J Mol Evol (2002) 55:431–444
DOI: 10.1007/s00239-002-2339-8

JOURNAL OF MOLECULAR EVOLUTION

Structure Prediction and Phylogenetic Analysis of a Functionally Diverse Family of Proteins Homologous to the MT-A70 Subunit of the Human mRNA:m^{6}A Methyltransferase

Janusz M. Bujnicki,[1] Marcin Feder,[1,2] Monika Radlinska,[2] Robert M. Blumenthal[3]

[1] Bioinformatics Laboratory, International Institute of Molecular and Cell Biology, ul. ks. Trojdena 4, 02-109 Warsaw, Poland
[2] Institute of Microbiology, Warsaw University, ul. Miecznikowa 1, 02-093 Warsaw, Poland
[3] Department of Microbiology & Immunology and Program in Bioinformatics & Proteomics/Genomics, Medical College of Ohio, 3055 Arlington Avenue, Toledo, OH 43614-5806, USA

Received: 18 August 2001 / Accepted: 2 April 2002

Abstract. MT-A70 is the *S*-adenosylmethionine-binding subunit of human mRNA:m^{6}A methyltransferase (MTase), an enzyme that sequence-specifically methylates adenines in pre-mRNAs. The physiological importance yet limited understanding of MT-A70 and its apparent lack of similarity to other known RNA MTases combined to make this protein an attractive target for bioinformatic analysis. The sequence of MT-A70 was subjected to extensive *in silico* analysis to identify orthologous and paralogous polypeptides. This analysis revealed that the MT-A70 family comprises four subfamilies with varying degrees of interrelatedness. One subfamily is a small group of bacterial DNA:m^{6}A MTases. The other three subfamilies are paralogous eukaryotic lineages, two of which have not been associated with MTase activity but include proteins having substantial regulatory effects. Multiple sequence alignments and structure prediction for members of all four subfamilies indicated a high probability that a consensus MTase fold domain is present. Significantly, this consensus fold shows the permuted topology characteristic of the β class of MTases, which to date has only been known to include DNA MTases.

Key words: N6-Methyladenine — RNA modification — DNA methylation — *S*-Adenosylmethionine — Methyltransferases — Circular permutation — Fold recognition — Protein structure prediction — Homology modeling

Correspondence to: Janusz M. Bujnicki; *email*: iamb@genesilico.pl

Introduction

In eukaryotic cells, mRNA precursors (pre-mRNAs) undergo several modifications. Two of these modifications, respectively, yielding the mature 5′ cap structure and internal methylation, require enzymes from the *S*-adenosyl-L-methionine (AdoMet)-dependent methyltransferase (MTase) superfamily. Methylation of internal adenosines [as opposed to those in the 3′ poly(A) tail] generates N6-methyladenosine (m^{6}A), which may affect the efficiency of pre-mRNA splicing or export from the nucleus (Narayan and Rottman 1992). m^{6}A is found in mRNA from all multicellular eukaryotes, including plants, mammals, and *Drosophila*, and is also found in viruses that replicate in the nuclei of these organisms; however, to date m^{6}A has not been found in mRNA from *Saccharomyces cerevisiae* or from several other fungi. Even in mammalian cells some mRNAs do not have detectable m^{6}A, suggesting that this modification is not essential to mRNA function in general (Narayan and Rottman 1992). The internal mRNA adenine methylation is sequence specific, but the specificity is not well understood—m^{6}A is found only in the

degenerate consensus sequence RR-m^{6}A-CH (R = purine, H = C, A, or U), but only a fraction of these sequences is methylated (Schibler et al. 1977).

In mammalian (HeLa) cells, the mRNA:m^{6}A MTase activity requires at least two separate proteins, MT-A and MT-B (Bokar et al. 1994). The 200-kDa MT-A is a multimeric protein that contains a 70-kDa AdoMet-binding subunit referred to as MT-A70. MT-A70 has been purified and partially sequenced; its cDNA has been also isolated and characterized. The amino acid sequence of MT-A70 was noted to be similar to those of the DNA MTase M.*Mun*I and the yeast sporulation protein IME4 (formerly called SPO8) but not to other known RNA MTases (Bokar et al. 1997). A combination of the physiological importance of MT-A70, the limited understanding of its biochemistry, and, especially, its lack of obvious similarity to other known RNA MTases made this protein a worthwhile subject of bioinformatic analysis.

Comparative analysis of MTase sequences faces three complications. First, though the AdoMet-dependent MTases virtually all share a highly conserved structural core involved in AdoMet binding and methyltransfer, this structural conservation is not reflected in sequence conservation (Fauman et al. 1999; Bujnicki 1999). Second, many MTases contain insertions into the core structure that mediate interaction with the various methylation targets (Schluckebier et al. 1998). Third, the DNA MTases exhibit circular permutation of the core catalytic domain (Malone et al. 1995; Jeltsch 1999; Bujnicki 2002). Why circular permutation was found only among DNA MTases has been a riddle.

We predict the structure and the mechanism of action of the mRNA:m^{6}A MTase subunit MT-A70 and identify its orthologues and paralogues to guide experimental analysis of this interesting protein family. In so doing, we found what appears to be the first example of circular permutation among MTases that act on substrates other than DNA. In addition, we find that the family of proteins related to MT-A70 includes not only probable MTases, but also a number of proteins that regulate mRNA levels via unknown mechanisms apparently not involving methylation.

Methods

Sequence Analysis

BLAST and its position-specific, iterative version, PSI-BLAST (Altschul et al. 1997), were used to search the nonredundant version of current sequences databases (nr) and the publicly available complete and incomplete genome sequences at the NCBI (http://www.ncbi.nlm.nih.gov/). Low-complexity sequence regions were left unmasked. The sequence of the human MT-A70 protein [Gene Identification (gi) No. 2460037] was used as an initial PSI-BLAST query with the default cutoff value. Following convergence, a preliminary phylogenetic tree was calculated from the alignment using the neighbor-joining approach (see below) to reveal representatives of branches other than the closest homologues (likely orthologues) of MT-A70. Using such sequences as seeds, the following divisions of the GenBank database were searched using TBLASTN: nonredundant (nr), expressed sequence tag (EST), sequence-tagged site (STS), high throughput genomic (HTG), and genome survey sequence (GSS). Partial sequences were assembled into contiguous fragments using intact sequences as guides and realigned using CLUSTALX (Thompson et al. 1997). This phylogenetic analysis and search was reiterated until no new MT-A70 homologues could be found.

The sequences assembled by us from multiple overlapping fragments included an *X. laevis* homologue of *Homo sapiens* BAB13453 [partial sequences with gi numbers 7393283 (amino acids 1–163), 7402421 (aa 153–371), and 7398047 (aa 316–501), an *L. esculentum* homologue of *Arabidopsis thaliana* T04002 [gi 8171357 (aa 1–160), 10902198 (aa 185–339), and 7410881 (aa 286–423)], and a *Glycine max* homologue of *A. thaliana* AT4g10760 [gi 9565361 (aa 1–160), 6665975 (aa 103–263), and 9564292 (aa 193–351)]. The "unfinished" sequences were corrected for frameshifts and splicing sites, where necessary, based on the results of reciprocal BLAST searches against the database comprising sequences of m^{6}A MTase homologues.

All sequences were subsequently realigned using the CLUSTALX program to the degapped profiles obtained from the multiple sequence alignments reported by BLAST. Manual adjustments were introduced based on the BLAST pairwise comparison, secondary structure prediction, fold recognition results, and, finally, superposition of modeled structures (see below).

Phylogenetic Analysis

The number of amino acid replacements per sequence position in the alignment was estimated using the JTT model (Jones et al. 1992). The sampling variance of the distance values was estimated from 1000 bootstrap resamplings of the alignment columns. Gaps were treated as missing characters. The evolutionary inference was carried out according to the maximum parsimony (Eck and Dayhoff 1966), maximum likelihood (Felsenstein 1973), and distance-based neighbor-joining (Saitou and Nei 1987) and Fitch–Margoliash (1967) methods implemented in the 3.6 version of the PHYLIP package (http://evolution.genetics.washington.edu/phylip.html). To reduce the computational cost and make the analysis plausible within a reasonable time, the exhaustive maximum likelihood search was initiated with the neighbor-joining trees. Multiple runs were conducted with randomized sequence input order and the "global rearrangements" option was used to avoid the tree being caught in a local statistical minimum.

