Gene Expression Profile of RNA N1-methyladenosine methyltransferases

Mingzhu Rao¹

¹School of Pharmaceutical Science and Technology, Tianjin University, Tianjin China

Abstract. N1-methyladenosine (m1A) is a kind of common and abundant methylation modification in eukaryotic mRNA and long-chain non-coding RNA. Nucleoside methyltransferase (MTase) of m1A is a diverse protein family, which is characterized by the presence of methyltransferases like domains and conserved S-adenosylmethionine (SAM) binding domains formed by the central sevenstranded beta-sheet structure. However, comprehensive analysis of the gene expression profile of such enzymes has not been performed to classify them according to evolutionary criteria and to guide the functional prediction. Here, we conducted extensive searches of databases to collect all members of previously identified m1A RNA methyltransferases. And we report bioinformatics studies on gene expression profile based on evolutionary analysis, sequence alignment, expression in tissues and cells within the family of RNA methyltransferases. Our analysis showed that the base modification behavior mediated by m1A RNA methyltransferases evolved from invertebrate, and the active sites of m1A RNA methyltransferases were highly conserved during the evolution from invertebrates to human. And m1A RNA methyltransferases have low tissue and cell specificity.

1 Introduction

RNA is a key link between DNA and protein in the process of genetic information transmission, but the level of synthesized protein is not necessarily positively correlated with the level of mRNA, suggesting the importance of post-transcriptional modification of RNA. To date, more than 100 different types of base modification behavior have been identified. Internal modification occurs in many types of RNA, of which N6-methyladenosine (m6A), the most common modification in eukaryotes, refers to methylation of adenylate at the sixth nitrogen position. m1A modification is different from m6A in that it occurs on the first N atom of the adenine base and has a positive charge under physiological conditions. m1A was first found in non-coding RNA such as rRNA and tRNA and widely existed in prokaryotic and eukaryotic mRNA. Compared with m6A, the content of m1A in human and mouse tissues is very limited, and the content of m1A in mRNA is less than one-tenth of m6A, but it can greatly affect protein-RNA interaction and RNA secondary structure through electrostatic effect. In addition, m1A modification is related to translation initiation sites, and transcripts containing m1A have higher translation efficiency. m1A is a positively charged nucleotide modification, and the methyl group on m1A will block the Watson-Crick base complementary pairing of adenine, which also makes them have an important contribution to the formation of tRNA structure. In tRNA m1A is mainly present at position 58 of the tRNA, and it is also found in small amounts at positions 9, 14, 22, and 57 of the tRNA. Based on the latest single-base resolution m1A sequencing technology, the researchers found that m1A modification is mainly concentrated in mRNA cap and 5’UTR region, and m1A modification also exists in mitochondrial-encoded transcripts. Because m1A contains a methyl group at the N1 position that can interfere with Watson-Crick base pairs, m1A can cause reverse transcription to stop and can also lead to erroneous synthesis. Based on this, Dominissini and Li developed transcriptome-wide sequencing methods to identify and map m1A in mRNA. These studies indicate that m1A is reversible and is mainly concentrated near the start codon of eukaryotic mRNA. However, the binding proteins of m1A are still unclear, and its specific functions and mechanisms need to be further explored. Further identification of m1A binding proteins will help to understand the functional role of m1A. The discovery of m1A is a new RNA marker positively correlated with translation, adding an exciting new dimension to the regulation of post-transcriptional gene expression mediated by RNA modification.

There are a variety of RNA methyltransferases and demethylases in cells, as well as various methyl binding proteins. Under their combined action, different types of RNA undergo dynamic changes of methylation and demethylation to regulate various physiological processes. The number of studies about human RNA methyltransferases has been increasing rapidly in the past few years, and both structural characteristics and biological functions are the research hotspots. However, there are only few studies on the expression profile of RNA methyltransferases. Our project focuses on all m1A RNA methyltransferases. Based on the gene
expression level and sequence alignment of the enzymes obtained from some databases, the gene expression profile of m1A methyltransferase was studied by means of bioinformatics. To understand the origin of these enzymes in detail, we summarized the species without m1A methyltransferase orthologues together with the corresponding discussion of the gene sequences and structure to verify the evolution of RNA methylation modifications. Moreover, expression and localization of m1A RNA methyltransferases across tissues and cells better indicate the specificity of gene expression. The clustering analysis of m1A RNA methyltransferases expression profiles reveals their tissue and cell specificities. The results of the present studies provide a further understanding of the evolutionary profile of RNA modification and valuable insights into their functions.

