Sequence–structure–function studies of tRNA:m⁵C methyltransferase Trm4p and its relationship to DNA:m⁵C and RNA:m⁵U methyltransferases

Janusz M. Bujnicki, Marcin Feder, Chastity L. Ayres¹ and Kent L. Redman¹,*

Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland and ¹Fort Wayne Center for Medical Education, Indiana University School of Medicine, 2101 Coliseum Boulevard East, Fort Wayne, IN 46805, USA

Received January 22, 2004; Revised March 16, 2004; Accepted April 2, 2004

ABSTRACT

Three types of methyltransferases (MTases) generate 5-methylpyrimidine in nucleic acids, forming m⁵U in RNA, m⁵C in RNA and m⁵C in DNA. The DNA:m⁵C MTases have been extensively studied by crystallographic, biophysical, biochemical and computational methods. On the other hand, the sequence–structure–function relationships of RNA:m⁵C MTases remain obscure, as do the potential evolutionary relationships between the three types of 5-methylpyrimidine-generating enzymes. Sequence analyses and homology modeling of the yeast tRNA:m⁵C MTase Trm4p (also called Ncl1p) provided a structural and evolutionary platform for identification of catalytic residues and modeling of the architecture of the RNA:m⁵C MTase active site. The analysis led to the identification of two invariant residues that are important for Trm4p activity in addition to the conserved Cys residues in motif IV and motif VI that were previously found to be critical. The newly identified residues include a Lys residue in motif I and an Asp in motif IV. A conserved Gln found in motif X was found to be dispensable for MTase activity. Locations of essential residues in the model of Trm4p are in very good agreement with the X-ray structure of an RNA:m⁵C MTase homolog PH1374. Theoretical and experimental analyses revealed that RNA:m⁵C MTases share a number of features with either RNA:m⁵U MTases or DNA:m⁵C MTases, which suggested a tentative phylogenetic model of relationships between these three classes of 5-methylpyrimidine MTases. We infer that RNA:m⁵C MTases evolved from RNA:m⁵U MTases by acquiring an additional Cys residue in motif IV, which was adapted to function as the nucleophilic catalyst only later in DNA:m⁵C MTases, accompanied by loss of the original Cys from motif VI, transfer of a conserved carboxylate from motif IV to motif VI and sequence permutation.

INTRODUCTION

Methylation of nucleic acids is catalyzed by a large and diverse class of S-adenosyl-L-methionine (AdoMet)-dependent methyltransferases. The enzymes characterized to date include members of two unrelated superfamilies: ‘classical’ Rossmann-fold-like (1,2) and SPOUT (3). The relatively small SPOUT superfamily includes only a few characterized RNA-specific enzymes with 2′-O-ribose or guanosine-N¹ modification specificity that will not be discussed further in this article. The Rossmann-fold superfamily (hereafter referred to as ‘MTases’) groups together enzymes acting on RNA, DNA, proteins, lipids and various small molecules. MTases have a catalytic domain with a common structural core and AdoMet-binding site. Shared motifs are usually detectable at the sequence level, but in some cases motifs that have diverged beyond recognition by sequence comparison can be identified by structural comparisons. The most common nomenclature involves motifs I–X, initially assigned to MTases that generate 5-methylcytosine (m⁵C) in DNA and which correspond to the key structural and functional elements associated with the cofactor-binding site (I–III), the catalytic pocket (X, IV, VI and VIII) and motifs implicated in preservation of the common fold (V and VII) (2,4,5). However, DNA and RNA MTases exhibit sequence permutation, resulting in a variable linear order of the conserved motifs (6,7).

DNA and RNA differ with respect to the number of observed modifications. Only three modified bases are typically found in DNA: m⁵C, N⁴-methylcytosine (m⁴C) and N⁶-methyladenine (m⁶A). Methylated nitrogens (m⁶A and m⁴C) occur primarily in Prokaryota, but m⁵C is found in organisms from all three Domains. Crystal structures have been determined for DNA MTases that generate each of these common modifications and a plethora of DNA-specific MTases have been cloned and characterized biochemically (reviewed in 8). In sharp contrast to the well-studied DNA MTase families, RNA MTases remain poorly characterized from the perspective of structure–function relationships. To date, crystal structures have been solved for several known or putative...
RNA MTases (reviewed in 9,10), but typically without a substrate, which limits the ability to correlate reaction mechanisms with active site architectures.