Structure Prediction

To search for structurally characterized homologues of MT-A70 we used the MetaServer available at http://bioinfo.pl/meta/, which is a gateway to several secondary structure prediction and fold recognition methods (Bujnicki et al. 2001b) and which includes detailed reference and links to individual servers. Fold recognition methods compare the query (target) sequence with a library of structures (templates) and return the 10 best-scoring alignments according to the implemented criterion of compatibility. The results are collected by the MetaServer and submitted to the Pcons server (Lundstrom et al. 2001) (http://www.sbc.su.se/~arne/pcons), which compares the models and the associated scores and produces a ranking of potentially best predictions. Pcons differs from other

consensus-based methods since it does not simply list the most frequently reported templates but provides additional assessment of the quality of all models. According to our extensive benchmarking of fold recognition servers, Pcons confidence scores correspond much better to the observed prediction reliability than confidence scores reported by any individual server (Bujnicki et al. 2001a). This is especially advantageous in cases where several alternative folds are reported, if the correct fold is reported by most servers but the alignments differ or if one needs to choose the best template from several similar structures.

Modeling

Homology modeling was carried out following a modified version of the "multiple models" approach described previously (Bujnicki and Rychlewski 2001). Using the SWISS-MODEL/PROMOD II server (Guex and Peitsch 1997), we generated a set of preliminary models based on fold recognition-derived pairwise target–template alignments obtained from the MetaServer. The preliminary models were then superimposed using SWISS-PDB VIEWER (Guex and Peitsch 1997) and the best fragments were merged into the final structure. The choice of fragments was based on the evaluation of their stereochemical and energetic parameters by WHATCHECK (Hooft et al. 1996) and PROSA II software embedded within PROMOD II (Sippl 1993), consensus between the individual methods, and agreement with the independently predicted pattern of secondary structures and solvent accessibility (see above). We also used SWISS-PDB VIEWER for calculations of molecular surfaces and the electrostatic potential and to generate all figures of protein structures.

Results and Discussion

Phylogenomic Analysis Reveals Four Subfamilies of MT-A70-Related Proteins

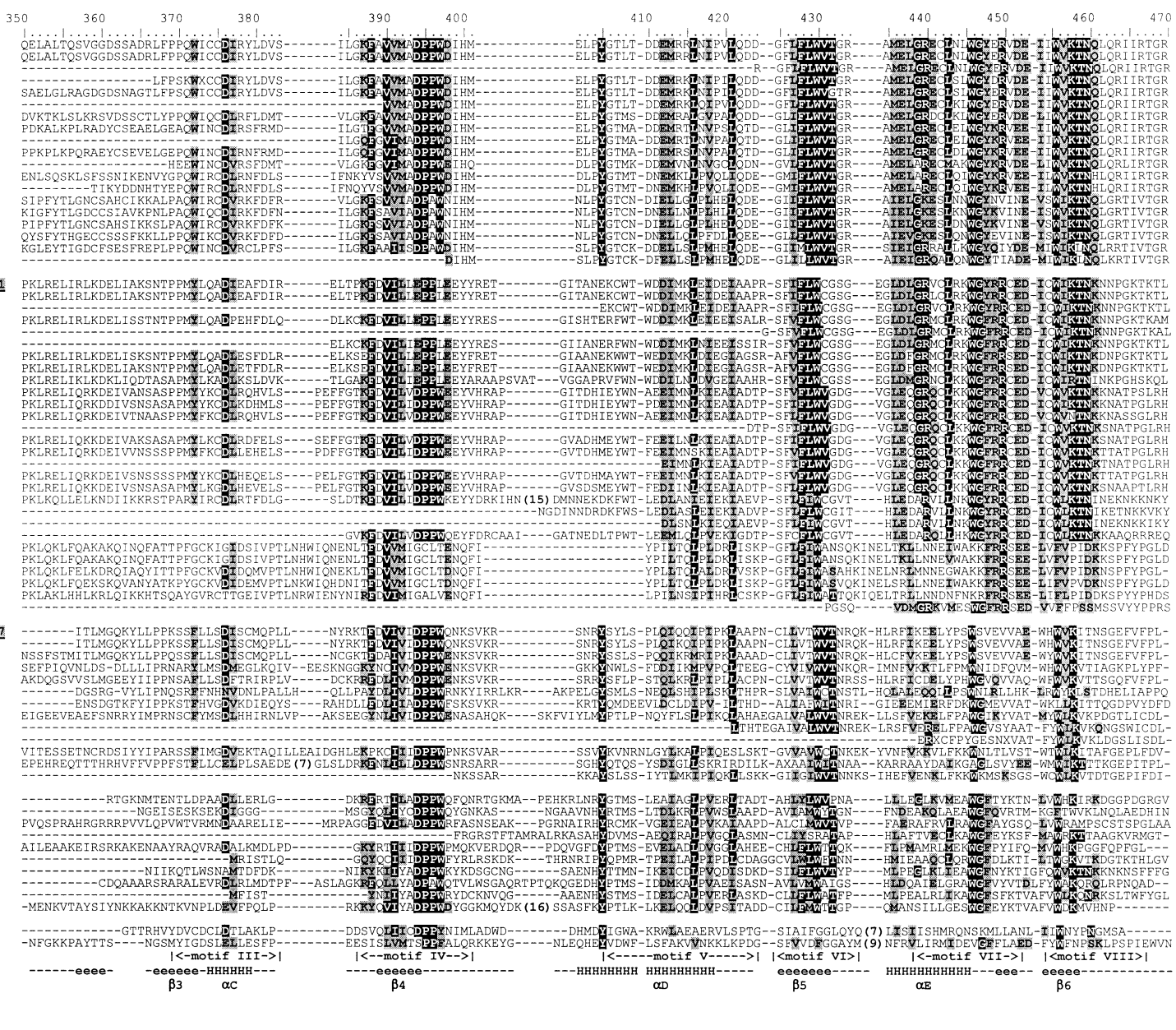
Members of protein families can often be identified by their characteristic amino acid sequence motifs. However, the AdoMet-dependent MTase superfamily shows very limited sequence conservation and considerable plasticity even within conserved motifs. Furthermore, the limited size of these motifs and the possibility that their order may be permuted (see below) make the significance of even good matches difficult to assess. Since simple motif analysis of the MT-A70 sequence provided little useful information (not shown), we made "phylogenomic" comparisons.

A PSI-BLAST search of the sequence databases for significant similarity to MT-A70 (see Methods) identified over 40 bacterial and eukaryotic proteins. The resulting family, shown in Figs. 1 and 2, is defined by sequence similarity in the carboxyl-proximal regions of the respective proteins. The amino-proximal regions of the eukaryotic proteins are highly diverse, often Pro-rich, and are conserved only within individual subfamilies (data not shown). Corresponding regions are not present in prokaryotic members of the family. In general terms this situation is reminiscent of the relationship between bacterial and mammalian DNA:m^5C MTases, where some of the mammalian enzymes differ from the bacterial ones in having a large amino-proximal region that plays regulatory roles (review by Robertson 2001). We could not detect any known motifs in the amino-proximal regions except, among MT-A70 orthologues (subfamily A; see below), for a bipartite nuclear localization signal (PROSITE profile PS50079).

To identify relationships between the MT-A70 family members, phylogenetic inference was carried out using distance matrix, maximum likelihood and parsimony approaches (see Methods). For this analysis, we masked columns of the sequence alignment (Fig. 1) in which more than 40% of characters corresponded to gaps. Sequences obtained from unfinished genomic data and cDNA clones may contain errors that could affect evolutionary inferences, so we also repeated the analysis after removing all incomplete or preliminary sequences from the alignment. Figure 2 shows the neighbor-joining tree. Confidence in this tree is greatly strengthened by the fact that its topology agrees perfectly with trees generated by the maximum likelihood and parsimony methods. Moreover, the same topology resulted when only the full-length, finished sequences were analyzed (data not shown). The branching pattern is also strongly supported by bootstrap analysis—most nodes show bootstrap values >900 (the few nodes with bootstrap support <400 have been collapsed). Finally, the predicted phylogenetic subfamilies correlate well with the presence of sequence signatures that can be regarded as synapomorphies (shared features derived from a common ancestor; see below).

Our analyses demonstrate that the MT-A70 family members can be grouped into four subfamily lineages (arbitrarily labeled A–D in Figs. 1 and 2). Lineages A–C are unique to eukaryotes and, respectively, include orthologues to the human proteins MT-A70, BAB13453, and BAB15520. BAB13453 (gi 10047331) is a hypothetical protein, the sequence for which is derived from a human brain cDNA clone (Nagase et al. 2000). BAB15520 (gi 10439572) is another hypothetical protein found during annotation of human cDNA sequences. Lineage D contains a small cluster of bacterial DNA:m^6A MTases including M.*Mun*I from *Mycoplasma* spp. (Siksnys et al. 1994), M.*Ava*V from *Anabaena* PCC 7120 [previously noted to be similar to MT-A70 (Matveyev et al. 2001)], and open reading frames identified in the unfinished sequences of various bacterial genomes.

