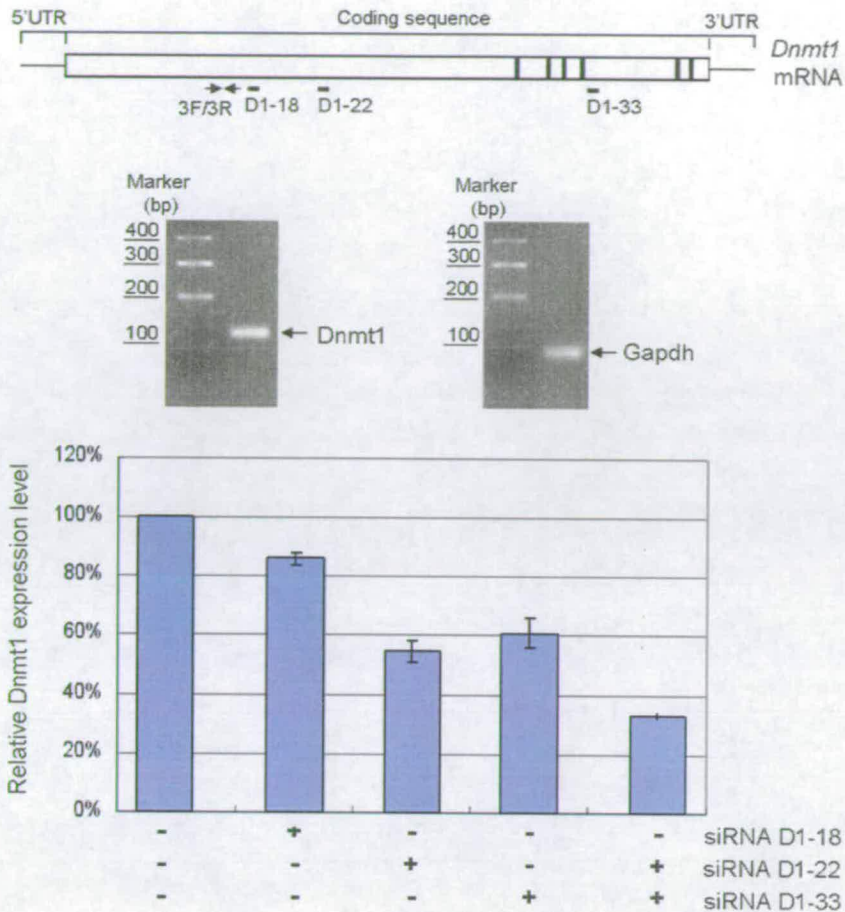
Figure 4.3: Analysis of siRNA mediated silencing of *Dnmt1* expression in wildtype male mouse fibroblasts. *Upper panel*: Schematic outline of the *Dnmt1* mRNA with the positions of PCR primers, and siRNAs, D1-18, D1-22, and D1-33. The position of primer set 3F/3R is indicated by arrows. The positions of siRNAs are indicated by thick lines. Vertical bars indicate the locations of the catalytic motifs. *Middle panel*: Gel images of RT-PCR products. Untreated wildtype mouse cells were harvested for RNA extraction and reverse transcription. The resulting cDNA was subjected to PCR using primers for *Dnmt1* or *Gapdh*. The expected PCR product is 124 bp for *Dnmt1* cDNA (left) and 99bp for *Gapdh* (right), as indicated by arrows. *Lower panel*: Quantitative analysis of *Dnmt1* expression level using real-time PCR. Three siRNAs, D1-18, D1-22 and D1-33, were individually tested for their silencing effect at 40 nM for 96 hours. The combined treatment of D1-22 and D1-33 was also examined at 40 nM final concentration for 96 hours. Plus and minus signs indicate cells receiving (+) or not receiving (-) the assigned siRNA. The level of *Dnmt1* mRNA in untreated cells was set to 100%. All samples were examined in quadruplicates and normalised against *Gapdh* before comparison to the untreated sample. Error bars represent the standard error of the mean.

the only amplicons that resulted from the reverse-transcribed *Dnmt1* mRNA and *Gapdh* mRNA, respectively. No primer dimers or cross-intronic DNA products were observed from the result of RT-PCR, demonstrating the PCR specifically amplifies the target cDNA only and the recorded fluorescent signals in the quantitative analysis will faithfully reflect the expression of the target gene. The transcript level of *Gapdh* was used as an internal control and the silencing effects of *Dnmt1* siRNAs were presented as expression relative to the *Dnmt1* mRNA of untreated samples.

As shown in the lower panel of Figure 4.3, cells treated with D1-18 siRNA only result in a 10% reduction of the *Dnmt1* mRNA, whereas the treatment of D1-22 or D1-33 siRNA each reduced *Dnmt1* to 50% and 40% of the control level. Furthermore, the effect of siRNAs used in combination was tested. Although enhanced silencing using a pool of siRNAs has not been clearly shown, this is often implied by the long dsRNA-based silencing systems that give multiple siRNAs after cellular cleavage, such as observed in plants and *D. melanogaster* (reviewed in Dorsett and Tuschl, 2004b). Combining the two more effective siRNAs, D1-22 and D1-33 resulted in depletion of *Dnmt1* mRNA by 70%, compared to the untreated control. Although the level of *Dnmt1* down regulation varies between individual experiments, the mixture of D1-22 and D1-33 siRNAs is always the most effective of all four siRNAs treatment and gives a 60-70% depletion of the *Dnmt1* mRNA. Therefore, the following experiments focused on the silencing effect using siRNA mixtures of D1-22 and D1-33.

4.3.3 Regulation of *Dnmt1* expression and its effects

4.3.3.1 Effects of *Dnmt1* siRNA and TSA on *Dnmt1* expression in

Mbd2$^{-/-}$ cells

In order to confirm that the observed effect of mRNA depletion was not cell line-specific, the most efficient treatment of siRNAs was applied to both wildtype and *Mbd2$^{-/-}$* male cells. The use of *Mbd2$^{-/-}$* male cells is also advantageous for the examination of potential methylation-dependent factors other than Mbd2 involved in the repression of the *Xist* gene. The treatment was carried out in triplicate, in order to use one sample each for the examination of the corresponding mRNA, protein and DNA methylation levels. As shown in Figure 4.4, *Dnmt1* expression was down-regulated in both cell types after siRNA treatments. In this experiment, a reduction of *Dnmt1* mRNA to about 40% of the control cells was observed in the siRNA-treated wildtype cells. Surprisingly, a low level (35%) of *Dnmt1* mRNA in untreated Mbd2 null cells was observed, suggesting that the expression of *Dnmt1* may be cell-line dependent. Following siRNA treatment, *Dnmt1* expression in *Mbd2*-null cells reached an even lower level, but this was only a 0.7-fold down-regulation in comparison to the untreated *Mbd2$^{-/-}$* male cells. This result showed that a depletion effect by siRNA was generally observed, although there may be a cell line-specific efficiency of the siRNA treatment as shown by the 60% and 30% depletion of *Dnmt1* mRNA in wildtype and *Mbd2*-null cells, respectively. However, the reduced depletion efficiency may result from a lower tolerance to a low *Dnmt1* level that selectively retains cells with a less efficient treatment.

As *Xist* expression will be examined for the effect of DNA demethylation, the synergistic effect of DNA methylation and histone deacetylation on *Xist* repression has drawn attention to see if there is also a synergistic effect of *Dnmt1* siRNA and TSA on *Xist* expression (Csankovszki et al., 2001; Keohane et al., 1998; Panning and Jaenisch, 1996). Both *Mbd2$^{+/+}$* and *Mbd2$^{-/-}$* cell lines were first examined for their

.
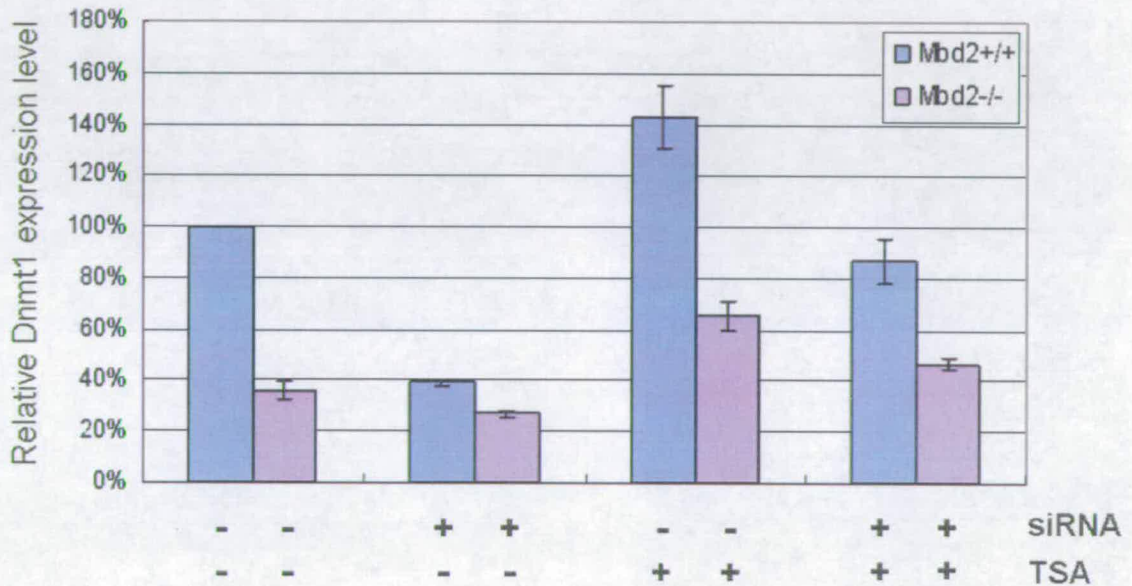
Figure 4.4: Quantitative real-time PCR analysis of the effects of *Dnmt1* siRNAs and TSA in wildtype (*Mbd2$^{+/+}$*; blue) and *Mbd2$^{-/-}$* (pink) male mouse cells. Both mouse cell lines were treated with (+) or without (-) siRNAs and/or TSA. Cells were transfected with mixture of D1-22 and D1-33 at 40 nM final concentration for 96 hours. TSA was added to a final of 1 µg/ml for 18 hours before being harvested for RNA extraction and reverse transcription. The resulting cDNA was subjected to real-time PCR. The relative level of *Dnmt1* mRNA was normalised against *Gapdh* before comparison to untreated wildtype cells. The data represents a quadruplicated measurement of one sample. Error bars represent the standard error of the mean.

Dnmt1 expression after treatment with TSA, as shown in Figure 4.4. The level of *Dnmt1* mRNA was upregulated by 1.5-to-2 fold in TSA-treated wildtype and Mbd2 null cells in comparison to the non-TSA treated cells. This is consistent with previous reports showing that the promoter activity of the Dnmt1 gene increased approximately 2-fold after TSA treatment (Kishikawa et al., 2002). When cells were double-treated with TSA and siRNAs, both cell lines showed an intermediate *Dnmt1* expression between the down-regulation by siRNA and the upregulation by TSA. This result suggests that the siRNA efficiently down-regulated *Dnmt1* mRNA before treatment of TSA. However, when *Dnmt1* mRNA synthesis is stimulated by TSA, the strength of the siRNA treatment is not sufficient to maintain an efficient depletion and results in an intermediate level of *Dnmt1* mRNA. Due to this cancelling-out effect between siRNA and TSA treatments on the expression of *Dnmt1*, the double treatment of TSA and siRNA may not genuinely represent a synergism of DNA methylation and histone deacetylation on *Xist* expression.

4.3.3.2 Reduction of *Dnmt1* expression leads to Dnmt1 depletion

Having demonstrated a decrease in *Dnmt1* mRNA levels, depletion of the protein product was subsequently examined by Western blotting using the antibody PATH52 which recognises mouse Dnmt1 (Bestor, 1992; Li et al., 1992). To assign the position of Dnmt1, protein extracts from $Dnmt1^{+/+}$ and $Dnmt1^{n/n}$ cells were first analysed. As shown in Figure 4.5, the Dnmt1 band was present at the 190 kDa position in $Dnmt1^{+/+}$ cells but absent in the extract of $Dnmt1^{n/n}$ cells. Using protein extracts from one of the triplicate experiments, the reduction of Dnmt1 protein in the siRNA-treated cells was clearly visible (Fig. 4.5, lanes 1 and 2, 3 and 4, 5 and 6, 7 and 8). In addition, it was evident that the endogenous level of Dnmt1 protein in
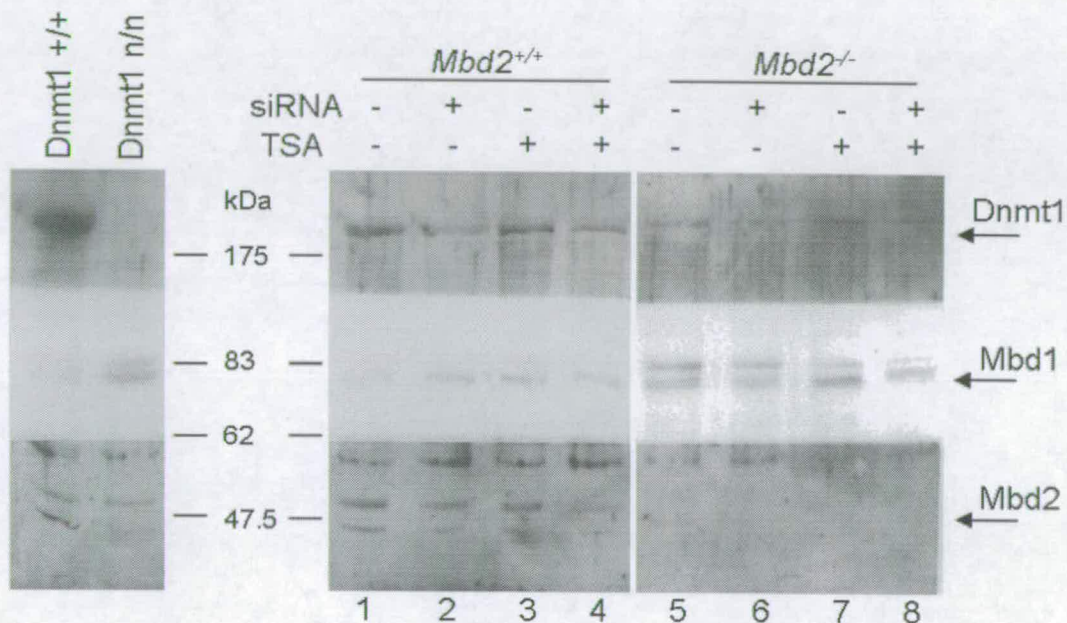
Figure 4.5: Western blot analysis to detect reduction in the Dnmt1 protein level. Whole cell extracts were prepared from cells receiving the designated treatments. Plus and minus signs indicate the presence or absence of siRNA and/or TSA. Total protein (40 μg) was loaded in each lane and separated on 8% SDS-PAGE gels for immunolot analysis with mouse Dnmt1, Mbd1, Mbd2 antibodies. Arrows indicate the positions of each protein on the blots. Protein extracts from $Dnmt1^{+/+}$ and $Dnmt1^{n/n}$ cell lines were used as positive and negative control for Dnmt1 protein. Mbd1 protein level is used as a loading control. Wildtype and Mbd2 null cells were distinguished by the appearance of anti-Mbd2 reactive bands.

Mbd2$^{-/-}$ cells was lower than in the wildtype cells. The lowest protein level observed in siRNA-treated *Mbd2$^{-/-}$* cells (Fig. 4.5, lane 6) was also in agreement with the lowest mRNA level shown in Figure 4.4. In TSA-treated cells (Fig. 4.5, lanes 3 and 7), the level of Dnmt1 protein was similar to the level observed in untreated wild type and higher than the level seen in *Mbd2$^{-/-}$* cells. This suggests that the TSA-induced upregulation of *Dnmt1* transcription may increase the cellular Dnmt1 protein level but not exceed a certain amount, even when the *Dnmt1* mRNA level had increased to around 1.5-fold of the control. From this experiment, it is confirmed that the knockdown of *Dnmt1* mRNA did lead to decreases in Dnmt1 protein level. Nevertheless, it is also clear that Dnmt1 was only reduced and not completely absent, by comparison to the negative control from *Dnmt1$^{n/n}$* cells.

4.3.3.3 Reduction of *Dnmt1* and DNA demethylation

One may argue that the remaining Dnmt1 protein is sufficient to maintain DNA methylation. To investigate the effect of the reduced Dnmt1 expression on DNA methylation, Southern blotting was carried out to analyse the methylation status of the major satellite DNA sequences as a general indicator. DNA acquired from the third triplicate of treated cells was digested with the methylation-sensitive enzyme TaiI, which restricts at ACGT sites. The resulting blot probed with a major satellite sequence is shown in Figure 4.6 and suggests that DNA from neither *Mbd2$^{+/+}$* nor *Mbd2$^{-/-}$* cell lines with any treatment underwent extensive genome-wide demethylation. All DNA samples exhibited a digestion pattern like the methylated control DNA from *Dnmt1$^{+/+}$* cells, but distinct from that of the non-methylated genomic DNA obtained from *Dnmt$^{n/n}$* cells. However, the methylation level of satellite DNA can only reflect extensive changes in genomic DNA methylation.
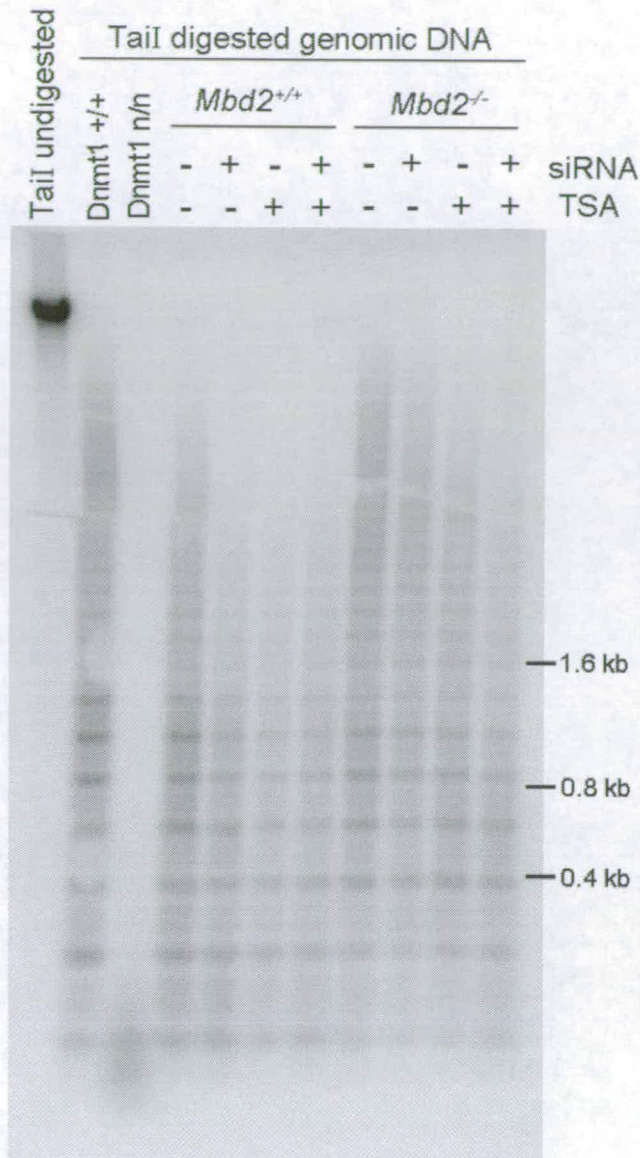
Figure 4.6: Southern blot of genomic DNA from cells with various levels of Dnmt1 mRNA. DNA was obtained from cells receiving the same treatments used in figure 4.4. Plus and minus signs indicate the presence or absence of siRNA and/or TSA. TaiI digested genomic DNA was blotted and probed with the major satellite repeat sequence. TaiI cleaves at the sequence ACGT and is blocked by CpG methylation. DNA from the untreated wildtype cells was used as undigested control. Methylated and nonmethylated controls were DNA from *Dnmt1$^{+/+}$* and *Dnmt1$^{n/n}$* cell lines, respectively.

