SPOUT: a Class of Methyltransferases that Includes spoU and trmD RNA Methylase Superfamilies, and Novel Superfamilies of Predicted Prokaryotic RNA Methylases

Vivek Anantharaman, Eugene V. Koonin and L. Aravind*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Nucleotide modification is a major source of structural and biochemical diversity of RNAs, particularly of tRNAs and rRNAs (Rozenski et al., 1999). Base-specific methylation is one of the prevalent modes of post-transcriptional modification along with pseudouridylation and thio-uridylation. The catalytic domains of the majority of RNA methylases belong to the vast large superclass of Rossmann-fold enzymes (Lo Conte et al., 2000). These Rossmann-fold methylases (RFM) bind S-adenosyl-L-methionine (SAM) via the nucleotide-binding loop typical of this fold and catalyze the transfer of its methyl group to substrates such as DNA, RNA, proteins or small molecules. However, two families of RNA methylases, typified, respectively, by Escherichia coli proteins SpoU (TrmH) and TrmD, appear to be unrelated to the RFMs. SpoU family proteins have been characterized as tRNA (G\textsuperscript{18}) methyltransferases in E. coli and S. cerevisiae (Cavaille et al., 1999; Koonin et al., 1993; Persson et al., 1997), whereas TrmD is responsible for the bacteria-specific tRNA methylation at the G\textsuperscript{37} position (Bjork et al., 1989; Li et al., 1999). Here, we provide evidence for a common evolutionary origin of the TrmD and SpoU methylase superfamilies that were previously considered unrelated. We show that the SpoU superfamily encompasses a greater diversity of (predicted) methylases than previously appreciated and contains two previously undetected families that are specific to the archaea and thermophilic bacteria. Additionally, we identify two previously unnoticed superfamilies of proteins from archaea and bacteria, respectively, that are predicted to be methylases related to both the TrmD and SpoU superfamilies. These four superfamilies together define a new structural scaffold for RNA methylase activity.

As part of a systematic survey of the enzymes involved in RNA metabolism (VA, EVK and LA, unpublished), we investigated the SpoU and TrmD families by iterative searches of the Non-redundant database (National Center for Biotechnology Information, NIH, Bethesda) using the PSI-BLAST program (Altschul et al., 1997). Such a search (profile inclusion expectation (E) value threshold of 0.01) seeded with the E. coli SpoU sequence detected not only the obvious SpoU orthologs and previously described homologs (Koonin, 1996) from bacteria, archaea, and eukaryotes, but also several uncharacterized proteins from archaea (typified by MJ1385) and the thermophilic bacteria Aquifex and Thermotoga (typified by aq_054). These proteins showed significant E-values (10\textsuperscript{-3}–10\textsuperscript{-7}) on first detection and retained the conservation pattern typical of the SpoU proteins (Figure 1), suggesting that they were bona fide members of this superfamily.

Additionally, in these searches we detected several uncharacterized archaeal proteins, typified by AF2226, and the TrmD methylases with marginally significant E-values (.06–1). A reciprocal search started with the AF2226 sequence recovered the TrmD methylases and SpoU superfamily members with similar borderline E-values. Another confirmatory PSI-BLAST search initiated with the TrmD ortholog from Aquifex aeolicus (2983865) detected not only the AF2226-like archaeal proteins, but also members of a highly conserved superfamily of predominantly bacterial proteins, typified by YbeA from E. coli, with significant E-values (<0.001). The TrmD methylases, the SpoU methylases, and the newly detected AF2226-like, and YbeA-like proteins showed striking conservation in the region predicted to be the SAM-binding loop in the SpoU superfamily (Koonin, 1996) (Figure 1). Furthermore, in the archaeal AF2226-like proteins, the putative methylase domain is fused with a recently described RNA-binding domain, THUMP, that has also been found fused to classic, Rossmann-fold RNA methylases, thiouridine synthases and pseudouridine synthases (Aravind et al., 2001).