Among those eukaryotes for which nearly completely sequenced genomes are available, *H. sapiens*, *Takifugu rubripes*, *D. melanogaster*, and *A. thaliana* each specify representatives of all three eukaryotic lineages (A, B, and C), while *S. pombe* and *C. elegans* each seem to have only one MT-A70 family member (of the C lineage). Lineages A and B can each be


-A-
H.sapiens MTA70-10179946
M.musculus-10179948
R.norvegicus-11677739
D.reiro-6402630
T.rubripes-JGI_27413
X.laevis-17422799
D.melanogaster-7301087
A.thaliana-7267774
S.tuberosum-13611773
G.max-9565361
C.reinhardtii-15699587
P.falciparum-unf'
P.yoelii-17299820
S.cerevisiae IME4-6321246
S.kluyveri-1216866
S.bayanus-1216077
K.lactis-12213029
C.albicans-265009C08
D.hansenii-1222002
-B-
H.sapiens BAB13453-10047331
M.musculus-11045943
B.taurus-10867711
T.rubripes-JGI_18418
O.latipes-17368825
D.reiro-5934927
S.tropicalis-16790685
X.laevis-398047
D.melanogaster-7297377
O.sativa-6630688
O.sativa-12039440
T.aestivum-9416864
G.max-16105345
E.esula-8404674
L.esculentum-10902198
G.hirsutum-5050368
S.tuberosum-11035273
A.thaliana-7487728
P.falciparum-9797737
P.vivax-13983682
P.berghei-13964863
T.gondii-17393664
S.cerevisiae KAR4-6319795
S.bayanus-12152560
Z.rouxii-12145686
S.servazzii-10954013
C.albicans-384163F03
Y.lipolytica-12178804
-C-
H.sapiens BAB15520-13647177
M.fascicularis-928028
M.musculus-15654942
C.intestinalis-16858991
T.rubripes-JGI_4885
D.melanogaster-969093
C.elegans-7496245
A.thaliana-8778413
G.max-15815542
M.grapefruit-8789158
S.pombe_C22G7.07C-1175468
A.fumigatus-unf
P.carinii_S46F12
-D-
M.capsulatus-unf
Y.enterocolitica-unf
R.sphaeroides-unf
R.meliloti-unf
Magnetococcus-unf
M.AvaV-8453090
M.haemolytica-4743483
N.aromaticivorans-unf
E.faecalis orf18-8100661
M.MunI-1171057
-structural matches-
M.RsrI-1eg2
M.PvuII-1boo
|<-motif III->|
|<--motif IV-->|
|<-----motif V----->|
|<motif VI>|
|<-motif VII->|
|<motif VIII>|
β3
αC
β4
αD
β5
αE
β6

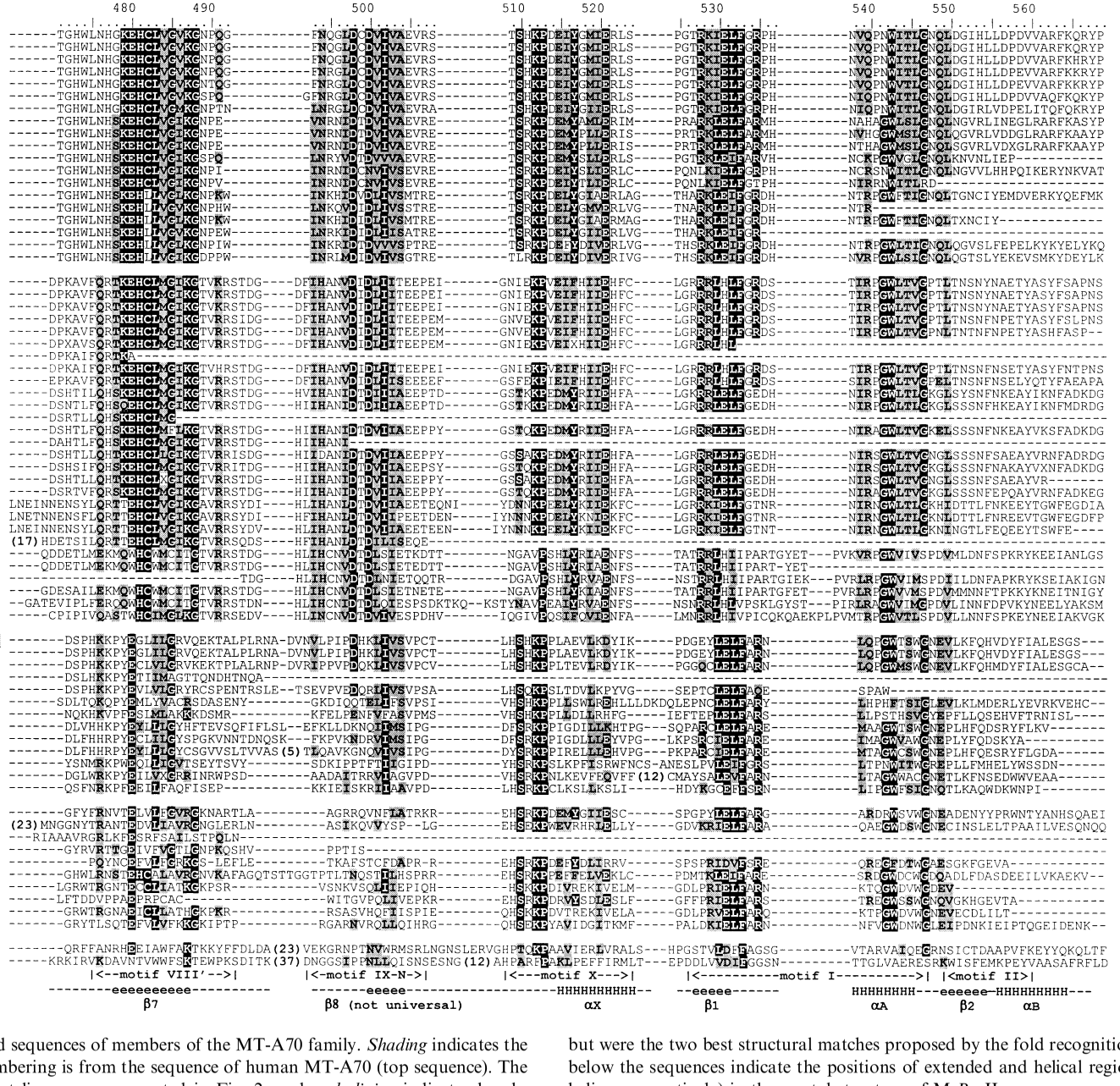


Fig. 1. Aligned amino acid sequences of members of the MT-A70 family. *Shading* indicates the degree of conservation. Numbering is from the sequence of human MT-A70 (top sequence). The letters A–D refer to different lineages as presented in Fig. 2, and *underlining* indicates breaks between sublineages. The bottom two sequences were not matched by direct sequence comparison but were the two best structural matches proposed by the fold recognition analysis, "e" and "H" below the sequences indicate the positions of extended and helical regions (i.e., β strands and α helices, respectively) in the crystal structure of M.*Pvu*II.