2 Methods

2.1 m1A RNA Methyltransferases Searches

Human enzymes with RNA m1A methylation function were collected from the PubMed database of the National Center for Biotechnology Information (https://pubmed.ncbi.nlm.nih.gov/). All the results were used as queries to carry out the second round of searching, and the same procedure was followed for the retrieval of key residues of the RNA methyltransferases. The information on RNA methyltransferases’ active residues, substrates, and binding sites with ligands was collected from previously published sources.

2.2 Analysis of Species Tree

The species tree including 259 species from Ensembl database (http://asia.ensembl.org/info/about/speciestree.html, Ensembl Release 100) was downloaded for further analysis. Ensembl is a genome browser for vertebrate genomes. The species tree from Ensembl describes the evolutionary relationship of primates, rodents, laurasiatheria, afrotheria, xenarthra, birds, reptiles, amphibians, fish, cyclostomes, two chordates, and three other eukaryotes. And the species of m1A RNA methyltransferases that don't have any orthologue with the enzyme’s gene are labeled in the evolutionary tree. The species tree was displayed and edited by a tool called Interactive Tree of Life (iTOL).

2.3 Structure Analysis

Crystal structures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.4.1, Schrodinger, LLC). Structure of m1A RNA methyltransferase in complex with SAM has been deposited in the Protein Data Bank with PDB IDs: TRMT6-TRMT61A (5CCB).

2.4 Sequence Alignment

Twenty representative species were selected to do sequence alignment based on various stages of vertebrate evolution including Saccharomyces cerevisiae, Drosophila melanogaster, Caenorhabditis elegans, Ciona intestinalis, Ciona savignyi, Lamprey (Petromyzon marinus), Hagfish (Eptatretus burgeri), Elephant shark (Callorhinichus miliii), Coelacanth (Latimeria chalumnae), Zebrafish (Danio rerio), Tropical clawed frog (Xenopus tropicalis), Tuatara (Sphenodon punctatus), Mainland tiger snake (Notechis scutatus), Painted turtle (Chrysemys picta bellii), Australian saltwater crocodile (Crocodylus porosus), Platypus (Ornithorhynchus anatinus), Mouse (Mus musculus), Great Tit (Parus major), Sooty mangabey (Cercocebus atys), and Human (Homo sapiens). The protein sequences of the different m1A RNA methyltransferases from the above mentioned 20 species were downloaded from the Ensembl database. Multiple sequence alignments of the methyltransferases were generated using MEGA and displayed the results using ClustalX. Creates a multiple alignment of all database hits for each query sequence.

2.5 Analysis of Gene Expression in Tissues and Cells

To analyze the RNA expression of m1A RNA methyltransferases in different human cells and tissues, we used the RNA data from the Human Protein Atlas (HPA) database (Version: 19) which contains six atlas parts, among which the Tissue Atlas shows the expression of proteins across major tissues in the human body and the Cell Atlas shows expression and localization of proteins in single cells. Normalized eXpression (NX) values in major tissues and cell lines were used to perform the analysis, which was created by combining the data from the three transcriptomics datasets (HPA, GTEx, and FANTOM5). Heatmap visualization was implemented with Heml (Heatmap Illustrator, version 1.0) to compare the expression level of m1A RNA methyltransferases in tissues and cells. Heml clustering method is the default average linkage, and the similarity metric is calculated by the default Pearson distance.

