

## SURVEY AND SUMMARY

# 5-methylcytosine in RNA: detection, enzymatic formation and biological functions

Yuri Motorin<sup>1</sup>, Frank Lyko<sup>2</sup> and Mark Helm<sup>3,4,\*</sup>

<sup>1</sup>Laboratoire ARN-RNP Maturation-Structure-Fonction, Enzymologie Moléculaire et Structurale (AREMS), UMR 7214 CNRS-UHP Faculté des Sciences et Techniques, Université Henri Poincaré, Nancy 1, Bld des Aiguillettes, BP 70239, 54506 Vandoeuvre-les-Nancy, France, <sup>2</sup>Division of Epigenetics, German Cancer Research Center (DKFZ), <sup>3</sup>Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Im Neuenheimer Feld 364, 69120 Heidelberg, Germany and <sup>4</sup>Institute of Pharmacy, University of Mainz, Staudingerweg 5, 55099 Mainz, Germany

Received October 11, 2009; Revised November 10, 2009; Accepted November 12, 2009

### ABSTRACT

The nucleobase modification 5-methylcytosine ( $m^5C$ ) is widespread both in DNA and different cellular RNAs. The functions and enzymatic mechanisms of DNA  $m^5C$ -methylation were extensively studied during the last decades. However, the location, the mechanism of formation and the cellular function(s) of the same modified nucleobase in RNA still remain to be elucidated. The recent development of a bisulfite sequencing approach for efficient  $m^5C$  localization in various RNA molecules puts ribo- $m^5C$  in a highly privileged position as one of the few RNA modifications whose detection is amenable to PCR-based amplification and sequencing methods. Additional progress in the field also includes the characterization of several specific RNA methyltransferase enzymes in various organisms, and the discovery of a new and unexpected link between DNA and RNA  $m^5C$ -methylation. Numerous putative RNA: $m^5C$ -MTases have now been identified and are awaiting characterization, including the identification of their RNA substrates and their related cellular functions. In order to bring these recent exciting developments into perspective, this review provides an ordered overview of the detection methods for RNA methylation, of the biochemistry, enzymology and molecular biology of the corresponding modification enzymes, and discusses perspectives for the emerging biological functions of these enzymes.

### INTRODUCTION

Nucleotide modifications, a ubiquitous feature of life, have first been described in 1948, when the occurrence of  $m^5C$  was detected in DNA (1). Today, about half a dozen DNA modifications and more than 100 chemically distinct RNA nucleotide modifications, including  $m^5C$ , are known to be enzymatically synthesized. While research into nucleotide modifications has been hampered by limited performance of analytical methods, the field has recently encountered renewed interest. This is related to the discovery of nucleotide modifications in regulatory RNAs including siRNA, piRNAs, miRNAs and precursors (2–7). Among the various modifications, a privileged position is held by  $m^5C$ . This residue has been described in several cellular RNAs and a complex enzymatic machinery for its synthesis was found in representative organisms from all kingdoms of life. In addition, recent data strongly suggest its implication in the regulation of various biological processes.

### METHODS FOR $m^5C$ DETECTION IN RNAs

Detection of nucleotide modifications is based on either of three basic principles: (i) separation according to physicochemical properties, (ii) differential enzymatic turnover and (iii) differential chemical reactivity.

Separation according to physico-chemical properties typically relies on chromatography, i.e. thin layer chromatography (TLC), high-performance liquid chromatography (HPLC), or mass spectrometry, and sometimes on a combination of both (LC-MS). Most widespread is the use of two-dimensional TLC. Because the migration characteristics of nearly 70 chemically distinct

\*To whom correspondence should be addressed. Tel: +6131 3925731; Fax: +6131 3923779; Email: mhelm@uni-mainz.de

modifications have been mapped in three standard solvent systems, identification and quantification of  $^{32}\text{P}$ -labeled nucleotides down to the femtomolar level is straightforward (8). Although the use of *in vitro* substrates containing a single type of labeled nucleotide by nearest neighbor analysis permits the determination of the nucleotide 3' to the modification, the obligatory nuclease degradation step before TLC analysis destroys most of the modification's sequence context. For known modification sites, sequence specificity of the analysis can be reintroduced by the use of DNAzymes (9) or oligonucleotide-directed RNase H cleavage (10) prior to labeling and TLC analysis. It has also been shown that chemical conjugation of nucleotides to a fluorescent dyes and subsequent separation and detection by CE-LIF allows quantifications of  $\text{m}^5\text{C}$  in DNA (11) in the femtomolar range. A similar approach has also been used for the detection of  $\text{m}^5\text{C}$  in RNA (12).

Similar to TLC and CE techniques, a typical LC-MS analysis requires a complete nuclease digestion, which is followed by dephosphorylation prior to separation on a C18 reverse phase column. In addition to the retention times, determination of the nucleosides' molecular mass, and/or their fragmentation pattern is used to confirm the chemical identity of a given modification. Applications of current LC/MS/MS techniques allow the detection of modified nucleosides even in the sub-picomolar range (13). Modifications can also be detected by LC-MS in oligonucleotides obtained, e.g. by digestion with RNase T1 (14), but sequence information is typically limited to ~10 nucleotides near the modification site.

Since direct identification through physico-chemical properties prevents the use of PCR amplification it also prevents high throughput applications. While the ability of certain nucleotide modifications to impede enzymatic reactions such as ligations (15) or primer elongations (16,17) can be used for sequence-specific detection, this does not apply to  $\text{m}^5\text{C}$ , which behaves normally in terms of Watson-Crick base pairing. However, a promising perspective is offered by differential reactivity of  $\text{m}^5\text{C}$  towards chemical reagents, which allows the application of PCR-based amplification steps, or affinity-based enrichment of  $\text{m}^5\text{C}$ -containing nucleic acids.

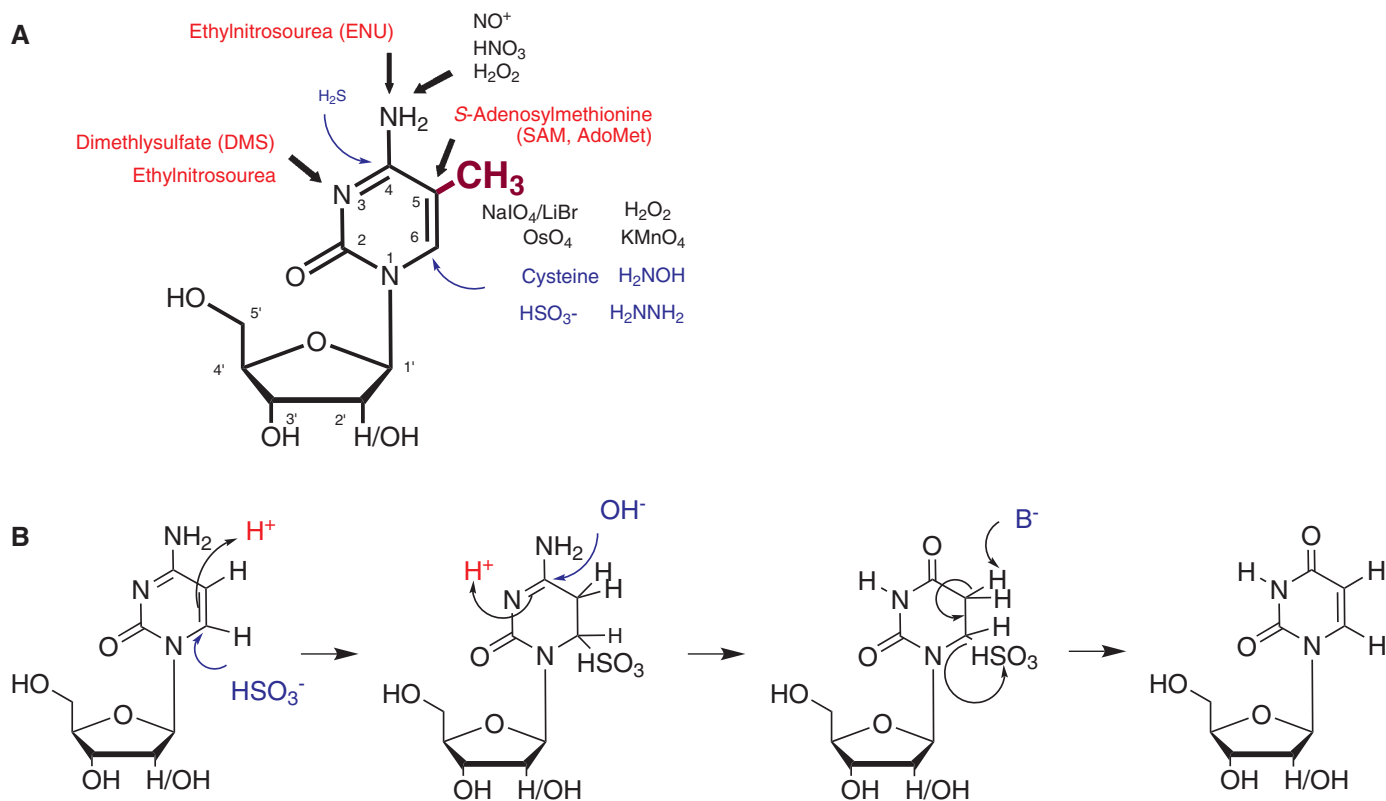
### Distinct chemical reactivity of unmodified C and $\text{m}^5\text{C}$

Figure 1A illustrates the relevant reactive sites investigated in 5-methyl and 2'-deoxy-derivatives of cytidine. Of foremost interest are the exocyclic N4-amino function and the C5-C6 double-bond, because reactivities of both sites are significantly affected by the introduction of a methyl group on C5. Methylation generally increases electron density in the aromatic ring by an inductive effect, which can be monitored by chemical shifts in  $^1\text{H-NMR}$ . Thus, C5-methylation of pyrimidines causes an upfield shift of ~0.15 p.p.m. in the H6 of both uridine and cytidine (see the SDBSWeb: <http://riodb01.ibase.aist.go.jp/sdbs/> (National Institute of Advanced Industrial Science and Technology, 14.7.2009)). This increase in electron density favors reaction types such as oxidation and electrophilic attacks, while disfavoring nucleophilic

attacks. There is clearly a strong interest to exploit these features for the detection of methylated pyrimidines by a combination of chemical and biochemical techniques. Some of corresponding approaches will be outlined below.

Oxidation of the 5-6 double-bond in pyrimidines to the corresponding glycol has been reported to result from exposure to ionizing radiation or reagents such as  $\text{KMnO}_4$ ,  $\text{OsO}_4$ , or Fenton reagents (18-22). The concomitant loss of aromaticity renders the nucleic acid chain susceptible to cleavage by aniline or piperidine in the case of RNA or DNA, respectively (23-25). These reactions are exploited for sequencing, structural probing, or footprinting applications of end-labeled nucleic acids (18,19). Although the permanganate reaction produces insoluble  $\text{MnO}_2$ , and further oxidation of the glycol to the bis-aldehyde may occur under certain conditions, it has been successfully employed in the discrimination of deoxy- $\text{m}^5\text{C}$  versus deoxycytidine (26). The osmium tetroxide reactions do not involve the formation of insoluble products, but the reagent is highly toxic and mutagenic. In the past years, a variety of osmate-based bioconjugate reagents have been developed for the selective detection of  $\text{m}^5\text{C}$  in DNA. These reagents exploit the additional electron density brought about by the methyl group for highly selective formation of a stable complex with deoxy- $\text{m}^5\text{C}$  but not with deoxycytidine. However, applications to RNA are still lacking (27).

Compared to cytidine,  $\text{m}^5\text{C}$  is more prone to spontaneous deamination of the N4 than cytidine *in vivo* (28). This reaction is promoted by oxidizing conditions, and is thought to involve the formation of an electron-deficient species, such as  $\text{NO}^+$ , which converts the N4-anilin function into a leaving group which is then hydrolyzed (29,30). Likewise, electrophilic alkylating agents preferentially attack the nitrogen nucleophilic sites. Both the N4- and the N3-aminofunctions are alkylated by highly reactive agents such as ethylnitrosourea (ENU) (31,32). The N3-methylation by dimethylsulfate (DMS) has gained importance in Maxam and Gilbert style sequencing techniques (24,25), as well as in structural probing and footprinting of RNA and DNA (33,34). In these biochemical applications, the N3-methylation renders the aromatic ring electron poor and thus susceptible to nucleophilic addition of hydrazine, which is destroying aromaticity prior to chain scission (24,25). Such nucleophilic additions are otherwise unusual in pyrimidines, and require comparatively more stringent conditions (24). Sodium bisulfite, also a nucleophilic reagent, causes the deamination of cytosine residues in a single-strand DNA through the initial formation a 5,6-dihydrocytosine-6-sulfonate intermediate at acidic pH (Figure 1B). This implies a protonation step at the 5-position in conjunction with a Michael addition of the nucleophile at the 6 position. Subsequent deamination leads to 5,6-dihydrouracil-6-sulfonate from which the 5-6 double-bond is restored by beta-elimination during the subsequent treatment under basic conditions. Because this reaction occurs with selectivity for cytosine over  $\text{m}^5\text{C}$ , its application to the detection of  $\text{m}^5\text{C}$  sites in DNA has gained outstanding importance in an approach called bisulfite sequencing (27).