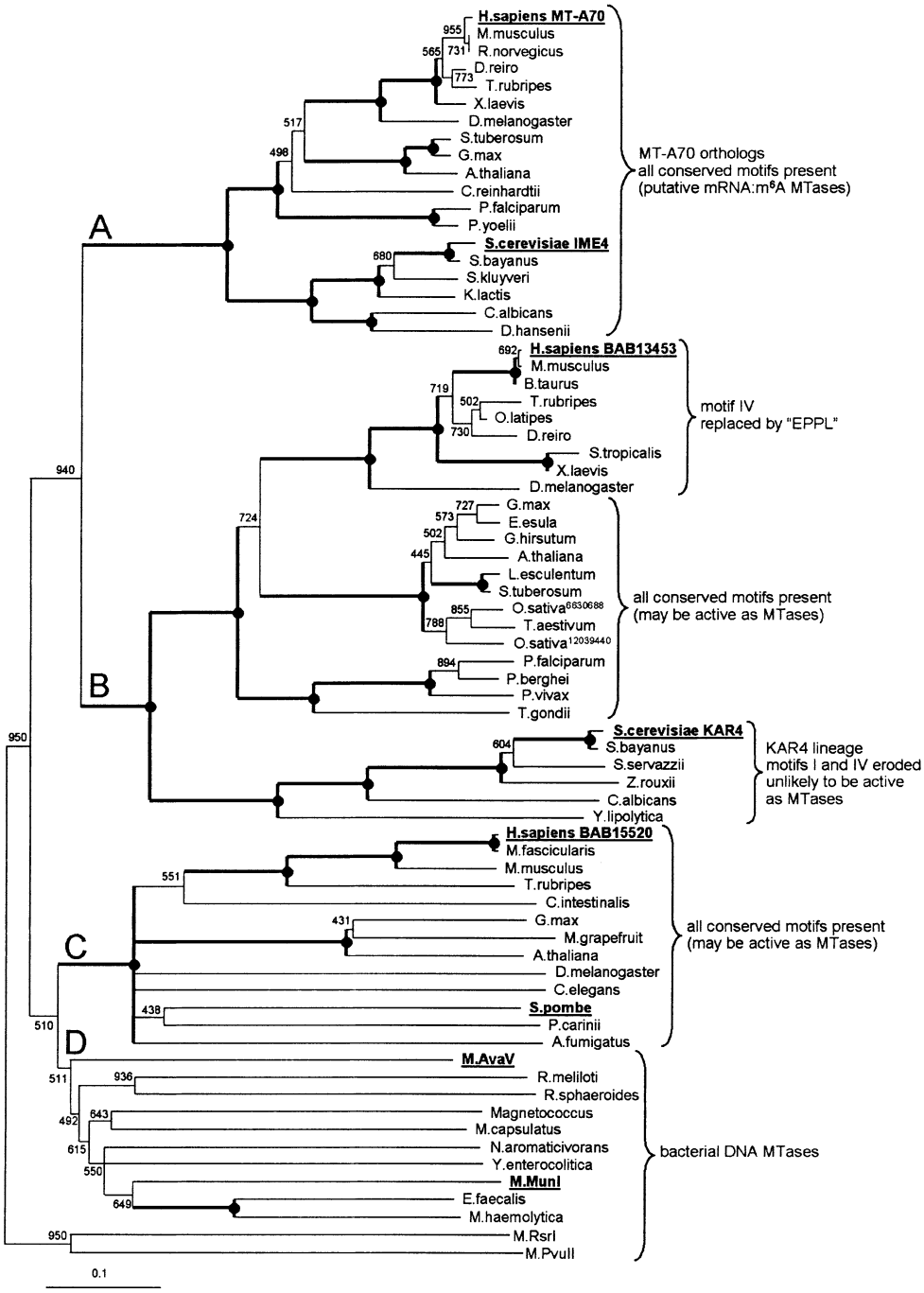


Fig. 2. Phylogenetic tree of the MT-A70 family. The lettering of the four main branches (A–D) is arbitrary and is included to facilitate comparison to the aligned sequences in Fig. 1. In each cluster, the best-studied or the most typical members are shown in *boldface* and *underlined*. The analysis was carried out as described under Methods. The tree topology is unchanged whether distance-based (neighbor-joining), parsimony, or maximum likelihood methods are used. Numbers at the nodes indicate bootstrap values from the neighbor-joining analysis (1000 replicas). Nodes with bootstrap values >99% are shown as *thick lines*, while nodes with support <40% were regarded as unresolved and have been collapsed. The following nodes received the percentage bootstrap support indicated in parentheses (via parsimony, via maximum likelihood): A (100, 56), B (78, 76), C (97, 91), D (47, 23), A/B (63, 45), C/D (57, 49), and M.*Pvu*II/M.*Rsr*I (100, 92).

subdivided into branches corresponding to proteins from (i) protozoans and multicellular organisms including green plants and metazoans and (ii) budding yeasts (Fig. 2). In lineage A this refers to the groups including MT-A70 (i) and IME4 (ii), and in lineage B to groups including BAB13453 (i) and KAR4 (ii). In contrast, the C lineage (BAB15520 orthologues) includes proteins from green plants, metazoans, and the fungi *S. pombe* and *Pneumocystis carinii* (subphylum Taphrinomycotina) and *Aspergillus fumigatus* (sub-

phylum Pezizomycotina) but not from budding yeasts or protozoans.

The inferred branching topologies of the A and B lineages are identical, given the resolution of the methods used. It is peculiar that the earliest-radiating branches in these two subfamilies involve the yeasts (IME4 and KAR4). In the Tree of Life, based on rRNA and protein sequences, the yeasts group with metazoa, while protists and green plants are the deeply branched outgroups (Gouy and Li 1989; see also http://tolweb.org/tree?group = Eukaryotes). We suspect that this discrepancy may reflect a higher rate of evolution of MT-A70 homologues in yeasts. Neighbor-joining analysis, using Kimura's correction for multiple substitutions, results in a topology consistent with the Tree of Life, however with weaker bootstrap support (data not shown). The availability of additional sequences from early-branching Eukaryota may help clarify this issue.

The deep-branching pattern within lineage C could not be resolved with confidence, however, there is no obvious incongruity with the universal Tree of Life (Gouy and Li 1989). Hence, it is reasonable to infer that two gene duplication events in the ancient eukaryote gave rise to the A, B, and C lineages. The representatives of the A and B lineages have most likely been lost in *C. elegans* and in the genomes of those yeasts species that retained the member of the C lineage, while representatives of the C lineage have been lost in protozoa and in budding yeasts.

The relationship of the bacterial subfamily D to lineages A–C was less clear. PSI-BLAST did not detect significant sequence similarity of any of the MT-A70 homologues to members of other known MTase families, whose sequences could have been used as an outgroup to root the tree. Quite unexpectedly, even the search initiated with the sequences of genuine DNA MTases in lineage D (M.*Mun*I and M.*Ava*V) revealed significant similarity only to the eukaryotic MT-A70 family members and not to other DNA MTases (not shown), suggesting that these enzymes are outliers in the DNA MTase superfamily that could have evolved convergently from RNA MTases.

To identify a suitable outgroup, we employed fold recognition methods that allow detection of remote relationships between protein families (see the following section). We detected an unambiguous relationship of the MT-A70 family to the β class of DNA MTases, represented by the structurally characterized enzymes M.*Pvu*II and M.*Rsr*I. The sequences of these two MTases were added to the alignment (Fig. 1, bottom rows) and used as an outgroup in phylogenetic calculations. All approaches used (distance matrix, parsimony, and maximum likelihood) confidently placed the root between the A/B and the C/D branches (Fig. 2). This result is consistent with either of two scenarios. First, duplication of the last common ancestor of the MT-A70 family preceded the split between Eukarota and Prokaryota, and all paralogous copies were lost in Prokaryota (at least those represented in the sequence databases) with the exception of the D subfamily. Second, the origin of the MT-A70 family is strictly eukaryotic and the D subfamily diverged from the C lineage after an ancient eukaryote-to-prokaryote horizontal transfer. The topology of the D lineage is inconsistent with established bacterial phylogeny and suggests a history of horizontal gene transfer among bacteria, as is commonly observed among RM systems (Jeltsch and Pingoud 1996). The D lineage MTase M.*Mun*I is in fact part of a RM system.

Fold Recognition Analysis Reveals a Common Structure for MT-A70 and the β-Class DNA MTases

The initial sequence analysis of MT-A70 (Bokar et al. 1997) revealed broad similarity to the known DNA MTase M.*Mun*I, however neither the previous analysis nor our iterative searches unambiguously linked the MT-A70 family to other known MTases. This led us to fold recognition analysis. Fold recognition often detects evolutionary relationships between protein families, including MTases, even when their sequences have diverged beyond detection by standard sequence comparison algorithms (Bujnicki and Rychlewski 2001).

The great majority of AdoMet-dependent MTases share a consensus core structure (reviews by Fauman et al. 1999; Cheng and Roberts 2001), as presented in the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.d.gg.b.html). Crystallographic and NMR studies have shown that the catalytic domain structures from numerous MTases are very similar, suggesting a common origin (Bujnicki 1999; Fauman et al. 1999). However the crystal structure of CbiF MTase (Schubert et al. 1998) is unrelated to the "classical" superfamily fold, and sequence analyses have suggested that the SET family of protein-lysine MTases (Jenuwein 2001) and the SPOUT superfamily of RNA MTases (Anantharaman et al. 2002a) might also have nonconsensus folds. Thus structure prediction for MTases showing no clear sequence similarity to structurally characterized members is not a trivial exercise with a foregone conclusion.