3 Results and Discussion

3.1 The Properties of m1A RNA Methyltransferases

We found that only four enzymes were identified with m1A methylation activity for human RNA through information collection (Table 1). These enzymes, including TRMT6, TRMT61A, TRMT61B, and TRMT10C. TRMT61A and TRMT6 are responsible for N1-methyladenine at position 58 (m1A58) in cytoplasmic tRNAs. TRMT6 together with the TRMT61A catalytic subunit mediates methylation of adenosine residues at the N1 position of a small subset of mRNAs, and N1 methylation takes place in tRNA T-loop-like structures of mRNAs and is only present at low stoichiometries.
TRMT61B is a kind of methyltransferase that catalyzes the formation of N1-methyladenine at position 58 (mA58) in various tRNAs in the mitochondrion, including tRNA(Leu) (deciphering codons UUA or UUG), tRNA(Lys) and tRNA(Ser) (deciphering codons UCA, UCU, UCG or UCC)\(^{16}\). And TRMT61B catalyzes the formation of 1-methyladenosine at position 947 of mitochondrial 16S rRNA and this modification is most likely important for mitoribosomal structure and function\(^ {17} \). In addition to tRNA N1-methyltransferase activity, it also acts as an mRNA N1-methyltransferase by mediating methylation of adenosine residues at the N1 position of MT-ND5 mRNA, leading to interference with mitochondrial translation\(^ {7} \). TRMT10C is also mitochondrial tRNA N1-methyltransferase involved in mitochondrial tRNA maturation\(^ {18} \). Both TRMT61B and TRMT10C are mitochondrial RNA methyltransferases.

3.2 Origin and Evolution of m\(^{1} \)A RNA Methyltransferases

To investigate the evolutionary history of RNA demethylases and methyltransferases genes, we performed species evolution analysis on representative species. The evolution of vertebrates has gone through the evolutionary process of chordates, cyclostomes, fish (cartilaginous fish and bony fish), amphibians, reptiles, mammals, birds, and primates\(^ {19} \). ENSEMBL is not only a eukaryotic genome annotation project that focuses on vertebrate genomic data but also includes other organisms such as yeast, nematodes, and so on. The database contains three invertebrate eukaryotes, two invertebrate chordates, two cyclostomes, and around two hundred other vertebrates. In this study, we selected all the species in ENSEMBL mentioned above to construct the species trees (Figure 1). Some species don't have any orthologues with their targets. SAM-binding sites and RNA binding sites provide clues to the range of tRNA conformations recognized by other tRNA-modifying enzymes to access their targets.

3.3 Verification of Active Site

Twenty representative species are sorted in the sequence alignment according to the evolutionary process. In eukaryotes, the m\(^{1} \)A58 MTase located in the cytosol is composed of a catalytic protein unit from the TRM61 subfamily (TRMT61A) and an RNA-binding protein unit from the TRM6 subfamily (TRMT6). The eukaryotic complex of TRMT6–TRMT61A has been reported as a heterotetramer\(^ {20} \). Here we display the structure of human m\(^{1} \)A58 MTase located in the cytosol is shown in the figure. The only hydrogen-bonding side chains close to the target-binding site is Asp181, Asp181 is best positioned to interact with the tRNA backbone. Asp181 makes hydrogen bonds with residues from motifs. TRMT6 has an active site of TRMT61A is shown in Figure 2B. SAM-binding sites and RNA binding sites have been previously reported\(^ {21} \). The stereo plot of the active site of TRMT61A is shown in Figure 2B. SAM-binding MTases are characterized by several conserved sequence motifs, four of which line the SAM-binding site, and the cofactor-binding site is highly conserved in TRMT61A (Figure 2C). The TRMT61A subunit could be distinguished from TRMT6 by density for SAM in the cofactor site. In TRMT61A, SAM makes equivalent hydrogen bonds with residues from motifs. TRMT6 has an 83-residue insert in its N-terminal domain and a 98-residue insert in the middle of the C-terminal domain, both of which contribute to tRNA binding. The cofactor-binding pocket is occluded by inserted residue Arg377, which hydrogen bonds to the tRNA backbone. TRMT6 lacks the conserved sequence motifs that line the SAM-binding pocket in SAM-dependent MTases (Figure 2D). The conserved Asp181 in TRMT61A, which has been
implicated in catalysis, is an Ala in TRMT6. Thus, TRMT6 cannot have methylase activity and only has a tRNA-binding role. TRMT6 in Sphenodon punctatus lacks RNA binding sites, indicating that the species does not have m\(^1\)A catalytic activity.