The RNA:m^5C MTases are a fascinating group of RNA modification enzymes, for which some useful information has been obtained by separate structural, biochemical and evolutionary studies. Following cloning of the first representative, 16S rRNA:m^5C MTase RsmB (previously called Sun or Fmu) from Escherichia coli by two groups (11,12), homologous sequences were identified and additional paralogous RNA:m^5C MTase subfamilies were predicted (13). Among these putative m^5C MTases, two eukaryotic proteins were characterized on the sequence–function level: a multisite-specific tRNA:m^5C MTase Trm4p (14,15) and apparent rRNA MTase Nop2p (15–17). Despite the apparent size and wide distribution of this protein family, RsmB (Fmu) and Trm4p remain the only proteins with biochemically confirmed RNA m^5C methyltransferase activity.

A particularly interesting aspect of RNA:m^5C MTases is their relationship to two distinct classes of enzymes that generate 5-methylpyrimidine in nucleic acids, namely RNA:m^5U MTases and DNA:m^5C MTases. The enzymatic mechanism of DNA:m^5C methylation has been extensively studied by crystallography, mutagenesis, biophysical methods and molecular dynamics simulations (18–25). Briefly, it involves an attack by the thiol of an invariant Cys residue from motif IV on the 6 position of the cytosine base to form a covalent complex, thereby activating the 5 position for methyl group transfer, which is followed by deprotonation and β-elimination to restore the free enzyme and release the methylated product (reviewed in 8,26). For RNA:m^5U MTases an analogous mechanism has been proposed, albeit involving an unrelated Cys from motif VI (27,28).

Remarkably, RNA:m^5C MTases possess counterparts of both DNA:m^5C-like and RNA:m^5U-like cysteine residues (13,15). Mutational analysis has suggested that in RNA:m^5C methylation the RNA:m^5U-like thiol acts in a classical fashion by forming a covalent link to carbon 6 of the pyrimidine base (29), while the DNA:m^5C-like thiol assists breakdown of the covalent adduct (15). However, no other residues in the active site of RNA:m^5C MTases have been studied. Therefore, the details of the catalytic mechanism remain obscure, as do the potential evolutionary relationships between the three types of 5-methylpyrimidine MTases. These enzymes could be the result of divergent evolution from a common ancestor, the product of progressive changes where one of the three mechanisms is an intermediate between the other two or the result of convergence where the MTase fold was independently adapted to perform three chemically similar reactions.

During the course of the work two crystal structures were solved for members of the RNA:m^5C MTase family. PH1374 is a putative RNA:m^5C MTase of unknown specificity from Pyrococcus horikoshii. A partial PH1374 structure is available from the Protein Data Bank (PDB) under accession number 1ixk, but as of January 2004, the work has not been published. The second structure was reported for RsmB from E.coli (30). It has been deposited in the PDB under accession numbers 1SQG and 1SQF, but not released prior to publication of this work. Neither of these structures was solved in the presence of a RNA substrate. Moreover, the limited functional analyses reported for RsmB did not include mutagenesis of the presumed active site, hence our knowledge of the residues required for catalysis of the RNA:m^5C methylation reaction remains incomplete.

To learn more about the mechanism of RNA:m^5C methylation and as an aid to resolution of the relationships between 5-methylpyrimidine MTases we have carried out extensive sequence analysis. The results of the sequence analysis suggest that these findings can be extrapolated to other members of the RNA:m^5C MTase family. Finally, the potential evolutionary relationships between different types of 5-methylpyrimidine MTases are discussed in the light of the available structural and biochemical data.

### MATERIALS AND METHODS

Recombinant RNasin was obtained from Promega and the Talon affinity resin was purchased from Clontech. BL21-Codon Plus(DE3)-RIL cells and the QuikChange mutagenesis kits are products of Stratagene. Calf liver tRNA was a product of Boehringer Mannheim. Oligonucleotides were synthesized by Integrated DNA Technologies (www.idtdna.com). DNA sequence analysis was carried out by the Biochemistry Biotechnology Facility at the Indiana University School of Medicine, Indianapolis IN.