These observations suggested that the SpoU superfamily, the TrmD superfamily, and the newly detected AF2226-like and YbeA-like superfamilies, belong to a class of methyltransferases with a common structural fold that does not include any structurally characterized proteins and is distinct from the classic Rossmann fold. To investigate this relationship further, we searched the entire set of members of the SpoU, TrmD, YbeA and AF2226 superfamilies for potential conserved motifs using the Gibbs sampling procedure (Neuwald et al., 1997). Two motifs were detected, one corresponding to the predicted SAM-binding loop and the other located directly C-terminal of it, with the
Further Reading

Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at caister.com

- MALDI-TOF Mass Spectrometry in Microbiology  
  Edited by: M Kostrzewa, S Schubert (2016)  
  www.caister.com/malditof

- Aspergillus and Penicillium in the Post-genomic Era  
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)  
  www.caister.com/aspergillus2

- The Bacteriocins: Current Knowledge and Future Prospects  
  Edited by: RL Dorfl, SM Roy, MA Riley (2016)  
  www.caister.com/bacteriocins

- Omics in Plant Disease Resistance  
  Edited by: V Bhadauria (2016)  
  www.caister.com/opdr

- Acidophiles: Life in Extremely Acidic Environments  
  Edited by: R Quatrini, DB Johnson (2016)  
  www.caister.com/acidophiles

- Climate Change and Microbial Ecology: Current Research and Future Trends  
  Edited by: J Marxsen (2016)  
  www.caister.com/climate

- Biofilms in Bioremediation: Current Research and Emerging Technologies  
  Edited by: G Lear (2016)  
  www.caister.com/biorem

- Microalgae: Current Research and Applications  
  Edited by: MN Tsaiologlou (2016)  
  www.caister.com/microalgae

- Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives  
  Edited by: H Shintani, A Sakudo (2016)  
  www.caister.com/gasplasma

- Virus Evolution: Current Research and Future Directions  
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)  
  www.caister.com/virusevol

- Arboviruses: Molecular Biology, Evolution and Control  
  Edited by: N Vasilakis, DJ Gubler (2016)  
  www.caister.com/arbo

- Shigella: Molecular and Cellular Biology  
  Edited by: WD Picking, WL Picking (2016)  
  www.caister.com/shigella

- Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment  
  Edited by: AM Romani, H Guasch, MD Balaguer (2016)  
  www.caister.com/aquaticbiofilms

- Alphaviruses: Current Biology  
  Edited by: S Mahalingam, L Herrero, B Herring (2016)  
  www.caister.com/alphha

- Thermophilic Microorganisms  
  Edited by: F Li (2015)  
  www.caister.com/thermophile

- Flow Cytometry in Microbiology: Technology and Applications  
  Edited by: MG Wilkinson (2015)  
  www.caister.com/flow

- Probiotics and Prebiotics: Current Research and Future Trends  
  Edited by: K Venema, AP Carmo (2015)  
  www.caister.com/probiotics

- Epigenetics: Current Research and Emerging Trends  
  Edited by: BP Chadwick (2015)  
  www.caister.com/epigenetics2015

- Corynebacterium glutamicum: From Systems Biology to Biotechnological Applications  
  Edited by: A Burkovski (2015)  
  www.caister.com/cory2

- Advanced Vaccine Research Methods for the Decade of Vaccines  
  Edited by: F Bagnoli, R Rappuoli (2015)  
  www.caister.com/vaccines

- Antifungals: From Genomics to Resistance and the Development of Novel Agents  
  Edited by: AT Coste, P Vandeputte (2015)  
  www.caister.com/antifungals

- Bacteria-Plant Interactions: Advanced Research and Future Trends  
  www.caister.com/bacteria-plant

- Aeromonas  
  Edited by: J Graf (2015)  
  www.caister.com/aeromonas

- Antibiotics: Current Innovations and Future Trends  
  Edited by: S Sánchez, AL Demain (2015)  
  www.caister.com/antibiotics