Analysis of the total DNA methylation level should be carried out to obtain a precise measure of DNA demethylation. Bisulfite sequencing of specific regions of a gene, such as the CpG island of the *Xist* promoter, would assist in gaining an insight into significant local changes in DNA methylation. Further siRNA depletion is also required to lower DNA methylation level significantly.

4.3.3.4 Effects of *Dnmt1* down-regulation on *Xist* expression

Although the effect of *Dnmt1* depletion did not cause an intense change in global DNA methylation, it could be sufficient to affect *Xist* expression in the wildtype cells. Moreover, Dnmt1 has been reported to associate with HDACs and may act as a transcriptional repressor (Fuks et al., 2000; Robertson et al., 2000; Rountree et al., 2000). Therefore, the reduction in the protein level may cause derepression of the *Xist* gene. The same cDNA used to examine the *Dnmt1* transcript level presented in Figure 4.4 was used to measure the level of *Xist* mRNA. The PCR primer set 7/11 was used for the amplification of *Xist*. Shown in Figure 4.7 are the changes in the transcription level of *Xist*. In wildtype male mouse fibroblasts (Fig. 4.7, blue columns), treatment with *Dnmt1* siRNA resulted in an increase of *Xist* transcription by 2-to-3 fold when the level of *Dnmt1* was reduced to 40% (Fig. 4.4). This suggests that a 60% down-regulation of *Dnmt1* mRNA may have affected DNA methylation, although not at a genome-wide scale. Alternatively, the reduction of Dnmt1 protein level may have alleviated the repression of the *Xist* gene. It is worth noting that the level of *Xist* expression in treated wildtype cells was upregulated from 2-to-5 fold in two repeated experiments. This pattern supports the upregulation of the *Xist* gene as a reproducible observation.

Subsequently, *Xist* expression in *Mbd2*$^{-/-}$ cells was examined to see whether a
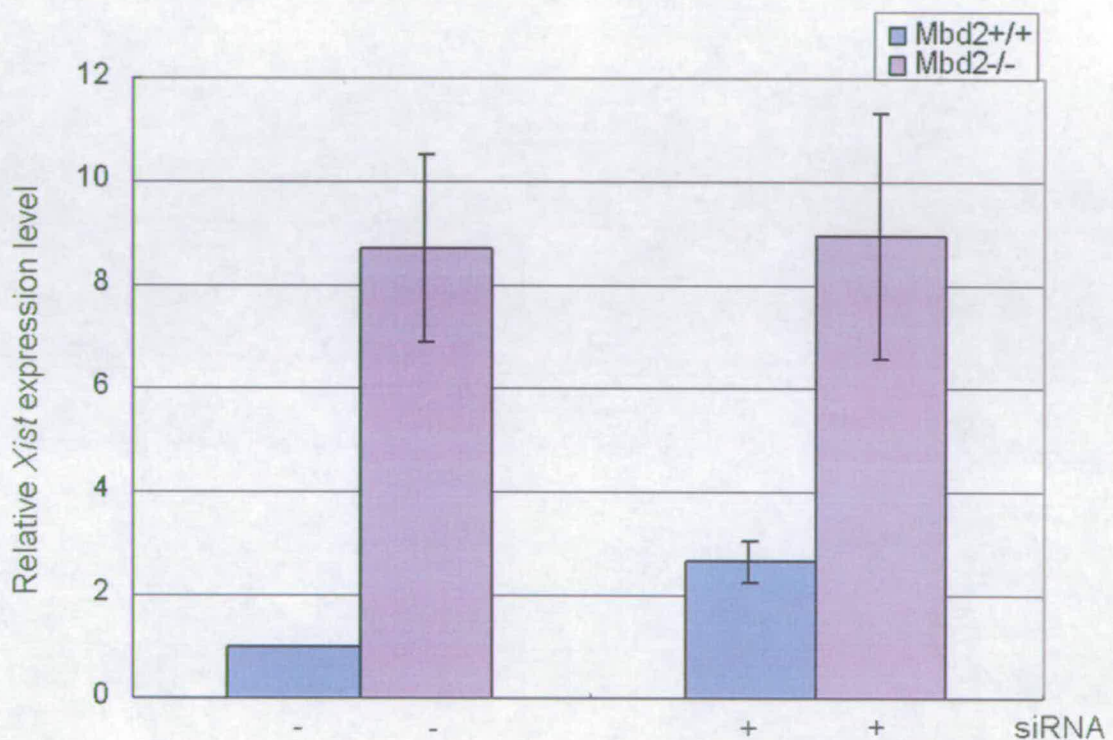
Figure 4.7: Quantitative PCR analysis of the effects of *Dnmt1* siRNAs and TSA treatments on the transcription of the *Xist* gene. Wildtype and *Mbd2*$^{-/-}$ male mouse cells were treated with (-) or without (+) siRNAs. D1-22 and D1-33 siRNAs were transfected at a final of 40 nM for 96 hours before being harvested for RNA extraction and reverse transcription. The resulting cDNA samples were first tested for the level of *Dnmt1* mRNA (Fig. 4.4) and then for the *Xist* gene expression. Relative *Xist* expression was normalised against *Gapdh* before comparison to the level of *Xist* in untreated wildtype cells, which is set to 1. The levels of Xist expression in wildtype cells (*Mbd2*$^{+/+}$; blue) and *Mbd2*-null cells (*Mbd2*$^{-/-}$; pink) are representative of a quadruplicated measurement resulting from one sample. Error bars represent the standard error of the mean.

30% down-regulation of *Dnmt1* mRNA and the consequent reduction in the protein level have caused changes in *Xist* expression. The absence of Mbd2 has been previously shown to derepress *Xist* expression (Barr and Bird, unpublished observations), and further upregulation of *Xist* expression would imply that some other protein mediates repression via DNA methylation or that Dnmt1 itself is involved in *Xist* repression. The effect of Mbd2 on *Xist* expression was analysed and shown in Figure 4.7. This figure also shows the data for *Xist* expression in wildtype cells in order to compare differences with and without Mbd2. In $Mbd2^{-/-}$ cells (Fig. 4.7, pink columns), *Xist* expression was at least 8 fold higher than untreated wildtype cells. This is consistent with the previous finding that the absence of Mbd2 alone resulted in increased *Xist* transcription (Barr and Bird, unpublished observation). The treatment with *Dnmt1* siRNA in the $Mbd2^{-/-}$ cells did not further elevate *Xist* expression regardless of the lower level of *Dnmt1* expression. However, it is difficult to conclude whether this was due to the limited reduction in the levels of *Dnmt1* mRNA and Dnmt1 protein between siRNA receiving and non-treated $Mbd2^{-/-}$ cells. Moreover, both *Dnmt1* mRNA and protein levels in untreated and siRNA-treated $Mbd2^{-/-}$ cells (Fig. 4.4 and Fig. 4.5, lane 6) were already low in comparison to untreated wildtype cells. Alternatively, this observation may result from the absence of Mbd2, if this protein acts as the sole interpreter of DNA methylation for *Xist* expression. Further down-regulation in the DNA methylation level needs to be demonstrated before reaching a comprehensive conclusion on the effect of DNA methylation on *Xist* repression. The synergistic effect of DNA demethylation and histone acetylation was not examined due to the contradictory effect of TSA and siRNA on the regulation of *Dnmt1* expression, as shown in Figure 4.4.
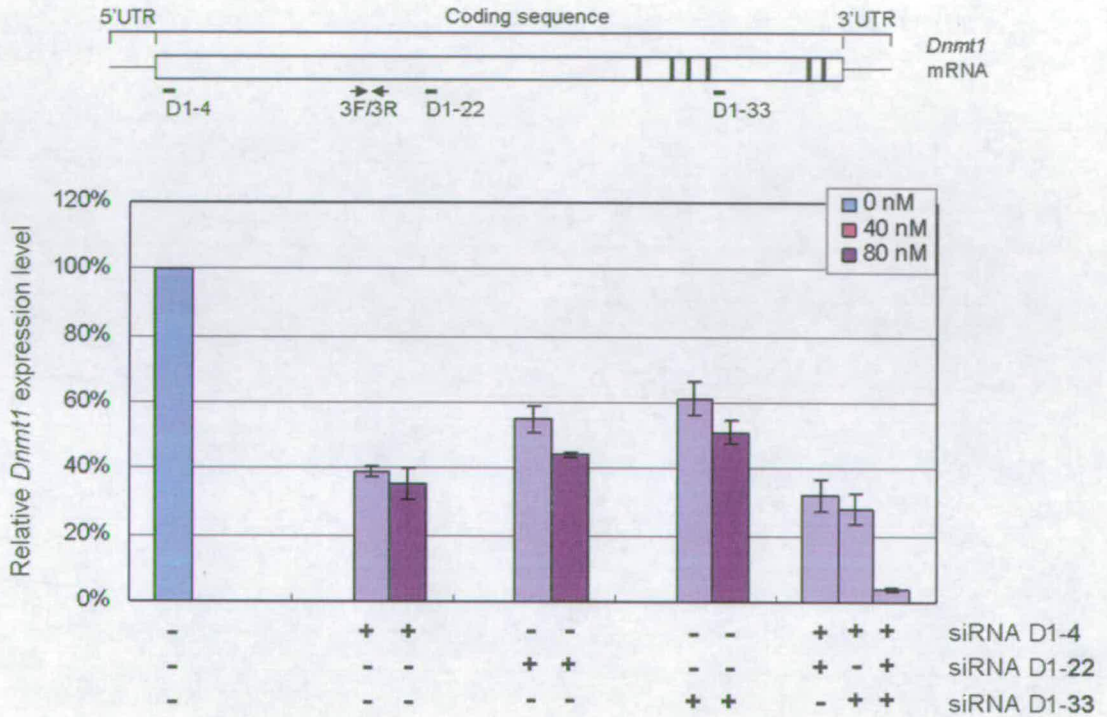
Figure 4.8: The silencing effects of various siRNAs on *Dnmt1* levels in wildtype male mouse fibroblasts. *Upper panel*: Schematic representation of the *Dnmt1* gene structure and the positions of PCR primers, and siRNAs, D1-4, D1-22, and D1-33. The position of the primer set 3F/3R is indicated by arrows. The positions of siRNAs are indicated by thick lines. Vertical bars indicate the locations of the catalytic motifs. *Lower panel*: Quantitative analysis of *Dnmt1* mRNA levels using real-time PCR. The siRNA D1-4 was tested for its silencing effects at a final concentration of 40 nM (light purple) and 80 nM (dark purple). The siRNAs D1-22 and D1-33 were re-examined at 40 nM and further tested at 80 nM. Combinations of siRNAs were also examined at the final concentration of 40 nM. Plus and minus signs indicate the type of siRNA(s) present or absent in the treatment. The relative levels of *Dnmt1* mRNA were normalised against that of *Gapdh* before comparing to the *Dnmt1* expression of the untreated cells (blue).

4.3.4 Optimisation of *Dnmt1* down-regulation

Meanwhile, attempts to further reduce DNA methylation were made by increasing the efficiency of the siRNA treatment. The siRNA D1-4 was designed to target exon 4 of *Dnmt1* and analysed for its efficiency on *Dnmt1* "knock-down" (Fig. 4.8, *upper panel*). As shown in the lower panel of Figure 4.8, treatment with *Dnmt1* siRNA D1-4 alone at 40 nM or 80 nM resulted in a decrease at around 60 and 70%, respectively. Dnmt1 siRNA D1-22 and D1-33 were also tested at the higher concentration (80 nM) which caused a better silencing effect, although the effect was not as good as the equivalent treatments using siRNA D1-4. Combinations of siRNA D1-4 with either D1-22 or D1-33 lowered Dnmt1 expression by another 5-10%. The optimal effect of *Dnmt1* siRNA was observed when combining all three siRNAs. The resulting Dnmt1 mRNA level was reduced to as little as 3% of the wildtype expression level which may be sufficient to induce a demethylated environment in cells. However, due to limited time, further analysis of the effect of reduced *Dnmt1* mRNA levels on the subsequent DNA demethylation and Mbd2-mediated *Xist* derepression will need to be carried out in the future.

4.4 Discussion

The data presented in this chapter included two different approaches that were aimed to modulate the level of DNA methylation without the use of 5-azaC, a cytotoxic demethylating agent. Although neither investigation system has yet lead to conclusive findings, further investigation based on the encouraging results will be extended.

4.4.1 DNA demethylation by the application of SAMase

As mentioned earlier, the preliminary calculation of codon usage has suggested that the T3 SAMase gene can be expressed in mammalian cells and the exogenous RNA from bacteriophage T3 had been translated in a cell-free mammalian system (Anderson et al., 1976). This information suggested that T3 RNA stayed stable in mammalian cells before being translated by mammalian ribosome. However, the result presented here using the T3 DNA sequence showed that T3 SAMase protein is not readily detected in mammalian cells.

Several explanations for the failure to detect the FLAG-SAMase protein are conceivable. One possibility is that the amount of T3 SAMase produced *in vivo* is low and therefore cannot be visualised with the immunological detection methods used in this study. It may thus be necessary to carry out enzyme activity assays similar to the one utilised in chapter 3, in order to show that the protein is expressed. In case low translation efficiency is the problem, rarely used codons can be substituted by silent mutations to the native DNA sequence, in order to reflect the intended host codon usage and enhance the translation efficiency. In addition, instability of SAMase in a mammalian environment may be a reason of the failure in SAMase detection. To evaluate this, the degradation of SAMase should be examined by measuring the turnover of the previously purified SAMase under physiological conditions.

The amount of SAMase mRNA detected in RT-PCR was very low in comparison to the amount of EGFP mRNA. This suggests SAMase mRNA may be unstable or not highly transcribed in a mammalian system. Further investigations on the transcript stabilising factors, such as poly(A) signals, 5′-capping and RNA

structures are required. In addition, changing expression system such as GST or EGFP fusion using the optimal codon may help to alleviate an unfavourable transcription and translation. By the incorporation of large tags, SAMase may be stabilised during expression without losing its enzyme activity, as shown in the bacterial expression described in the previous chapter.

Ultimately, the undetectable expression of SAMase may result from the toxicity of SAM depletion, which will be discussed in chapter 6.

4.4.2 DNA demethylation by the application of siRNA targeting *Dnmt1*

Using siRNA targeting Dnmt1, the level of *Dnmt1* transcripts has been successfully modulated to levels ranging from 5-95% down-regulation. This allows us to carry out further analysis of the differential effects of varying Dnmt1 levels. According to a study on the stability of Dnmt1 protein *in vivo*, Dnmt1 degradation starts between 10-24 hours after inhibition of protein synthesis and less than 15% of Dnmt1 was observed 28 hours after inhibition (Ding and Chaillet, 2002). A comprehensive DNA demethylation should therefore occur after treatment with siRNA for 96 hours, as the Dnmt level is almost depleted for 2-3 rounds of replication allowing passive demethylation to occur. Although a genome-wide demethylation was not clearly demonstrated using Southern blot analysis of the major satellite DNA, the upregulation of *Xist* expression has indirectly suggested an effect of DNA demethylation. However, further investigation using bisulfite sequencing is required to clearly demonstrate the change in the methylation at *Xist* promoter. Using TLC or high-performance liquid chromatography (HPLC) after

methylation-sensitive digestion, a precise measurement of genomic mC content will also allow a better understanding of the level of genome-wide DNA demethylation (Cedar et al., 1979; Ramsahoye, 2002).

The correlation between the three factors of *Dnmt1* mRNA, Dnmt1 protein and DNA methylation has never been fully clarified; neither is the corresponding timing of changes of these factors *in vivo* fully understood. Applying siRNA to achieve various degrees of down-regulation may help to correlate the Dnmt1 mRNA levels with the subsequent protein level (and/or Dnmt1 activity) and the consequent DNA methylation level. From the low expression of *Dnmt1* mRNA in *Mbd2⁻/⁻* cells, cell-line variations in the expressions of *Dnmt1* should also be examined. It has been shown that *Dnmt1* expression varies among different tissues and developmental stages (Ding and Chaillet, 2002; La Salle et al., 2004; Robertson et al., 1999). In addition, the genomic DNA methylation in the *Mbd2⁻/⁻* cells was largely maintained (Fig. 4.6) and the bisulfite sequencing of the CpG island of the *Xist* promoter in these cells showed that DNA methylation was not disturbed at that region (Barr and Bird, unpublished observations). These results suggest that the low level of *Dnmt1* expression is not likely to be a direct effect due to the absence of Mbd2 and does not affect the level of DNA methylation. To elucidate the critical Dnmt1 expression level that will result in a demethylated genome, assorted cell lines are required to be examined for their endogenous Dnmt1 mRNA levels and consequent enzyme activity and DNA methylation. Utilising the resulting correlations, we may further define the severity of DNA demethylation by screening the spectrum of gene expression. This may also help to separate the toxic effect of 5-azaC from the actual effects of DNA demethylation, by examining cells that are demethylated to the same levels by either *Dnmt1* siRNA or 5-azaC.

Dnmt1 and Dnmt3b has been shown to cooperatively maintain DNA methylation (Leu et al., 2003; Rhee et al., 2002). Although these reports are different from the DNA hypomethylation commonly observed in Dnmt1 null cells (Chan et al., 2001; Lei et al., 1996; Li et al., 1992), the growing evidence for Dnmt cooperation, as mentioned in section 1.3, suggests a more versatile regulation of the maintenance of DNA methylation. This may provide another explanation for the remaining DNA methylation shown in this study. Therefore, further analysis may need to include other cooperative factors, such as Dnmt3b, to ensure that DNA demethylation is appropriately achieved.

With respect to the siRNA treatment, a previous publication showed that the amount of mRNA depletion increased in direct correlation to the oligomer concentration (Robert et al., 2003). A similar dose-dependent effect was observed in this study, as a higher concentration of 80 nM reduced the mRNA level more than treatment with 40 nM siRNA. However, a combination of siRNAs gives the optimal silencing effect in comparison to treatments with single-siRNA even at the same concentration. This observation may constitute supportive evidence for an enhanced silencing effect by multiple siRNAs.