**Cloning and site-directed mutagenesis of the NCLI/TRM4 gene**

Amplification of the NCLI/TRM4 gene from yeast genomic DNA, cloning of the amplified product to generate pEMNCL1 and transfer of the cloned gene into a modified version of the pET28b expression plasmid were previously described (15). Site-directed mutations were generated by the use of Stratagene’s QuikChange kit that utilizes two complementary oligonucleotides for each mutation. Table 1 shows the sequence of the positive strand oligonucleotide used for each mutation, the restriction site generated by the change and

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Coding strand oligonucleotide (new restriction site underlined)</th>
<th>Restriction site formed</th>
<th>Sites used to move mutant fragment to expression plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q150A</td>
<td>CCGTTGGTAATATCTCGAGAGCCGAAGCTTTTCAATGATTCC</td>
<td>XhoI</td>
<td>SacI–BglII</td>
</tr>
<tr>
<td>D257A</td>
<td>GACGAAATCCTGGCCGATTCATGGATGATGCCATGTTTCTCTGTTGATGATG</td>
<td>FspI</td>
<td>BglII–HindIII</td>
</tr>
<tr>
<td>K179M</td>
<td>GTGTCGTGCTTCTGGATGCCATGACTGCTCAAATTATCGAAGC</td>
<td>BamHI</td>
<td>SacI–BglII</td>
</tr>
</tbody>
</table>

Downloaded from academic.oup.com/nar/article-abstract/32/8/2453/2904532 by guest on 18 February 2019
the pair of restriction sites used to move the altered region from the modified pGEMNCL1 plasmid to the pET28-based expression plasmid. The BglII–HindIII region of the D257A expression construct was sequenced in both directions using the primers previously reported for that purpose (15), but the SacI–BglII regions of the Q150A and K179M expression constructs were sequenced using a T7 promoter primer that initiates within the expression plasmid in the forward direction and the NCL1SEQR2 primer (GAAGAATTGGGCG-TCATGGTT) in the reverse direction.

Expression and purification of Trm4p

BL21-Codon Plus(DE3)-RIL cells carrying TRM4 plasmids were grown in Luria Broth to an OD 600 of 0.4, then Trm4p expression was induced with 1 mM isopropyl β-d-1-thiogalactopyranoside for 2 h. The cell pellet from each 50 ml culture was suspended in 3 ml of extraction buffer (see below) containing 0.75 mg/ml lysozyme for 30 min at 24°C. Samples were then ice chilled, shaken with 200–300 µm glass beads for 30 s using a Mini Bead Beater at 3800 r.p.m. and centrifuged at 16 000 g for 10 min at 8°C. Trm4p was purified from the supernatant with 1 ml of Talon affinity resin using a combined batch–column procedure. Affinity resin equilibrated with extraction buffer (50 mM sodium phosphate pH 8, 300 mM NaCl and 5 mM benzamidine) was mixed for 20 min with the extract from one or two 50 ml cultures diluted to a concentration of 0.5 g/ml. The mixture was then incubated at 8°C for 5 min. The resin was washed twice with 10 bed vol of wash buffer (identical to extraction buffer except that the pH was 7.0). Washed resin was transferred to a minicolumn and eluted with pH 7 buffer containing 50 mM sodium phosphate, 300 mM NaCl and 150 mM imidazole. Protein concentrations were determined with the Coomassie Protein Assay reagent from Pierce using cytochrome C as the standard. Enzyme preparations were stored in Teflon capped glass vials that were flushed with N2 and the assays were done standard. Enzyme preparations were stored in Te®on capped glass vials that were¯ushed with N2 and the assays were done standard. Enzyme preparations were stored in Te¯on capped concentrations were determined with the Coomassie Protein and eluted with pH 7 buffer containing 50 mM sodium