**Figure 1.** (A) Chemical reactivity of cytosine and  $m^5C$ . The methyl group of  $m^5C$  is shown in purple. Attack sites on the nucleobase are indicated by arrows, where black indicates oxidizing agents, red indicates alkylating electrophiles and blue arrows indicate nucleophiles. Numbering of the nucleobase atoms 1–6 is indicated inside the ring. Ribose carbon atom numbering is indicated by primed (') numbers (B) Bisulfite mediated conversion of cytosine to uracil. Nucleophiles, bases and nucleophilic attacks are indicated in blue, acids are indicated in red.

### Adaptation of DNA bisulfite methylation mapping to RNA

The principle of bisulfite sequencing is outlined in Figure 2. A critical parameter is the fraction of intact analyte nucleic acids surviving the sequential treatment under acidic bisulfite and alkaline conditions (Figure 1B). In these polymers, all cytidines are deaminated to uridines, and will thus behave like thymidines in polymerase-based sequencing or PCR-amplification reactions while  $m^5C$  residues remain unchanged and behave like cytidines. Typically, the resulting fragments are PCR-amplified before cloning and/or sequencing. Sequence alignments are facilitated by sequencing of untreated samples. In alignments of multiple bisulfite sequence reads, significant methylation is indicated by multiple occurrences of cytosine signals at a given position. The sequencing depth, i.e. the number of individual sequence reads in the alignment, is thus an important quality parameter of the analysis.

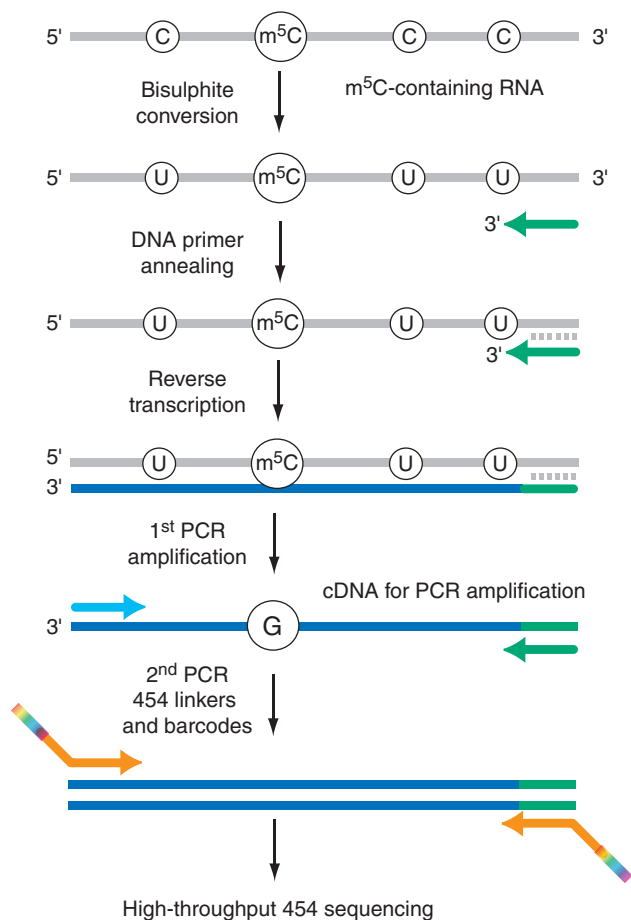
DNA bisulfite sequencing (35,36) has become an indispensable tool for the mapping of  $m^5C$  in DNA, and its continued development represents a key approach towards characterizing the human epigenome (37,38). An adaptation of the bisulfite treatment protocol to the requirements of RNA showed that individual  $m^5C$  residues in tRNA could be mapped by primer extension (39). The same study also suggested that complex tRNA modification patterns could be analyzed by bisulfite

treatment of RNA samples, followed by reverse transcription into cDNA, PCR amplification, molecular cloning and DNA sequencing of cloned PCR products. However, no experimental details were provided for this approach.

A stringent analysis of cytosine methylation patterns by bisulfite sequencing requires very high cytosine to uridine conversion rates of  $>95\%$ . This is usually achieved only after prolonged ( $>1$  h) incubation with sodium bisulfite, which promotes degradation of DNA and, particularly, RNA molecules and therefore reduces the efficacy of the ensuing PCR amplification step. Sustaining high deamination rates, and, at the same time, limiting the degradation of RNA molecules thus represented the major challenge for the establishment of an RNA bisulfite sequencing protocol. This problem has recently been solved by reducing the bisulfite reaction temperature from 95 to 60°C, which permitted an extension of the reaction time to 3 h and more, without any major reduction in the efficacy of the RT-PCR reaction (40). Under these conditions, deamination rates of 97% and more could be repeatedly achieved for single-stranded RNA sequences.

### High-throughput techniques of $m^5C$ mapping and quantification in RNA

Initial tRNA bisulfite sequencing experiments indicated a substantial variation in the cytosine methylation



**Figure 2.** Principle of  $m^5C$  detection by RNA bisulfite sequencing and combination with next-generation sequencing. The DNA primer sequence for reverse transcription is adapted for C to U conversions in the analyzed RNA.

patterns of tRNAs. The robust quantification of cytosine methylation levels at specific positions therefore requires a considerable sequencing depth of 100 (or more) sequence reads. This can be achieved by next-generation sequencing technologies, which permit the massively parallel sequencing of hundreds of thousands single DNA molecules. The sequencing depth provided by next-generation sequencing also permits the analysis of complex methylation patterns. Indeed, 454 sequencing has recently been used for the methylation analysis of various tRNA and rRNAs and has facilitated the quantitative analysis of RNA methylation patterns in these RNA molecules (40).

For 454 RNA bisulfite sequencing, PCR amplicons are generated from deaminated RNA following reverse transcription and then re-amplified with 454 primers (Figure 2). The latter primers contain linker sequences for the 454 sequencing technology and barcode sequences for the bioinformatical identification of individual amplicons. Amplicons are then mixed, sequenced and later subjected to bioinformatical analysis (40).

Importantly, the results obtained with these sequencing approaches can also be used to robustly quantify the methylation level of individual cytosine residues. It is conceivable that RNA methylation can be used as a signal for

various cellular processes. In this context, the methyl group would be a dynamic mark that can be added or removed depending on intrinsic or extrinsic cues. Deep sequencing of bisulfite amplicons therefore provides an opportunity to use RNA methylation analysis as a quantifiable assay for various biological questions.

Lastly, deep sequencing of cDNA libraries also permits an additional quality control step that addresses the PCR bias problem associated with most bisulfite-based analytical tools. The biased PCR amplification of certain (c)DNA sequences represents a major problem in the quantitative analysis of bisulfite sequencing experiments, because it can result in a dramatic over- or under-representation of particular PCR products (41). A focused analysis of sequences with unique deamination artifacts or the introduction of random sequence permutations in an additional barcode patch in the reverse transcription primer allows a 1:1 representation of RNA molecules and sequence reads. These are important steps towards the further optimization of the method that will significantly broaden its application potential.

## THE PRESENCE OF $m^5C$ IN VARIOUS CELLULAR RNAs

$m^5C$  is a prevalent modification of many cellular RNAs. Stable and highly abundant RNAs, like tRNAs and rRNAs were extensively analyzed in the past by various techniques. These studies allowed the precise mapping of  $m^5C$  residues in tRNAs and also in ribosomal RNAs from all three living domains.

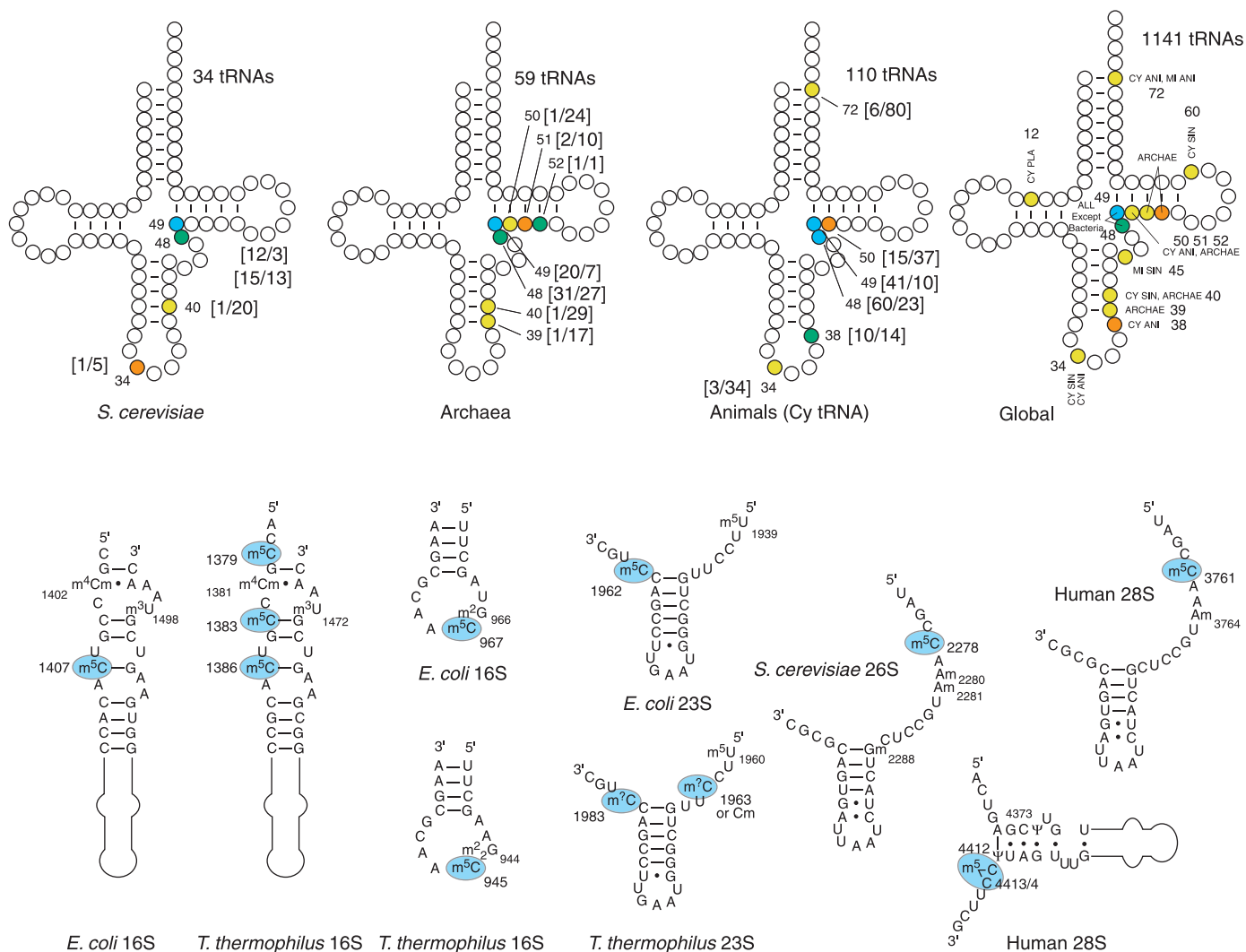
### tRNAs

While tRNAs from eubacterial organisms do not contain  $m^5C$ , this modification was detected in numerous archaeal and eukaryotic tRNAs. For the best studied eukaryotic tRNAs,  $m^5C$  residues seem to be clustered in the junction between variable region and T $\Psi$ C-stem, and positions 48 and 49 are most frequently modified. However, other locations in the anticodon loop were also identified. The same tendency exists also for archaeal tRNAs, where the positions 48/49 are most frequently methylated, but other methylation sites have also been reported. In contrast to other organisms, higher eukaryota frequently have an additional  $m^5C$  residue in the tRNA acceptor stem, at position 72. The overall distribution of  $m^5C$ -modified positions in yeast, archaeal and higher eukaryotic tRNAs is shown in Figure 3 (42).