Fig2. Conserved structures of the m\(^1\)A methyltransferases. A: TRMT6–TRMT61A complex bound to tRNA\(^{\text{Lys}}\) (PDB 5CCB); B: Binding site for SAM in the TRMT61A subunit; C: Sequence alignment of m\(^1\)A methyltransferase TRMT61A. The red box surrounding the aligned sequences identify conserved sequence motifs that surround the cofactor-binding site in SAM-dependent methyltransferases. The red star identifies the Asp that is important for catalysis; D: Sequence alignment of m\(^1\)A methyltransferase TRMT6. The red box surrounding the aligned sequences indicates the RNA-binding motif.
TRMT61B and TRMT10C are all mitochondrial tRNA methyltransferases. The identity of the enzyme catalyzing mitochondrial mRNA N1-methyltransferase is unclear. According to a report, mitochondrial mRNA N1-methyltransferase activity is catalyzed by TRMT61B. According to a second report, it is mediated by TRMT10C. As both reports only tested one protein (either TRMT61B or TRMT10C), it is possible that both proteins have this activity. Vertebrate TRMT61B proteins were predicted to localize to the mitochondria. TRMT61B catalyzed the formation of m'1A at position 58 of mitochondrial tRNA in a SAM-dependent manner. Sequence alignment of TRMT61B from different species shows that they possess several conserved motifs specific to SAM-dependent methyltransferases (Figure 3A), including the GxGxG sequence for SAM binding, and the conserved carboxylate motif for recognition of ribose hydroxyls. Each TRMT10C is also bound by a SAM molecule. The SAM pocket is enclosed by four extended loops highly conserved in sequence, harboring motifs I (aa 288–294), II (aa 307–323), III (aa 335–338), and IV (aa 350–360). Sequence alignment of TRMT10C shows that the four motifs corresponding to loops enclosing the SAM ligand-binding pocket are highly conservative in different species (Figure 3B).

3.4 RNA Expression Analysis in Tissues and Cells

The expression of RNA methyltransferases of m'1A in normal tissues and cells was reviewed through data retrieval in the Human Protein Atlas. Clustering of transcript data in tissues and cells are shown in Figure 4A and 4B, respectively. In the Human Protein Atlas the specificity category must meet the requirement that expression level in a particular tissue/cell type is at least four times than any other tissue/cell type. It can be seen from Figure 4 that m'1A RNA methyltransferases have low tissue specificity and cell specificity. Even though TRMT10C is highly expressed in Daudi cell lines, it still fails to meet the value requirements of cell specificity. Daudi is a kind of human lymphoma cell, but it is not clear whether the expression of TRMT10C is related to lymphoma.
4 Conclusion

We summarize the types of m1A RNA methyltransferases, all of which belong to TRMT family. Moreover, we showed the gene expression profile of the RNA methyltransferases based on species evolution and analysis of gene sequence. The extensive presence of TRMT6 and TRMT61A in early invertebrates and vertebrates suggests that the modification of m1A in RNA may have evolved from the invertebrate period. Active sites of m1A RNA methyltransferases are highly conserved, it is suggested that the m1A modification in RNA is highly conserved in different evolutionary stages. In addition, the expression of RNA methyltransferases of m1A in normal tissues and cells showed low specificity. In summary, our study improves the understanding about the gene expression profile of m1A RNA methyltransferases and provides valuable information for future studies.

ACKNOWLEDGMENTS

Financial support from National Natural Science Foundation of China (NSFC 21772143), the Natural Science Foundation of Tianjin (17JCYBJC42200, 18JCYBJC94000), Tianjin Youth 1000-Plan Talent Program and Startup Funding of Tianjin University is gratefully acknowledged.

REFERENCES


14. Sarah Ozanick, Annette Krecic, Joshua Andersland and James T Anderson (2005). The bipartite structure of the tRNA m1A58 methyltransferase from S. cerevisiae is conserved in humans. RNA, 11(8), 1281-1290.


20. Louis Droogmans, Martine Roovers, Janusz M. Bujnicki, Catherine Tricot, Thomas Hartsch, Victor Stalon and Henri Grosjean (2003). Cloning and characterization of tRNA (m1A58) methyltransferase (Trm1) from Thermus thermophilus HB27, a protein required for cell growth at extreme temperatures. Nucleic Acids Research, 31(8), 2148-2156.