A cell-line specific efficiency of siRNA silencing is also suggested by the siRNA-treated wildtype and $Mbd2^{-/-}$ cells. Cell-specific factors of siRNA treatments are not well understood. Inherent differences in transfection efficiency between different cell lines may influence RNAi efficiency (Walters and Jelinek, 2002). A difference in the concentration of the molecular components involved in silencing might also contribute to the siRNA efficiency between cell lines (Harborth et al., 2001; Harborth et al., 2003). Thus, the efficiency of the *Dnmt1* siRNA should be further monitored in other cell lines to verify such an effect. Another reason for the less

efficient depletion of *Dnmt1* mRNA may also result from cells being intolerant to highly depleted *Dnmt1* mRNA. The *Mbd2^{-/-}* cells have a low endogenous *Dnmt1* expression and may be more sensitive to highly down-regulated *Dnmt1* mRNA/DNA methylation than wildtype cells. This should be resolved by the examination of siRNA efficiency using other *Mbd2^{-/-}* cell lines.

Concerning the preliminary result of *Xist* derepression in *Mbd2*-null cells, the role of Mbd2 was confirmed from the significantly increased *Xist* expression in the absence of Mbd2. However, the level of DNA methylation needs to be further reduced in the *Mbd2*-null cells and a comparable reduction of Dnmt1 protein in these two cell lines may be required, in order to elucidate the question whether other potential DNA methylation-dependent factors are involved in *Xist* repression. In addition, it is important to note that the double treatment of TSA and siRNA may not genuinely reflect the synergism of DNA methylation and histone deacetylation on *Xist* expression, due to the opposing effect of siRNA and TSA treatments on the expression of *Dnmt1*. To further study the synergistic regulation of DNA methylation and histone deacetylation on *Xist* expression, the double treatment of siRNA with TSA should be avoided.

Chapter 5

Demethylation *in vivo*: studies of DNA demethylation in preimplantation mouse embryos

5.1 Introduction

In the light of the recently described oxidative demethylation pathway, as discussed in section 1.6.3, we were interested to see whether such a mechanism plays a role in the active demethylation of 5-methylcytosine (5-mC). If this does happen, a hydroxymethyl intermediate, 5-hydroxmethylcytosine (5-HmC), may be detectable along with reaction byproducts, such as succinate and formaldehyde (Fig. 5.1). However, the oxidative demethylation pathway has only been verified for 1-methyladenine and 3-methylcytosine DNA damage. There is no report of such a mechanism in demethylation of 5-mC in any organism studied so far. Thus, the natural generation of 5-HmC from 5-mC *in vivo* is still open to question.

Synthetically, 5-HmC has been processed from 5-mC in a few occasions. In solution, 5-mC can be converted to 5-HmC by reactions involving hydroxyl radicals, gamma and UV irradiation, which are considered as common causes of DNA damage (Castro et al., 1996; Privat and Sowers, 1996; Tardy-Planechaud et al., 1997). Naturally occurring HmC has been reported in calf thymus DNA, and the brain
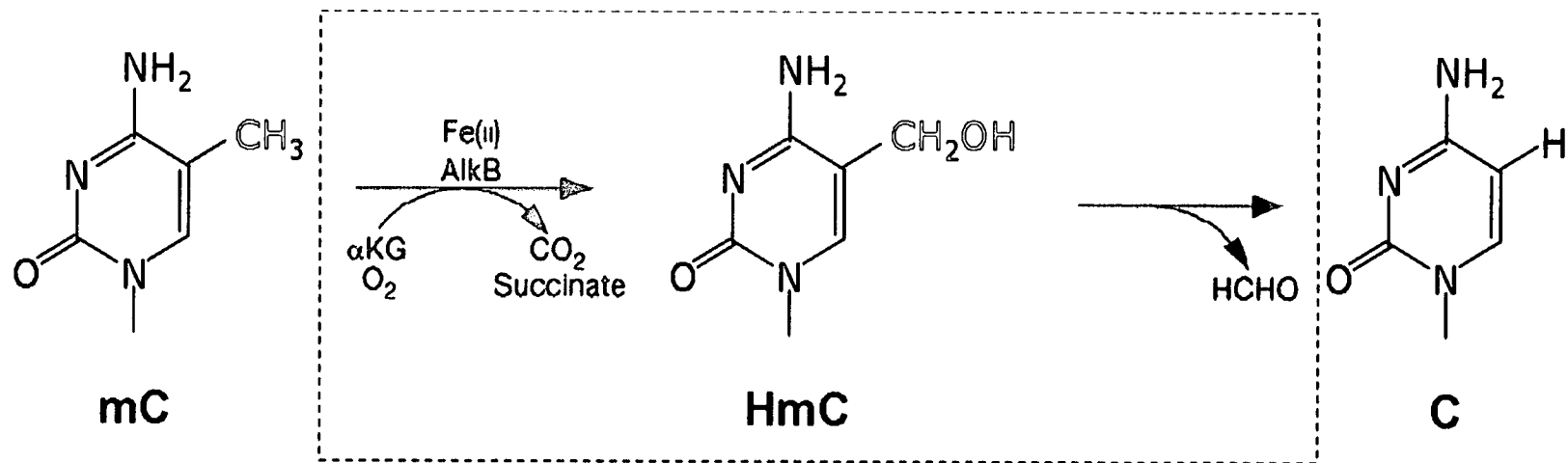
Figure 5.1: Hypothetical oxidative demethylation of 5-methylcytosine. If oxidative demethylation of 5mC would take place *in vivo*, the formation of a hydroxmethyl intermediate, i.e. HmC, and reaction byproducts, i.e. CO_2, succinate, and formaldehyde should occur after the consumption of oxygen and α-ketoglutarate (α-KG) by an oxidative repair enzyme using Fe (II) as a cofactor. However, there is as yet no evidence of such a reaction. Reactions within the dashed box is a brief model following the AlkB demethylation pathway (see Figure 1.7; Trewick et al., 2002). Side chains involved in the hypothetical demethylation are shown in red.

DNA of rat, mouse, and frog (Penn et al., 1972; Penn, 1976; Steinberg et al., 1992). However, there is a lack of further investigation on the usage of HmC in these animals. A hydroxymethylcytosine DNA glycosylase activity has also been reported in calf thymus extract (Cannon et al., 1988; Cannon-Carlson et al., 1989), but the cleavage of oligonucleotides containing HmC was not observed in extracts from human fibroblast (Rusmintratip and Sowers, 2000).

The only confirmed occurrence of 5-HmC in nature, is found in T-even bacteriophage. As mentioned in section 1.4.2, bacteriophage T2, T4, and T6 utilise an unusual base to counteract bacterial host restriction. There are slight variations in base source and antirestriction system components among these phages. Taking the best-studied bacteriophage T4 as an example, the source of the unusual base is the bacterial host (reviewed in Carlson et al., 1994). Host cytosine monophosphate (CMP) is first hydroxymethylated to produce 5-hydroxymethylcytidine precursors of HmCMP. After phosphorylation of HmCMP, HmCTP is synthesised in the host and incorporated into T4 genome via DNA replication. Finally, 5-HmCTP residues are glucosylated by a phage enzyme that transfers glucose from uridine diphosphoglucose (UDPG) provided by the bacterial host. Single-stranded DNA, free HmC and cytidine-DNA do not act as glucosylase acceptors. This reaction specifically incorporates HmC in double-stranded DNA and results in glucosylated HmC DNA (Glu-HmC DNA). However, host or phage defects do lead to failures in glucosylation. For example, mutations in the *E. coli galU* gene, which encodes UDPG pyrophosphorylase, prevent glucosylation because the reaction substrate, UDPG, is not synthesised. Mutations of the two phage glucosyltransferases, *αgt* and *βgt*, inactivate glucose transfer from UDPG to DNA. In these circumstances, T4 DNA containing unglucosylated HmC (HmC DNA) is targeted for degradation by the

host Mcr (modified cytosine restriction, also termed Rgl for restricts glucoseless phage) system.

There is little evidence for presence of the reaction intermediate and by-product of a presumptive oxidative demethylation in mammals. Nonetheless, the presence of HmC in bacteriophage T4 and the ample evidence of HmC appearance both *in vivo* and *in vitro* do not allow us to rule out the possible incorporation of HmC into mammalian DNA during oxidative demethylation. DNA demethylation has been reported previously in various cells or developmental stages, as described in section 1.6. However, it is preferable to study in a system with *bona fide* active demethylation, such as preimplantation embryos, to minimise potential confusions while verifying occurrence of active demethylation from passive demethylation. In one-cell embryos, it is not easy to monitor changes in the formation of formaldehyde and succinate using limited sample size and number, whereas the presence of 5-HmC can be easily detected within any DNA samples under examination, since 5-HmC is not a common compound in nucleic acids. Therefore, we chose to start out monitoring the presence of 5-HmC to see if such a hydroxymethyl intermediate appears in embryos undergoing active demethylation.

Immunocytochemistry has been successfully applied in the preimplantation embryos for the detection of a rapid loss of the paternally inherited methylation using an anti-mC antibody (Mayer et al., 2000a). In order to detect 5-HmC in the same fashion (Mayer et al., 2000a; Santos et al., 2002), an antibody specific to the antigen 5-HmC is required along with materials to test the antibody specificity, such as unglucosylated HmC containing DNA from bacteriophage T4.

5.2 Synthesis of HmC containing DNA

In order to carry out DNA slot blot analysis for antibody specificity, HmC DNA used for positive signals was synthesised from *E. coli* strain W4597, a UDPG-deficient bacterial strain, infected by bacteriophage T4 (Hattman and Fukusawa, 1963a). When wild type bateriophage T4 infects W4597 in minimal medium, newly synthesised phage DNA is left unglucosylated because glucose transfer is inhibited by the UDPG deficiency of the host. Packaging of HmC DNA produces "mutant" T4 containing unglucosylated HmC residues, instead of the wildtype T4 which has Glu-HmC DNA. Here, "mutant" T4 does not refer to any gene mutation or amino acid mutation, but indicates a contrast to the wildtype Glu-HmC containing T4. Extraction of HmC DNA from "mutant" T4 allows us to test the specificity of the anti-HmC antibody in DNA slot blot assays.

The production of HmC DNA and Glu-HmC DNA has to be clearly distinguished between sources. For this purpose, a plaque assay and enzyme restriction digestions were carried out. The plaque assay was performed by mixing bacteria with phage in top agar. When a single phage particle encounters a permissive bacterium, the phage infects and later lyses the bacterial cell with the concomitant release of newly formed phage particles. Bacterial cells resume growth until phage reaches lysogenic phase, lyses bacterial cells and releases more phage. The progeny phage again infects neighbouring bacteria, repeats lytic cycles, which results in a growing zone full of liberated phage, and eventually becomes visible to the naked eyes as a transparent "plaque." If the phage is absent or killed by the bacterial defence systems, the bacteria will grow to stationary phase forming a smooth opaque layer or "lawn." Accordingly, when T4 containing unglucosylated

HmC DNA grows in the presence of Mcr positive bacteria, such as the bacterial strain NM654, the "mutant" phage will be degraded by the host restriction system and is not able to form plaques of progeny phage. On the other hand, the "mutant" phage can survive and form plaques in the presence of Mcr negative bacterial strain NM494, which does not perform Mcr restriction.

The plaque assay was performed to distinguish HmC and Glu-HmC DNA. Starting with an equivalent amount, phage was diluted in a 100-fold series by volume and dropped on top agars containing bacterial hosts. As shown in Figure 5.2A, bacteriophage containing either glucosylated (wildtype) or unglycosylated (mutant) DNA infected the bacterial host NM494, and formed plaques in dilutions as low as 10^{-6}. This shows that both phages proliferate to an equal extent in a restriction-free environment. On the Mcr positive strain NM654, wildtype phage no longer proliferated as well as on the plate seeded with NM494 (Fig. 5.2A). Instead, 100-fold more phage was required to form a clear plaque. Meanwhile, the growth condition was even more stringent for the mutant phage to grow on NM494, where the phage required 10,000-fold more units of phage to form a mutant phage plaque in a restriction-positive plate. In this assay, the change in plaque-forming efficiency demonstrates that phage containing the normal modification is indeed more viable when encountering host restriction systems. Due to a lack of glucosylated HmC DNA, mutant phage is sensitive to the host Mcr restriction and proliferates less well than wild type.

Differences in Glu-HmC DNA and HmC DNA can also be observed using modification-sensitive restriction enzymes (Kutter et al., 1994). Phage DNA was extracted from the phage stock, used in Figure 5.2A, and digested with a selection of
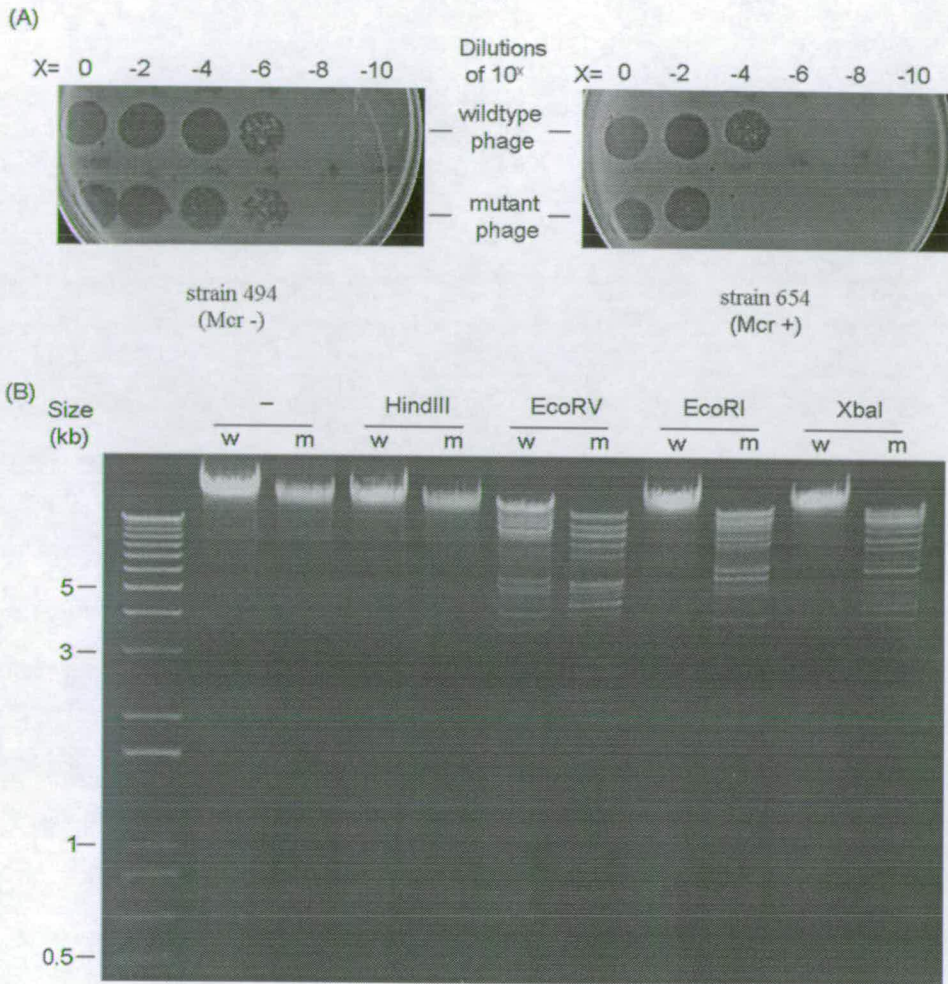
Figure 5.2: Glu-HmC DNA was distinguished from HmC DNA by phage plaque assay (A) and enzyme restriction digestions (B). (A) In the phage plaque assay, the difference between wildtype phage containing Glu-HmC DNA and mutant phage containing non-glucosylated HmC DNA were verified by bacterial Mcr restriction. Phages of the same titre (10^9 PFU/ml) were diluted in a 100-fold series (as designated above). An equal volume (10µl) of the diluted phages was dropped onto agar plates seeded with bacteria strains NM494 (left panel) or NM654 (right panel). (B) Glu-HmC DNA from wildtype phage (w) and HmC DNA from mutant phage (m) was examined by enzyme restriction. Enzymes used for each DNA digestion are specified on top. The control digestion was incubated without enzyme (-). DNA size marker is labelled on the left. 10 µg of total DNA was loaded in each lane.

enzymes. As seen in Figure 5.2B, control digestion without enzymes, or using Hind III digestion at AAGCTT site or EcoR V digestion at GATATC sequence, illustrated some characters of the DNA. Uncleaved DNA in the control without enzymes demonstrated the integrity of the input DNA, whereas the unrestricted bulk DNA retrieved after Hind III digestion confirmed that the cytosines in T4 DNA were modified because Hind III digestion is blocked by all kinds of modification to cytosine. The digestion pattern of EcoR V, on the other hand, verified that the extracted DNA was genuine T4 DNA, which was digested regardless of types of base modification. However, the digestion patterns of Glu-HmC and HmC DNA varied after incubation with EcoR I or Xba I. Glucosylation of HmC DNA blocks EcoR I and Xba I digestion at GAATTC and TCTAGA sequences, respectively. The incubation with EcoR I or Xba I resulted in undigested bulk Glu-HmC DNA running at the top of the gel, while digested fragments from HmC DNA were observed after restrictions with EcoR I or Xba I. In addition, the resulting patterns of DNA digested by EcoR V, EcoR I, and Xba I, matched that of T4 DNA in previously published data (Kutter et al., 1994). At this stage, a protocol was successfully established to isolate unglucosylated HmC DNA that is distinct from the wildtype glucosylated DNA and to use in DNA slot blotting tests.