### Prokaryotic and eukaryotic rRNAs

The presence of  $m^5C$  in ribosomal RNAs from different organisms was demonstrated by total digestion of highly purified rRNAs followed by RP-HPLC and MS analysis of the resulting nucleoside mixtures (43–46). Bacteria have a variable number of  $m^5C$  residues; only three such modifications are present in *Escherichia coli* rRNAs, while rRNAs from *Thermus thermophilus* has five or six  $m^5C$  residues (Figure 3B). Analysis of small and large subunit rRNAs from bacteria and eukaryota similarly demonstrated the presence of several  $m^5C$  residues in





**Figure 3.** Occurrence of  $m^5C$  residues in tRNAs and rRNAs. The frequency of  $m^5C$  occurrence is color-coded in tRNA: yellow, <10%; orange, between 10 and 30%; green, between 30 and 60%; blue >60%. Numbers in square brackets show the number of  $m^5C$  and number of unmodified or other modified C at a given position [ $m^5C/C$ ]. The same color code is used for global distribution in all sequenced tRNAs, numbers are not shown for clarity. Abbreviations: CY, cytoplasm; MI, mitochondria; ANI, animals; SIN, single cell eukaryota; PLA, plants; ARCHAE, Archaea. The lower panel shows rRNA fragments from different species with  $m^5C$  residues highlighted in blue.

these molecules (see The RNA Modification database <http://library.med.utah.edu/RNAMods/>). More recently,  $m^5C$  was also detected in archaeal rRNAs (47). In addition, the locations of this modification in rRNA are rather well conserved from bacteria to humans.

Eukaryota have comparably fewer  $m^5C$  residues in rRNA. *Saccharomyces cerevisiae* rRNA methylation has not been exhaustively analyzed, but one  $m^5C$  residue was mapped in 25S rRNA from *S. carlsbergensis* (48). Similarly, human 28S rRNA contains only two  $m^5C$  nucleotides (49) and no  $m^5C$  was detected in several eukaryotic 18S rRNAs (50). Although very little is known about eukaryotic mitochondrial rRNA methylation, the presence of  $m^5C$  was reported for 18S mt rRNA from the Syrian hamster *Mesocricetus auratus* (the SSU rRNA Modification Database, WEB <http://134.58.19.20/ssu/ssu.htm>).

Analysis of archaeal rRNAs confirmed the presence of  $m^5C$  residues in SSU and LSU rRNAs, but demonstrated

that the number of modifications present in archaeal rRNAs is highly variable. For example, no  $m^5C$  was detected in *Haloflex volcanii* 16S rRNA, while its counterpart from *Sulfolobus solfataricus* has at least one residue in 16S rRNA and one or two  $m^5C$  in 23S rRNA (47,51). For a comparative methylation analysis of selected prokaryotic and eukaryotic rRNAs see Ref. (52).

### $m^5C$ in other cellular RNAs

Shortly after the discovery of nucleotide methylation in the cap-structure of eukaryotic mRNAs, numerous investigations of the presence of methylated nucleotides mRNAs and viral RNAs were reported. These were typically based on pre-labeling with  $^3H$ -methionine, which is converted to [ $^3H$ ]S-adenosyl-L-methionine (SAM) in living cells followed by incorporation of  $^3H-CH_3$  into RNA. Several studies reported the detection of  $m^5C$  in mRNA from cultured BHK-21 hamster cells (53), in S26 sindbis virus RNA from infected hamster

**Table 1.** Known and some putative m<sup>5</sup>C-MTRs and their RNA substrates

Enzyme family	Enzyme name	Other names	Organism	Life domain	Accession	Identification type	RNA substrate	Modified position(s)	X-ray structure	Reference
RsmB family										
RsmB/Nol1	RsmB	Fmu/Fmv	<i>Escherichia coli</i>	B	AP_004502	Biochemical	16S rRNA	967	1SQF, 1SQG	(68,70)
RsmB/Nol1	P120	NSUN1	<i>Homo sapiens</i>	E	P46087	NO	Unknown	NA		(92)
RsmB/Nol1	Nop2		<i>Saccharomyces cerevisiae</i>	E	YNL061W	NO	Unknown	NA		(87)
RsmF/YebU family										
YebU			<i>Haloferax volcanii</i>	A	HVO_1594	Bioinformatics	tRNA	39/40/48/49		(75)
YebU			<i>Pyrococcus abyssi</i>	A	PAB1947	Biochemical	tRNA	Multiple (mostly 49)		(79)
YebU	aTrm4		<i>Pyrococcus horikoshii</i>	A	PH1374	Bioinformatics	Unknown	NA	1HXK	(78)
YebU	RsmF	YebU	<i>Escherichia coli</i>	B	P76273	Biochemical	16S rRNA	1407	2FRX	(71,72)
YebU	hTrm4	NSUN2/Misu	<i>Homo sapiens</i>	E	NM_017755	Biochemical	pre-tRNA <sup>Leu</sup>	34		(86)
YebU	FLJ22609	NSUN3	<i>Homo sapiens</i>	E	Q9H649	NO	Unknown	NA		(65)
YebU	MGC22920	NSUN4	<i>Homo sapiens</i>	E	Q96CB9	NO	Unknown	NA		(65)
YebU	Trm4	Ncl1	<i>Saccharomyces cerevisiae</i>	E	YBL024W	Biochemical	tRNA and pre-tRNA	48/49 tRNA 34/40 pre-tRNA		(80,81)
Dnmt2 family										
DNMT2	Dnmt2		<i>Drosophila melanogaster</i>	E	Q9U6H7	Biochemical	tRNA	38		(102,107)
DNMT2	Dnmt2	trdmt1	<i>Drosophila rerio</i>	E	Q588C1	Biochemical	tRNA	38?		(122)
DNMT2	Dnmt2		<i>Homo sapiens</i>	E	O14717	Biochemical	tRNA	38	1G55	(100,102)
DNMT2	Dnmt2		<i>Mus musculus</i>	E	O55055	Bioinformatics				
DNMT3	pmt1	pmt1	<i>Saccharomyces pombe</i>	E	P40999	Bioinformatics				
RlmI family										
COG1092	RlmI	YccW	<i>Escherichia coli</i>	B	P75876	Biochemical	23S rRNA	1962	3C0K	(73,74)
Ynl022 family										
Ynl022c	WBSCR20A	NSUN5A	<i>Homo sapiens</i>	E	Q96P11	NO	Unknown	NA	2B9E	
Ynl022c	WBSCR20B	NSUN5B	<i>Homo sapiens</i>	E	Q3KNT7	NO	Unknown	NA		
Ynl022c	WBSCR20C	NSUN5C	<i>Homo sapiens</i>	E	Q63ZY6	NO	Unknown	NA		
Ynl022c	WBSCR22		<i>Homo sapiens</i>	E	O43709	NO	Unknown	NA		
Ynl022c		NSUN7	<i>Homo sapiens</i>	E	Q8NE18	NO	Unknown	NA		
Ynl022c	Ynl022c		<i>Saccharomyces cerevisiae</i>	E	YNL022C	NO	Unknown	NA		
NSUN6 family										
NSUN6		NSUN6	<i>Homo sapiens</i>	E	Q8TEA1	NO	Unknown	NA		

Abbreviations: A: archaea; B: bacteria; E eukaryota; NA not analyzed.

cells (54,55), and in adenovirus RNA from infected HeLa cells (56). However, no m<sup>5</sup>C was found in mRNA from HeLa cells (57,58), Novikoff hepatoma cells (59), mouse myeloma (60) and SV 40 viral RNA (61).

Certain findings suggest that classical tRNA modification enzymes may also methylate tRNA-like structures in other RNA molecules. Thus, RNA derived from tRNA-related short interspersed nuclear elements (SINES) was found to be methylated to m<sup>5</sup>C *in vitro* (62), although a recent analysis failed to detect m<sup>5</sup>C methylation of this RNA *in vivo* (63). Moreover, the tRNA-like 3'-end of the Turnip Yellow Mosaic Virus (TYMV) RNA was found to be methylated to m<sup>5</sup>C upon injection into *Xenopus* oocytes (64). It will be interesting to further investigate the possibility that other cellular RNAs might be methylated in order to fully understand the biological function of this modification (see below).

## FAMILIES OF RNA: m<sup>5</sup>C-MTases

The known enzymes transferring methyl groups from SAM to nucleic acids belong to the SPOUT and MTase superfamilies, the latter containing a Rossmann fold for the accommodation of the cofactor (65). RNA:m<sup>5</sup>C-MTases represent a large protein family which can be subdivided into several distinct groups, on the basis of sequence similarity and RNA substrate specificity. These subgroups are presented in Table 1.

The first tentative approach towards a whole-genome analysis of m<sup>5</sup>C-MTases was initiated immediately after the characterization of RsmB as a bacterial m<sup>5</sup>C-MTase (66). More than 30 homologous proteins were identified by a BLAST search and a classification into eight distinct subgroups was proposed. However, only two *E. coli* genes (RsmB and YebU) were listed, while the third bacterial protein (YccW = RlmI) showed only little sequence similarity to RsmB.

An exhaustive bioinformatical analysis of structure–function relationships in the m<sup>5</sup>C-MTase family was later performed by Bujnicki *et al.* (65). The known m<sup>5</sup>C-MTases were subdivided into four major sub-families (related to Nop2/Nol1, YebU/Trm4, a large group related to RsmB or Ynl022c and a small group represented by *P. horikoshii* PH1991 and human NSUN6). This classification is also used in Table 1.

Recently, a careful analysis of m<sup>5</sup>C-MTase homologs in higher eukaryota (67) suggested the existence of a new, exclusively eukaryotic, subgroup of m<sup>5</sup>C-MTase-related proteins called RCMT9. These members are distantly related to Trm4 and were detected only in Viridiplantae, Alveolata, Euglenozoa and Mycetozoa taxa.

## *E. coli* m<sup>5</sup>C-MTases: activity, structure and substrates

All three known m<sup>5</sup>C-MTases from *E. coli* were extensively characterized, both biochemically and structurally. Since m<sup>5</sup>C in bacteria is only found in ribosomal,

RNA, all three enzymes are rRNA specific, but recognize their RNA target in a differential manner.

*Fmu/fmv* (*RsmB*). The Fmu/fmv protein from *E. coli* was the first identified and characterized RNA-specific m<sup>5</sup>C-MTase (68,69). A partially purified protein fraction possessing 16S rRNA:m<sup>5</sup>C967-MTase activity was analyzed by N-terminal sequencing and was found to match to the first 23 amino acids of Fmu/fmv (later renamed to RsmB). This is the first and so far still the only example for a biochemical identification of an RNA m<sup>5</sup>C-MTase. With the exception of YccW(RlmI) and Dnmt2-related proteins, all enzymes discussed below have been identified by sequence similarity with the Fmu protein and their activity was further confirmed by various *in vivo* and *in vitro* tests. RsmB from *E. coli* efficiently catalyzes the methylation of naked 16S rRNA substrates, but assembled 30S ribosomal subunits are not methylated (68). In addition, a small 16-nt stem-loop minisubstrate derived from *E. coli* 16S rRNA was methylated by RsmB *in vitro*, albeit less efficiently than the full-size rRNA transcript (69).

The crystal structures of RsmB alone and in complex with SAM have been obtained at high resolution (1.65 and 2.1 Å, respectively) (70). RsmB folds into three almost independent domains, a C-terminal domain that is responsible for the m<sup>5</sup>C-MTase activity, and two other domains (the N-terminal domain and the central N1 domain) that show significant sequence similarity to known RNA-binding proteins and that may mediate the interaction with the 16S rRNA substrate. The crystal structure of the RsmB-16S rRNA complex has not been published yet, but docking of a 56-nt fragment derived from the 16S rRNA substrate clearly demonstrated structure and charge complementarity between the two molecules. In this predicted conformation, the target cytosine comes into close proximity with the catalytic Cys375 and the activated methyl group of SAM. This overall structure of the complex is consistent with the binding of a completely folded rRNA substrate (70).