Even within the MTase superfamily, furthermore, there are variations on this consensus structure. The core of a "classical" AdoMet-dependent MTase is an $\alpha/\beta/\alpha$ sandwich (Fig. 3). The central β-sheet comprises seven strands, and a topological switchpoint between strand 1 and strand 4 results in a deep cleft into which AdoMet binds. Strand 7 is antiparallel to the other six strands and forms a signature reversed β hairpin ($6\uparrow 7\downarrow 5\uparrow 4\uparrow 1\uparrow 2\uparrow 3\uparrow$). However, in protein

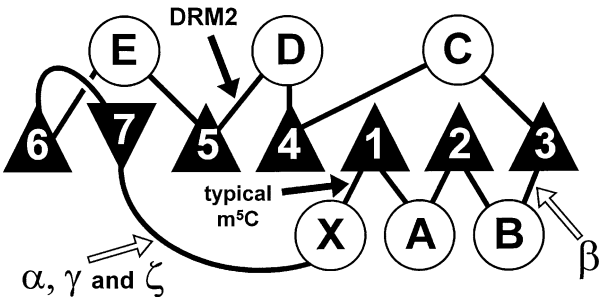


Fig. 3. Circular permutation of MTases. In this representation of the MTase consensus fold, *triangles* represent β strands and *circles* represent α helices. This is a schematic view from "above," looking at the edge of the β-sheet (for simplicity, the polarity of the helices is not indicated). The *gray arrows* indicate topological breakpoints at which the structure is opened in various DNA:m^4C and DNA:m^6A MTase classes, generating the amino and carboxyl termini. For the β class, which up to now has included only N-MTases such as M.*Pvu*II and M.*Rsr*I, this opening occurs before strand 3. For other N-MTases, the opening occurs on either side of helix X (so named as it includes conserved motif X). *Black arrows* indicate two additional breakpoints, which have been identified in DNA:m^5C MTases but not in N-MTases.

isoaspartyl methyltransferase the β hairpin formed by strands 6 and 7 is inverted (7↑ 6↓ 5↑ 4↑ 1↑ 2↑ 3↑) (Skinner et al. 2000), while in protein arginine MTases the β hairpin is absent altogether (5↑ 4↑ 1↑ 2↑ 3↑) (Weiss et al. 2000). Other MTases exhibit variant connectivity of structural elements in the consensus fold, resulting from permutation of the gene (Fauman et al. 1999). For instance, in the β subfamily of DNA MTases the "1↑ 2↑" and "6↑ 7↓ 5↑ 4↑ 3↑" regions have exchanged places in the relative to most MTases, with strands 1 and 2 at the carboxyl rather than the amino terminus (Fig. 3). This has been confirmed in the crystal structures of two DNA MTases (Gong et al. 1997; Scavetta et al. 2000). Fold predictions must take these possible variations into account.

We subjected all four MT-A70 subfamilies, separately, to three-dimensional fold recognition analysis using the Structure Prediction Meta Server [(Bujnicki et al. 2001b); the relative performance of the individual servers was reviewed by Bujnicki et al. (2001a)]. The amino-proximal regions of all eukaryotic proteins in the family are predicted to have little secondary structure (not shown), suggesting intrinsic disorder at least in the unliganded state, and these regions were omitted from further structural analyses.

For all four subfamilies, the MTase fold scored best. More specifically, the crystal structures of M.*Pvu*II (PDB 1boo) and M.*Rsr*I (PDB leg2) were always among the top-ranked matches, with the Pcons consensus scores (Lundstrom et al. 2001) consistently well above the established false-positive threshold. In benchmarking tests, 87% of models having a Pcons score of 3 were correct, while 98% of the models with a score of 5 were correct (Lundstrom et al. 2001). For the human or bacterial representatives in each lineage, Pcons scores for the match to M.*Pvu*II were as follows: (A) hu-MT-A70, 6.22; (B) BAB13453, 5.21; (C) BAB15520, 5.28; and (D) M.*Mun*I, 4.78. For *S. cerevisiae* KAR4 the M.*Pvu*II MTase fold was also ranked as the most favorable, though with a somewhat lower Pcons score (2.35), which may reflect the fact that KAR4 lacks hallmark MTase sequence motifs (see below).

Other MTase structures were also reported by fold recognition algorithms for all MT-A70 family members, however with lower scores and with alignments spanning only the AdoMet-binding domain or the catalytic domain (not shown). In contrast, the M.*Pvu*II and M.*Rsr*I alignments always spanned the entire protein sequence (aside from the excluded amino-proximal regions not present in bacterial enzymes). This result is intriguing, as M.*Pvu*II and M.*Rsr*I both belong to the β class of DNA MTases, having a permuted linear order of structural elements, and previously this permutation had been seen only among DNA MTases (Fig. 3). The congruence between methods and similarity of the results obtained for all four MT-A70 subfamilies greatly increase the probability that the structures of these proteins are truly similar to those of DNA MTases from the β class.

This apparent structural similarity to β DNA MTases was further explored by building and evaluating all-atom models of the M.*Mun*I, MT-A70, BAB13453, BAB15520, and *S. cerevisiae* KAR4 MTase domains. We used a variation of the "multiple models" approach described under Methods, based on the pairwise sequence-to-structure alignments obtained from fold recognition servers and the initial multiple sequence alignment. The models passed the quality tests implemented in PROSA II (Sippl, 1993), the pG server http://guitar.rockefeller.edu/pg/, and WHATCHECK (Hooft et al. 1996), with scores similar to those for low-resolution experimental structures or homology models based on largely correct alignments to templates with a sequence identity $<30\%$ (*i.e.*, PROSA II/pG scores between 0.69 and 0.90, WHATCHECK Z-scores between -1.35 and -1.73). These results provide compelling evidence that M.*Pvu*II and M.*Rsr*I are legitimate structural templates for MT-A70 homologues, that our fold recognition alignments and structural models (web site Fig. 5) are accurate enough to justify the use of M.*Pvu*II and M.*Rsr*I as an outgroup in the phylogenetic analysis (previous section), and that assignment of sequence motifs can be productively guided by the fold recognition alignments (next section).

Structure-Guided Sequence Analysis of the MT-A70 Family

When comparing MTase families that recognize substrates as different as glycine and DNA, only one region is readily recognized at the sequence level: the Rossmann-fold-like AdoMet-binding site. This region includes the glycine-rich motif I, originally discovered among DNA MTases that generate 5-methylcytosine (m^5C) (Posfai et al. 1989), that forms the binding pocket for the methionine moiety of AdoMet. The remainder of the core MTase has obviously conserved motifs only within groups of MTases that act on chemically related targets. Motif IV is the most readily recognized of these motifs. Among MTases that act on exocyclic amino nitrogens of DNA or RNA bases, which would include MT-A70, motif IV has the degenerate consensus (N/D/S)–P–P–(F/Y/W/H) (Malone et al. 1995; Fauman et al. 1999; Bujnicki 2000). However, NPPY motifs are also present in amino MTases acting on the secondary metabolite aminophenylalanine (Blanc et al. 1997) and on glutamine in proteins (Nakahigashi et al. 2002), suggesting that this motif is a hallmark of amino-methylating enzymes and not specifically of nucleic acid N-MTases.

Candidates for motifs I and IV of MT-A70 were previously proposed by comparison to the sequence of M.*Mun*I (Bokar et al. 1997). The multisequence alignment shown in Fig. 1 reveals that the MT-A70 DP(P/A)W^{397} candidate for motif IV is highly conserved (except in lineage B, discussed below). However, the MT-A70 family shows remarkable and evenly distributed sequence conservation, and the previously proposed motif I is not conserved significantly better than neighboring sequences. We therefore used the fold recognition alignment with M.*Pvu*II and M.*Rsr*I to map all 11 MTase-specific motifs in the MT-A70 family (Fig. 1, bottom). These 11 motifs correspond to the 9 (X and I–VIII) common to the catalytic domains of nearly all N-MTases (Malone et al. 1995; Bujnicki 2001) and 2 additional motifs (VIII′ and IX-N) specific to the α and β classes of m^4C MTases (Bujnicki and Radlinska 1999).

The insufficiency of sequence information alone, for assigning motifs in the MT-A70 family, is consistent with a previous structure-guided alignment of a more disparate group of MTases that found just three positions highly conserved at the sequence level (Fauman et al. 1999). First was a Gly from motif I, at or near the start of helix αA. The Gly residue in motif I (G^{534} in *H. sapiens* MT-A70) is only partially conserved in the MT-A70 family and is at the second position of the (F/G)xGxG motif, a position poorly conserved in the MTase superfamily. The second highly conserved position was an acidic side chain (D/E) at the end of strand β2 (motif II) that H-bonds the ribose hydroxyls in AdoMet This residue is substituted in most MT-A70 family members by an amide side chain (Q or N) or by His (H^{554} in *H. sapiens* MT-A70). The third highly conserved position from the earlier study was a branched aliphatic side chain (I/V) in the middle of strand β4 (motif IV), involved in strand packing to form the β sheet. This aliphatic side chain is conserved in all MT-A70 homologues.