5.3 Synthesis of monoclonal anti-HmC antibodies

The production of a monoclonal antibody is outlined in Figure 5.3A and the underlying principle of the immuno assays for antibody selection is illustrated in Figure 5.3B. Synthesis of the monoclonal anti-HmC antibody was in collaboration with the Sowers lab, Lomo Linda University, and the monoclonal antibody unit of
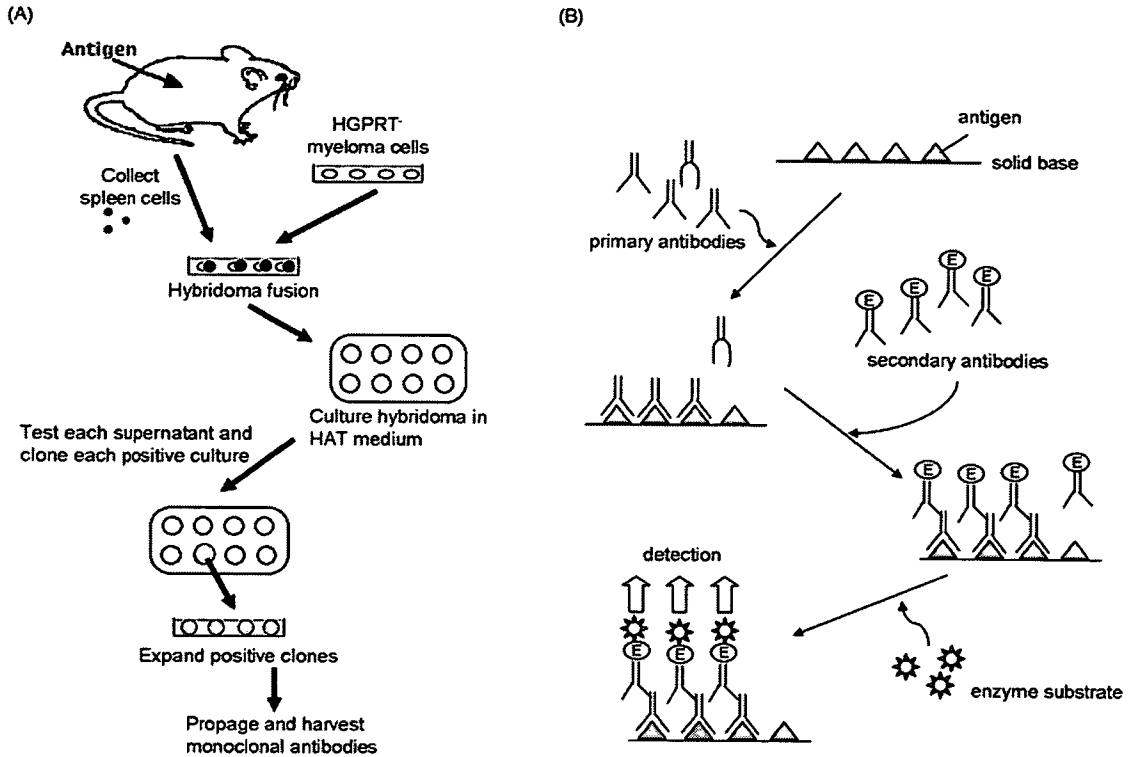
Figure 5.3: Methodology of antibody production and selection. (A) A general procedure for monoclonal antibody production. In the production of antibody recognizing a DNA base, nucleotide-protein conjugates are first synthesised by linking 3' and 4' carbon of the base with a lysine residue of a protein carrier, such as BSA. The resulting conjugate is injected into animals as the immunogen to induce an immune reaction. Selection of animals with a positive response is followed by fusing immunised spleen cells with myeloma cells. Hybridomas are selected by growing in selection medium such as HAT medium, and single hybridoma clones are isolated and subcultured to secrete only one antibody molecule, directed against a single epitope on the targeted antigen. (B) The principle of indirect immunoassays. First, the antigen is coated onto a solid base by molecule adsorption of the support material, such as an optical plate. The antibody under examination is then applied to form immune complexes with the antigen. Subsequent addition of an enzyme-coupled immunoglobin is used as a detection intermediate after binding to the antibody. Finally, when mixed with appropriate substrates, the linked enzyme reveals the antigen-antibody interaction by yielding a product that can be visualized and/or measured by optical means. (Modified from Brown and Ling, 1988; Catty and Raykundalia, 1988)

Babraham Institute. The immunogens of HmC-BSA and HmC-KLH conjugates were synthesised in Lomo Linda University by Ms. Victoria Valinluck. The following immunisation by the two immunogens was done in three rats each. Indirect enzyme-linked immunosorbent assay (ELISA) using optical readout was performed to select rats with positive immune response and as a primary screen for positive clones after the production of hydridoma fusions and subsequent cloning. The animal injection and clone isolation through ELISA analysis was carried out in the Babraham Institute by Ms. Amanda Hutchings. Following, the second stage of antibody selection involving DNA slot blot analysis was performed by myself as a means of confirming the antibody specificity.

After selection by ELISA, positive clones of the monoclonal anti-HmC antibody were further tested for the specificity by DNA slot blots. To demonstrate the specificity between cytosine and its derivatives of Glu-HmC, HmC, and mC, DNA containing Glu-HmC and HmC was prepared and tested as described in section 2.6.4 and above, whereas nonmethylated cytosine DNA was obtained from *dcm* mutant phage λ, and DNA containing 5-methylcytosine was prepared by SssI methylation of the nonmethylated λ DNA as described in section 2.5.9. In the examination of primary clones (Fig. 5.4A), both clones 33 and 118 gave a signal for HmC DNA only, whereas clone 131 recognised all DNA samples. Clone 117 did not recognise DNA with any type of cytosine. Diluted incubation (1:100) of all clones, except clone 117, was also tested to see if the cross reaction of clone 131 will be reduced. However, no signal was detected with diluted incubation, suggesting that the antibody needs to be applied at a high undiluted concentration.

Subclones from clones 63, 118 and 131 were isolated, and supernatants from clones 63-3, 118-5 and 131-10 all showed good reactivity to HmC and low

Figure 5.4: Further analysis of the specificity for 5-HmC by DNA slot blotting. (A) Denatured DNA (200 ng) from nonmethylated bacteriophageλ (-, i), methylated bacteriophageλ (m, ii), glucosylated bacteriophage T4 DNA (Glu, iii), and unglucosylated bacteriophage T4 DNA (Hm, iv) was immobilised on nitrocellulose membranes. These membranes were incubated with each anti-HmC antibody at the specified working dilution. (B) Antibody subclones 63-3, 118-5, 131-10, were tested with denatured DNA (200 ng) from nonmethylated bacteriophageλ (-, i), methylated bacteriophageλ (m, ii), glucosylated HmC T4 DNA (Hm, iii), and unglucosylated HmC T4 DNA (Glu, iv). Denatured DNA was immobilised on nitrocellulose membranes and incubated with antibodies before (1-3) and after (4-7) a 10-fold concentration by ultracentrifugation. Most blots were incubated with undiluted antibodies (1-6), except for the concentrated clone 63-3, which was also incubated at a 1:100 dilution (7). The blot incubated with an anti-mC antibody (purchased from Oncogene) was used as a positive control signal for the blot procedure.

background (Fig. 5.4B, blots 1-3). Subsequently, the volume of each antibody supernatant was reduced 10-fold by ultracentrifugation, in turn to increase antibody concentration. However, concentrated antibodies of clone 63-3 and 118-5 gave high background that interfered with HmC signal reading (Fig. 5.4 B, blots 4 and 5). Although the 100-fold dilution of the concentrated clone 63-3 did result in a reduced background, the HmC signal was also weaker when used in undiluted incubation and the reduced background still interferes with HmC signal (Fig. 5.4B, blot 7). It is also worth noticing that the 100-fold dilution of concentrated 63-3 is equivalent to a 10-fold dilution from the original supernatant, but the resulting background is stronger than the undiluted, unconcentrated antibody (Fig. 5.4B, blot 1). Among the concentrated antibodies, clone 131-10 gave an HmC-specific signal and low background (Fig. 5.4B, blot 6), although the original clone 131 showed non-specific binding to all cytosine modifications. As a whole, concentration of the antibodies does not enhance HmC signal recognition. In contrast, concentration increased background signals, which may result from the simultaneous concentration of non-specific binding proteins.

In summary, using both ELISA and slot blot analysis, we screened two positively-immunised rats out of six injected animals and obtained 10 positive hybridoma clones after selection by growing in HAT (hypoxanthine, aminotpterin, and thymidine) containing medium. Subcloning of these 10 clones resulted in 21 sublones, from which we selected a final three positive clones named clone 63-3, clone 118-5 and, clone 131-10, to use as the principle antibody sources in this study. Clone 131-10 is preferred for its high specificity and low background.

5.4 Characterisation of monoclonal anti-HmC antibodies

Due to the genome sizes and nucleotide compositions of different organisms, the occurrence of C, mC or HmC varies in the genome of bacteriophage λ, SssI-methylated λ or T4. This may cause a signal bias in previous analysis as the amount of DNA used was determined by weight. To address this concern, a more precise measure of the specificity of the anti-HmC antibodies was tested based on the molarity of each modified base. In bacteriophage λ, a cytosine base occurs once every two base pairs in a genome size of 48502 bp. There are 3112 potential Sss I-methylatable CpG sites in the λ genome, resulting in one 5-mC base in every eight base pairs. As for 5-HmC, the average ratio of A·T:G·HmC in bacteriophage T4 is 2.1:1 (Greenberg et al., 1994). Therefore, the 5-HmC base occurs at a frequency of one in every three pairs in the 168-kb genome of non-glucosylated T4. Using this information, we calculated for the total DNA weight containing 1 pmole base of interest (C, mC or HmC) present in a particular genome according to the following equation:

$$w\,(g) = \frac{X}{p} \times 650 \quad (1)$$

where w is the weight of X mol base of interest and p is the frequency of one target base occurring in 1 molecule of the DNA.

The equation is derived from the knowledge that (i) In X_a mole of total DNA a, the molarity of base present is X, and $X = X_a \times p \times N$, where N is the DNA size in base pairs; (ii) The weight of X_a mole total DNA is $w\,(g) = X_a \times 650 \times N$. The frequency for C, mC and HmC is 1/2, 1/8 and 1/3, respectively; therefore, for each pmole of C, 1.3 ng non-methylated λ DNA is required; for each pmole of mC, 5.2 ng

SssI methylated λ DNA is required; for each pmole of HmC, 1.95 ng non-glucosylated T4 DNA is required.

The result of antibody specificity analysed according to molarity of bases is shown in Figure 5.5. When clone 131-10 was tested on a membrane with immobilised HmC, mC and C DNA, the antibody readily recognised as little as 1 pmole HmC and the signal increased with the molarity up to 0.5 nmole (Fig. 5.5, lane a). As a control, a membrane containing the same DNA samples was incubated with the anti-mC antibody acquired from Eurogentec. A specific signal for mC was evident with sample containing more than 50 pmole mC (Fig. 5.5, lane e) and not in HmC and C-containing DNA (Fig 5.5, lanes d and f). This shows that the anti-HmC antibody is at least 50 times more sensitive to HmC than the anti-mC antibody to mC. However, signals from anti-HmC antibody 131-10 appeared clearly in slots with more than 500 pmole mC (Fig. 5.5, lane b), whereas no signal was detected with any of the non-methylated λ DNA which contained only C (Fig. 5.5, lane c). This suggests that anti-HmC antibody clone 131-10 is about 100-times more sensitive to HmC than to mC and this clone recognises HmC more than 1500-fold better than unmodified C. Similar observations were seen in the incubations with antibody clone 63-3 (Fig. 5.5, lane g-i) or clone 118-5 (Fig. 5.5, lane j-l). However, using the same antibody dilution and procedure, clone 63-3 only recognised HmC input of more than 10 pmole (Fig. 5.5, lane g) and clone 118-5 required more than 50 pmole HmC residues (Fig. 5.5, lane J) for the detection of a weak signal. Meanwhile, the non-specific binding to mC was equivalent for these two antibody clones, both seen when more than 500 pmoles was applied (lanes h and k). This shows that clone 63-3 and 118-5 are not as sensitive and specific to HmC as clone 131-10. Therefore, clone 131-10 is the most sensitive antibody against HmC available to us.
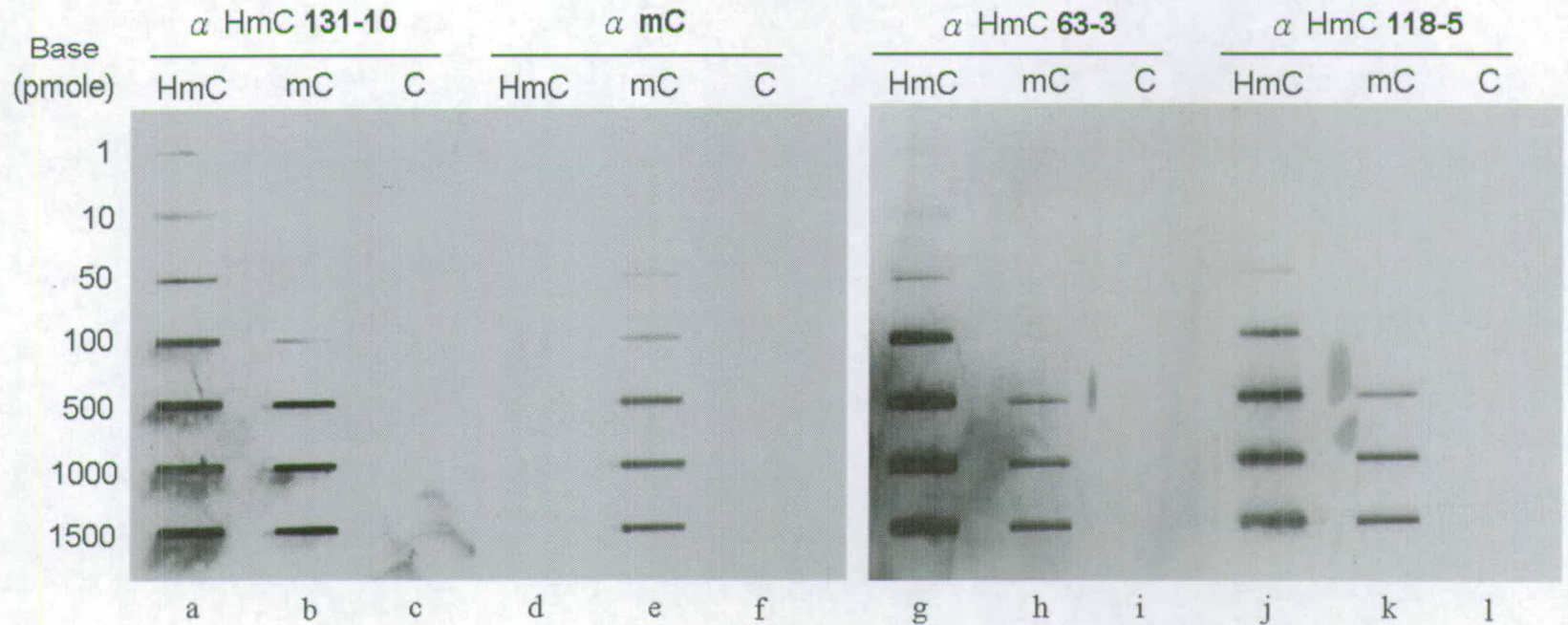
Figure 5.5: The specificity of anti-HmC antibodies tested verses antigen base molarity. Denatured DNA containing unglucosylated bacteriophage T4 DNA (HmC; a, d, g, j), SssI-methylated bacteriophage λ DNA (mC; b, e, h, k) or nonmethylated bacteriophage λ DNA (C; c, f, i, l) was blotted on nitrocellulose membranes in increasing base molarities, ranging from 1-1500 pmol (designated on the left). Antibody clone 131-10 (lanes a-c), anti-mC antibody from Eurogentec (lanes d-f) and two other anti-HmC antibody clones 63-3 (lanes g-i) and 118-5 (lanes j-l) were tested in independent incubations.

From this quantitative assay, a non-specific binding of the anti-HmC antibodies is revealed. However, instead of being proportional to the HmC sensitivity, the appearance of non-specific mC signals all arose strongly from 500 pmoles or more (Fig. 5.5, lane b, h, k). This may be a consequence of the large amount of DNA used, due to the relatively low frequency of mC in SssI-methylated λ DNA. The 500 pmole-mC sample contained a total DNA input of 2.6 μg, which was 2.7-fold more than HmC-containing DNA and four-fold more than C-containing λ DNA used for the same base molarity.

To further characterise the anti-HmC antibody specificity, a base competition assay was carried out by avoiding excessive DNA amounts that may cause antibody cross-reaction. DNA samples of 50-500 ng were tested, in which the upper limit corresponds to the DNA input given signals seen at 100 pmoles mC. In this assay, denatured DNA antigens were incubated with the anti-HmC antibody and free bases as competitors. If the antibody is able to bind to the competing bases, the
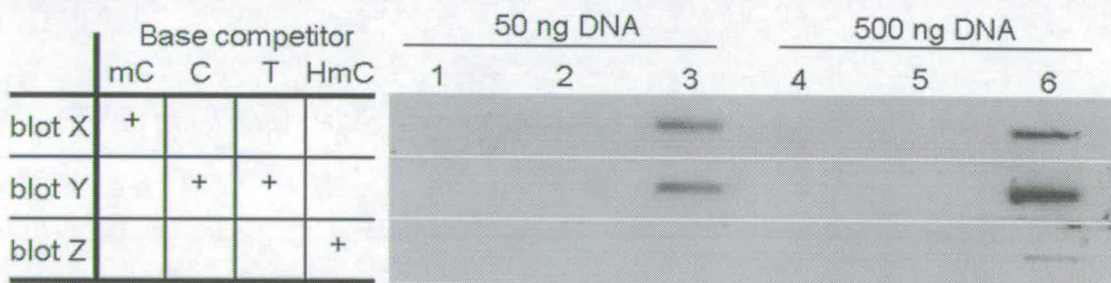
Figure 5.6: The specificity of anti-HmC antibody to 5-HmC tested by base competition. Denatured DNA of 50 ng or 500 ng in weight was extracted from nonmethylated bacteriophage λ (lanes 1 and 4), SssI-methylated bacteriophage λ (lanes 2 and 5), or unglucosylated bacteriophage T4 DNA (lanes 3 and 6), and immobilized on nitrocellulose membranes. These membranes were incubated with monoclonal anti-HmC antibody clone 131. (Blot X) Antibody incubation with 0.2 mM 5-methylcytidine (mC). (Blot Y) Antibody incubation with 0.1 mM of cytidine (C) and 0.1 mM thymidine (T). (Blot Z) Antibody incubation with 0.2 mM HmC. Blot Y is used as control incubation without the potential cross-reacting base, mC.

detectable signal of HmC DNA will be competed out, whereas HmC signal remains on blots where additional bases cannot compete with HmC DNA for antibody binding. It is important to use excess amount of competitor bases, so a signal is a genuine reflection of the antibody specificity. This assay allows us to find out the individual specificity of each competitor base added to the incubation.

The blots resulting from antibody clone 131-10 with competitions of 5-methylcytidine (mC), cytidine (C) and thymidine (T), or 5-hydroxymethylcytidine (HmC) are shown in Figure 5.6. When using 50 ng total DNA, 10-50 fmole of modified base was present in each slot. Inclusion of mC (Fig. 5.6, blot X), C and T (Fig. 5.6, blot Y), or HmC (Fig. 5.6, blot Z) in between 8-40 fold excess did not affect anti-HmC antibody in response to DNA containing cytosine (lane 1) or methylcytosine (lane 2) since this antibody does not recognise C or mC under these conditions. The reactivity of the antibody to HmC was not affected by the addition of mC, C and T, where slots of the hydroxymethylated genome gave clear signals (Fig. 5.6, slots 3-X and 3-Y, respectively). However, when 400 pmole of HmC was added, the signal for DNA containing HmC disappeared (Fig. 5.6, slot 3-Z). This shows that the anti-HmC antibody binding to the immobilised HmC DNA was specifically competed out by the addition of HmC base only. When a total of 500 ng DNA was used to reduce the magnitude of competition by 10-fold, incubation in the presence of HmC again reduced the activity, which resulted in only a weak HmC band (Fig. 5.6, slot 6-Z) compared to the strong HmC bands evident in the presence of mC, C, and T competitors (Fig. 5.6, slots 6-X and 6-Y, respectively). Although 400 pmole of competitor HmC was not able to entirely compete out the HmC signal from 500 ng immobilised DNA, the reduced signal still indicates that free HmC competes with immobilised HmC for binding, whereas the competition was not

relevant in the non-reactive DNA containing methylcytosine or cytosine (lanes 4 and 5, respectively).