*YebU* (*RsmF*). Sequence similarity between RsmB and an *E. coli* ORF encoded by the *yebU* gene suggested that YebU also catalyzes m<sup>5</sup>C formation in *E. coli* rRNA. YebU was subsequently cloned and expressed, and its activity was studied *in vitro* and *in vivo* (71). The results demonstrated that the loss of active YebU protein led to the concomitant loss of m<sup>5</sup>C1407 in 16S rRNA, while m<sup>5</sup>C1962 in 23S rRNA remained unaffected. The recombinant YebU protein (later renamed to RsmF) retained its activity, but, in contrast to RsmB, methylated only assembled 30S ribosomal subunits and not naked 16S rRNA or complete 70S *E. coli* ribosomes.

Crystallized RsmF protein shows two distinct structural domains, the m<sup>5</sup>C-MTase domain, which resides in the N-terminal part, and the C-terminal domain, which shows significant similarity to the PUA domain frequently found in other RNA modification enzymes, like RNA: pseudouridine synthases and TGT-transglycosylases (72). *In silico* docking of the 30S ribosomal subunit to the RsmF structure suggests that it interacts not only with

16S rRNA but also with the ribosomal protein S12, providing a plausible explanation as to why naked 16S rRNA is not a substrate for methylation.

*YccW* (*RlmI*). The last putative m<sup>5</sup>C-MTase in *E. coli* (YccW, now renamed to RlmI) has recently been characterized. It is worth mentioning that RlmI only distantly relates to two other *E. coli* m<sup>5</sup>C-MTases (RsmB and RsmF), and its sequence is closer to known m<sup>5</sup>U-MTases. Consequently, RlmI does not contain recognizable PC- and TC- sequence motifs present in other m<sup>5</sup>C-MTases and the catalytic Cys residue is located in the SCS motif close to the C-terminus. On this basis, RlmI was predicted to have pyrimidine-C5 MTase activity. However, mass-spectrometric analysis of 23S rRNA extracted from a YccW-deficient *E. coli* strain clearly linked RlmI activity to the unique m<sup>5</sup>C1962 found in *E. coli* 23S rRNA (73). Recombinant RlmI methylates naked 23S rRNA, but not assembled 50S ribosomal subunits, in spite of the rather high accessibility of the target m<sup>5</sup>C1962 at the subunit interface.

The crystal structure of RlmI has also become available recently (74). The protein folds into three independent domains, an N-terminal domain resembling the PUA domain of other RNA modification enzymes, a central EEHEE domain that is common to other modification enzymes including m<sup>5</sup>U-MTases and a C-terminal catalytic m<sup>5</sup>C-MTase domain. A model of the ternary complex containing RlmI, SAM and an rRNA fragment was analyzed by *in silico* modeling and subsequent minimization of the structure. The results predicted an interaction between RlmI and single-stranded 23S rRNA, rather than the folded and highly constrained two-dimensional structure. This observation provides an explanation for the inability of RlmI to methylate 50S ribosomal subunits, even if the target cytosine residue is highly accessible.

### Archaeal m<sup>5</sup>C-MTases

The number of putative m<sup>5</sup>C MTases in Archaea is quite variable. Only one related protein seems to be present in mesophilic archaeobacteria *H. volcanii* (75), while many hyperthermophilic organisms display several homologous proteins. The presence of m<sup>5</sup>C in bulk tRNA fractions extracted from *P. furiosus* and other hyperthermophilic Archaea was first demonstrated by HPCL/MS-ESI analysis (76). Later on, activity tests performed with cell-free *P. furiosus* extracts *in vitro* allowed the detection of at least m<sup>5</sup>C49- and, probably, m<sup>5</sup>C40-forming activities acting on heterologous tRNA<sup>Ile</sup> from *H. volcanii* and yeast tRNA<sup>Asp</sup> and tRNA<sup>Phe</sup> substrates (77).

The analysis of numerous archaeal genomic sequences demonstrated the presence of multiple putative m<sup>5</sup>C-MTases in many genomes. One of these proteins, the ORF PH1374-encoded protein from *P. horikoshi* showed significant sequence similarity to the putative human p120 m<sup>5</sup>C-MTase (see below). The protein was then recombinantly expressed and its structure was determined at 1.9 Å resolution (78). The RNA substrate of this archaeal m<sup>5</sup>C-MTase is not known, but its overall fold is



very similar to the central domain of bacterial RsmB and RsmF, suggesting similar activity (72).

In the genome sequence of hyperthermophilic *P. abyssi*, the ORF PAB1947 encodes a potential tRNA: m<sup>5</sup>C-MTase and forms a bicistronic operon with the gene encoding archaease, a protein known to stabilize several other archaeal proteins. ORF PAB1947 has been expressed in *E. coli* and its activity was characterized *in vitro*. The major methylation sites corresponded to the expected positions 48 and 49 in a *P. furiosus* tRNA<sup>Asp</sup>, but surprisingly, other cytosines were also methylated, especially at high incubation temperatures. Lower temperatures or the presence of archaease increased the specificity of tRNA substrate recognition (79). These data strongly suggest that archaeal m<sup>5</sup>C-MTases may have a rather broad specificity, compared to their yeast or human counterparts.

### Classes of eukaryotic m<sup>5</sup>C-MTases

At least four distinct groups of RNA:m<sup>5</sup>C-MTases, differing in their sequence and RNA substrate specificity, were found in different eukaryotic organisms. Only three of them were detected in *S. cerevisiae*, but all four groups seem to be present in most eukaryotic genomes.

*Yeast and human Trm4 (Ncl1)*. The *S. cerevisiae* Ncl1 protein was initially characterized as a nuclear protein of unknown function (80). Further biochemical studies demonstrated its role in the modification of yeast tRNAs and some tRNA precursors at positions 34, 40, 48 and 49. Positions 34 and 40 are modified only in the intron-containing pre-tRNA and are specific for tRNA<sup>Leu</sup>(CAA) and tRNA<sup>Phe</sup>(GAA), respectively (81). On the basis of this finding, Ncl1 (encoded by ORF YBL024) was renamed as Trm4 for tRNA-specific MTase 4.

Structure–function relationships in yeast Trm4 were further studied by various approaches like site-directed mutagenesis and limited trypsin digestion. By these methods, the catalytic role of Cys310 in motif IV was confirmed and the importance of the N-terminal domain for cytosine methylation was identified. The C-terminal domain is not required for tRNA binding and for catalysis and seems to fold independently from the N-terminal part of the protein. However, the presence of the C-terminal domain of Trm4 considerably stimulates the MTase activity of the N-terminal catalytic domain (82,83).

Sequence similarity with yeast Trm4 also allowed the cloning and characterization of a human homologue called hTrm4 or Misu/NSUN2. Misu/NSUN2 was found to be involved in Myc-mediated proliferation of cancer cells. It is expressed at low levels in normal epidermis but is up-regulated upon Myc activation (84). Misu/NSUN2 is phosphorylated at Ser139 by the cell-cycle protein kinase Aurora-B, and this phosphorylation affects its m<sup>5</sup>C-MTase activity and its association with nucleophosmin and nucleolin, two nucleolar interaction partners (85). Despite its high similarity with the yeast protein, the tRNA-MTase activity of hTrm4 seems to be restricted to position 34 of the intron containing human

tRNA<sup>Leu</sup>(CAA) (86). Human tRNA methyltransferases mediating m<sup>5</sup>C48/49 methylation still remain to be identified.

*Yeast and human Nop2/p120*. The second member of the family, the yeast *S. cerevisiae* nucleolar protein Nop2 (encoded by ORF YNL061) was shown to be involved in biogenesis of 60S ribosomal subunit and in the maturation of 26S rRNA (87,88). These functions of Nop2, and not the putative RNA-MTase activity, are probably important for the viability of yeast cells. Furthermore, an RNA:m<sup>5</sup>C-MTase activity or an RNA substrate for Nop2 remain to be identified. Site-directed mutagenesis of yeast Nop2 demonstrated that it has a Cys residue (Cys424) essential for viability. All other tested amino acid substitutions did not affect Nop2 functions *in vivo* (89). Further analysis showed that this Cys residue is important for enzyme regeneration at the second step of the reaction and its substitution by alanine or serine led to the accumulation of a covalent intermediate attached to the second catalytic cysteine (90) (see below). Genetic analysis of Nop2 thermosensitive (ts) alleles allowed to distinguish between the function as a 26S rRNA maturation factor and a potential RNA:m<sup>5</sup>C-MTase. These data further confirmed the role of Nop2 in 26S rRNA maturation, even though the absence or reduced levels of Nop2 did not affect 18S rRNA maturation or 40S subunit biogenesis (91).

The strong sequence similarity of human proliferation associated nucleolar protein p120 with yeast Nop2 and the already characterized bacterial and eukaryotic m<sup>5</sup>C-MTases, together with its implication in rRNA biogenesis, strongly suggested that p120 possesses a similar activity (92,93). Experimental analysis of p120 delineated its N-terminal Arg-rich region as a high-affinity binding domain for rRNA. In addition, human p120 associates with 60S–80S pre-ribosomal particles in HeLa nuclear extracts (93). However, the significance of this physical association and the exact functions of human p120 protein in rRNA maturation and processing remain elusive.

The family of NSUN/NOP2/NOL1 related proteins in humans is represented by at least nine different members (NSUN1 to NSUN7, genes NSUN5A, B and C probably resulted from a recent gene duplication). Most genes of this group are highly conserved in mammals. At the protein level, the number of related protein variants is very important since multiple mRNA splicing isoforms were characterized and some variants may lack important catalytic or RNA-binding domains. One of the NSUN proteins (NSUN5A fragment 127–429) was crystallized and its structure was determined at high resolution (PDB accession 2B9E). In most cases, these NSUN proteins retained the putative m<sup>5</sup>C-MTase domain bearing two catalytic cysteines. However, with the exception of hTrm4/NSUN2, neither the activity nor the substrate specificity of these potential m<sup>5</sup>C-MTases have been investigated so far. NSUN5A, B and C are deleted in Williams–Beuren syndrome, which is a complex developmental disorder with multisystemic manifestations including supravalvular aortic stenosis and a specific



cognitive phenotype (94). Of note, NSUN5 also shows considerable sequence similarity to the third putative m<sup>5</sup>C-MTase in *S. cerevisiae* (encoded by ORF YNL022C), which has not been characterized in detail yet. Lastly, the human NSUN7 gene provides another illustration for the implication of putative m<sup>5</sup>C-MTases in various disorders, because a point mutation in this gene has been shown to cause reduced sperm motility and infertility in male mice (95).

*Higher eukaryotic Dnmt2 and homologs.* A relatively new addition to the eukaryotic RNA:m<sup>5</sup>C-MTase family are the Dnmt2-related proteins. The first Dnmt2-like protein (pmt1) was described in the fission yeast *S. pombe* (96) and recognized by its substantial sequence homologies to DNA:m<sup>5</sup>C-MTases. It was also noted that the pmt1 protein sequence contained a serine insertion in the catalytic proline–cysteine (PC) motif, which was interpreted to represent an inactivating mutation, consistent with the absence of DNA methylation in fission yeast (96). Several years later, human and mouse Dnmt2 were discovered in screens for sequences with homologies to known DNA:m<sup>5</sup>C-MTases (97,98). The corresponding protein sequences contained all the catalytic signature motifs of DNA:m<sup>5</sup>C-MTases, but no corresponding enzymatic activity could be described initially. Dnmt2 has been widely conserved during evolution and the use of more sensitive methods has allowed the identification of a residual DNA-MTase activity for Dnmt2. However, the biological relevance of this activity has remained contentious (99).

When the three-dimensional structure of human Dnmt2 was first analyzed, the results demonstrated a high similarity to the protein structure of the bacterial M.HhaI DNA:m<sup>5</sup>C-MTase (100), thus again suggesting that Dnmt2 might have DNA-MTase activity. However, an independent examination of the structural data also identified a tyrosine residue in the target recognition domain of Dnmt2 that was not present in other DNA:m<sup>5</sup>C-MTases (101). It was proposed that this tyrosine residue might impede the binding of DNMT2 to DNA substrates and that the enzyme might thus favor alternative substrates. This notion was confirmed when it was shown that Dnmt2 from various species methylates C38 of tRNA<sup>Asp</sup> (102). The tRNA-MTase activity has now been confirmed in several other laboratories and appears to be substantially higher than the residual DNA-MTase activity described initially. In light of the findings discussed above, it appears reasonable to assume that Dnmt2 is predominantly a tRNA-MTase and that the weak and highly distributive DNA-MTase activity is a consequence of a secondary enzyme activity with potentially little biological relevance. A more recent publication suggesting a role of Dnmt2-mediated DNA methylation in the epigenetic regulation of transposable elements in *Drosophila* (103) seems to contradict this notion. However, the results from this study indicate a robust and highly processive catalytic activity of Dnmt2 in the methylation of genomic retroelements. These findings are surprising because they suggest that a substantial DNA-MTase activity of Dnmt2 has been

overlooked previously. It will therefore be important to confirm these results in independent assays and in other model systems.