...that was used as a query in PSI-BLAST (33) searches of the NRDB...
Our analysis revealed more members of the RNA: m^5C MTase family from higher eukaryotes than reported previously (13). Previous analyses identified only three eukaryotic lineages, corresponding to the yeast proteins Trm4p, Nop2p and Ynl022c. The alignment in Figure 1 shows the representative members of all major lineages of the RNA: m^5C MTase family identified in this work, with special emphasis on seven human members. In conjunction with the phylogenetic tree (Fig. 2), we predict that new human RNA: m^5C MTase candidates FLJ22609 and MGC22960 are most closely related to the Trm4p lineage (including the human Trm4p ortholog FLJ20303). We confirm that the human protein Nol1p is orthologous to the yeast protein Nop2p. However, our results suggest that WBSCR20, a predicted Nol1p ortholog that is one of several genes encoded within a chromosomal region implicated in the pathogenesis of Williams–Beuren syndrome (WBS) (50), is orthologous to Ynl022c rather than to Nol1p. We have also identified a lineage paralogous to Ynl022c/WBSCR20, which includes a human protein FLJ14001 and a few other proteins exclusively from metazoa. The topology of the tree reveals that the Ynl022c/WBSCR20/FLJ14001 lineage clusters together with the uncharacterized archaeal lineage previously denoted ‘subfamily IV’ (13) and with the bacterial RsmB lineage. That these lineages are orthologous is supported by the fact that the bacterial and archaeal proteins share the N-terminally fused NusB domain. Finally, our analysis revealed that the uncharacterized human protein FLJ23743 (NopD1p) is an ortholog of the archaeal lineage earlier denoted as ‘subfamily VI’ (13), since both these lineages share the unique feature of a PUA domain (51) inserted between motifs N1 and X.

The results of our phylogenetic analysis provide a convenient platform for experimental characterization of the putative human m^5C MTases. For instance, human proteins FLJ22609 and MGC22960 with no orthologs in yeast could share a general specificity for tRNA with their close paralog Trm4p and be responsible for generation of higher Eukaryote-specific m^5C modifications at positions 50 and 72 in tRNA. Likewise, that Ynl022c and WBSCR20 are predicted to be orthologous to RsmB suggests that they share its specificity for rRNA methylation. Interestingly, an RNA: m^5C MTase homolog FLJ14001 seems to lack a few important residues of the active site identified in this work, which suggests that it may be inactive. Finally, the clustering of Nol1p and FLJ23743 with their sister archaeal lineages suggests that these subfamilies of
enzymes may share specificity for an RNA:m\textsubscript{5}C modification present in Eukaryota and Archaea but absent from Bacteria.

**Structure prediction of Trm4p**

In order to provide a structural platform for sequence–function studies of Trm4p, we carried out computational structure modeling (see the Materials and Methods). The protein fold recognition analysis revealed that the Trm4p sequence is compatible with the Rossmann-like MTase fold (9). All fold recognition algorithms reported various MTase structures with very high scores, in particular the only member of the RNA:m\textsubscript{5}C MTase family available in the PDB (putative archaeal MTase PH1374) (data not shown). For over 50 alternative sequence–structure alignments we have generated preliminary models and used their best scoring fragments to generate a hybrid model (46) (see also Materials and Methods for details). The N- and C-terminal extensions with no counterparts in the template structures were predicted to be at least partially disordered and therefore were not included in the final model. Only the structural core comprising residues 22–411 was homology modeled. The final model passed the quality test implemented in VERIFY3D (48). The model of the Trm4p–AdoMet–cytidine complex was constructed by transferring the ligand from the superimposed coordinates of the DNA:m\textsubscript{5}C MTase M.HhaI crystal structure (see Supplementary Material for images of the Trm4p model superimposed on the crystal structures of M.HhaI and PH1374).

**Structure-based mutagenesis**

Examination of the RNA m\textsubscript{5}C MTase model and the multiple sequence alignment revealed two highly conserved amino acids within the active site (Figs 1 and 3). These correspond to Trm4p residues Gln150 and Asp257, which could aid catalysis in addition to the pair of cysteines already implicated in catalysis (15). Gln150 is positioned behind the cytosine base relative to the conserved cysteines (Cys260 and Cys310) and is adjacent to the AdoMet cofactor (Fig. 3). Thus, Gln150 could help form substrate-binding sites or it could have a more direct catalytic role. We found that mutation of Gln150 to Ala had little or no effect on the initial enzymatic activity of the expressed enzyme (Fig. 4A), but the activity of the mutant enzyme declines more rapidly than that of wild-type Trm4p upon storage (data not shown). Therefore, the invariant...
Gln150 is not essential for catalysis, but its conversion to Ala appears to slightly alter Trm4p. Increased sensitivity to oxidation might result from a reduced affinity for tRNA (see below), which could make the active site cysteines more accessible to oxidizing agents.