From the results of antibody selection, quantitative assay and base competition assay, it is clear that the anti-HmC antibody recognises HmC specifically at a level of less than 500 ng total T4 DNA, which represents about 250 pmoles HmC, and has a detection limit as little as 1 pmole HmC according to the range tested in this study. There are about 3×10^9 bp in a haploid mouse cell (Hogan et al., 1994), which is equivalent to 10 fmole total bases and may suggest that the detection limit of 100 pmole for the anti-HmC antibody is not sensitive enough to detect the presence of HmC. However, anti-mC antibody has been applied in preimplantation embryos and resulted in clear mC signal when the detection limite was assayed to be ~50pmol in ELISA. As our anti-HmC antibody is more sensitive than the anti-mC antibody, and the less sensitive anti-mC antibody has shown successful detection of mC, anti-HmC antibody was used in immunofluoresence as a measure of HmC presence in the early mouse embryos regardless of the inadequet detection limit determined from DNA blotting. Nevertheless, further characterisation of this anti-HmC antibody should be carried out under physiological conditions.

5.5 HmC in preimplantation mouse embryos

To perform embryonic staining in the search of an HmC intermediate during the period of active demethylation, embryos at preimplantation stages were collected by Dr. Jim Selfridge within 5-9 hours after the expected time of fertilisation. The staining procedure was adopted from Santos *et al.* (2002), which involves denaturation of embryonic DNA by HCl treatment, followed by indirect

immunostaining with antibodies targeting different bases. Denaturation of DNA opens up double helix and helps base exposure to antibodies. Three different dilutions of anti-HmC antibody were examined in groups each including more than 10 fertilised embryos and the experiment was performed in duplicate. Low dilutions of antibody are not used here as initial analysis showed strong background signals (data not shown). The resulting images of indirect immunofluoresence are presented in Figure 5.7. Anti single-stranded DNA (ssDNA) antibody was used to facilitate the detection of embryonic DNA after denaturation. The green staining of ssDNA antibody showed that both pronuclei (p and m) and the polar body (PB) were present in the preimplantation embryos, in which haploid oocytes were successfully fertilized by the haploid sperm (Figure 5.7, A, E, I and M). Sizes of the pronuclei and the relative positions between each pronucleus and the polar body allow us to identify the parental origins of the pronuclei. The female pronucleus (m) is the smaller of the two and often lies closer to the polar body than the paternal pronucleus (p). The identity of the maternal pronucleus was confirmed by virtue of its highly methylated genome which resulted in clear staining by the anti-mC antibodies, whereas mC signal in the paternal pronucleus was weak or not detectable under the same conditions (Fig. 5.7, B, F, J and N). These two sets of staining demonstrated that the embryos had undergone proper fertilisation. Furthermore, the sperm DNA had entered the oocyte and was undergoing demethylation, while the maternal DNA methylation was retained. However, when analysing for HmC in the actively demethylating paternal pronucleus, no signal was observed regardless of the anti-HmC antibody working dilution (Fig. 5.7, C, K and O). A strong signal of HmC was detected in a few of the embryos (four out of a total
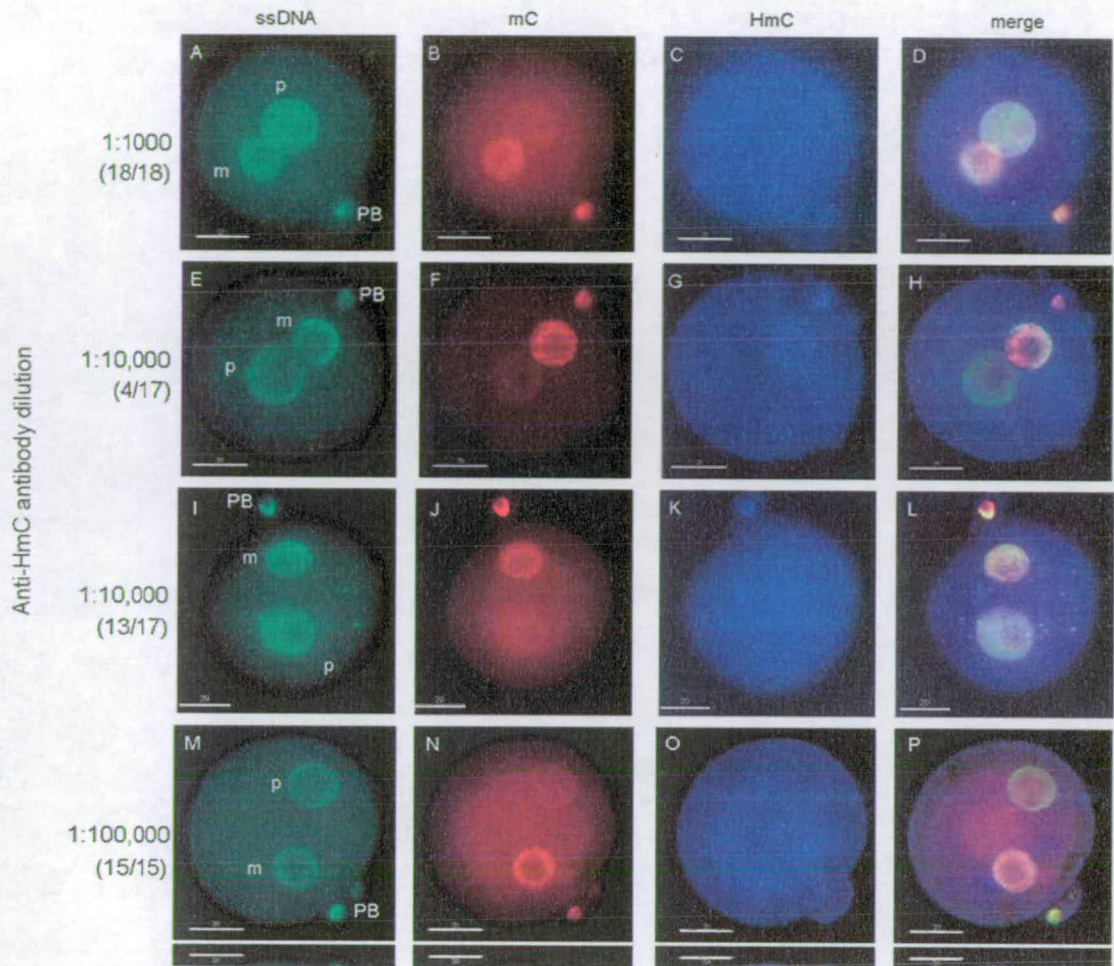
Figure 5.7: Distribution of cytosine methylation and hydroxymethylation in preimplantation mouse embryos. Embryonic DNA was first denatured to form single-straned DNA to expose DNA bases. Indirect immunofluoresence of the target base was resolved by triple-staining using anti-ssDNA (green; A, E, I, M), anti-mC (red; B, F, J, N) and anti-HmC (blue; C, G, K, O) antibodeis and merged in D, H, L, P. One-cell embryos were stained with anti-HmC antibody at dilutions of 1:1000 (A-D), 1:10,000 (E-H and I-L), or 1:100,000 (M-P). The working dilution was fixed for all anti-mC (1:50) and anti-ssDNA (1:100) staining as controls. The total number of fertilised oocytes scored was: A-D= 18; E-L= 17; M-P= 15. However, the number of fertilised oocytes showing profile as E-H was 4, whereas that of I-L was 13. The brackets indicate embryos numbers as observed over total. Components in the embryos are labelled as follow: (p) paternal pronucleus; (m) maternal pronucleus; (PB) polar body. All images were taken under 20x magnifications. Scale bar represents 20 μm.

17 of embryos) stained with anti-HmC antibody at the dilution of 1 in 10,000 (Fig. 5.7, G), but this was an atypical result in comparison (Fig. 5.7, K). More importantly, the HmC signal was found in the maternal pronucleus, where no active or passive demethylation is thought to happen. Therefore, the signal is more likely to be a non-specific binding of the antibody to mC. To conclude, HmC is not observed in the embryos undergoing active demethylation.

5.6 Discussion

As described in section 5.1, the presence of HmC residues in actively demethylating cells may provide a supportive line of evidence for an oxidative demethylation mechanism involving proteins with functions similar to the dioxygenase AlkB. Using immunostaining of HmC, early mouse embryos were examined for the presence of candidate intermediate, HmC.

From the result of embryonic staining, there is no evidence for the presence of HmC in the expected compartment of the embryo, where the paternal pronucleus undergoes active demethylation. One explanation is that the samples are taken wrongly. In this study, embryos were collected at a period covering the time of active demethylation, which had been reported to occur within 6-8 hours after fertilisation (Mayer et al., 2000a). The resulting staining of 5-mC shown in Figure 5.7 is also consistent with the reported pattern of ongoing active demethylation. When further examined my result in the system defined by *in vitro* fertilisation (IVF) system, a similar conclusion is drawn. In Santos *et al.* (2002), early mouse embryos taken between 3-10 hours after IVF were sorted into six stages according to the positional association between the two pronuclei. The most dramatic reduction of

mC staining occurs at 4-6 hours post fertilisation, when the maternal and the paternal pronuclei were 10-25 μ m apart. Active demethylation reached completion just before syngamy, which was around 8 hours post fertilisation. Following this classification, some of the embryos used in this study appeared to fall out of the time window for active demethylation. As shown in Figure 5.7 A and E, these two embryos were at a stage close to the completion of active demethylation and may provide a reason of the undetectable HmC signals as the intermediate would be long removed. Nonetheless, about three quarters of the embryos collected were within the period of rapid active demethylation, and no HmC signal was observed in these embryos either, such as the example shown in Figure 5.7, I. In addition to the analysis of collected embryos presented here, *in vitro* fertilised oocytes were also examined in collaboration with Dr. Fatima Santos of the Babraham Institute. These embryos were harvested 2.5 and 5 hours post fertilisation and the result was also negative. Therefore, it is convincing that HmC intermediate is not detected using this method at the time point of our examination.

However, the accuracy of the detection method may not be satisfactory given the half life of the HmC intermediate may be extremely short during active demethylation. The incorporation of HmC in the DNA of bacteriophage T4 suggests that HmC is a stable base *in vivo*. In terms of chemical stability, the hydroxymethyl group of HmC forms a C-C bond with the C-5 of cytosine, instead of the less stable N-C bond which is in the case of 1-HmA and 3-HmC and causes a weaker binding of the hydroxymethyl group. Therefore, release of the hydroxymethyl group from the C-5 position of cytosine is not likely to occur without a catalysing enzyme. Nevertheless, enzyme candidates such as a hydroxymethyl glycosylase has been suggested to carry out the removal of HmC (Cannon et al., 1988; Cannon-Carlson et

al., 1989). Therefore, we cannot yet rule out the possibility that HmC is a very short-term intermediate during the rapid process of active demethylation and this may be the reason we can not detect it.

Moreover, it is worth trying to confirm the antibody specificity under physiological conditions, using methods such as injection of the HmC-containing DNA in cells. The observation of HmC signals in the maternal pronucleus raised the alarm that the antibody has a varying specificity *in vivo*, although this signal was not observed in the majority of the embryos. As a diploid mouse cell contains 10 fmol bases in total, non-specific signals of the anti-HmC antibody should not arise when the expected antigen is within the upper limit of 100-fmol specificity. Therefore, further evaluation of the antibody efficiency should be carried out. Following a better understanding of this antibody *in vivo*, we may apply the anti-HmC antibody on early mouse embryos containing introduced HmC DNA to monitor HmC DNA metabolism. This may also provide information on the ability of the early embryos removing hydroxymethyl groups from modified cytosine residues.

In the context of oxidative demethylation, hydroxymethyluracil (HmU) and thymine glycol (Tg) can also be candidates for demethylation intermediates, because it has been shown that mC has a higher susceptibility to oxidation compared to thymine *in vitro* (Burdzy et al., 2002). Oxidation of mC generates HmC or 5-formylcytosine, while an mC oxidation coupled with deamination generates thymine glycol or HmU (Burdzy et al., 2002). Interestingly, the repair enzymes human TDG and MBD4 have been reported to remove thymine glycol *in vitro* in glycosylase assays using thymine glycol-incorporated oligonucleotides, although the activity is only half of what was observed for the removal of thymine (Yoon et al., 2003). A HmU DNA glycosylase activity has been reported in murine, hamster and

human cell lines and has been argued to be limited to higher eukaryotic organisms that have mC (Boorstein et al., 1987; Hollstein et al., 1984). More strikingly, an unexpected high excision rate for HmU in human HeLa cell extracts has been reported (3 x 10^{-9} μmol oligo/μg protein/min), although the estimated rate of spontaneous conversion of mC:G to HmU:G base pair is only once per 2,000 years and the appearance of a HmU:G base pair does not disturb the overall DNA conformation (Carbonnaux et al., 1990; Rusmintratip and Sowers, 2000). Further characterisation of the mispaired HmU repair activity has shown that the activity is distinct from previously reported glycosylases including TDG and uracil glycosylase activities, and the partially purified activity is highly selective for the 5-hydroxymethyl group (Baker et al., 2002). Therefore, other oxidative DNA damage products, especially HmU, may be considered as alternative candidates for the intermediates of active DNA demethylation.

For the purpose of immunostaining, we produced a highly sensitive anti-HmC antibody with a detection limit to fmole levels. This provides a useful tool to approach HmC nucleotides from a different angle. Applying the antibody to DNA from various sources, we may be able to identify HmC in various organisms and tissues and thus gain further insights into the role of HmC. Also, this may help to further investigate the related HmC DNA glycosylase activity (Cannon et al., 1988; Cannon-Carlson et al., 1989; Penn et al., 1972; Penn, 1976). However, calibrations of the antibody specificity such as signal intensities against antibody concentration and/or antigen concentration should be carried out before further applications.

Discussion

Chapter 6

Discussion

The data presented in this thesis address the role and mechanism of DNA demethylation from various perspectives. From an *in vitro* point of view, SAMase was studied in the context of its function on SAM depletion, while the unique activity of SAM cleaving was utilised *in vivo* to test its potential application as a demethylating agent. An alternative approach to achieve demethylation using siRNA was adopted and used successfully to reduce the *Dnmt1* transcript to various extents. Finally, a hypothetical pathway for active DNA demethylation *in vivo* was examined, and the observed result indicated that a candidate intermediate, 5-hydroxymethylcytidine, was undetectable during the period of active DNA demethylation.

The *in vitro* SAMase studies were designed to increase the understanding of SAMase structure and function, in order to facilitate the later *in vivo* application to reduce levels of DNA methylation. Functional SAMase without fusion tags has been expressed successfully, although further optimisation of the purification procedure will be required before extending this study. The successful expression of the GST-SAMase fusion protein has allowed us to obtain large amounts of homogeneous SAMase and has in turn permitted crystallisation trials. However, the

fusion tag may be refractory to the protein crystallisation, and as yet, no crystal has been observed in the precipitants tested to date. Adjustments of the components of the protein storage buffer are required to obtain optimal conditions for the production of stable native protein after GST tag cleavage. For example, using Mg^{++} may provide a suitable ionic strength for the enzyme storage without abolishing it activity; the use of glycerol may also help to stabilise the protein. Additionally, it may be worth trying to purify native protein following conventional methods and to use the resultant protein for crystallisation and for further studies on the kinetic behaviour of SAMase. This analysis may shed light on the potential application of SAMase in mammalian systems.

From the preliminary expression of the native SAMase DNA sequence, it is not clear whether SAMase expression in mammalian systems is feasible. Substitution of codons that are highly biased in prokaryotic systems may greatly improve the expression in mammals, but the potentially detrimental effect of SAM depletion should always be kept in mind. As mentioned in section 4.1, the degradation of SAM did not appear lethal in some bacteria; nevertheless, the observed low expression levels of native SAMase using pBAD-SH and pJF-SH constructs had implied a toxicity of the SAMase activity. Also, the growth rate of cells expressing SAMase was slower than that of non-expressing cells. The low transcription level of SAMase in human cells also suggests such a possibility (Fig. 4.3C). Hence, GST fusion of SAMase could be used to investigate if the protein has an undesired toxic effect, while attempts at SAMase expression are made in mammalian systems. In addition, the normal function of SAMase is to counteract the host bacterial restriction-modification cleavage system. Using GST-SAMase to explore the mechanism of T3 antirestriction may also increase our understanding about the

roles of DNA methylation in prokaryotes, and in turn this may bring about advances towards the application of SAMase in eukaryotic systems.

More importantly, SAM acts as the methylation donor for various methylation reactions. SAM is involved in various metabolic pathways such as folate and vitamin B12 metabolism. In addition to DNA methylation, RNA methylation plays an important role in the maturation and stabilisation of RNA synthesis (Smith and Steitz, 1997). Protein methylation is a post-translational modification that occurs during mechanisms such as the stress response, cell growth control and signal transduction (Aletta et al., 1998). More specifically, growing evidence has shown that methylation of histones plays an important role in the regulation of gene activity. Various arginine and lysine sites on the tails of histones H3 and H4 can be methylated to different degrees resulting in different regulatory consequences (Ahmad and Henikoff, 2002; Fahrner and Baylin, 2003; reviewed in Zhang and Reinberg, 2001). For example, histone H3 lysine 9 methylation is associated with gene silencing; whereas histone H3 lysine 4 tri-methylation is a marker for active transcription. Consequently, SAM depletion may simultaneously affect gene regulation both by the level of histone methylation and the level of DNA methylation. Although SAMase can be applied to a wide range of cells, the resulting system is not likely to be useful for the examination of the exclusive effect of DNA methylation, since SAMase also contributes to the demethylation of histones.

In comparison, siRNA targeting *Dnmt1* can act more specifically to bring about DNA demethylation. Application of *Dnmt1* siRNA in cells has successfully down-regulated the level of *Dnmt1* expression. However, precise measurements of the level of DNA demethylation will be important in its correlation with the silencing effect of *Dnmt1* siRNA. Although gene silencing using siRNA may have a

cell-line specific efficiency, further examination of specific siRNA treatment in different cell lines is required to determine the critical conditions at which reduced *Dnmt1* expression leads to DNA demethylation. As a whole, with the observation of a general *Dnmt1* down-regulation in siRNA-treated cells, siRNA has achieved the purpose of decreasing Dnmt1 protein expression and may therefore be an easier and more practical means to bring about DNA demethylation.