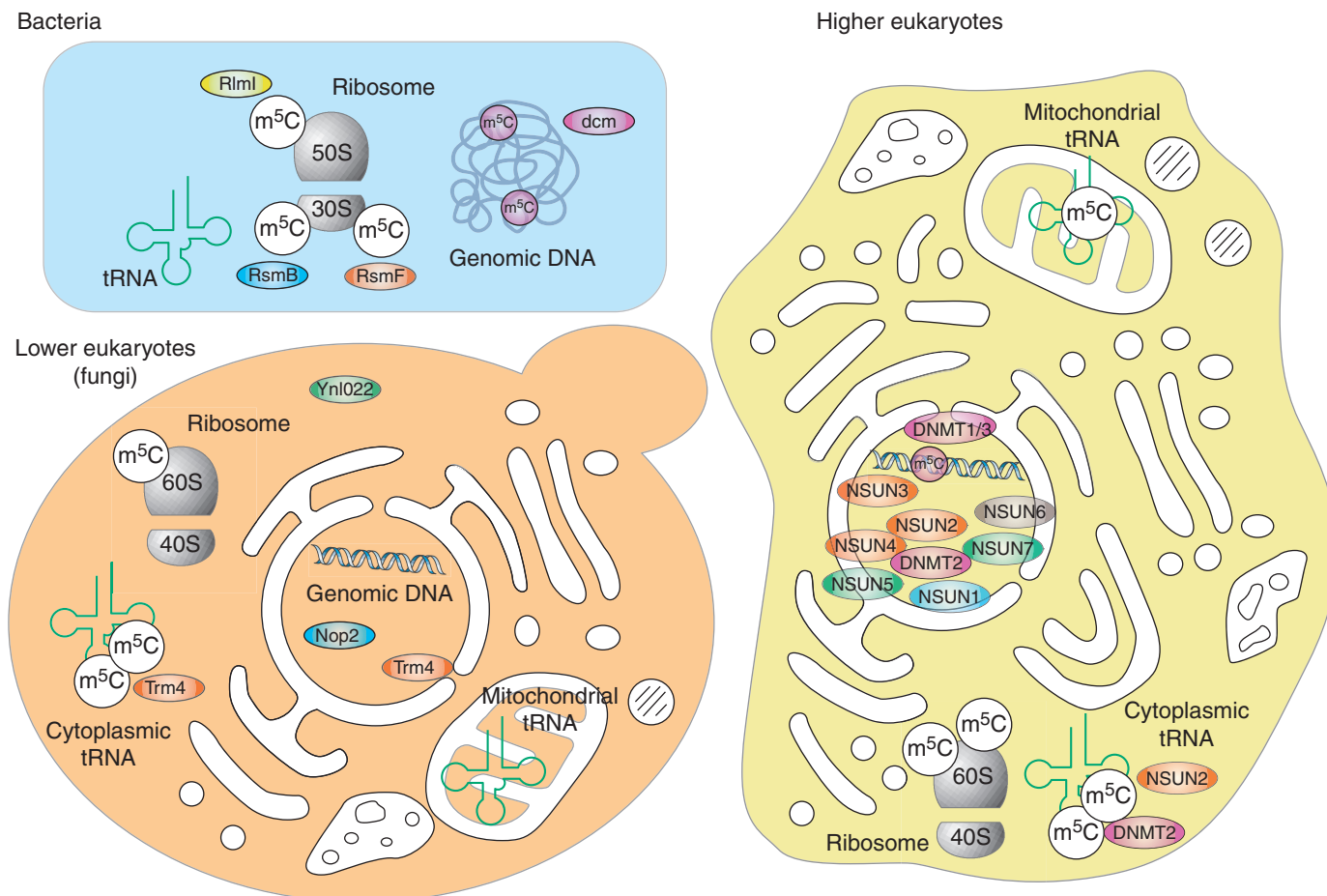
### Subcellular localization of eukaryotic m<sup>5</sup>C MTases

The subcellular localization of RNA modification enzymes can provide important indications about their biological functions. Figure 4 gives an overview of the current knowledge on the subcellular localization of methylated RNAs and modification enzymes. Most m<sup>5</sup>C-MTases are predicted to be nuclear or nucleolar proteins which corresponds well to their functions in tRNA and rRNA processing. Subcellular enzyme localization has been studied experimentally, by using various immunostaining and direct fluorescence approaches, but can also be inferred from the modification pattern of RNA in different cellular compartments. This latter method is particularly useful for mitochondrial localization, since these small and very dynamic organelles are difficult to observe by microscopy and because mitochondrial tRNAs and rRNAs are produced in the mitochondrial matrix and may be modified only by mitochondria-associated RNA modification enzymes. For example, yeast mitochondrial tRNAs do not contain m<sup>5</sup>C even though they contain many potential target cytosine residues for Trm4 activity, thus indicating the absence of Trm4 from this compartment.

Little is known about precise localization of various m<sup>5</sup>C-MTases in eukaryota. Yeast Ncl1 (Trm4) was localized at nuclear periphery (80), while Nop2 is a well-known nucleolar protein (88). Similarly, human p120 is also nuclear and nucleolar (104). The subcellular localization of Dnmt2 has been discussed controversially with various studies providing indications for either a nuclear or a cytoplasmic localization. A recent analysis of the subcellular localization of Dnmt2 in *Drosophila* suggested that the enzyme localizes both to the nucleus and to the cytoplasm (105). This may be reflective of the complex activities of Dnmt2.

### THE CATALYTIC MECHANISM OF m<sup>5</sup>C-MTases

The catalytic mechanism of m<sup>5</sup>C-MTases bears strong parallels to that of m<sup>5</sup>U-MTases, including in particular Trm2 which generates m<sup>5</sup>U (also known as rT) at position 54 of most tRNAs. As opposed to most other RNA-MTases, which methylate nitrogen atoms, the target position of m<sup>5</sup>-pyrimidine-MTases is a non-nucleophilic carbon atom. Activation of this position proceeds *via* Michael addition of a catalytic cysteine to the 6-position, which, after protonation, results in an enamin in the case of m<sup>5</sup>C-MTases (shown in Figure 5A), or an enol in the case of m<sup>5</sup>U-MTases (not shown). The 5-carbon in these intermediates, frequently referred to as ‘enolate’, is a nucleophile and as such easily methylated by the electrophilic SAM. Methyl-group transfer leads to another key intermediate, a covalent adduct of enzyme and nucleic acid, from which the enzyme is regenerated by beta-elimination of cysteine and the hydrogen on C5, which is thought



**Figure 4.** Subcellular localization of m<sup>5</sup>C residues in RNAs and m<sup>5</sup>C-MTases in bacteria (upper left), lower eukaryota (lower left) and higher eukaryota (right). Known m<sup>5</sup>C residues in RNAs are shown in white circles, together with the corresponding enzyme (if known). DNA-MTases and Dnmt2 are colored in purple, the RsmB/Nol1 family in blue, the YebU (RsmF) family in orange, the Ynl022 family in green and NSUN6 in grey.

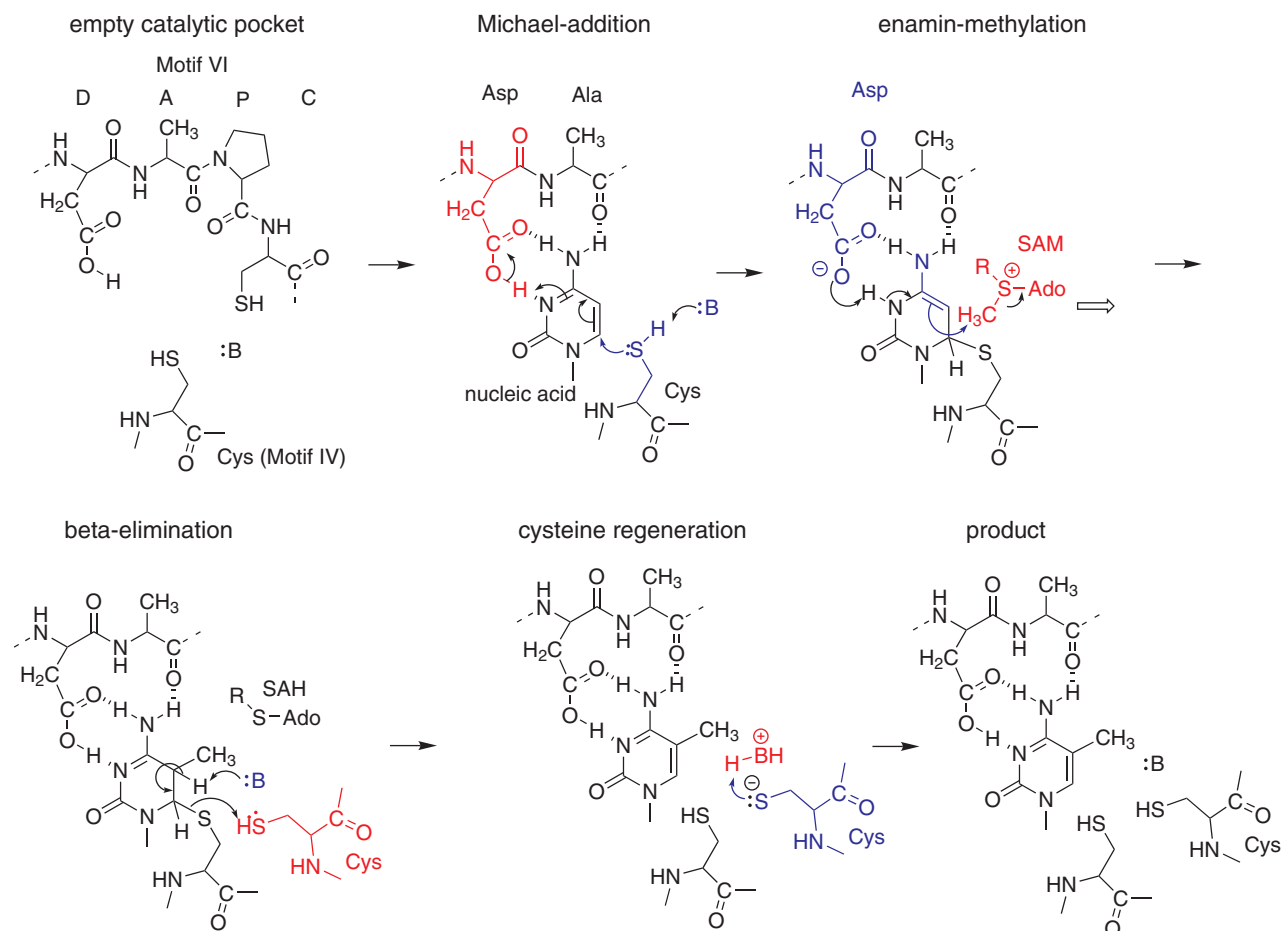
to be base-mediated. The similarities among pyrimidine-MTases extend to inhibition of the elimination step by 5-fluoro-pyrimidine targets. Nucleic acids containing 5-fluorouracil or 5-fluorocytidine frequently form covalent adducts with cognate MTases (83).

The elucidation of the catalytic mechanism(s) of m<sup>5</sup>C-MTases continues to be critically important for understanding the substrate specificity and biological function of these enzymes. As outlined below, structural biology with particular emphasis on the residues in active sites has been combined with insights from bioinformatics and biochemistry. The resulting picture discriminates typical DNA- versus RNA-MTases, and identified Dnmt2 as a hybrid RNA-MTase, which uses a catalytic cysteine typical of DNA-MTase (see below).