- Leishmania: Current Biology and Control  
  Edited by: S Adak, R Datta (2015)  
  www.caister.com/leish2

- Acanthamoeba: Biology and Pathogenesis (2nd edition)  
  Author: NA Khan (2015)  
  www.caister.com/acanthamoeba2

- Microarrays: Current Technology, Innovations and Applications  
  Edited by: Z He (2014)  
  www.caister.com/microarrays2

- Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications  
  Edited by: D Marco (2014)  
  www.caister.com/n2
probability of occurring by chance <10^{-14}. A multiple alignment was constructed for each of these four groups using the T-Coffee program (Notredame et al., 2000) and the secondary structure was predicted separately for each group using the JPRED program (Cuff et al., 1999). The strong correlation of the structural elements in the individually predicted secondary structure elements for these four groups (Figure 1) also supported the prediction of a common structural fold for all these proteins. Hereinafter we refer to this monophyletic assemblage of proteins as the SPOUT (SpoU-TrmD) class.

An overall multiple alignment of the SPOUT proteins was constructed using the T-Coffee program and adjusted on the basis of the PSI-BLAST alignments and individual secondary structure predictions (Figure 1). Sequence conservation concentrates in the C-terminal region of the SPOUT domain where the predicted SAM-binding loop is located. This loop is predicted to have a strand-loop-helix configuration that resembles the SAM-binding loop in RFMs, but the SPOUT fold differs from the classic Rossmann fold in that the conserved loop is located near the C-terminus rather than at the N-terminus of the methylase domain. The overall secondary structure prediction for the SPOUT domain reveals a roughly alternating α-helix, β-strand pattern (Figure 1). Consistent with this, sequence-structure threading performed using the hybrid fold prediction method (Fischer, 2000) gave generic, moderate scoring hits to several 3-layered α/β proteins (data not shown). Thus, the SPOUT domain is predicted to have a 3-layered α/β fold with a central sheet of 5-6 strands sandwiched between helices on either side, with the SAM-binding site located in the C-terminal part of domain.

On the basis of sequence conservation, the SPOUT class of (predicted) methylases is divided into four distinct superfamilies, namely SpoU, TrmD, YbeA, and AF2226 (Figure 1, Table 1). Of these, TrmD is the YbeA superfamily includes highly conserved proteins that are present in a single copy in diverse bacteria (with the exception of several lineages such as Aquifex, cyanobacteria, actinomycetes and spirochaetes) and Arabidopsis thaliana.

The SpoU superfamily was divided into several ancient conserved families using clustering by sequence similarity with the BLASTCLUST program (I. Dondoshansky, Y. I. Wolf, and EVK, unpublished) (Table 1). Each of these families has a distinct phylogenetic pattern (Table 1). The majority of typical SpoU-like methylases belong to family 1, which consists of several orthologous groups such as spoU/Trm3, YifF, YifH, and YsgA, (Cavaille et al., 1999; Koonin, 1996; Persson et al., 1997), most of which show bacteria-eukaryotic distribution, with a single archaeal representative in A. fulgidus. In contrast, family 2 of SpoU-like methylases shows an archaeal-bacterial distribution. Thus, unlike many RNA-modifying and processing enzymes that are either archaeo-eukaryotic-specific or bacterial-specific, the SpoU methylases show the opposite phylogenetic trend (Aravind et al., 1999; Makarova et al., 1999). The underlying evolutionary scenario might involve acquisition of Family 1 SpoU methylases by eukaryotes from bacteria, which could have been accompanied by displacement of SpoU family members (probably of family 2) that the eukaryotes inherited from their common ancestor with the archaea. Several families of the SpoU superfamily show restricted phylogenetic distribution. Specifically, Family 3 is found only in the bacteria, Family 4 only in the archaea, and family 5 only in thermophilic bacteria (Table 1). It appears that members of this superfamily substantially contribute to the diversity of lineage-specific methylation patterns. The SpoU family 4 is one of the most divergent families of this superfamily and its representative from Thermoplasma is fused to a HD phosphoesterase domain (Aravind et al., 1998) that is highly conserved as a standalone protein throughout the archaea. Conceivably, the methylases of this family function together with HD hydrolases (possibly, RNAses) (Aravind et al., 1998) in a novel, archaea-specific RNA processing pathway.