Considering the role of DNA methylation in the control of *Xist* expression, Mbd2 has been previously shown to be a mediator protein that represses *Xist* expression through DNA methylation (Barr and Bird, unpublished observation). Whether other factors are involved in *Xist* repression in a similar manner could be demonstrated using the most efficient *Dnmt1* siRNA mixture. If other methylation-dependent factors are involved, DNA hypomethylation resulting from *Dnmt1* siRNA treatment would further elevate the level of *Xist* derepression in the absence of Mbd2. On the contrary, *Xist* expression will remain at the same level in Mbd2 deficient cells, when Mbd2 is the sole interpreter of DNA methylation for *Xist* expression. Interestingly, preliminary results from bisulfite sequencing using $Dnmt1^{+/+}$ and $Dnmt1^{n/n}$ cells show that three CpG sites in the promoter region of the *Xist* gene are demethylated in more than 75% of the Dnmt1 null clones, whereas other nearby CpG sites remain methylated in most clones (data now shown). These three demethylated CpG sites may be the prime candidates involved in the control of *Xist* repression and deserve closer examination of their association with methylation dependent regulatory proteins such as Mbd2.

In the study of active DNA demethylation *in vivo*, a hydroxmethylated intermediate, 5-HmC, was hypothesised to be the early demethylation product of a putative demethylase belonging to the α-ketoglutarate-Fe^{2+}-dependent dioxygenases

family. The hypothesis led us to search for DNA containing the 5-HmC base as an indicator that a such demethylation mechanism is present *in vivo*, prior to identifying any actual demethylase enzymes. Although 5-HmC containing DNA was not observed in embryos undergoing active DNA demethylation, we cannot yet discount the possibility that it is a short half-life intermediate, nor could we discount the candidate mechanism utilising the aforementioned iron-dependent dioxygenase. However, it is also important to consider a coupled enzyme that releases a less formidable side chain from the pyrimidine ring, while searching for an intermediate containing a potential leaving group like the hydroxymethyl structure. The hydroxyl group will pull electrons away from the neighbouring carbon. This in turn causes a less stable bond between the carbon on the hydroxymethyl side chain and the ring carbon in comparison to the bonding between a methyl group and the ring carbon. Nevertheless, a hydroxymethyl group on the carbon-5 position of the cytosine ring is still more stable than one such moiety on the 3' of nitrogen, as the nitrogen is electrophilic. Therefore, it is likely that enzymes such as an aldolase (E.C. 4.1.2) are involved in bringing about the breakage of the C-C bond. Enzymes of this family can perform the reverse of aldol condensation and are mostly prokaryotic enzymes. Similar mammalian enzymes, such as aldolase A (E.C.4.1.2.13), have been detected in muscle tissues and are involved in the breakdown of sugars to generate energy in cells (DiMauro and Bruno, 1998). Nevertheless, a mammalian enzyme that catalyses a carbon-carbon bond cleavage precisely between a hydroxyalkyl group and a ring structure still awaits discovery, together with the obscure identity of a demethylase. Alternatively, other candidate intermediates of active DNA demethylation may be considered. For example, oxidative DNA damage products, especially hydroxylmethyluracil (HmU),

are compounds with a potential transition from mC and may be repaired by a correspondent glycosylase (Rusmintratip and Sowers, 2000). HmU has been shown to have an unexpected high excision rate and an HmU DNA glycosylase activity has been reported to remove HmU (Rusmintratip and Sowers, 2000).

The mechanism of active demethylation in 5-mC has been largely considered as an enzymatic reaction, which has included potential catalysts such as the direct demethylase (Bhattacharya et al., 1999), glycosylase (Jost, 1993; Vairapandi and Duker, 1993) and, more recently, dioxygenase (Falnes et al., 2002; Trewick et al., 2002). The involvement of ribonucleic acids has also been proposed to be a substitute moiety in deoxyribonucleic acids during demethylation, although the evidence was never strong (Fremont et al., 1997; Saluz et al., 1986; Swisher et al., 1998; Weiss et al., 1996). Interestingly, a recent report demonstrated DNA demethylation may be directed by non-coding RNA (Imamura et al., 2004). The CpG island of the differentially methylated region (T-DMR) of sphingosine kinase-1 gene, *Sphk1*, underwent demethylation after the expression of Sphk1 antisense transcript, Khps1, whereas two non-CpG cytosines surprisingly became methylated at the same time. However, more information is required to identify details such as whether the demethylation drives active transcription and whether or not the proposed mechanism is a primary cause of the demethylation. The idea of an RNA molecule being instrumental in the occurrence of DNA demethylation is controversial, bearing in mind the proposed mechanism of gene silencing via RNA interference at either transcriptional or post-transcriptional level (reviewed in Meister and Tuschl, 2004; Paszkowski and Whitham, 2001). Further investigation will hopefully shed light on whether RNA can indeed bring about DNA demethylation.

Regarding a model system exhibiting active DNA demethylation, it has been suggested that sperm DNA decondensation and protamine-histone exchange prior to pronucleus formation may provide a good opportunity for demethylation in the paternal pronucleus (Barton et al., 2001). Nonetheless, a recent study using interspecies intracytoplasmic sperm injection demonstrated that murine oocytes have significantly higher demethylating activity than those of the sheep or bovine oocytes, regardless of sperm species origin (Beaujean et al., 2004). On the other hand, mouse sperm injected into ovine oocytes, which do not normally demethylate, underwent significant demethylation. This study suggests that there is species-specificity in the performance of active DNA demethylation and that the composition of oocyte and sperm may both contribute to the demethylation activity. Hence, future studies may not focus only on the formation and changes in the paternal pronucleus, but also in the maternal domain.

In addition, systems described in section 1.5.1, such as myoblast and erythroleukemia cells, were not chosen as a platform during the hunt for an active demethylase, due to a limited numbers of demethylating sites that may not be detectable using immunocytochemistry. However, these cells may be useful when using detection methods involving specific demethylated sites, such as the CpG sites in the promoter-enhancer of the interleukin-2 (Il2) gene (Bruniquel and Schwartz, 2003). The ease of growing a large number of cultured cells can compensate for the limited number of demethylation sites and the accurate timing of stimulating differentiation will allow a better monitoring of the removal of the methyl group.

In summary, the work presented in this thesis constitutes an investigation of DNA demethylation and has achieved preliminary successes in the acquisition of

homogeneous SAMase protein, siRNA silencing of *Dnmt1* expression and the production of a new antibody against HmC residues in DNA. Although the ultimate aim of identifying a potential mammalian demethylase was not achieved, further application of these new tools may shed light on the mechanism and role of DNA demethylation.

Bibliography

Bibliography

Aas,P.A., Otterlei,M., Falnes,P.O., Vagbo,C.B., Skorpen,F., Akbari,M., Sundheim,O., Bjoras,M., Slupphaug,G., Seeberg,E., and Krokan,H.E. (2003). Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature *421*, 859-863.

Ahmad,K. and Henikoff,S. (2002). Epigenetic Consequences of Nucleosome Dynamics. Cell *111*, 281-284.

Akashi,H. (2001). Gene expression and molecular evolution. Curr. Opin. in Genet. & Dev. *11*, 660-666.

Aletta,J.M., Cimato,T.R., and Ettinger,M.J. (1998). Protein methylation: a signal event in post-translational modification. Trends in Biochemical Sciences *23*, 89-91.

Allaman-Pillet,N., Djemai,A., Bonny,C., and Schorderet,D.F. (1998). Methylation status of CpG sites and methyl-CpG binding proteins are involved in the promoter regulation of the mouse Xist gene. Gene Expr. *7*, 61-73.

Anderson,C.W., Atkins,J.F., and Dunn,J.J. (1976). Bacteriophage T3 and T7 early RNAs are translated by eukaryotic 80S ribosomes: active phage T3 coded S-adenosylmethionine cleaving enzyme is synthesized. Proc. Natl. Acad. Sci. U. S. A *73*, 2752-2756.

Aravind,L. and Koonin,E.V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol 2, RESEARCH0007.

Atanasiu,C., Byron,O., McMiken,H., Sturrock,S.S., and Dryden,D.T. (2001). Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. Nucleic Acids Res. *29*, 3059-3068.

Atanasiu,C., Su,T.J., Sturrock,S.S., and Dryden,D.T. (2002). Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme.

Nucleic Acids Res. *30*, 3936-3944.

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,D., Liu,P., Burdzy,A., and Sowers,L.C. (2002). Characterization of the substrate specificity of a human 5-hydroxymethyluracil glycosylase activity. Chem. Res Toxicol. *15*, 33-39.

Barr, H and Bird, A. Role of Mbd2 in repression of the mouse *Xist* gene. 2004. Unpublished observations.

Barton,S.C., Arney,K.L., Shi,W., Niveleau,A., Fundele,R., Surani,M.A., and Haaf,T. (2001). Genome-wide methylation patterns in normal and uniparental early mouse embryos. Hum. Mol Genet. *10*, 2983-2987.

Beard,C., Li,E., and Jaenisch,R. (1995). Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev. *9*, 2325-2334.

Beaujean,N., Taylor,J.E., McGarry,M., Gardner,J.O., Wilmut,I., Loi,P., Ptak,G., Galli,C., Lazzari,G., Bird,A., Young,L.E., and Meehan,R.R. (2004). The effect of interspecific oocytes on demethylation of sperm DNA. Proc. Natl. Acad. Sci. U. S. A *101*, 7636-7640.

Bell,A.C. and Felsenfeld,G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

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., Hellewell,S.B., and Ingram,V.M. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. Mol Cell Biol *4*, 1800-1806.

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

2395-2402.

Bhattacharya,S.K., Ramchandani,S., Cervoni,N., and Szyf,M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. Nature *397*, 579-583.

Bird,A.P. and Taggart,M.H. (1980). Variable patterns of total DNA and rDNA methylation in animals. Nucl. Acids. Res. *8*, 1485-1497.

Bird,A.P., Taggart,M.H., and Smith,B.A. (1979). Methylated and unmethylated DNA compartments in the sea urchin genome. Cell *17*, 889-901.

Bird,A. (1997). Does DNA methylation control transposition of selfish elements in the germline? Trends in Genetics *13*, 469-470.

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

Bird,A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends in Genetics *3*, 342-347.

Boeke,J., Ammerpohl,O., Kegel,S., Moehren,U., and Renkawitz,R. (2000). The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J. Biol Chem. *275*, 34963-34967.

Boorstein,R.J., Levy,D.D., and Teebor,G.W. (1987). 5-Hydroxymethyluracil-DNA glycosylase activity may be a differentiated mammalian function. Mutat. Res *183*, 257-263.

Bourc'his,D., Le Bourhis,D., Patin,D., Niveleau,A., Comizzoli,P., Renard,J.P., and Viegas-Pequignot,E. (2001a). Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos. Curr. Biol *11*, 1542-1546.

Bourc'his,D., Xu,G.L., Lin,C.S., Bollman,B., and Bestor,T.H. (2001b). Dnmt3L and the Establishment of Maternal Genomic Imprints. Science *294*, 2536-2539.

Brown,G. and Ling,N.R. (1988). Murine monoclonal antibodies. In Antibodies: A Practical Approach, D.Catty, ed. (Washington DC: IRL Press), pp. 81-104.

Bruniquel,D. and Schwartz,R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat. Immunol. *4*,

235-240.

Burdzy,A., Noyes,K.T., Valinluck,V., and Sowers,L.C. (2002). Synthesis of stable-isotope enriched 5-methylpyrimidines and their use as probes of base reactivity in DNA. Nucleic Acids Res *30*, 4068-4074.

Cameron,E.E., Bachman,K.E., Myohanen,S., Herman,J.G., and Baylin,S.B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat. Genet. *21*, 103-107.

Cannon,S.V., Cummings,A., and Teebor,G.W. (1988). 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. Biochem. Biophys. Res Commun. *151*, 1173-1179.

Cannon-Carlson,S.V., Gokhale,H., and Teebor,G.W. (1989). Purification and characterization of 5-hydroxymethyluracil-DNA glycosylase from calf thymus. Its possible role in the maintenance of methylated cytosine residues. J. Biol Chem. *264*, 13306-13312.

Carbonnaux,C., Fazakerley,G.V., and Sowers,L.C. (1990). An NMR structural study of deaminated base pairs in DNA. Nucleic Acids Res *18*, 4075-4081.

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

Carlson,K., Raleigh,E.A., and Hattman,S. (1994). Restriction and Modification. In Molecular Biology of Bacteriophage T4, J.D.Karam, J.W.Drake, K.N.Kreuzer, G.Mosig, D.H.Hall, F.A.Eiserling, L.W.Black, E.K.Spicer, E.Kutter, K.Carlson, and E.S.Miller, eds. (Washington, D.C.: American Society for Microbiology), pp. 369-381.

Castro,G.D., Diaz Gomez,M.I., and Castro,J.A. (1996). 5-Methylcytosine attack by hydroxyl free radicals and during carbon tetrachloride promoted liver microsomal lipid peroxidation: structure of reaction products. Chem. Biol Interact. *99*, 289-299.

Catty,D. and Raykundalia,C. (1988). ELISA and related enzyme immunoassays. In Antibodies: A Practical Approach, D.Catty, ed. (Washington DC: IRL Press), pp. 97-152.

Cedar,H., Solage,A., Glaser,G., and Razin,A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. Nucleic Acids Res *6*,

2125-2132.

Cervoni,N., Bhattacharya,S., and Szyf,M. (1999). DNA demethylase is a processive enzyme. J. Biol Chem. *274*, 8363-8366.

Chan,M.F., van Amerongen,R., Nijjar,T., Cuppen,E., Jones,P.A., and Laird,P.W. (2001). Reduced rates of gene loss, gene silencing, and gene mutation in Dnmt1-deficient embryonic stem cells. Mol. Cell Biol *21*, 7587-7600.

Charache,S., Dover,G., Smith,K., Talbot,C.C., Jr., Moyer,M., and Boyer,S. (1983). Treatment of sickle cell anemia with 5-azacytidine results in increased fetal hemoglobin production and is associated with nonrandom hypomethylation of DNA around the gamma-delta-beta-globin gene complex. Proc. Natl. Acad. Sci. U. S. A *80*, 4842-4846.

Chedin,F., Lieber,M.R., and Hsieh,C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. PNAS *99*, 16916-16921.

Chesnokov,I.N. and Schmid,C.W. (1995). Specific Alu Binding Protein from Human Sperm Chromatin Prevents DNA Methylation. J. Biol. Chem. *270*, 18539-18542.

Chu,W.M., Ballard,R., Carpick,B.W., Williams,B.R., and Schmid,C.W. (1998). Potential Alu Function: Regulation of the Activity of Double-Stranded RNA-Activated Kinase PKR. Mol. Cell. Biol. *18*, 58-68.

Cohen,D.E. and Lee,J.T. (2002). X-chromosome inactivation and the search for chromosome-wide silencers. Curr. Opin. Genet. Dev. *12*, 219-224.

Collier,G.B., Mattson,T.L., Connaughton,J.F., and Chirikjian,J.G. (1994b). A novel Tn10 tetracycline regulon system controlling expression of the bacteriophage T3 gene encoding S-adenosyl-L-methionine hydrolase. Gene *148*, 75-80.

Collier,G.B., Mattson,T.L., Connaughton,J.F., and Chirikjian,J.G. (1994a). A novel Tn10 tetracycline regulon system controlling expression of the bacteriophage T3 gene encoding S-adenosyl-L-methionine hydrolase. Gene *148*, 75-80.

Comeron,J.M. and Aguade,M. (1998). An evaluation of measures of synonymous codon usage bias. J. Mol Evol. *47*, 268-274.

Cooper,D.N., Taggart,M.H., and Bird,A.P. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res *11*, 647-658.

Csankovszki,G., Nagy,A., and Jaenisch,R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. J. Cell Biol *153*, 773-784.

Dean,W., Santos,F., Stojkovic,M., Zakhartchenko,V., Walter,J., Wolf,E., and Reik,W. (2001). Conservation of methylation reprogramming in mammalian development: Aberrant reprogramming in cloned embryos. PNAS *98*, 13734-13738.

Deplus,R., Brenner,C., Burgers,W.A., Putmans,P., Kouzarides,T., Launoit,Y.d., and Fuks,F. (2002). Dnmt3L is a transcriptional repressor that recruits histone deacetylase. Nucl. Acids. Res. *30*, 3831-3838.

Detich,N., Bovenzi,V., and Szyf,M. (2003). Valproate induces replication-independent active DNA demethylation. J. Biol Chem. *278*, 27586-27592.

Di Croce,L., Raker,V.A., Corsaro,M., Fazi,F., Fanelli,M., Faretta,M., Fuks,F., Coco,F.L., Kouzarides,T., Nervi,C., Minucci,S., and Pelicci,P.G. (2002). Methyltransferase Recruitment and DNA Hypermethylation of Target Promoters by an Oncogenic Transcription Factor. Science *295*, 1079-1082.

DiMauro,S. and Bruno,C. (1998). Glycogen storage diseases of muscle. Curr. Opin. Neurol. *11*, 477-484.

Ding,F. and Chaillet,J.R. (2002). In vivo stabilization of the Dnmt1 (cytosine-5)-methyltransferase protein. Proc. Natl. Acad. Sci U. S. A *99*, 14861-14866.

Dinglay,S., Trewick,S.C., Lindahl,T., and Sedgwick,B. (2000). Defective processing of methylated single-stranded DNA by E. coli AlkB mutants. Genes Dev. *14*, 2097-2105.

Dorsett,Y. and Tuschl,T. (2004). siRNAs: applications in functional genomics and potential as therapeutics. Nat Rev Drug Discov *3*, 318-329.

Dryden,D.T., Murray,N.E., and Rao,D.N. (2001). Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res *29*, 3728-3741.

Duncan,T., Trewick,S.C., Koivisto,P., Bates,P.A., Lindahl,T., and Sedgwick,B. (2002). Reversal of DNA alkylation damage by two human dioxygenases. Proc. Natl. Acad. Sci. U. S. A *99*, 16660-16665.

Eden,A., Gaudet,F., Waghmare,A., and Jaenisch,R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. Science *300*, 455.

Elbashir,S.M., Martinez,J., Patkaniowska,A., Lendeckel,W., and Tuschl,T. (2001a). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J. *20*, 6877-6888.

Elbashir,S.M., Harborth,J., Lendeckel,W., Yalcin,A., Weber,K., and Tuschl,T. (2001b). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Elbashir,S.M., Harborth,J., Weber,K., and Tuschl,T. (2002). Analysis of gene function in somatic mammalian cells using small interfering RNAs. Methods *26*, 199-213.

Fahrner,J.A. and Baylin,S.B. (2003). Heterochromatin: stable and unstable invasions at home and abroad. Genes Dev. *17*, 1805-1812.

Falnes,P.O., Johansen,R.F., and Seeberg,E. (2002). AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature *419*, 178-182.

Fatemi,M., Hermann,A., Gowher,H., and Jeltsch,A. (2002). Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. Eur. J. Biochem. *269*, 4981-4984.

Fatemi,M., Hermann,A., Pradhan,S., and Jeltsch,A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J. Mol Biol *309*, 1189-1199.

Feng,Q. and Zhang,Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev. *15*, 827-832.

Fremont,M., Siegmann,M., Gaulis,S., Matthies,R., Hess,D., and Jost,J.P. (1997). Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA. Nucleic Acids Res *25*, 2375-2380.

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.