#### Mechanism with two cysteines

Comparative studies of RNA- and DNA-specific m<sup>5</sup>C-MTases resulted in the conclusion that DNA and RNA m<sup>5</sup>C-methylation are being mediated by similar yet distinct enzymatic mechanisms, which involve different amino acid residues (106). DNA and RNA-MTases exhibit sequence permutations, resulting in a variable linear order of the conserved motifs, which are

numbered I through X. Typical catalytic pockets of RNA and DNA MTases are shown in Figure 6. Typical RNA:m<sup>5</sup>C-MTases contain two Cys-containing characteristic motifs (motif IV and VI, referred to as PC- and TC-sequences, respectively). However, only the PC-sequence (motif IV) is found to be conserved in DNA:m<sup>5</sup>C-MTases, while motif VI contains a conserved ENV-sequence. Site-directed mutagenesis of the *E. coli* Fmu (RsmB) protein showed that only the replacement of Cys375 (in motif VI, TC-sequence), but not Cys325 (in motif IV, PC-sequence) abolishes the RNA m<sup>5</sup>C-MTase activity *in vitro* and the capacity to form a covalent adduct with 5-fluoro-cytidine containing RNA. These *in vitro* data came in apparent contradiction with the results of site-directed mutagenesis of Nop2 *in vivo* (89). Indeed, in this case, only a substitution of Cys424 (motif IV, PC-sequence) affected the viability of yeast cells. All other substitutions in Nop2 (including replacement of Cys478 in motif VI) did not affect Nop2 functions *in vivo*. This apparent contradiction was resolved by further studies (90) demonstrating that Nop2 and Trm4 (and probably other RNA:m<sup>5</sup>C-MTases, like RsmB) use both conserved cysteine residues for catalysis. The initial activation of the cytosine ring is performed by the cysteine



**Figure 5.** Catalytic mechanism of  $m^5C$ -MTases. Nucleophiles, bases and nucleophilic attacks are indicated in blue; acids and electrophiles are indicated in red.

residue of motif VI (TC-sequence, Cys478 in the case of Nop2), while the release of the RNA substrate from the covalent reaction intermediate requires the presence of the second Cys in motif IV (PC-sequence, Cys424 in Nop2). In the absence of the 'regeneration' Cys residue, the RNA substrate is not released from the intermediate covalent Nop2-RNA complex as indicated by the presence of covalent adducts of higher molecular mass accumulating in yeast cells and in *E. coli*. In the crystal structure of *E. coli* RsmB, both Cys325 and Cys375 are located in close proximity to each other and Cys325 aligns well with Cys81, which is located in motif IV of the *HpaI* DNA: $m^5C$ -MTase. The spatial locations of these conserved residues are thus fully compatible with the proposed two-cysteine catalytic mechanism of RNA: $m^5C$ -MTases (70). More recently, the catalytic mechanism involving both conserved Cys residues was also confirmed for yeast Trm4, which forms a stable covalent intermediate with 5-fluoro-cytidine containing RNA (83).

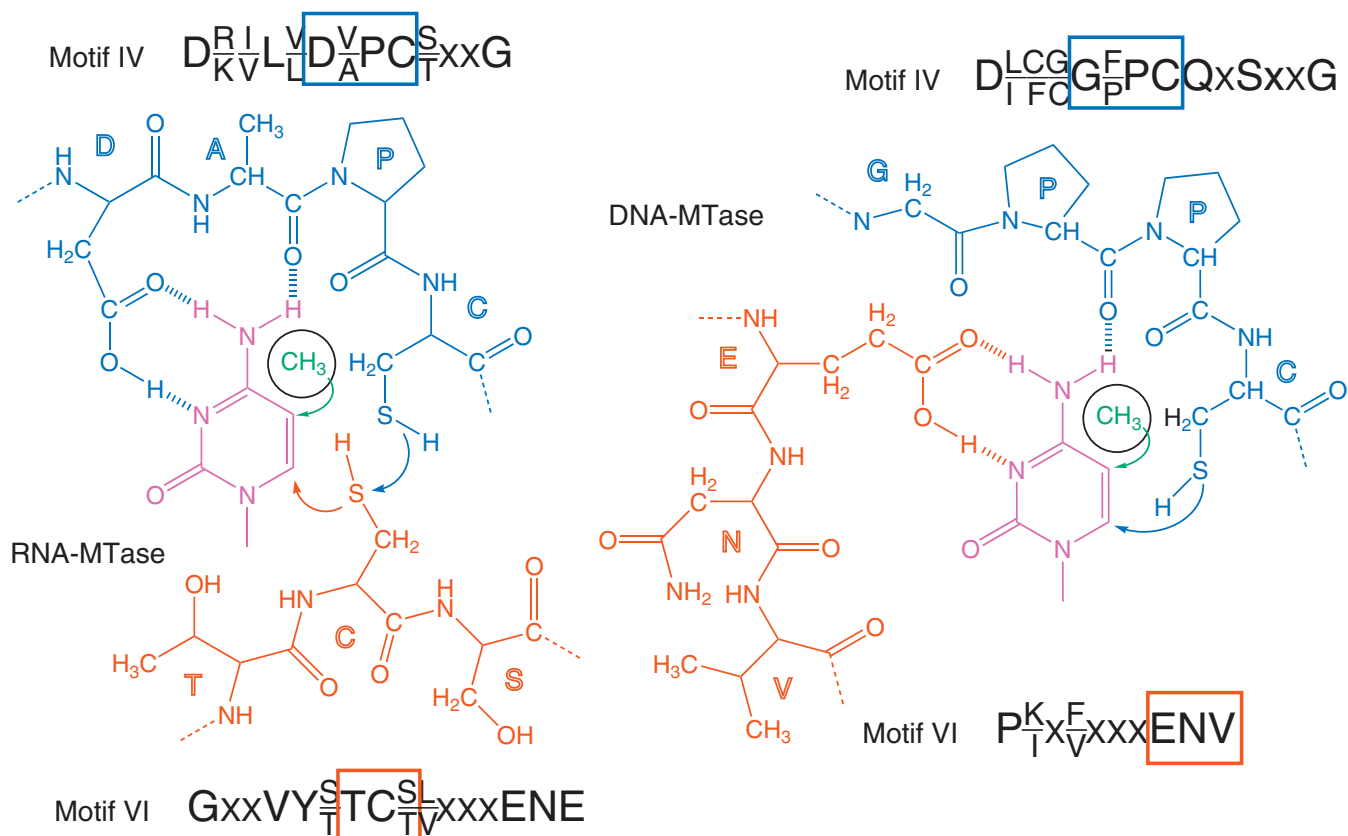
A recent and intriguing addition to this field is the catalytic mechanism of the Dnmt2  $m^5C$ -MTase. Dnmt2 is an enzyme that has long been regarded as pure DNA-MTase, based on sequence similarity with other DNA-MTases, including a linear arrangement of

motifs I through X, which is characteristic for Dnmt enzymes. The discovery of RNA-MTase activity in Dnmt2 was thus a surprise (102). In subsequent studies of human Dnmt2, mutational analysis identified essential residues for analysis which include E119 (ENV-motif VI), as well as R160 and R162 (motif VIII). The essential catalytic cysteine, which conducts the initial nucleophilic attack for tRNA methylation at position C38 was identified as C79, which is situated in the PC-motif IV. It was thus concluded, that C38 methylation of tRNA by Dnmt2 utilizes a catalytic mechanism typical for DNA-MTases (107). These findings uncover a previously unexpected link between RNA and DNA methylation which may be important for understanding the biological function(s) of Dnmt2 (see below).

#### Covalent complex formation with the RNA substrate

Covalent complexes formed between RNA substrate and mutant forms of RNA: $m^5C$ -MTases have been observed in several studies, but their detailed biochemical characterization was not reported until recently. Site-directed mutagenesis of yeast Trm4 confirmed the importance of motif VI Cys310 (TC-sequence) for catalysis and motif IV Cys260 (PC-sequence) for substrate release. Trm4 C260A





**Figure 6.** Catalytic pocket of RNA- (left) and DNA-MTases (right). The target cytidine residue is indicated in purple, the sequences of conserved motifs IV and VI are given for both classes of enzymes. Motif IV is in blue and motif VI is in orange. Boxes highlight residues belonging to the respective sequence motifs.

or C260S mutants efficiently form covalent complexes both *in vitro* and *in vivo*, when expressed in *E. coli*. SDS-PAGE analysis showed that SAM is required for covalent complex formation and that the RNA substrate is present in a covalent complex that contains the CH<sub>3</sub>-group, an indication that the release of the final reaction product is indeed impeded by the mutation in the PC-sequence (108).

## THE BIOLOGICAL SIGNIFICANCE OF m<sup>5</sup>C IN RNA

### General stabilization of tRNA molecules and Mg<sup>2+</sup> binding

A known general role of modified nucleotides in tRNA is their structural and metabolic stabilization (109). Structural stabilization of RNAs by m<sup>5</sup>C was demonstrated by various physico-chemical techniques, like circular dichroism and NMR. One of the m<sup>5</sup>C residues present in the *S. cerevisiae* tRNA<sup>Phe</sup> at position 40 was reported to be important for appropriate Mg<sup>2+</sup> binding and appears to induce conformational transitions of the whole anticodon loop (110–112).

Modified residues in tRNAs are also known to influence the stability of these molecules *in vivo*, and there are at least two pathways that are known to degrade hypomodified tRNA (113–115). An exosome-independent pathway was found to degrade m<sup>5</sup>C-deficient tRNA, but

only in double mutants lacking both TRM4 and the methyltransferase activity responsible for the formation of 7-methylguanosine. The accelerated degradation of hypomodified tRNA thus appears not to be specific to the lack of single m<sup>5</sup>C residues. Rather, it is hypothesized that the combined lack of several modifications lowers the conformational stability of the tRNA below a certain threshold, which leads to its degradation by a mechanism that is yet to be elucidated in detail (114,115).

### Importance of m<sup>5</sup>C for mRNA translation

Very little is known about specific role of m<sup>5</sup>C residues in tRNA and rRNA during mRNA decoding on the ribosome. The presence of m<sup>5</sup>C at position 34 in eukaryotic tRNA<sup>Leu</sup> seems to be important for its function in translation, specifically for suppressor function *in vivo* (116).

Deletion of the non-essential yeast *TRM4* gene does not affect cell growth or steady-state levels of large and small ribosome subunits, monoribosomes and polyribosomes. However, disruption of *TRM4* leads to increased sensitivity to the antibiotic paromomycin (80). Paromomycin is known to affect the precision of the ribosome decoding, thus the presence of m<sup>5</sup>C in yeast tRNA may be important for maintenance of low error translation.

On the other hand, m<sup>5</sup>C in ribosomal RNA may also participate in tRNA recognition and peptidyl transfer. Despite their distant locations in the rRNA sequence, all three m<sup>5</sup>C residues present in bacterial 16S and 23S rRNAs are localized in close proximity to each other in the three-dimensional structure of the ribosome. In addition, the paromomycin binding site has been mapped precisely to this region (117). It is not clear if yeast rRNA contains a m<sup>5</sup>C residue at a similar position, but the sensitivity of the ΔYBL024 mutant to paromomycin may indicate the relations between ribosomal decoding and m<sup>5</sup>C functions in rRNA (80).

### Other biological functions of m<sup>5</sup>C residues

Because the phenotypes of RNA:m<sup>5</sup>C-MTase mutants are comparably weak, it has generally been difficult to pinpoint the biological function of RNA cytosine methylation. However, there is some direct evidence that the methylation of RNA molecules can have a modulatory effect on the innate human immune system. Unmodified RNA molecules have been shown to strongly stimulate the immune system through activation of toll-like receptors (118). This effect was greatly reduced when either of several methylated nucleotides, including m<sup>5</sup>C, was incorporated into the stimulatory RNAs, suggesting that the methylation prevents recognition of endogenous RNAs by the innate immune system.

Another, currently more speculative role for RNA methylation is the regulation of epigenetic inheritance patterns. While it is clear that most phenotypic traits are inherited through DNA, there is also evidence for RNA-dependent inheritance of certain phenotypes (119). The methylation of cytosine residues in DNA has long been known to play an important role in the modulation of genetic information (120). It will be fascinating to explore the possibility whether RNA methylation plays a similar role.

### CONCLUSION AND PERSPECTIVES

Recent developments in the field of RNA methylation open a large perspective for further analysis of this important biological process. Enzymatic activity, catalytic mechanisms and RNA-recognition specificity of the corresponding RNA:m<sup>5</sup>C-MTases clearly require better characterization. The biological role of m<sup>5</sup>C residues in RNA has been largely ignored and further studies using genetically engineered models are needed to investigate this point. High-throughput sequencing techniques will be extremely helpful in the methylation profiling of low abundance species [e.g. mRNA (53)], or even newly discovered RNA species, including various regulatory RNAs. While some of the latter are known to be methylated at other positions such as the 2'-OH, (6,121), verification of the presence or absence of m<sup>5</sup>C in these species should further deepen our understanding of the biological meaning of m<sup>5</sup>C in RNA.