The only other residues conserved in all or nearly all members of the MT-A70 family (W^{430}, W^{456}, P^{513}, and W^{543} in *H. sapiens* MT-A70) are important for structural stability of the protein core or hydrophobicity of the binding pocket, rather than catalytic activity *per se*. Even most of the moderately conserved amino acids are predicted to help form the proteins' hydrophobic cores. Thus the MT-A70 family comprises proteins with similar architecture but different surfaces, suggesting a range of distinct substrate classes.

The fold recognition alignment of MT-A70 makes testable predictions. The side chains of F^{533}, D^{376}, E^{531}, and H^{543} should participate in AdoMet binding, while the main-chain carbonyl of P^{395} and side chains of D^{394}, W^{397}, E^{480}, and K^{512} should play roles in catalysis (Fig. 4). Our results also suggest candidate regions for nucleic acid sequence recognition. The "variable" region between motif VIII′ and motif IX-N, proposed as the locus of sequence specificity in the β family of DNA MTases (Malone et al. 1995), is missing from MT-A70 homologues (including *bona fide* DNA MTases M.*Mun*I and M.*Ava*V). However, small DNA MTases may recognize their target using one or more loops protruding from the catalytic face of the protein rather than a separate subdomain (Bujnicki and Radlinska 1999). We propose that moderately conserved residues in loops β4/αD, β6/β7, and β4/α10 are involved in sequence recognition by members of the MT-A70 family. Most of these residues are conserved in M.*Mun*I and M.*Ava*V, though a large part of the β4/αD loop in M.*Mun*I is not present in M.*Ava*V (Fig. 1). To understand substrate recognition by members of the MT-A70 family that act on DNA or RNA, it could be very useful to determine which portions of CAATTG-specific M.*Mun*I and GATC-specific M.*Ava*V are responsible for sequence specificity.

Possible Loss of MTase Activity in the KAR4 Lineage

S. cerevisiae Kar4p is a phosphoprotein required for expression of karyogamy-specific genes during mating and, also acts during mitosis and meiosis (Gammie et al. 1999). It is not yet known if Kar4p interacts with DNA, RNA, or other proteins. We propose that the KAR4 proteins (subset of lineage B in Fig. 2) diverged from one of the MT-A70-like nucleic acid MTases, losing the ability to bind AdoMet and/or

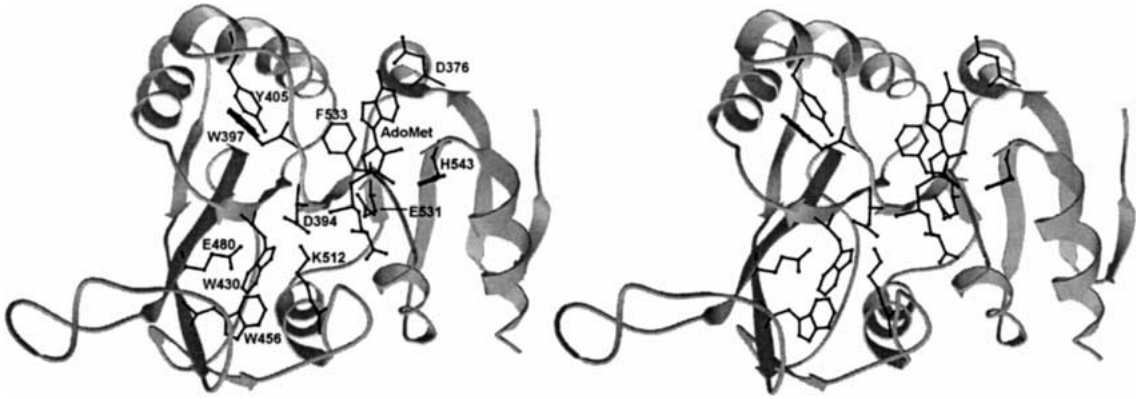


Fig. 4. Homology-based structural model for the MTase domain of MT-A70 shown in a ribbons representation. The residues predicted to be important for structural stability and to participate in cofactor binding and catalysis are shown in *wireframe*.

catalyze methyl transfer but maintaining the ability to bind nucleic acids. We base this hypothesis on the status of motifs I and IV in the KAR4 lineage.

Motif IV (DPPW^{397} in MT-A70; Fig. 1) contains catalytic residues essential for methyltransfer, as noted above. The DPPW tetrapeptide is generally retained in motif IV (Fig. 1), however, in the metazoan members of lineage B it is replaced by EPPL, and in the KAR4 lineage by GCLT (GALV in one case). By analogy to solved DNA MTase structures, the N/D/S side chain (D^{394} in MT-A70) and the main-chain carbonyl of the subsequent Pro (P^{395}) are expected to form hydrogen bonds to the target amino group. Though no proven N-MTase has been found to have the D→E substitution in motif IV, this is a conservative change. In contrast, the D→G substitution in the KAR4 proteins is likely to cause loss of MTase function. One major caveat, however, is that a family of RNA:m^{5}C MTases was proven to use an equivalent of motif IV from a different region of the protein (Liu and Santi 2000).

In addition to the aberrant motif IV, the KAR4 lineage has particularly deviant motif I sequences. Like Motif IV, motif I is conserved among most MT-A70 homologues, though in all cases it deviates from the (D/E)x(F/G)xGxG consensus (ELFGRPH^{537} in MT-A70). However, in the KAR4 lineage motif I is altered at the acidic first position and contains an insertion relative to other motif I sequences. The acidic side chain at this position, at the end of strand 1, is conserved in nearly all MTase families analyzed to date (Fauman et al. 1999). Yet in KAR4 and "EPPL" lineages, Glu is replaced by His and we could not identify acidic residues in appropriate alternative positions. Furthermore, in KAR4 proteins motif I is disrupted by an 8- to 11-aa insertion that seems likely to occlude the cofactor-binding site (Figs. 5 and 6).

Taken together, the sequence and structural predictions suggest that the KAR4 proteins are inactive for methyltransfer and may not even bind AdoMet. If this is correct, it would be interesting to know how gene expression is activated by a nonmethylating protein with the consensus MTase fold or if this activation is purely a function of the "non-MTase" amino-proximal region. Another possibility is that MTases incapable of methyltransfer act via base flipping. Base flipping is the rotation of a single nucleotide out of a DNA or RNA double helix so that the nucleotide can enter an enzyme's catalytic pocket; MTases are among the enzymes that carry out base flipping (Cheng and Roberts 2001). There have been suggestions, and some evidence, that base flipping proteins may act as chaperones in promoting proper folding of complex RNA structures (Gutgsell et al. 2001).

Another group of MTase-related regulatory proteins has recently been identified that, like the KAR4 group, might not act by catalyzing methyltransfer. The eukaryotic mitochondrial transcription factor B (mtfB) mediates interaction of RNA polymerase with mitochondrial promoters. MtfB is related to the Erm family of rRNA:m^{6}A MTases at the sequence and structural levels and in vitro binds AdoMet and nonspecifically binds RNA and mtDNA (Schubot et al. 2001; McCulloch et al. 2002). Reminiscent of the relationship between lineage A and lineage C of the MT-A70 family, human mtfB exhibits all sequence features typical of active adenine MTases, while the yeast mtfB sequences lack the "catalytic" motif IV (NVTG^{140} in *S. cerevisiae*) and exhibit an insertion [in this case between motif VII and motif VIII (Schubot et al. 2001)]. The MTase activity of mtfB proteins has not been demonstrated to date.