The most parsimonious explanation of the phylogenetic distribution and evolutionary affinities of the SPOUT domains is that the last common ancestor of all life forms encoded a SpoU-like tRNA methylase. The TrmD, YbeA, and AF2226 superfamilies were subsequently derived through independent duplications; the former two in the bacterial and the latter in the archaea-eukaryotic lineages. The YbeA-like protein from A. thaliana is particularly closely related to its orthologs in Gram-positive bacteria, which suggests a horizontal transfer from this bacterial lineage to plants. The SpoU superfamily members underwent further
independent duplication in archaea and bacteria, giving rise to lineage-specific methyltransferase families. The eukaryotes appear to have acquired their SpoU family members horizontally from bacteria, which probably resulted in displacement of the ancestral archaeo-eukaryotic version.

The unification of the TrmD and SpoU superfamilies into the SPOUT class points to a major radiation of RNA-specific methyltransferases independent of the RFMs. The previously unnoticed SPOUT methylase families are expected to contribute substantially to the diversity of lineage-specific RNA modifications that might be necessary for particular microbial adaptations, for example, thermophily. The prediction of these novel methylases may help in experimental studies that will advance our understanding of these modifications.

References


Cavaille, J., Chetouani, F. and Bachelier, J.P. 1999. The yeast Saccharomyces cerevisiae YDL112w ORF encodes the putative 2’-O-ribose methyltransferase catalyzing the formation of Gm18 in tRNAs. RNA 5: 66–81.


Table 1. Sequence-based classification and phyletic distribution of SPOUT methyltransferases∗

<table>
<thead>
<tr>
<th>Orthologous sets</th>
<th>Bacteria</th>
<th>Archaea</th>
<th>Eukaryotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spou/Tm3p</td>
<td>Bb, Ec, Mt, Dr, Aae</td>
<td>–</td>
<td>At(2), Ce, Hs, Sc</td>
</tr>
<tr>
<td>YfH</td>
<td>All (except Ct)</td>
<td>Aa</td>
<td>At, Ce, Hs, Sc, Sp</td>
</tr>
<tr>
<td>YsgA</td>
<td>Ec, Hi</td>
<td>–</td>
<td>At(2), Dm, Hs</td>
</tr>
<tr>
<td>LAST/MJ1476</td>
<td>Ec(2), Hi, Pa, Xi, Nm, Mt, Ssp, Dr</td>
<td>Alla</td>
<td></td>
</tr>
<tr>
<td>YbK</td>
<td>Ct, Nm, Mlo(2), Ssp, Mt(2), Dr, Uu, Bs</td>
<td>–</td>
<td>At</td>
</tr>
<tr>
<td>MJ1385</td>
<td>–</td>
<td>All</td>
<td>Ta (SPOUT+HD)</td>
</tr>
<tr>
<td>AF2226- like Archaeal superfamily</td>
<td>–</td>
<td>Af, Ph, Mj, Ape</td>
<td>–</td>
</tr>
<tr>
<td>TrmD superfamily</td>
<td>All bacteria</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>YbeA superfamily</td>
<td>Cj, Hf, Ec, Hi, Pa, Xi, Nm, Mt, Dr, Bs, Uu, Tm</td>
<td>–</td>
<td>At</td>
</tr>
</tbody>
</table>

*The species abbreviations are as in Figure 1.

*a*Pyrococcus sp. has a member but horikoshii and abyssi do not.