### ACKNOWLEDGEMENTS

M.H. thanks Andres Jäschke for constant support.

### FUNDING

M.H. and F.L. acknowledge funding by the DFG Dnmt2 (FOR 1082). Funding for open access charge: DFG HE 3397/6-1.

*Conflict of interest statement.* None declared.

### REFERENCES

- Hotchkiss,R.D. (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.*, **175**, 315–332.
- Luciano,D.J., Mirsky,H., Vendetti,N.J. and Maas,S. (2004) RNA editing of a miRNA precursor. *RNA*, **10**, 1174–1177.
- Blow,M.J., Grocock,R.J., van Dongen,S., Enright,A.J., Dicks,E., Futreal,P.A., Wooster,R. and Stratton,M.R. (2006) RNA editing of human microRNAs. *Genome Biol.*, **7**, R27.
- Yang,Z., Ebright,Y.W., Yu,B. and Chen,X. (2006) HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.*, **34**, 667–675.
- Yu,B., Yang,Z., Li,J., Minakhina,S., Yang,M., Padgett,R.W., Steward,R. and Chen,X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, **307**, 932–935.
- Horwich,M.D., Li,C., Matranga,C., Vagin,V., Farley,G., Wang,P. and Zamore,P.D. (2007) The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.*, **17**, 1265–1272.
- Saito,K., Sakaguchi,Y., Suzuki,T., Siomi,H. and Siomi,M.C. (2007) Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.*, **21**, 1603–1608.
- Grosjean,H., Droogmans,L., Roovers,M. and Keith,G. (2007) Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods Enzymol.*, **425**, 55–101.
- Hengesbach,M., Meusbürger,M., Lyko,F. and Helm,M. (2008) Use of DNazymes for site-specific analysis of ribonucleotide modifications. *RNA*, **14**, 180–187.
- Zhao,X. and Yu,Y.T. (2004) Detection and quantitation of RNA base modifications. *RNA*, **10**, 996–1002.
- Stach,D., Schmitz,O.J., Stilgenbauer,S., Benner,A., Dohner,H., Wiessler,M. and Lyko,F. (2003) Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Res.*, **31**, E2.
- Cornelius,M.G. and Schmeiser,H.H. (2007) RNA analysis by MEKC with LIF detection. *Electrophoresis*, **28**, 3901–3907.
- Li,H.Y., Wang,S.M., Liu,H.M., Li,J., Han,D., Bu,S.S. and Zhang,M.Z. (2008) Analysis of modified nucleosides in the urine of patients with malignant cancer by liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.*, **22**, 3161–3171.
- Wagner,T.M., Nair,V., Guymon,R., Pomerantz,S.C., Crain,P.F., Davis,D.R. and McCloskey,J.A. (2004) A novel method for sequence placement of modified nucleotides in mixtures of transfer RNA. *Nucleic Acids Symp. Ser.*, **48**, 263–264.
- Dai,Q., Fong,R., Saikia,M., Stephenson,D., Yu,Y.T., Pan,T. and Piccirilli,J.A. (2007) Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N6-methyladenosine. *Nucleic Acids Res.*, **35**, 6322–6329.
- Yu,Y.T., Shu,M.D. and Steitz,J.A. (1997) A new method for detecting sites of 2'-O-methylation in RNA molecules. *RNA*, **3**, 324–331.
- Motorin,Y., Müller,S., Behm-Ansmant,I. and Branlant,C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.*, **425**, 21–53.

18. Lambrinakos, A., Humphrey, K.E., Babon, J.J., Ellis, T.P. and Cotton, R.G. (1999) Reactivity of potassium permanganate and tetraethylammonium chloride with mismatched bases and a simple mutation detection protocol. *Nucleic Acids Res.*, **27**, 1866–1874.
19. Bui, C.T., Rees, K. and Cotton, R.G. (2003) Permanganate oxidation reactions of DNA: perspective in biological studies. *Nucleosides Nucleotides Nucleic Acids*, **22**, 1835–1855.
20. Cotton, R.G., Rodrigues, N.R. and Campbell, R.D. (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl Acad. Sci. USA*, **85**, 4397–4401.
21. Rubin, C.M. and Schmid, C.W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res.*, **8**, 4613–4619.
22. Zuo, S., Boorstein, R.J. and Teebor, G.W. (1995) Oxidative damage to 5-methylcytosine in DNA. *Nucleic Acids Res.*, **23**, 3239–3243.
23. Tabone, T., Sallmann, G., Chiotis, M., Law, M. and Cotton, R. (2006) Chemical cleavage of mismatch (CCM) to locate base mismatches in heteroduplex DNA. *Nat. Protoc.*, **1**, 2297–2304.
24. Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl Acad. Sci. USA*, **74**, 560–564.
25. Peattie, D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl Acad. Sci. USA*, **76**, 1760–1764.
26. Fritzsche, E., Hayatsu, H., Igloi, G.L., Iida, S. and Kossel, H. (1987) The use of permanganate as a sequencing reagent for identification of 5-methylcytosine residues in DNA. *Nucleic Acids Res.*, **15**, 5517–5528.
27. Okamoto, A. (2009) Chemical approach toward efficient DNA methylation analysis. *Org. Biomol. Chem.*, **7**, 21–26.
28. Rideout, W.M. 3rd, Coetzee, G.A., Olumi, A.F. and Jones, P.A. (1990) 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science*, **249**, 1288–1290.
29. Suzuki, T., Nakamura, T., Yamada, M., Ide, H., Kanaori, K., Tajima, K., Morii, T. and Makino, K. (1999) Isolation and characterization of diazoate intermediate upon nitrous acid and nitric oxide treatment of 2'-deoxycytidine. *Biochemistry*, **38**, 7151–7158.
30. Ali, M.M., Alam, M.R., Kawasaki, T., Nakayama, S., Nagatsugi, F. and Sasaki, S. (2004) Sequence- and base-specific delivery of nitric oxide to cytidine and 5-methylcytidine leading to efficient deamination. *J. Am. Chem. Soc.*, **126**, 8864–8865.
31. Singer, B. (1976) All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature*, **264**, 333–339.
32. Singer, B. (1975) The chemical effects of nucleic acid alkylation and their relation to mutagenesis and carcinogenesis. *Prog. Nucleic Acid Res. Mol. Biol.*, **15**, 219–284.
33. Peattie, D.A. and Gilbert, W. (1980) Chemical probes for higher-order structure in RNA. *Proc. Natl Acad. Sci. USA*, **77**, 4679–4682.
34. Peattie, D.A. and Herr, W. (1981) Chemical probing of the tRNA-ribosome complex. *Proc. Natl Acad. Sci. USA*, **78**, 2273–2277.
35. Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831.
36. Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.*, **22**, 2990–2997.
37. Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A. et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.*, **38**, 1378–1385.
38. Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S. and Jaenisch, R. (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.*, **33**, 5868–5877.
39. Gu, W., Hurto, R.L., Hopper, A.K., Grayhack, E.J. and Phizicky, E.M. (2005) Depletion of *Saccharomyces cerevisiae* tRNA<sup>His</sup> guanylyltransferase Thg1p leads to uncharged tRNA<sup>His</sup> with additional m<sup>7</sup>C. *Mol. Cell Biol.*, **25**, 8191–8201.
40. Schaefer, M., Pollex, T., Hanna, K. and Lyko, F. (2009) RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res.*, **37**, e12.
41. Warnecke, P.M., Storzaker, C., Melki, J.R., Millar, D.S., Paul, C.L. and Clark, S.J. (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.*, **25**, 4422–4426.
42. Sprinzl, M. and Vassilenko, K.S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **33**, D139–D140.
43. Smith, J.E., Cooperman, B.S. and Mitchell, P. (1992) Methylation sites in *Escherichia coli* ribosomal RNA: localization and identification of four new sites of methylation in 23S rRNA. *Biochemistry*, **31**, 10825–10834.
44. Kowalak, J.A., Pomerantz, S.C., Crain, P.F. and McCloskey, J.A. (1993) A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res.*, **21**, 4577–4585.
45. Kowalak, J.A., Bruenger, E. and McCloskey, J.A. (1995) Posttranscriptional modification of the central loop of domain V in *Escherichia coli* 23S ribosomal RNA. *J. Biol. Chem.*, **270**, 17758–17764.
46. Suzuki, T., Ikeuchi, Y., Noma, A. and Sakaguchi, Y. (2007) Mass spectrometric identification and characterization of RNA-modifying enzymes. *Methods Enzymol.*, **425**, 211–229.
47. Noon, K.R., Bruenger, E. and McCloskey, J.A. (1998) Posttranscriptional modifications in 16S and 23S rRNAs of the archaeal hyperthermophile *Sulfolobus solfataricus*. *J. Bacteriol.*, **180**, 2883–2888.
48. Veldman, G.M., Klootwijk, J., de Regt, V.C., Planta, R.J., Branlant, C., Krol, A. and Ebel, J.P. (1981) The primary and secondary structure of yeast 26S rRNA. *Nucleic Acids Res.*, **9**, 6935–6952.
49. Maden, B.E. (1988) Locations of methyl groups in 28S rRNA of *Xenopus laevis* and man. Clustering in the conserved core of molecule. *J. Mol. Biol.*, **201**, 289–314.
50. Maden, B.E. (1986) Identification of the locations of the methyl groups in 18S ribosomal RNA from *Xenopus laevis* and man. *J. Mol. Biol.*, **189**, 681–699.
51. Kowalak, J.A., Bruenger, E., Crain, P.F. and McCloskey, J.A. (2000) Identities and phylogenetic comparisons of posttranscriptional modifications in 16S ribosomal RNA from *Haloferax volcanii*. *J. Biol. Chem.*, **275**, 24484–24489.
52. Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. (2008) The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.*, **36**, D178–D183.
53. Dubin, D.T. and Taylor, R.H. (1975) The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic Acids Res.*, **2**, 1653–1668.
54. Dubin, D.T. and Stollar, V. (1975) Methylation of Sindbis virus “26S” messenger RNA. *Biochem. Biophys. Res. Commun.*, **66**, 1373–1379.
55. Dubin, D.T., Stollar, V., Hsueh, C.C., Timko, K. and Guild, G.M. (1977) Sindbis virus messenger RNA: the 5'-termini and methylated residues of 26 and 42S RNA. *Virology*, **77**, 457–470.
56. Sommer, S., Salditt-Georgieff, M., Bachenheimer, S., Darnell, J.E., Furuichi, Y., Morgan, M. and Shatkin, A.J. (1976) The methylation of adenovirus-specific nuclear and cytoplasmic RNA. *Nucleic Acids Res.*, **3**, 749–765.
57. Salditt-Georgieff, M., Jelinek, W., Darnell, J.E., Furuichi, Y., Morgan, M. and Shatkin, A. (1976) Methyl labeling of HeLa cell hnRNA: a comparison with mRNA. *Cell*, **7**, 227–237.
58. Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-Georgieff, M. and Darnell, J.E. (1975) Methylated, blocked 5' termini in HeLa cell mRNA. *Proc. Natl Acad. Sci. USA*, **72**, 1904–1908.
59. Desrosiers, R., Friderici, K. and Rottman, F. (1974) Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl Acad. Sci. USA*, **71**, 3971–3975.
60. Adams, J.M. and Cory, S. (1975) Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature*, **255**, 28–33.
61. Lavi, S. and Shatkin, A.J. (1975) Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells. *Proc. Natl Acad. Sci. USA*, **72**, 2012–2016.