Furste,J.P., Pansegrau,W., Frank,R., Blocker,H., Scholz,P., Bagdasarian,M., and Lanka,E. (1986). Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. Gene *48*, 119-131.

Gaudet,F., Hodgson,J.G., Eden,A., Jackson-Grusby,L., Dausman,J., Gray,J.W., Leonhardt,H., and Jaenisch,R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489-492.

Gefter,M., Hausmann,R., Gold,M., and Hurwitz,J. (1966). The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. X. Bacteriophage T3-induced S-adenosylmethionine cleavage. J. Biol Chem. *241*, 1995-2006.

Gonzalez-Zulueta,M., Bender,C.M., Yang,A.S., Nguyen,T., Beart,R.W., Van Tornout,J.M., and Jones,P.A. (1995). Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. Cancer Res *55*, 4531-4535.

Good,X., Kellogg,J.A., Wagoner,W., Langhoff,D., Matsumura,W., and Bestwick,R.K. (1994). Reduced ethylene synthesis by transgenic tomatoes expressing S-adenosylmethionine hydrolase. Plant Mol. Biol *26*, 781-790.

Greenberg,G.R., He,P., Hilfinger,J., and Tseng,M. (1994). Deoxyribonucleoside Triphosphate Synthesis and Phage T4 DNA Replication. In Molecular Biology of Bacteriophage T4, J.D.Karam, J.W.Drake, K.N.Kreuzer, G.Mosig, D.H.Hall, F.A.Eiserling, L.W.Black, E.K.Spicer, E.Kutter, K.Carlson, and E.S.Miller, eds. (Washington, D.C.: American Society for Microbiology), pp. 14-27.

Guy,J., Hendrich,B., Holmes,M., Martin,J.E., and Bird,A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat. Genet. *27*, 322-326.

Guzman,L.M., Belin,D., Carson,M.J., and Beckwith,J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. *177*, 4121-4130.

Haaf,T. (1995). The effects of 5-azacytidine and 5-azadeoxycytidine on chromosome structure and function: implications for methylation-associated cellular processes. Pharmacol. Ther. *65*, 19-46.

Haaf,T. and Schmid,M. (1989). 5-Azadeoxycytidine induced undercondensation in

the giant X chromosomes of Microtus agrestis. Chromosoma *98*, 93-98.

Hajkova,P., Erhardt,S., Lane,N., Haaf,T., El Maarri,O., Reik,W., Walter,J., and Surani,M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech. Dev. *117*, 15-23.

Han,L., Lin,I.G., and Hsieh,C.L. (2001). Protein binding protects sites on stable episomes and in the chromosome from de novo methylation. Mol Cell Biol *21*, 3416-3424.

Hansen,R.S., Stoger,R., Wijmenga,C., Stanek,A.M., Canfield,T.K., Luo,P., Matarazzo,M.R., D'Esposito,M., Feil,R., Gimelli,G., Weemaes,C.M., Laird,C.D., and Gartler,S.M. (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Hum. Mol Genet. *9*, 2575-2587.

Hansen,R.S., Wijmenga,C., Luo,P., Stanek,A.M., Canfield,T.K., Weemaes,C.M.R., and Gartler,S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. PNAS *96*, 14412-14417.

Harbers,K., Schnieke,A., Stuhlmann,H., Jahner,D., and Jaenisch,R. (1981). DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning. Proc. Natl. Acad. Sci. U. S. A *78*, 7609-7613.

Harborth,J., Elbashir,S.M., Bechert,K., Tuschl,T., and Weber,K. (2001). Identification of essential genes in cultured mammalian cells using small interfering RNAs. J. Cell Sci *114*, 4557-4565.

Harborth,J., Elbashir,S.M., Vandenburgh,K., Manninga,H., Scaringe,S.A., Weber,K., and Tuschl,T. (2003). Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. Antisense Nucleic Acid Drug Dev. *13*, 83-105.

Hark,A.T., Schoenherr,C.J., Katz,D.J., Ingram,R.S., Levorse,J.M., and Tilghman,S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486-489.

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.

Hattman,S. and Fukusawa,T. (1963). Host-induced modification of t-even phages due to defective glucosylation of their DNA. Proc. Natl. Acad. Sci U. S. A *50*, 297-300.

Hendrich,B., Abbott,C., McQueen,H., Chambers,D., Cross,S., and Bird,A. (1999a). Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. Mamm. Genome *10*, 906-912.

Hendrich,B. and Bird,A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell Biol *18*, 6538-6547.

Hendrich,B., Guy,J., Ramsahoye,B., Wilson,V.A., and Bird,A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. *15*, 710-723.

Hendrich,B., Hardeland,U., Ng,H.H., Jiricny,J., and Bird,A. (1999b). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

Hogan,B., Beddington,R., Costantini,F., and Lacy,E. (1994). Developmental Genetics and Embryology of the Mouse. In Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Habor Laboratory Press), p. 3.

Hollstein,M.C., Brooks,P., Linn,S., and Ames,B.N. (1984). Hydroxymethyluracil DNA glycosylase in mammalian cells. Proc. Natl. Acad. Sci U. S. A *81*, 4003-4007.

Howell,C.Y., Bestor,T.H., Ding,F., Latham,K.E., Mertineit,C., Trasler,J.M., and Chaillet,J.R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell *104*, 829-838.

Howlett,S.K. and Reik,W. (1991). Methylation levels of maternal and paternal genomes during preimplantation development. Development *113*, 119-127.

Hsieh,C.L. (1994). Dependence of transcriptional repression on CpG methylation density. Mol Cell Biol *14*, 5487-5494.

Hsieh,C.L. (1999a). Evidence that protein binding specifies sites of DNA demethylation. Mol Cell Biol *19*, 46-56.

Hsieh,C.L. (1999b). In Vivo Activity of Murine De Novo Methyltransferases, Dnmt3a and Dnmt3b. Mol. Cell. Biol. *19*, 8211-8218.

Hughes,J.A., Brown,L.R., and Ferro,A.J. (1987a). Expression of the cloned coliphage T3 S-adenosylmethionine hydrolase gene inhibits DNA methylation and polyamine biosynthesis in Escherichia coli. J. Bacteriol. *169*, 3625-3632.

Hughes,J.A., Brown,L.R., and Ferro,A.J. (1987b). Nucleotide sequence and analysis of the coliphage T3 S-adenosylmethionine hydrolase gene and its surrounding ribonuclease III processing sites. Nucleic Acids Res. *15*, 717-729.

Huntriss,J., Lorenzi,R., Purewal,A., and Monk,M. (1997). A Methylation-Dependent DNA-Binding Activity Recognising the Methylated Promoter Region of the MouseXistGene. Biochemical and Biophysical Research Communications *235*, 730-738.

Imamura,T., Yamamoto,S., Ohgane,J., Hattori,N., Tanaka,S., and Shiota,K. (2004). Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem. Biophys. Res Commun. *322*, 593-600.

Jablonka,E., Goitein,R., Marcus,M., and Cedar,H. (1985). DNA hypomethylation causes an increase in DNase-I sensitivity and an advance in the time of replication of the entire inactive X chromosome. Chromosoma *93*, 152-156.

Jaenisch,R., Schnieke,A., and Harbers,K. (1985). Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. Proc. Natl. Acad. Sci. U. S. A *82*, 1451-1455.

Jones,P.A. (2002). DNA methylation and cancer. Oncogene *21*, 5358-5360.

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

Jones,P.A., Taylor,S.M., Mohandas,T., and Shapiro,L.J. (1982). Cell cycle-specific reactivation of an inactive X-chromosome locus by 5-azadeoxycytidine. Proc. Natl. Acad. Sci. U. S. A *79*, 1215-1219.

Jorgensen,H.F., Ben Porath,I., and Bird,A.P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol *24*, 3387-3395.

Jost,J.P. (1993). Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. Proc. Natl.

Acad. Sci. U. S. A *90*, 4684-4688.

Jost,J.P., Fremont,M., Siegmann,M., and Hofsteenge,J. (1997). The RNA moiety of chick embryo 5-methylcytosine- DNA glycosylase targets DNA demethylation. Nucleic Acids Res. *25*, 4545-4550.

Jost,J.P., Oakeley,E.J., Zhu,B., Benjamin,D., Thiry,S., Siegmann,M., and Jost,Y.C. (2001). 5-Methylcytosine DNA glycosylase participates in the genome-wide loss of DNA methylation occurring during mouse myoblast differentiation. Nucleic Acids Res. *29*, 4452-4461.

Jost,J.P., Siegmann,M., Sun,L., and Leung,R. (1995). Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. J. Biol Chem. *270*, 9734-9739.

Jost,J.P., Siegmann,M., Thiry,S., Jost,Y.C., Benjamin,D., and Schwarz,S. (1999). A re-investigation of the ribonuclease sensitivity of a DNA demethylation reaction in chicken embryo and G8 mouse myoblasts. FEBS Lett. *449*, 251-254.

Juttermann,R., Li,E., and Jaenisch,R. (1994). Toxicity of 5-aza-2'-Deoxycytidine to Mammalian Cells is Mediated Primarily by Covalent Trapping of DNA Methyltransferase Rather than DNA demethylation. PNAS *91*, 11797-11801.

Kafri,T., Ariel,M., Brandeis,M., Shemer,R., Urven,L., McCarrey,J., Cedar,H., and Razin,A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev. *6*, 705-714.

Kass,S.U., Landsberger,N., and Wolffe,A.P. (1997). DNA methylation directs a time-dependent repression of transcription initiation. Curr. Biol *7*, 157-165.

Kawano,S., Miller,C.W., Gombart,A.F., Bartram,C.R., Matsuo,Y., Asou,H., Sakashita,A., Said,J., Tatsumi,E., and Koeffler,H.P. (1999). Loss of p73 Gene Expression in Leukemias/Lymphomas Due to Hypermethylation. Blood *94*, 1113-1120.

Ke,A. and Wolberger,C. (2003). Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. Protein Sci *12*, 306-312.

Keohane,A.M., Lavender,J.S., O'neill,L.P., and Turner,B.M. (1998). Histone

acetylation and X inactivation. Dev. Genet. *22*, 65-73.

Keohane,A.M., O'neill,L.P., Belyaev,N.D., Lavender,J.S., and Turner,B.M. (1996). X-Inactivation and histone H4 acetylation in embryonic stem cells. Dev. Biol *180*, 618-630.

Keshet,I., Lieman-Hurwitz,J., and Cedar,H. (1986). DNA methylation affects the formation of active chromatin. Cell *44*, 535-543.

Kishikawa,S., Ugai,H., Murata,T., and Yokoyama,K.K. (2002). Roles of histone acetylation in the Dnmt1 gene expression. Nucleic Acids Res Suppl 209-210.

Kondo,H., Nakabeppu,Y., Kataoka,H., Kuhara,S., Kawabata,S., and Sekiguchi,M. (1986). Structure and expression of the alkB gene of Escherichia coli related to the repair of alkylated DNA. J. Biol. Chem. *261*, 15772-15777.

Kozak,M. (1999). Initiation of translation in prokaryotes and eukaryotes. Gene *234*, 187-208.

Kruger,D.H., Reuter,M., Schroeder,C., Glatman,L.I., and Chernin,L.S. (1983). Restriction of bacteriophage T3 and T7 ocr+ strains by the type II restriction endonuclease EcoRV. Mol. Gen. Genet. *190*, 349-351.

Kruger,D.H., Schroeder,C., Santibanez-Koref,M., and Reuter,M. (1989). Avoidance of DNA methylation. A virus-encoded methylase inhibitor and evidence for counterselection of methylase recognition sites in viral genomes. Cell Biophys. *15*, 87-95.

Kuge,M., Fujii,Y., Shimizu,T., Hirose,F., Matsukage,A., and Hakoshima,T. (1997). Use of a fusion protein to obtain crystals suitable for X-ray analysis: crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. Protein Sci *6*, 1783-1786.

Kunert,N., Marhold,J., Stanke,J., Stach,D., and Lyko,F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development *130*, 5083-5090.

Kutter,E., Steidham,T., Guttman,B., Kutter,E., Batts,D., Peterson,S., Djavakhishvili,T., Arisaka,F., Mesyanzhinov,V., Rueger,W., and Mosig,G. (1994). Appendix A: Genomic Map of Bacteriophage T4. In Molecular Biology of Bacteriophage T4, J.D.Karam, J.W.Drake, K.N.Kreuzer, G.Mosig, D.H.Hall, F.A.Eiserling, L.W.Black, E.K.Spicer,

E.Kutter, K.Carlson, and E.S.Miller, eds. (Washington, D.C.: American Society for Microbiology), pp. 369-381.

La Salle,S., Mertineit,C., Taketo,T., Moens,P.B., Bestor,T.H., and Trasler,J.M. (2004). Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. Developmental Biology *268*, 403-415.

Lee,J., Inoue,K., Ono,R., Ogonuki,N., Kohda,T., Kaneko-Ishino,T., Ogura,A., and Ishino,F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development *129*, 1807-1817.

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.

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.

Leu,Y.W., Rahmatpanah,F., Shi,H., Wei,S.H., Liu,J.C., Yan,P.S., and Huang,T.H. (2003). Double RNA interference of DNMT3b and DNMT1 enhances DNA demethylation and gene reactivation. Cancer Res. *63*, 6110-6115.

Ley,T.J., DeSimone,J., Anagnou,N.P., Keller,G.H., Humphries,R.K., Turner,P.H., Young,N.S., Keller,P., and Nienhuis,A.W. (1982). 5-azacytidine selectively increases gamma-globin synthesis in a patient with beta+ thalassemia. N. Engl. J. Med. *307*, 1469-1475.

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

Liang,G., Chan,M.F., Tomigahara,Y., Tsai,Y.C., Gonzales,F.A., Li,E., Laird,P.W., and Jones,P.A. (2002). Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. Mol. Cell. Biol. *22*, 480-491.

Lin,I.G., Tomzynski,T.J., Ou,Q., and Hsieh,C.L. (2000). Modulation of DNA binding protein affinity directly affects target site demethylation. Mol Cell Biol *20*, 2343-2349.

Liu,K., Wang,Y.F., Cantemir,C., and Muller,M.T. (2003). Endogenous Assays of DNA Methyltransferases: Evidence for Differential Activities of DNMT1, DNMT2, and

DNMT3 in Mammalian Cells In Vivo. Mol. Cell. Biol. *23*, 2709-2719.

Liu,W.M., Chu,W.M., Choudary,P.V., and Schmid,C.W. (1995). Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. Nucleic Acids Res *23*, 1758-1765.

Liu,W.M., Maraia,R.J., Rubin,C.M., and Schmid,C.W. (1994). Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. Nucleic Acids Res *22*, 1087-1095.

Livak,K.J. and Schmittgen,T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2-[Delta][Delta]CT Method. Methods *25*, 402-408.

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.

Loenen,W.A., Daniel,A.S., Braymer,H.D., and Murray,N.E. (1987). Organization and sequence of the hsd genes of Escherichia coli K-12. J. Mol Biol *198*, 159-170.

Lyko,F., Ramsahoye,B.H., and Jaenisch,R. (2000). DNA methylation in Drosophila melanogaster. Nature *408*, 538-540.

Macleod,D., Charlton,J., Mullins,J., and Bird,A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. *8*, 2282-2292.

Macleod,D., Clark,V.H., and Bird,A. (1999). Absence of genome-wide changes in DNA methylation during development of the zebrafish. Nat. Genet. *23*, 139-140.

Margot,J., Ehrenhofer-Murray,A., and Leonhardt,H. (2003). Interactions within the mammalian DNA methyltransferase family. BMC Molecular Biology *4*, 7.

Marinus,M.G. (1984). Methylation of Prokaryotic DNA. In DNA methylation: Biochemistry and Biological Significance, A.Razin, H.Cedar, and A.Riggs, eds. Springer-Verlag), pp. 81-109.

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

Mayer,W., Smith,A., Fundele,R., and Haaf,T. (2000b). Spatial separation of parental

genomes in preimplantation mouse embryos. J. Cell Biol *148*, 629-634.

McDonald,L.E., Paterson,C.A., and Kay,G.F. (1998). Bisulfite genomic sequencing-derived methylation profile of the Xist gene throughout early mouse development. Genomics *54*, 379-386.

Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L., and Bird,A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499-507.

Meister,G. and Tuschl,T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343-349.

Mertineit,C., Yoder,J.A., Taketo,T., Laird,D.W., Trasler,J.M., and Bestor,T.H. (1998). Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development *125*, 889-897.

Millar,C.B., Guy,J., Sansom,O.J., Selfridge,J., MacDougall,E., Hendrich,B., Keightley,P.D., Bishop,S.M., Clarke,A.R., and Bird,A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science *297*, 403-405.

Mishina,Y., Lee,C.H., and He,C. (2004). Interaction of human and bacterial AlkB proteins with DNA as probed through chemical cross-linking studies. Nucleic Acids Res *32*, 1548-1554.

Mohandas,T., Sparkes,R.S., and Shapiro,L.J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science *211*, 393-396.

Monk,M. (2002). Mammalian embryonic development--insights from studies on the X chromosome. Cytogenet. Genome Res *99*, 200-209.

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.

Nakamura,Y., Gojobori,T., and Ikemura,T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucl. Acids. Res. *28*, 292.

Nan,X., Campoy,F.J., and Bird,A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.

Nan,X., Meehan,R.R., and Bird,A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res *21*, 4886-4892.

Nan,X., Ng,H.H., Johnson,C.A., Laherty,C.D., Turner,B.M., Eisenman,R.N., and Bird,A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Newman,E.B., Budman,L.I., Chan,E.C., Greene,R.C., Lin,R.T., Woldringh,C.L., and D'Ari,R. (1998). Lack of S-adenosylmethionine results in a cell division defect in Escherichia coli. J. Bacteriol. *180*, 3614-3619.

Newman,J.R. and Fuqua,C. (1999). Broad-host-range expression vectors that carry the arabinose-inducible Escherichia coli araBAD promoter and the araC regulator. Gene *227*, 197-203.

Ng,H.H., Jeppesen,P., and Bird,A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol *20*, 1394-1406.

Ng,H.H., Zhang,Y., Hendrich,B., Johnson,C.A., Turner,B.M., Erdjument-Bromage,H., Tempst,P., Reinberg,D., and Bird,A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat. Genet. *23*, 58-61.

Norris,D.P., Patel,D., Kay,G.F., Penny,G.D., Brockdorff,N., Sheardown,S.A., and Rastan,S. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41-51.

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. Nucl. Acids. Res. *26*, 2536-2540.

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. Cur. Biol. *10*, 475-478.

Palmer,B.R. and Marinus,M.G. (1994). The dam and dcm strains of Escherichia coli--a review. Gene *143*, 1-12.

Panning,B. and Jaenisch,R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev. *10*, 1991-2002.

Paroush,Z., Keshet,I., Yisraeli,J., and Cedar,H. (1990). Dynamics of demethylation and activation of the alpha-actin gene in myoblasts. Cell *63*, 1229-1237.

Paszkowski,J. and Whitham,S.A. (2001). Gene silencing and DNA methylation processes. Cur. Opin. in Plant Biol. *4*, 123-129.