Other Groups in the MT-A70 Family

Physiological roles are not known for any of the proteins in the BAB13453 group of lineage B, but these proteins probably vary in their capacity to catalyze methyltransfer. The KAR4 cluster, as noted above, shows major apparent defects in conserved

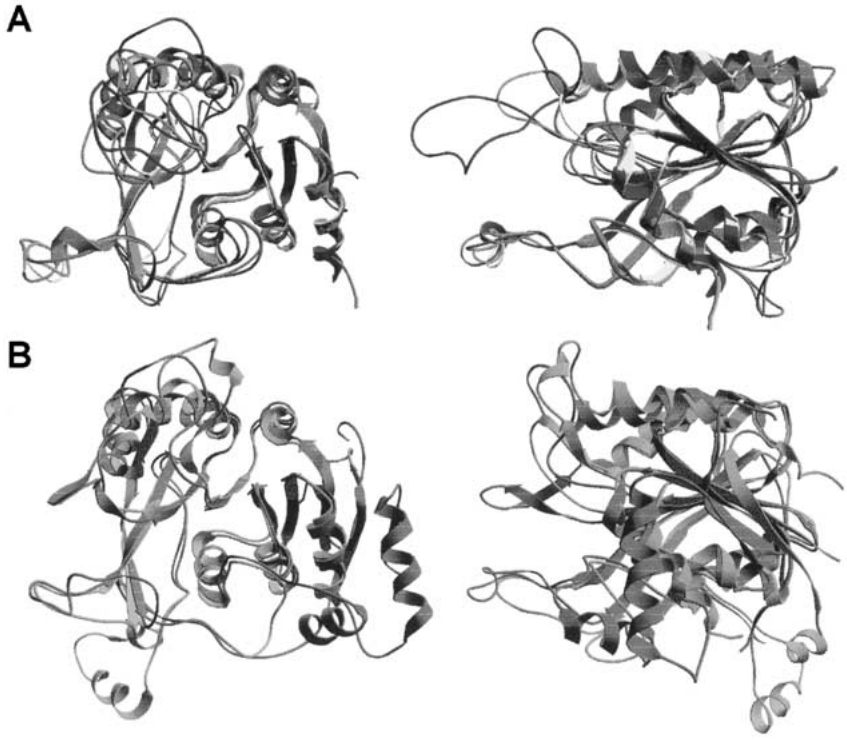


Fig. 5. (Full color version available at www.link.springer-ny.com/link/service/journals/00239.) Homology models for members of the MT-A70 family and template structures in front and side view. **(A)** Human proteins MT-A70 (in *yellow*), BAB13453 protein (*in green*), and BAB15520 (in *blue*), DNA MTase M.*Mun*I (in *red*) and *S. cerevisiae* KAR4 protein (in *magenta*). **(B)** Template structures of M.*Pvu*II (1boo in Protein Data Bank; in *magenta*) and M.*Rsr*I (1eg2 in PDB; in *blue*). Note the absence of additional helices present in both template structures in comparison with the common MTase core (schematically shown in Fig. 3) and the variable size of the β4/αD loop that immediately follows the conserved DPPW tetrapeptide (motif IV).

MTase motifs and probably lacks MTase activity. The metazoan orthologues have a more intermediate status with the EPPL version of motif IV feasibly functional, but these proteins share the KAR4 feature of a His in motif I replacing the conserved acidic residue. The remaining lineage B members are likely to have MTase activity.

There are two other groups in the MT-A70 family that appear to possess all features typical of active adenine MTases. One is the IME4 cluster from fungi, which, like MT-A70 itself, is within lineage A. IME4 (previously SPO8) from *S. cerevisiae* is the best-studied member of this cluster. IME4 increases accumulation of IME1 mRNA by unknown means, and IME1 in turn stimulates expression of several genes involved in initiation of meiosis (Shah and Clancy 1992). IME4 was previously identified as an MT-A70 homologue (Bokar et al. 1997) but there is no reported evidence for its possible MTase activity.

Finally, there are the BAB15520 orthologues of lineage C, wich are relatively closely related to the M.*Mun*I cluster of bacterial DNA MTases (lineage D). These include a putative protein specified by the bithorax complex (BX-C) of *Drosophila*, one of two complexes that act as master regulators of the body plan of the fly (Martin et al. 1995). These proteins may represent uncharacterized regulatory factors of substantial importance. It will be interesting to determine if their function is connected to their potential MTase activity. Among other possibilities, they might be part of the mRNA:m^{6}A MTase complex or could represent the U6 snRNA:m^{6}A MTase, which, to our knowledge, has neither been cloned nor had its sequence identified (Shimba et al. 1995).

Gene Permutation and MTase Evolution

It has been noted for some time (Malone et al. 1995) that the genes for one group of DNA MTases have undergone circular permutation while maintaining the overall protein fold (Fig. 3). Two structures have since confirmed this interpretation (Gong et al. 1997; Scavetta et al. 2000). This observation has led to questions about why such permutation was found only among DNA MTases (Fauman et al. 1999). It now appears that the fairly large MT-A70 family, including RNA MTases, M.*Mun*I-related DNA MTases, possible transcription factors, and other proteins of unknown function, all share the same permutation as the β family of DNA MTases exemplified by M.*Pvu*II and M.*Rsr*I. While this paper was under review another group reached a similar conclusion, though without structural modeling, as part of a broad survey of proteins involved in RNA metabolism (Anantharaman et al. 2002b). Thus circular permutation of MTases is apparently not unique to those that act on DNA. This important conclusion awaits experimental verification.

With both the MT-A70 family and a subgroup of "classical" bacterial DNA MTases sharing the same circular permutation, it is worth considering their evolutionary relationship to one another. The MT-A70/M.*Mun*I and M.*Pvu*II/M.*Rsr*I families of β

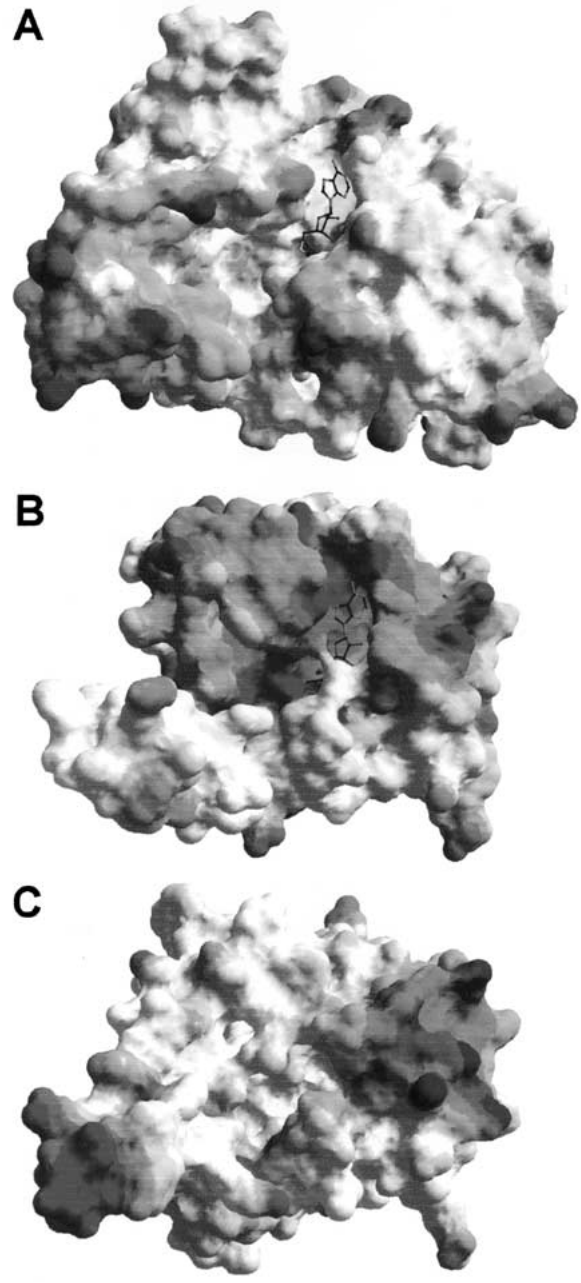


Fig. 6. (Full color version available at www.link.springer-ny.com/link/service/journals/00239.) Comparison of electrostatic surface potential. The surface potential is mapped onto the molecular surfaces of (**A**) M.*Pvu*II, (**B**) MT-A70, and (**C**) *S. cerevisiae* KAR4. The values of surface potentials are expressed as a spectrum ranging from −5 kT/e (*deep red*) to +5 kT/e (*deep blue*). The cofactor-binding site of KAR4 is completely blocked by the insertion. The AdoMet coordinates in the MT-A70 model have been modeled based on the crystal structure of the M.*Pvu*II–AdoHcy complex.

MTases could in theory have arisen either from a single permutation event ancestral to both families or from two independent but similar permutations. Sequence comparison and phylogenetic analyses indicate that these two families differ substantially in amino acid sequence. Moreover, the most parsimonious scenario inferred from our analysis suggests that the existence of the ancestor of the MT-A70 family predated the divergence of green plants, yeasts, protists, and animals. If there was in fact a single permuted ancestor for both β MTase families, it was probably quite ancient.