In the model, Asp257 is located adjacent to the cytosine base and near the AdoMet cofactor. Conversion of Asp257 to Ala inactivates Trm4p, with no observed methyl transfer activity above the background level of control extract prepared from cells not expressing Trm4p (Fig. 4B). The levels of $^3$H incorporation into RNA indicate that the D257A mutant has less than 1% of the activity of the wild-type protein.

The presence of a potentially charged aspartic acid residue within the active site drew our attention to Lys179 of motif I, because the side chain of this conserved residue extends into the active site with the terminal amino group residing near Asp257. An isosteric (shape and size conservative) mutation...
DISCUSSION

The role of conserved residues in the active site of RNA:m^5C MTases

Gln150 is located in motif X and superimposes well on the chemically similar invariant Asn residue of DNA:m^5C MTases (Asn304 in M.Hhal; data not shown). The crystallographic analysis has suggested that Asn304 of M.Hhal is involved in binding of the target cysteine, flipped into the enzyme active site (52). The role of this residue, however, has not been studied by mutagenesis. Our analysis has shown that Gln150 of RNA:m^5C MTase Trm4p is not essential for enzyme activity. Interestingly, this otherwise conserved Gln is substituted by hydrophobic residues in the two newly identified human paralogs of Trm4 (Fig. 4). It will be interesting to determine if the homologous Asn residue is required for catalysis in DNA:m^5C MTases.

Lys179 is located in motif I, which is traditionally regarded as a part of the AdoMet-binding site rather than a part of the active site. However, it maps not to the loop shown to bind AdoMet in many crystal structures (reviewed in 2), but to the solvent-exposed face of the subsequent α-helix. According to the crystal structure of PH1374 and the homology modeled structure of Trm4p, the side chain of Lys179 points towards the active site rather than the cofactor-binding site. It will be of interest to study the role of this residue using high resolution biophysical and/or structural methods, as it may be the first catalytically active residue of MTases located in such an unusual position. Alternatively, Lys179 may be an atypical component of the cofactor-binding site or be involved in substrate recognition and binding. While the severe loss of activity associated with the Lys179 mutation seems more characteristic of the loss of a catalytic residue than a ligand-binding residue, its position within the active site does not make these two potential roles mutually exclusive.

Asp257 is located in motif IV, which in RNA:m^5C MTases includes the invariant, essential Cys260 residue and typically assumes the ‘DAPC’ pattern. Residues 1 and 4 of motif IV often contribute to the catalytic activity and reaction specificity of various nucleic acid MTase families owing to key interactions of the respective functional groups with the target base (reviewed in 2). Specifically, Asp257 of Trm4p is homologous to the Asp residue conserved in many ‘amino-MTases’, i.e. RNA and DNA MTases that methylate exocyclic amino groups in adenine (DPPY in members of the α-class of DNA:m^5C MTases and DPPW in mRNA:m^6A MTases), guanine (DPPE in RNA:m^2G MTases RsmC/RsmD) and cytosine (DPHH in DNA:m^5C MTase M.NgoMXV) (reviewed in 53). It is noteworthy that the consensus sequence signature of amino-MTases is rather loosely defined as (S/D/N)-(I/P)-P-(Y/F/W/H), but it is neither specific nor unique to these enzymes, as there are known cases of amino-MTases with a different motif IV (54) and MTases acting on proteins rather than DNA, which exhibit the ‘NPPY’ tetrapeptide in motif IV (9). A more general (D/N/S)-X-P-X pattern (where X is any amino acid) is shared by the majority of nucleic acid MTases as well as some MTases acting on other substrates (2) and may be regarded as an ancestral form of one of the largest subgroups of MTases of the class I fold (51; J.M.Bujnicki, E.V.Koonin and Aravind, manuscript in preparation). An Asp

of Lys179 to Met was made to eliminate the amino group. Like the Asp257 mutant, the K179M mutant is completely inactive (Fig. 4C). The strong effect of the D257A and K179M mutations combined with their possible juxtaposition in the active site indicates that an interaction between these residues may be critical for catalysis or proper conformation of the active site region.