62. Sakamoto, K. and Okada, N. (1985) 5-Methylcytidylic modification of *in vitro* transcript from the rat identifier sequence; evidence that the transcript forms a tRNA-like structure. *Nucleic Acids Res.*, **13**, 7195–7206.
63. Rozhdetsvensky, T.S., Crain, P.F. and Brosius, J. (2007) Isolation and posttranscriptional modification analysis of native BC1 RNA from mouse brain. *RNA Biol.*, **4**, 11–15.
64. Brule, H., Grosjean, H., Giege, R. and Florentz, C. (1998) A pseudoknotted tRNA variant is a substrate for tRNA (cytosine-5)-methyltransferase from *Xenopus laevis*. *Biochimie*, **80**, 977–985.
65. Bujnicki, J.M., Feder, M., Ayres, C.L. and Redman, K.L. (2004) Sequence-structure-function studies of tRNA:m<sup>5</sup>C methyltransferase Trm4p and its relationship to DNA:m<sup>5</sup>C and RNA:m<sup>5</sup>U methyltransferases. *Nucleic Acids Res.*, **32**, 2453–2463.
66. Reid, R., Greene, P.J. and Santi, D.V. (1999) Exposition of a family of RNA m<sup>5</sup>C methyltransferases from searching genomic and proteomic sequences. *Nucleic Acids Res.*, **27**, 3138–3145.
67. Pavlopoulou, A. and Kossida, S. (2009) Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. *Genomics*, **93**, 350–357.
68. Tscherne, J.S., Nurse, K., Popienick, P., Michel, H., Sochacki, M. and Ofengand, J. (1999) Purification, cloning, and characterization of the 16S RNA m<sup>5</sup>C967 methyltransferase from *Escherichia coli*. *Biochemistry*, **38**, 1884–1892.
69. Gu, X.R., Gustafsson, C., Ku, J., Yu, M. and Santi, D.V. (1999) Identification of the 16S rRNA m<sup>5</sup>C967 methyltransferase from *Escherichia coli*. *Biochemistry*, **38**, 4053–4057.
70. Foster, P.G., Nunes, C.R., Greene, P., Moustakas, D. and Stroud, R.M. (2003) The first structure of an RNA m<sup>5</sup>C methyltransferase, Fmu, provides insight into catalytic mechanism and specific binding of RNA substrate. *Structure*, **11**, 1609–1620.
71. Andersen, N.M. and Douthwaite, S. (2006) YebU is a m<sup>5</sup>C methyltransferase specific for 16S rRNA nucleotide 1407. *J. Mol. Biol.*, **359**, 777–786.
72. Hallberg, B.M., Ericsson, U.B., Johnson, K.A., Andersen, N.M., Douthwaite, S., Nordlund, P., Beuscher, A.E.t. and Erlandsen, H. (2006) The structure of the RNA m<sup>5</sup>C methyltransferase YebU from *Escherichia coli* reveals a C-terminal RNA-recruiting PUA domain. *J. Mol. Biol.*, **360**, 774–787.
73. Purta, E., O'Connor, M., Bujnicki, J.M. and Douthwaite, S. (2008) YccW is the m<sup>5</sup>C methyltransferase specific for 23S rRNA nucleotide 1962. *J. Mol. Biol.*, **383**, 641–651.
74. Sumita, S., Tkaczuk, K.L., Purta, E., Kasprzak, J.M., Douthwaite, S., Bujnicki, J.M. and Sivaraman, J. (2008) Crystal structure of the *Escherichia coli* 23S rRNA:m<sup>5</sup>C methyltransferase RlmI (YccW) reveals evolutionary links between RNA modification enzymes. *J. Mol. Biol.*, **383**, 652–666.
75. Grosjean, H., Gaspin, C., Marck, C., Decatur, W.A. and de Crecy-Lagard, V. (2008) RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics*, **9**, 470.
76. Kowalak, J.A., Dalluge, J.J., McCloskey, J.A. and Stetter, K.O. (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry*, **33**, 7869–7876.
77. Constantinesco, F., Motorin, Y. and Grosjean, H. (1999) Transfer RNA modification enzymes from *Pyrococcus furiosus*: detection of the enzymatic activities *in vitro*. *Nucleic Acids Res.*, **27**, 1308–1315.
78. Ishikawa, I., Sakai, N., Tamura, T., Yao, M., Watanabe, N. and Tanaka, I. (2004) Crystal structure of human p120 homologue protein PH1374 from *Pyrococcus horikoshii*. *Proteins*, **54**, 814–816.
79. Auxilien, S., El Khadali, F., Rasmussen, A., Douthwaite, S. and Grosjean, H. (2007) Archease from *Pyrococcus abyssi* improves substrate specificity and solubility of a tRNA m<sup>5</sup>C methyltransferase. *J. Biol. Chem.*, **282**, 18711–18721.
80. Wu, P., Brockenbrough, J.S., Paddy, M.R. and Aris, J.P. (1998) NCL1, a novel gene for a non-essential nuclear protein in *Saccharomyces cerevisiae*. *Gene*, **220**, 109–117.
81. Motorin, Y. and Grosjean, H. (1999) Multisite-specific tRNA:m<sup>5</sup>C-methyltransferase (Trm4) in yeast *Saccharomyces cerevisiae*: identification of the gene and substrate specificity of the enzyme. *RNA*, **5**, 1105–1118.
82. Walbott, H., Auxilien, S., Grosjean, H. and Golinelli-Pimpaneau, B. (2007) The carboxyl-terminal extension of yeast tRNA m<sup>5</sup>C methyltransferase enhances the catalytic efficiency of the amino-terminal domain. *J. Biol. Chem.*, **282**, 23663–23671.
83. Walbott, H., Husson, C., Auxilien, S. and Golinelli-Pimpaneau, B. (2007) Cysteine of sequence motif VI is essential for nucleophilic catalysis by yeast tRNA m<sup>5</sup>C methyltransferase. *RNA*, **13**, 967–973.
84. Frye, M. and Watt, F.M. (2006) The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. *Curr. Biol.*, **16**, 971–981.
85. Sakita-Suto, S., Kanda, A., Suzuki, F., Sato, S., Takata, T. and Tatsuka, M. (2007) Aurora-B regulates RNA methyltransferase NSUN2. *Mol. Biol. Cell.*, **18**, 1107–1117.
86. Brzezicha, B., Schmidt, M., Makalowska, I., Jarmolowski, A., Pienkowska, J. and Szweykowska-Kulinska, Z. (2006) Identification of human tRNA:m<sup>5</sup>C methyltransferase catalysing intron-dependent m<sup>5</sup>C formation in the first position of the anticodon of the pre-tRNA<sup>Leu</sup>(CAA). *Nucleic Acids Res.*, **34**, 6034–6043.
87. de Beus, E., Brockenbrough, J.S., Hong, B. and Aris, J.P. (1994) Yeast NOP2 encodes an essential nucleolar protein with homology to a human proliferation marker. *J. Cell. Biol.*, **127**, 1799–1813.
88. Hong, B., Brockenbrough, J.S., Wu, P. and Aris, J.P. (1997) Nop2p is required for pre-rRNA processing and 60S ribosome subunit synthesis in yeast. *Mol. Cell. Biol.*, **17**, 378–388.
89. King, M., Ton, D. and Redman, K.L. (1999) A conserved motif in the yeast nucleolar protein Nop2p contains an essential cysteine residue. *Biochem. J.*, **337**, 29–35.
90. King, M.Y. and Redman, K.L. (2002) RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. *Biochemistry*, **41**, 11218–11225.
91. Hong, B., Wu, K., Brockenbrough, J.S., Wu, P. and Aris, J.P. (2001) Temperature sensitive *nop2* alleles defective in synthesis of 25S rRNA and large ribosomal subunits in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **29**, 2927–2937.
92. Koonin, E.V. (1994) Prediction of an rRNA methyltransferase domain in human tumor-specific nucleolar protein P120. *Nucleic Acids Res.*, **22**, 2476–2478.
93. Gustafson, W.C., Taylor, C.W., Valdez, B.C., Henning, D., Phippard, A., Ren, Y., Busch, H. and Durban, E. (1998) Nucleolar protein p120 contains an arginine-rich domain that binds to ribosomal RNA. *Biochem. J.*, **331**, 387–393.
94. Doll, A. and Grzeschik, K.H. (2001) Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams-Beuren syndrome. *Cytogenet. Cell Genet.*, **95**, 20–27.
95. Harris, T., Marquez, B., Suarez, S. and Schimenti, J. (2007) Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the Sun domain-containing family of putative RNA methyltransferases. *Biol. Reprod.*, **77**, 376–382.
96. Wilkinson, C.R., Bartlett, R., Nurse, P. and Bird, A.P. (1995) The fission yeast gene *pmt1+* encodes a DNA methyltransferase homologue. *Nucleic Acids Res.*, **23**, 203–210.
97. 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.
98. Okano, M., Xie, S. and Li, E. (1998) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.*, **26**, 2536–2540.
99. Jeltsch, A., Nellen, W. and Lyko, F. (2006) Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem. Sci.*, **31**, 306–308.
100. Dong, A., Yoder, J.A., Zhang, X., Zhou, L., Bestor, T.H. and Cheng, X. (2001) Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res.*, **29**, 439–448.
101. Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
102. Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H. (2006) Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science*, **311**, 395–398.

103. Phalke,S., Nickel,O., Walluscheck,D., Hortig,F., Onorati,M.C. and Reuter,G. (2009) Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.*, **41**, 696–702.
104. Ochs,R.L., Reilly,M.T., Freeman,J.W. and Busch,H. (1988) Intracellular localization of human proliferating cell nuclear antigen p120. *Cancer Res.*, **48**, 6523–6529.
105. Schaefer,M., Steringer,J.P. and Lyko,F. (2008) The *Drosophila* cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS ONE*, **3**, e1414.
106. Liu,Y. and Santi,D.V. (2000) m<sup>5</sup>C RNA and m<sup>5</sup>C DNA methyl transferases use different cysteine residues as catalysts. *Proc. Natl Acad. Sci. USA*, **97**, 8263–8265.
107. Jurkowski,T.P., Meusbürger,M., Phalke,S., Helm,M., Nellen,W., Reuter,G. and Jeltsch,A. (2008) Human DNMT2 methylates tRNA<sup>Asp</sup> molecules using a DNA methyltransferase-like catalytic mechanism. *RNA*, **14**, 1663–1670.
108. Redman,K.L. (2006) Assembly of protein-RNA complexes using natural RNA and mutant forms of an RNA cytosine methyltransferase. *Biomacromolecules*, **7**, 3321–3326.
109. Helm,M. (2006) Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.*, **34**, 721–733.
110. Chen,Y., Sierzputowska-Gracz,H., Guenther,R., Everett,K. and Agris,P.F. (1993) 5-Methylcytidine is required for cooperative binding of Mg<sup>2+</sup> and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA. *Biochemistry*, **32**, 10249–10253.
111. Basti,M.M., Stuart,J.W., Lam,A.T., Guenther,R. and Agris,P.F. (1996) Design, biological activity and NMR-solution structure of a DNA analogue of yeast tRNA<sup>Phe</sup> anticodon domain. *Nat. Struct. Biol.*, **3**, 38–44.
112. Stuart,J.W., Koshlap,K.M., Guenther,R. and Agris,P.F. (2003) Naturally-occurring modification restricts the anticodon domain conformational space of tRNA<sup>Phe</sup>. *J. Mol. Biol.*, **334**, 901–918.
113. Kadaba,S., Krueger,A., Trice,T., Krecic,A.M., Hinnebusch,A.G. and Anderson,J. (2004) Nuclear surveillance and degradation of hypomodified initiator tRNA<sup>Met</sup> in *S. cerevisiae*. *Genes Dev.*, **18**, 1227–1240.
114. Alexandrov,A., Chernyakov,I., Gu,W., Hiley,S.L., Hughes,T.R., Grayhack,E.J. and Phizicky,E.M. (2006) Rapid tRNA decay can result from lack of nonessential modifications. *Mol. Cell*, **21**, 87–96.
115. Chernyakov,I., Whipple,J.M., Kotelawala,L., Grayhack,E.J. and Phizicky,E.M. (2008) Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes Dev.*, **22**, 1369–1380.
116. Strobel,M.C. and Abelson,J. (1986) Effect of intron mutations on processing and function of *Saccharomyces cerevisiae* SUP53 tRNA in vitro and in vivo. *Mol. Cell. Biol.*, **6**, 2663–2673.
117. Vicens,Q. and Westhof,E. (2001) Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure*, **9**, 647–658.
118. Kariko,K., Buckstein,M., Ni,H. and Weissman,D. (2005) Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, **23**, 165–175.
119. Rassoulzadegan,M., Grandjean,V., Gounon,P., Vincent,S., Gillot,I. and Cuzin,F. (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*, **441**, 469–474.
120. Suzuki,M.M. and Bird,A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
121. Huang,Y., Ji,L., Huang,Q., Vassilyev,D.G., Chen,X. and Ma,J.B. (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature*, **461**, 823–827.
122. Rai,K., Chidester,S., Zavala,C.V., Manos,E.J., James,S.R., Karpf,A.R., Jones,D.A. and Cairns,B.R. (2007) Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev.*, **21**, 261–266.