mRNA:guanine-N7 cap methyltransferases: identification of novel members of the family, evolutionary analysis, homology modeling, and analysis of sequence-structure-function relationships

ABSTRACT

Computational methods were used to infer the evolutionary relationships and predict the structure of Eukaryotic cap MTase. Identification of novel cap MTase homologs suggests candidates for cloning and biochemical characterization, while the structural model will be useful in designing new experiments to better understand the molecular function of cap MTases.

INTRODUCTION

Background Transcripts produced by RNA polymerase II are modified at their 5' end by the addition of a methylated 5'-terminal cap structure m7G(5')ppp(5')N, which directs pre-mRNA to the processing and transport pathways in the cell nucleus and regulates both mRNA turnover and the initiation of translation. Cap is formed by a series of three enzymatic reactions as follows: an RNA triphosphatase (TPase) removes the γ-phosphate at the 5' end of the transcript, a GTP:RNA guanylyltransferase (GTase) adds a GMP residue to the 5' diphosphate end in a 5'-to-5' orientation, and an RNA:guanine-N7 (m7G) methyltransferase (cap 0 MTase, for simplicity referred to hereafter as cap MTase) adds the methyl group to the guanine. Mutations in the TPase, GTase, or cap MTase of the yeast capping apparatus that inhibit any of these activities are lethal in vivo. The capping apparatus differs significantly in fungi, metazoans, protozoa and viruses in respect to the evolutionary origin and structure of individual subunits and the subunit composition of the proteins that carry the three activities. Hence, the capping enzymes encoded by viral, fungal and protozoal pathogens are attractive targets for specific inhibitors that would exert limited effect on the host enzyme. The mechanisms and structures of cellular and viral capping enzymes have been extensively studied. The crystal structures of the GTase from Chlorella virus PBVCV-1 and the TPase from yeast have been solved and used to guide extensive site-directed mutagenesis experiments. However, there are a few important gaps in our understanding of capping enzymes. For instance, there is a large body of mutagenesis data on cap MTase; however, its structure remains unknown. Therefore, many important details of the cap binding and m7G methyltransfer reaction mechanism remain unexplained. Cap MTase belongs to the AdoMet-dependent MTase superfamily, which contains numerous remotely related families of DNA, RNA, protein, and small molecule-modifying enzymes. To date, three-dimensional structures have been determined for more than a dozen MTases. The common fold of the catalytic domain, which bears the AdoMet binding site and the active site, has been identified (reviewed in). Despite low sequence similarity, the catalytic domains of typical MTases display a common tertiary architecture, similar to the Rossmann-fold, but with a unique peripheral β-hairpin structure instead of a typical right-handed β-α turn. Another characteristic feature of many MTase families is the presence of an additional "variable" domain, which is primarily responsible for substrate recognition and binding. This domain has been initially characterized in DNA:cytosine-C5 (m5C) MTases and dubbed TRD (for target recognition domain). More recently, it was determined that the majority of TRDs of individual MTase families are unrelated. They occur in different locations in the primary structure of the protein and fold into different structures, suggesting that they have originated from independent gene fusions (. Nevertheless, it has been shown that the TRDs of m5C MTases are structurally similar, even

though only several common residues could be delineated in their sequences that are critical for stability of the hydrophobic core and interactions of the TRD with the substrate. Moreover, based on the sequence-to-structure threading, it has been predicted that the TRDs of type I DNA MTases (a subclass of enzymes that modify adenine in DNA) share the common fold with the TRD of m5C MTases. This prediction has been later supported by mutagenesis studies. Therefore, aside from the structural and evolutionary diversity among TRDs, some MTase families may share conserved homology in the catalytic and substrate binding domains, even though their sequences seem dissimilar. The prolonged unavailability of the atomic structure of cap MTase prompted us to predict its structure and construct a three-dimensional model, which is accompanied by an evolutionary study. The results form this report should aid in the interpretation and design of mutagenesis experiments and provide a framework for comparative sequence-structure-function analysis of members of the MTase family. Cap MTases exhibited limited similarities to other MTases in the common AdoMet-binding region, and the substrate-binding site could not be unambiguously identified, based on sequence analysis and mutagenesis results. Therefore, we resorted to the sequence-to-structure threading method to find a structural template for homology modeling. We report here that cap MTases are related in structure to the glycine N-MTase. In addition, we carried out extensive database searches to identify novel genes that exhibit homology to known cap MTases, which may encode yet unidentified RNA modification enzymes.

CONCLUSION

Conclusions In this report, we used computational methods to infer the evolutionary relationships and predict the structure of cap MTase. A tertiary model has been built for the Eukaryotic enzyme and used to interpret the available mutation data and guide the comparative sequence analysis. We propose that cap MTases share the catalytic domain and the "S" domain with glycine N-MTases, which raises the possibility that these two families of N-MTases are relatively closely related. Moreover, we have identified a novel family of putative MTases that are specific to green plants and share structure and mechanism with cap MTases. Therefore, the alignment presented in this work will be a good starting point for further analysis of other N-MTase subfamilies that may share the "molecular basket" structure. Our analysis of the AdoMet-binding site in cap MTases, combined with evolutionary considerations, highlighted a case of correlated mutation in viral enzymes, which may be important for design of specific antivirals. We also used the model to predict the guanine binding site and identify conserved residues that may serve catalytic or structural function, which can be tested by site-directed mutagenesis. A putative non-specific mRNA binding patch was also proposed. Prior to the experimental solution of the structure of cap MTase, our model will be useful in designing new experiments to better understand the molecular function of cap MTases, whereas the identification of a novel family of genes will aid in identifying candidates for cloning and biochemical characterization. We hope that the prediction of numerous structural and functional features presented in this paper will advance these studies.