A second question about the evolution of these two permuted MTase families arises from the fact that one consists entirely of DNA MTases, while the other contains a relatively small DNA MTase subfamily (D in Fig. 2). It is not clear how the interconversion of specificity between RNA and DNA occurs in the MTase superfamily, though such changes are by no means unexpected given evidence that a family of protein:glutamine-N^5 MTases evolved from typical DNA MTases (Nakahigashi et al. 2002). However, the substrate specificities of the eukaryotic B and C lineages remain unknown, which makes it difficult to judge whether the common ancestor of the MT-A70 family was more likely to be an RNA MTase like MT-A70 itself or a DNA MTase like M.*Mun*I and M.*Ava*V. Structural and functional analysis of MT-A70 family members may help in answering this question and could also shed light on the origin of nucleic acid MTases in the RNA–protein world that apparently existed before DNA became the carrier of genetic information.

Acknowledgments. This work was supported by the Polish State Committee for Scientific Research (KBN; Grants 8T11-F01019 to J.M.B. and 6P04-B00519 to M.R.) and by a grant from the U.S. National Science Foundation (MCB-9904523, to R.M.B.).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anantharaman V, Koonin EV, Aravind L (2002a) SPOUT: A class of methyltransferases that includes SpoU and TrmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. J Mol Microbiol Biotechnol 4:71–75

Anantharaman V, Koonin EV, Aravind L (2002b) Comparative genomics and evolution of protein involved in RNA metabolism. Nucleic Acids Res 30:1427–1464

Blanc V, Gil P, Bamas-Jacques N, Lorenzon S, Zagorec M, Schleuniger J, Bisch D, Blanche F, Debussche L, Crouzet J, Thibaut D (1997) Identification and analysis of genes from *Streptomyces pristinaespiralis* encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I. Mol Microbiol 23:191–202

Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F (1994) Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. J Biol Chem 269:17697–17704

Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM (1997) Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA 3:1233–1247

Bujnicki JM (1999) Comparison of protein structures reveals monophyletic origin of the AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation. In Silico Biol 1:1–8 (http://www.bioinfo.de/isb/1999-01/0016/)

Bujnicki JM (2000) Phylogenomic analysis of 16S rRNA:(guanine-N^2) methyltransferases suggests new family members and reveals highly conserved motifs and a domain structure similar to other nucleic acid amino-methyltransferases. FASEB J 14:2365–2368

Bujnicki JM (2001) Understanding the evolution of restriction-modification systems: Clues from sequence and structure comparisons. Acta Biochim Pol 48:1–33

Bujnicki JM (2002) Sequence permutations in the molecular evolution of DNA methyltransferases. BMC Evol Biol 2:3

Bujnicki JM, Radlinska M (1999) Molecular evolution of DNA-(cytosine-N^4) methyltransferases: Evidence for their polyphyletic origin. Nucleic Acids Res 27:4501–4509

Bujnicki JM, Rychlewski L (2001) Sequence analysis and structure prediction of aminoglycoside-resistance 16S rRNA:m^7G methyltransferases. Acta Microbiol Pol 50:7–17

Bujnicki JM, Elofsson A, Fischer D, Rychlewski L (2001a) LiveBench-2: Large-scale automated evaluation of protein structure prediction servers. Proteins 45:184–191

Bujnicki JM, Elofsson A, Fischer D, Rychlewski L (2001b) Structure prediction Meta Server. Bioinformatics 17:750–751

Cheng X, Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 29:3784–3795

Eck RV, Dayhoff MO (1966). Atlas of protein sequence and structure. Natl Biomed Res Found, Washington, DC

Fauman EB, Blumenthal RM, Cheng X (1999) Structure and evolution of AdoMet-dependent MTases. In: Cheng X, Blumenthal RM (eds) S-Adenosylmethionine-dependent methyltransferases: Structures and functions. World Scientific, Singapore, pp 1–38

Felsenstein J (1973) Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet 25:471–492

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Gammie AE, Stewart BG, Scott CF, Rose MD (1999) The two forms of karyogamy transcription factor Kar4p are regulated by differential initiation of transcription, translation, and protein turnover. Mol Cell Biol 19:817–825

Gong W, O'Gara M, Blumenthal RM, Cheng X (1997) Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. Nucleic Acids Res 25:2702–2715

Gouy M, Li WH (1989) Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi. Mol Biol Evol 6:109–122

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PDB Viewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723

Gutgsell NS, Del Campo MD, Raychaudhuri S, Ofengand J (2001) A second function for pseudouridine synthases: A point mutant of RluD unable to form pseudouridines 1911,1915, and 1917 in *Escherichia coli* 23S ribosomal RNA restores normal growth to an RluD-minus strain. RNA 7:990–998

Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272

Jeltsch A (1999) Circular permutations in the molecular evolution of DNA methyltransferases. J Mol Evol 49:161–164

Jeltsch A, Pingoud A (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. J Mol Evol 42:91–96

Jenuwein T (2001) Re-SET-ting heterochromatin by histone methyltransferases. Trends Cell Biol 11:266–273

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Liu Y, Santi DV (2000) m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. Proc Natl Acad Sci USA 97:8263–8265

Lundstrom J, Rychlewski L, Bujnicki JM, Elofsson A (2001) Pcons: A neural-network-based consensus predictor that improves fold recognition. Protein Sci 10:2354–2362

Malone T, Blumenthal RM, Cheng X (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J Mol Biol 253:618–632

Martin CH, Mayeda CA, Davis CA, Ericsson CL, Knafels JD, Mathog DR, Celniker SE, Lewis EB, Palazzolo MJ (1995) Complete sequence of the bithorax complex of *Drosophila*. Proc Natl Acad Sci USA 92:8398–8402

Matveyev AV, Young KT, Meng A, Elhai J (2001) DNA methyltransferases of the cyanobacterium *Anabaena* PCC 7120. Nucleic Acids Res 29:1491–1506

McCulloch V, Seidel-Rogol BL, Shadel GS (2002) A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. Mol Cell Biol 22:1116–1125

Nagase T, Kikuno R, Nakayama M, Hirosawa M, Ohara O (2000) Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 7:273–281

Nakahigashi K, Kubo N, Narita SS, Shimaoka T, Goto S, Oshima T, Mori H, Maeda M, Wada C, Inokuchi H (2002) HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and *hem*K knockout induces defects in translational termination. Proc Natl Acad Sci USA 99:1473–1478

Narayan P, Rottman FM (1992) Methylation of mRNA. Adv Enzymol 65:255–285

Posfai J, Bhagwat AS, Posfai G, Roberts RJ (1989) Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 17:2421–2435

Robertson KD (2001) DNA methylation, methyltransferases, and cancer. Oncogene 20:3139–3155

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Scavetta RD, Thomas CB, Walsh MA, Szegedi S, Joachimiak A, Gumport RI, Churchill ME (2000) Structure of *RsrI* methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases. Nucleic Acids Res 28:3950–3961

Schibler U, Kelley DE, Perry RP (1977) Comparison of methylated sequences in messenger RNA and heterogenous nuclear RNA from mouse L cells. J Mol Biol 115:695–714

Schluckebier G, Labahn J, Granzin J, Saenger W (1998) M.*TaqI*: Possible catalysis via cation-pi interactions in N-specific DNA methyltransferases. Biol Chem 379:389–400

Schubert HL, Wilson KS, Raux E, Woodcock SC, Warren MJ (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. Nature Struct Biol 5:585–592

Schubot FD, Chen CJ, Rose JP, Dailey TA, Dailey HA, Wang BC (2001) Crystal structure of the transcription factor sc-mtTFB offers insights into mitochondrial transcription. Protein Sci 10:1980–1988

Shah JC, Clancy MJ (1992) IME4, a gene that mediates MAT and nutritional control of meiosis in *Saccharomyces cerevisiae*. Mol Cell Biol 12:1078–1086

Shimba S, Bokar JA, Rottman F, Reddy R (1995) Accurate and efficient N^6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro. Nucleic Acids Res 23:2421–2426

Siksnys V, Zareckaja N, Vaisvila R, Timinskas A, Stakenas P, Butkus V, Janulaitis A (1994) CAATTG-specific restriction-modification *Mun*I genes from *Mycoplasma*: Sequence similarities between R.*Mun*I and R.*Eco*RI. Gene 142:1–8

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17:355–362

Skinner MM, Puvathingal JM, Walter RL, Friedman AM (2000) Crystal structure of protein isoaspartyl methyltransferase. A catalyst for protein repair. Struct Fold Des 8:1189–1201

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Weiss VH, McBride AE, Soriano MA, Filman DJ, Silver PA, Hogle JM (2000) The structure and oligomerization of the yeast arginine methyltransferase, HMT1. Nat Struct Biol 7:1165–1171