In previous studies, ethidium bromide staining of SDS-PAGE gels containing Trm4 revealed that small RNA molecules co-purify with the wild-type protein, although they are not covalently attached as observed with a Cys260 mutant (15). To determine if the Trm4p mutants developed here retain the ability to bind RNA, affinity-purified enzyme preparations were phenol extracted and the isolated material was resolved in denaturing gels (Fig. 5). Bacterial RNAs the size of tRNAs co-purify with all three mutant forms of Trm4p reported above, although it is not known if the mutants retain identical tRNA affinity. In fact, a reduced amount of RNA seems to be associated with the Q150A mutant despite its normal level of activity. This finding suggests essentially normal folding for all of the mutants, although minor structural perturbations that could influence docking of the target base into the active site that could significantly affect enzyme activity cannot be excluded.

The locations of the three residues mutated in this study are in very good agreement with the X-ray structure of a RNA:m^5C MTase homolog PH1374. They are all relatively near the AdoMet-binding site in our model and in the Fmu (RsmB) structure (30). Lys260 of Fmu corresponds to Trm4p Lys179 and the former has been proposed to be involved in cofactor binding, but to date we have not been able to obtain evidence as to the ability or inability of the D257A or K179M Trm4p mutants to bind the AdoMet cofactor (via electrostatic interactions and/or hydrogen bonds). Therefore, it cannot be excluded that the enzymatic defects in the Asp257 and Lys179 mutants are due to compromised cofactor binding, but the severity of the activity loss would also be consistent with a faulty catalytic mechanism.
Table 2. Similarities and differences between nucleic acid NH2-MTases and three pyrimidine-C5 MTase families

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NH2-MTases</td>
<td>RNA/DNA</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>variable</td>
</tr>
<tr>
<td>RNA:m5U</td>
<td>RNA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>N-terminal</td>
</tr>
<tr>
<td>RNA:m5C</td>
<td>RNA</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N-terminal</td>
</tr>
<tr>
<td>DNA:m5C</td>
<td>DNA</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>C-terminal</td>
</tr>
</tbody>
</table>

Relationships between 5-methylpyrimidine MTases

RNA:m5U MTases and RNA:m5C MTases share three characteristic features: (i) they act on RNA; (ii) they utilize the invariant Cys residue in motif VI as the catalytic nucleophile; (iii) they possess a conserved Asp residue in motif IV (like many other MTases acting on RNA) (51). RNA:m5C MTases and RNA:m5C MTases also share three characteristic features: (i) they methylate cytosine; (ii) they possess an invariant Cys residue in motif IV; (iii) they possess a conserved amide side chain (Gln or Asn, respectively) in motif X, which could make contact with the target base. RNA:m5U MTases and RNA:m5C MTases share none of the above-mentioned features; they act on different nucleic acids, methylate different bases, their invariant cysteine residues are located in different motifs and they do not share any other particularly similar residues known or predicted to interact with the target base (Table 2).

Combination of the biochemical data with bioinformatic analyses (including homology modeling and comparison of models with experimentally solved structures) and phylogenetic studies (sequence analysis and identification of characters shared between distinct enzyme families) allows us to infer a possible scenario of evolution for the large class of pyrimidine-C5 MTases. The `mosaic' similarity of RNA:m5C MTases to RNA:m5U MTases and DNA:m5C MTases (sharing features with both of the latter families) strongly suggests that RNA:m5C MTases may be the evolutionary intermediate. If this were the case, the question arises as to the direction of evolution, i.e. whether RNA:m5U MTases or DNA:m5C MTases are the best candidate for the `ancestral' form and which of them may be regarded as the latest evolutionary development. A few observations support the scenario in which RNA:m5U MTases may be ancestral: (i) motif IV of RNA:m5U MTases conforms to the ancestral ‘DXPX’ pattern shared by the majority of nucleic acid MTases, while motif IV of DNA:m5C MTases strongly deviates from this pattern and may be considered as derived; (ii) in searches of sequence databases, RNA:m5U MTases show significant similarity to many other nucleic acid MTase families, including amino- MTases, while DNA:m5C MTases appear rather ‘isolated’, which is most likely due to strong divergence and a relatively large phylogenetic distance from the common ancestor; (iii) RNA:m5U MTases exhibit the most typical (ancestral) order of sequence motifs shared by all MTases acting on substrates other than nucleic acids (from the N-terminus, X, followed by I–VIII), while most of DNA:m5C MTases exhibit a unique circular permutation, in which motif X is transferred to the C-terminus (6), which again may be considered as derived rather than ancestral. Moreover, it is believed that RNA is more...
ancient than DNA, thus RNA:m^5U MTases (as well as RNA:m^5C MTases) could have evolved in the hypothetical ‘ribonucleoprotein world’, with DNA:m^5C MTases having emerged once DNA took over as the principal carrier of cellular genetic information. The tentative scenario of relationships between the three pyrimidine C-5 MTases is shown in Figure 6. Alternative scenarios, including origin of the RNA MTases from DNA MTases or convergent evolution (independent origin of the major pyrimidine C-5 MTase families) cannot be completely excluded, but we find them rather unlikely.