Penn,N.W. (1976). Modification of brain deoxyribonucleic acid base content with maturation in normal and malnourished rats. Biochem. J. *155*, 709-712.

Penn,N.W., Suwalski,R., O'Riley,C., Bojanowski,K., and Yura,R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. Biochem. J. *126*, 781-790.

Poot,M., Koehler,J., Rabinovitch,P.S., Hoehn,H., and Priest,J.H. (1990). Cell kinetic disturbances induced by treatment of human diploid fibroblasts with 5-azacytidine indicate a major role for DNA methylation in the regulation of the chromosome cycle. Hum. Genet. *84*, 258-262.

Posnick,L.M. and Samson,L.D. (1999). Influence of S-adenosylmethionine pool size on spontaneous mutation, dam methylation, and cell growth of Escherichia coli. J. Bacteriol. *181*, 6756-6762.

Privat,E. and Sowers,L.C. (1996). Photochemical deamination and demethylation of 5-methylcytosine. Chem. Res Toxicol. *9*, 745-750.

Raleigh,E.A., Murray,N.E., Revel,H., Blumenthal,R.M., Westaway,D., Reith,A.D., Rigby,P.W., Elhai,J., and Hanahan,D. (1988). McrA and McrB restriction phenotypes of some E. coli strains and implications for gene cloning. Nucleic Acids Res *16*, 1563-1575.

Ramsahoye,B.H. (2002). Measurement of genome wide DNA methylation by

reversed-phase high-performance liquid chromatography. Methods *27*, 156-161.

Ratnam,S., Mertineit,C., Ding,F., Howell,C.Y., Clarke,H.J., Bestor,T.H., Chaillet,J.R., and Trasler,J.M. (2002). Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev. Biol *245*, 304-314.

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.

Rhee,I., Bachman,K.E., Park,B.H., Jair,K.W., Yen,R.W., Schuebel,K.E., Cui,H., Feinberg,A.P., Lengauer,C., Kinzler,K.W., Baylin,S.B., and Vogelstein,B. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552-556.

Rhee,I., Jair,K.W., Yen,R.W., Lengauer,C., Herman,J.G., Kinzler,K.W., Vogelstein,B., Baylin,S.B., and Schuebel,K.E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. Nature *404*, 1003-1007.

Rice,P., Longden,I., and Bleasby,A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. *16*, 276-277.

Robert,M.F., Morin,S., Beaulieu,N., Gauthier,F., Chute,I.C., Barsalou,A., and MacLeod,A.R. (2003). DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. Nat. Genet. *33*, 61-65.

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., Uzvolgyi,E., Liang,G., Talmadge,C., Sumegi,J., Gonzales,F.A., and Jones,P.A. (1999). The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res *27*, 2291-2298.

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.

Rusmintratip,V. and Sowers,L.C. (2000). An unexpectedly high excision capacity for mispaired 5-hydroxymethyluracil in human cell extracts. Proc. Natl. Acad. Sci. U. S. A *97*, 14183-14187.

Sado,T., Fenner,M.H., Tan,S.S., Tam,P., Shioda,T., and Li,E. (2000). X Inactivation in the Mouse Embryo Deficient for Dnmt1: Distinct Effect of Hypomethylation on Imprinted and Random X Inactivation. Developmental Biology *225*, 294-303.

Sado,T., Okano,M., Li,E., and Sasaki,H. (2004). De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. Development *131*, 975-982.

Saluz,H.P., Jiricny,J., and Jost,J.P. (1986). Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis. Proc. Natl. Acad. Sci. U. S. A *83*, 7167-7171.

Sambrook,J., Fritsch,E.F., and Maniatis,T. (1989). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory Press), p. 16.30-16.36.

Santos,F., Hendrich,B., Reik,W., and Dean,W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev. Biol *241*, 172-182.

Schmid,C.W. (1996). Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Prog. Nucleic Acid Res Mol Biol *53*, 283-319.

Schubeler,D., Lorincz,M.C., Cimbora,D.M., Telling,A., Feng,Y.Q., Bouhassira,E.E., and Groudine,M. (2000). Genomic Targeting of Methylated DNA: Influence of Methylation on Transcription, Replication, Chromatin Structure, and Histone Acetylation. Mol. Cell. Biol. *20*, 9103-9112.

Schwarz,S., Bourgeois,C., Soussaline,F., Homsy,C., Podesta,A., and Jost,J.P. (2000). A CpG-rich RNA and an RNA helicase tightly associated with the DNA demethylation complex are present mainly in dividing chick embryo cells. Eur. J. Cell Biol *79*, 488-494.

Scopes,R.K. (1993). Protein Purification: Principles and Practice, Springer-Verlag), p. p341.

Sharp,P.M. and Li,W.H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res *15*, 1281-1295.

Shi,W. and Haaf,T. (2002). Aberrant methylation patterns at the two-cell stage as an indicator of early developmental failure. Mol Reprod Dev. *63*, 329-334.

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.

Simpson,V.J., Johnson,T.E., and Hammen,R.F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.

Singer,J., Roberts-Ems,J., Luthardt,F.W., and Riggs,A.D. (1979). Methylation of DNA in mouse early embryos, teratocarcinoma cells and adult tissues of mouse and rabbit. Nucleic Acids Res *7*, 2369-2385.

Smith,C.M. and Steitz,J.A. (1997). Sno Storm in the Nucleolus: New Roles for Myriad Small RNPs. Cell *89*, 669-672.

Smyth,D.R., Mrozkiewicz,M.K., McGrath,W.J., Listwan,P., and Kobe,B. (2003). Crystal structures of fusion proteins with large-affinity tags. Protein Sci *12*, 1313-1322.

Spoerel,N. and Herrlich,P. (1979). Colivirus-T3-coded S-adenosylmethionine hydrolase. Eur. J. Biochem. *95*, 227-233.

Spoerel,N., Herrlich,P., and Bickle,T.A. (1979). A novel bacteriophage defence mechanism: the anti-restriction protein. Nature *278*, 30-34.

Steinberg,J.J., Cajigas,A., and Brownlee,M. (1992). Enzymatic shot-gun 5'-phosphorylation and 3'-sister phosphate exchange: a two-dimensional thin-layer chromatographic technique to measure DNA deoxynucleotide modification. Journal of Chromatography: Biomedical Applications *574*, 41-55.

Studier,F.W. and Movva,N.R. (1976). SAMase gene of bacteriophage T3 is

responsible for overcoming host restriction. J. Virol. *19*, 136-145.

Sullivan,C.H. and Grainger,R.M. (1987). Delta-crystallin genes become hypomethylated in postmitotic lens cells during chicken development. Proc. Natl. Acad. Sci. U. S. A *84*, 329-333.

Swisher,J.F., Rand,E., Cedar,H., and Marie Pyle,A. (1998). Analysis of putative RNase sensitivity and protease insensitivity of demethylation activity in extracts from rat myoblasts. Nucl. Acids. Res. *26*, 5573-5580.

Tada,M., Tada,T., Lefebvre,L., Barton,S.C., and Surani,M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. EMBO J. *16*, 6510-6520.

Tardy-Planechaud,S., Fujimoto,J., Lin,S.S., and Sowers,L.C. (1997). Solid phase synthesis and restriction endonuclease cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine. Nucleic Acids Res. *25*, 553-559.

Trewick,S.C., Henshaw,T.F., Hausinger,R.P., Lindahl,T., and Sedgwick,B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature *419*, 174-178.

Tweedie,S., Charlton,J., Clark,V., and Bird,A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol. Cell. Biol. *17*, 1469-1475.

Urieli-Shoval,S., Gruenbaum,Y., and Razin,A. (1983). Sequence and substrate specificity of isolated DNA methylases from Escherichia coli C. J. Bacteriol. *153*, 274-280.

Vairapandi,M. and Duker,N.J. (1993). Enzymic removal of 5-methylcytosine from DNA by a human DNA-glycosylase. Nucleic Acids Res *21*, 5323-5327.

Vasques,L.R., Klockner,M.N., and Pereira,L.V. (2002). X chromosome inactivation: how human are mice? Cytogenet. Genome Res *99*, 30-35.

Verona,R.I., Mann,M.R., and Bartolomei,M.S. (2003). Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu. Rev Cell Dev. Biol *19*, 237-259.

Wade,P.A., Gegonne,A., Jones,P.L., Ballestar,E., Aubry,F., and Wolffe,A.P. (1999). Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet. *23*, 62-66.

Walkinshaw,M.D., Taylor,P., Sturrock,S.S., Atanasiu,C., Berge,T., Henderson,R.M., Edwardson,J.M., and Dryden,D.T. (2002). Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. Mol. Cell 9, 187-194.

Walsh,C.P., Chaillet,J.R., and Bestor,T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 20, 116-117.

Walters,D.K. and Jelinek,D.F. (2002). The effectiveness of double-stranded short inhibitory RNAs (siRNAs) may depend on the method of transfection. Antisense Nucleic Acid Drug Dev. 12, 411-418.

Weiss,A., Keshet,I., Razin,A., and Cedar,H. (1996). DNA demethylation in vitro: involvement of RNA. Cell 86, 709-718.

Woodcock,D.M., Lawler,C.B., Linsenmeyer,M.E., Doherty,J.P., and Warren,W.D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. J. Biol Chem. 272, 7810-7816.

Wright,F. (1990). The 'effective number of codons' used in a gene. Gene 87, 23-29.

Yoder,J.A. and Bestor,T.H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum. Mol. Genet. 7, 279-284.

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 in Genetics 13, 335-340.

Yoon,J.H., Iwai,S., O'Connor,T.R., and Pfeifer,G.P. (2003). Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. Nucleic Acids Res 31, 5399-5404.

Zhang,Y., Ng,H.H., Erdjument-Bromage,H., Tempst,P., Bird,A., and Reinberg,D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 13, 1924-1935.

Zhang,Y. and Reinberg,D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev. 15, 2343-2360.

Zhu,B., Zheng,Y., Angliker,H., Schwarz,S., Thiry,S., Siegmann,M., and Jost,J.P. (2000a). 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. Nucleic Acids Res. *28*, 4157-4165.

Zhu,B., Zheng,Y., Hess,D., Angliker,H., Schwarz,S., Siegmann,M., Thiry,S., and Jost,J.P. (2000b). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. Proc. Natl. Acad. Sci. U. S. A *97*, 5135-5139.

Appendix

Appendix I:
T3 SAMase gene

```
  1 ATG ATT TTC ACT AAA GAG CCT GCG CAC GTC TTC TAT GTA CTG GTT 45
     M   I   F   T   K   E   P   A   H   V   F   Y   V   L   V
 46 TCC GCT TTC CGT TCT AAC CTC TGC GAT GAG GTG AAT ATG AGC AGA 90
     S   A   F   R   S   N   L   C   D   E   V   N   M   S   R
 91 CAC CGC CAC ATG GTA AGC ACT TTA CGT GCC GCA CCG GGT CTT TAT 135
     H   R   H   M   V   S   T   L   R   A   A   P   G   L   Y
136 GGC TCC GTT GAG TCA ACC GAT TTG ACC GGG TGC TAT CGT GAG GCA 180
     G   S   V   E   S   T   D   L   T   G   C   Y   R   E   A
181 ATC TCA AGC GCA CCA ACT GAG GAA AAA ACT GTT CGT GTA CGC TGC 225
     I   S   S   A   P   T   E   E   K   T   V   R   V   R   C
226 AAG GAC AAA GCG CAG GCA CTC AAT GTT GCA CGC CTA GCT TGT AAT 270
     K   D   K   A   Q   A   L   N   V   A   R   L   A   C   N
271 GAG TGG GAG CAA GAT TGC GTA CTG GTA TAC AAA TCA CAG ACT CAC 315
     E   W   E   Q   D   C   V   L   V   Y   K   S   Q   T   H
316 ACG GCT GGT CTG GTG TAC GCT AAA GGT ATC GAC GGG TAT AAG GCT 360
     T   A   G   L   V   Y   A   K   G   I   D   G   Y   K   A
361 GAA CGT CTG CCG GGT AGT TTC CAA GAG GTT CCT AAA GGC GCA CCG 405
     E   R   L   P   G   S   F   Q   E   V   P   K   G   A   P
406 CTG CAA GGC TGC TTC ACT ATT GAT GAG TTC GGT CGC CGC TGG CAA 450
     L   Q   G   C   F   T   I   D   E   F   G   R   R   W   Q
451 GTA CAA TAA   459
     V   Q   .
```

Figure A.1: Alignment of the bacteriophage T3 SAMase gene (black) and protein (red) sequences. This DNA sequence was utilised in the construction of SAMase expressing plasmids and the Codon Usage analysis.

Appendix II:
Codon Usage Table

Table A.1:Codon usages utilised in bacteriophage T3 and human.

| Amino acids	Codon	Usage in		Amino acids	Codon	Usage in	
		T3	human			T3	human
Gly	GGG	12%	25%	Trp	TGG	100%	100%
Gly	GGA	14%	25%	Stop	TGA	24%	49%
Gly	GGT	50%	16%	Cys	TGT	46%	45%
Gly	GGC	23%	34%	Cys	TGC	54%	55%
Glu	GAG	60%	58%	Stop	TAG	6%	23%
Glu	GAA	40%	42%	Stop	TAA	70%	28%
Asp	GAT	40%	46%	Tyr	TAT	37%	44%
Asp	GAC	60%	54%	Tyr	TAC	63%	56%
Val	GTG	29%	47%	Leu	TTG	14%	13%
Val	GTA	23%	12%	Leu	TTA	11%	7%
Val	GTT	29%	18%	Phe	TTT	29%	46%
Val	GTC	19%	24%	Phe	TTC	71%	54%
Ala	GCG	21%	11%	Ser	TCG	8%	6%
Ala	GCA	21%	23%	Ser	TCA	15%	15%
Ala	GCT	42%	26%	Ser	TCT	27%	18%
Ala	GCC	16%	40%	Ser	TCC	21%	22%
Arg	AGG	4%	21%	Arg	CGG	4%	21%
Arg	AGA	8%	21%	Arg	CGA	11%	11%
Ser	AGT	12%	15%	Arg	CGT	42%	8%
Ser	AGC	16%	24%	Arg	CGC	31%	19%
Lys	AAG	67%	57%	Gln	CAG	60%	74%
Lys	AAA	33%	43%	Gln	CAA	40%	26%
Asn	AAT	25%	47%	His	CAT	28%	42%
Asn	AAC	75%	53%	His	CAC	72%	58%
Met	ATG	100%	100%	Leu	CTG	35%	40%
Ile	ATA	9%	16%	Leu	CTA	9%	7%
Ile	ATT	46%	36%	Leu	CTT	15%	13%
Ile	ATC	45%	48%	Leu	CTC	16%	20%
Thr	ACG	16%	12%	Pro	CCG	40%	12%
Thr	ACA	14%	28%	Pro	CCA	20%	27%
Thr	ACT	33%	24%	Pro	CCT	34%	28%
Thr	ACC	36%	36%	Pro	CCC	6%	33%
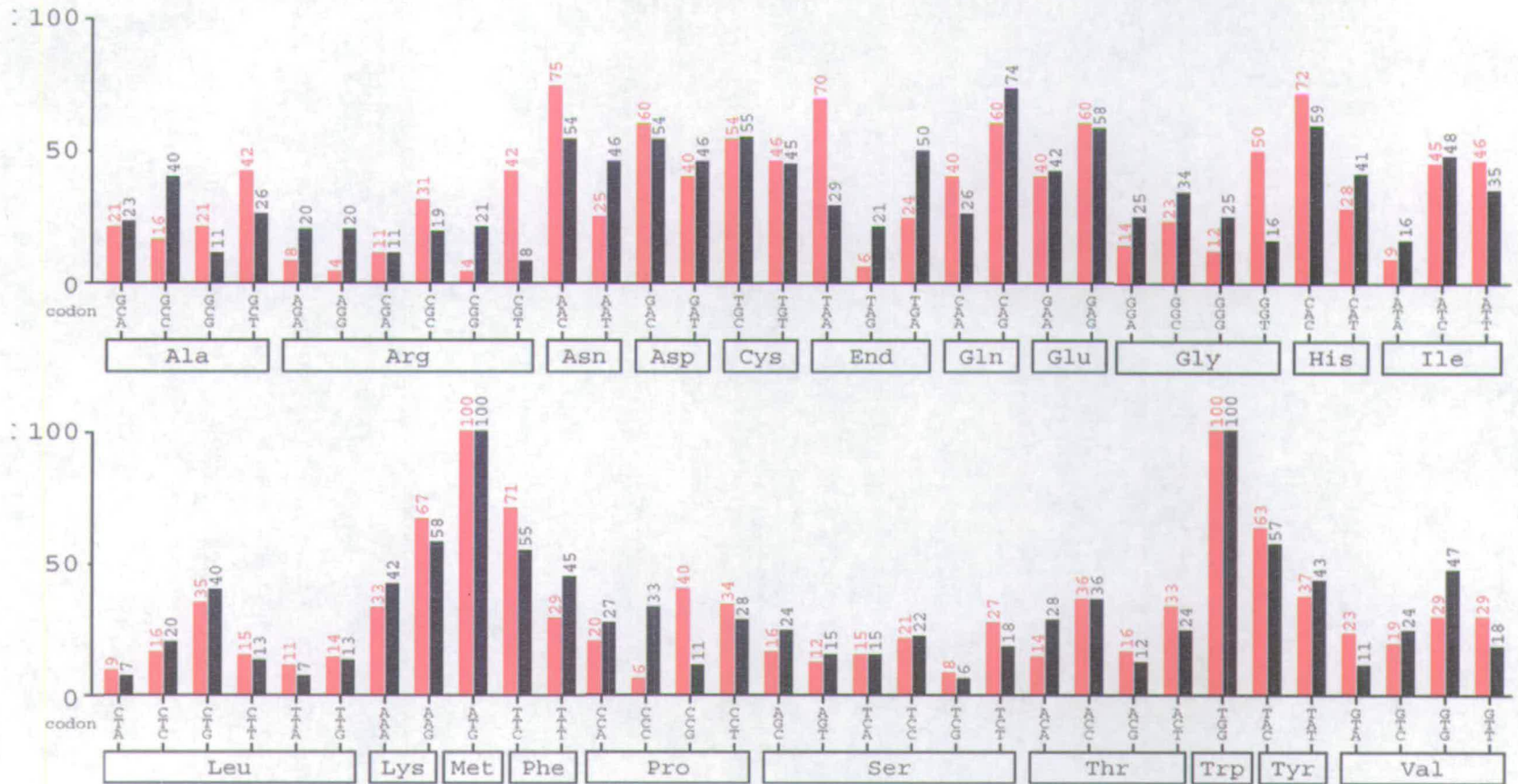
Figure A.2: Codon usage comparison between human and bacteriophage T3. The codon usages in the bacteriophage T3 (red) and human (black) genomes were calculated by the percentage of one codon appearing in the synonymous codons, as shown in Table A.1. Values of each codon frequency were presented on top of the columns. Codons and the consequent amino acid are specified at the bottom.