We propose that RNA:m^5U MTases are the most ancient of pyrimidine C-5 MTases, as they exhibit most of the primitive/ancestral features typical of most of the RNA Mtases, as well as many MTases acting on substrates other than nucleic acids (the common order of sequence motifs and the Asp residue in motif IV being most notable). They were probably the first to evolve the ability to catalyze the chemically complex reaction of pyrimidine-C5 methylation, involving an attack of the thiol of a Cys residue from motif VI on the 6 position of uracil to form a covalent complex, thereby activating the 5 position for methyl group transfer from AdoMet, to be followed by deprotonation and β-elimination to restore the free enzyme and release the methylated product (reviewed in 8,26). RNA:m^5C MTases might have evolved very early from RNA:m^5U MTases by developing a second Cys residue in motif IV to aid in breakdown of the covalent adduct. Alternatively, these two RNA MTase families might have evolved from a common ancestor that methylated both types of pyrimidine bases. The new ‘auxiliary’ Cys residue in motif IV becomes the principal catalyst in DNA:m^5C MTases, which prompted us to suggest a tentative evolutionary model of the enzyme–substrate complex and a mechanistic role for the motif IV Asp. We have found that RNA:m^5C MTases have evolved from a common ancestral type of DNA:m^5C MTases (especially from their early branching subfamilies) become available and especially as more experimental data on the function of representative members of all major subfamilies are obtained. We hope that our analysis will guide further comparative investigations of residues conserved between the RNA:m^5C MTases and either the RNA:m^5U or DNA:m^5C MTases, as such studies have the potential to provide insight into the origin of m^5C methylation, the most important type of DNA modification in Eukaryota, including humans.

CONCLUSIONS

To learn more about the mechanism of RNA:m^5C methylation and as an aid to the resolution of the relationships between 5-methylpyrimidine MTases we have carried out extensive bioinformatics analysis followed by mutagenesis of three potential catalytic residues in the yeast tRNA:m^5C MTase Trm4p. In addition to the previously reported Cys residues in motifs IV and VI (15), we have found two additional residues that are essential for enzyme activity: Lys in motif I and Asp in motif IV. Based on these analyses, we propose a structural model of the enzyme–substrate complex and a mechanistic role for the motif IV Asp. We have found that RNA:m^5C MTases share a number of features with RNA:m^5U MTases or DNA:m^5C MTases, which prompted us to suggest a tentative phylogenetic model of origin of the three classes of 5-methylpyrimidine MTases, in which DNA:m^5C MTases arose from RNA:m^5C MTases following ‘switching’ of the role of the nucleophilic catalyst from the now lost Cys in motif VI to Cys in motif IV and a number of other evolutionary innovations, such as sequence permutation.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

J.M.B. thanks Henri Grosjean for critical reading of the manuscript and for useful comments and stimulating discussions. Technical assistance for the production of Trm4p mutants was provided by Michelle King. Funding for portions
of this work was provided by a Biomedical Research Grant from the Indiana University School of Medicine. J.M.B. is supported by the EMBO/HHMI Young Investigator Programme award, by the Polish State Committee for Scientific Research (grant PBZ-KBN-088/P04/2003) and by a fellowship from the Foundation for Polish Science.

REFERENCES


