
RNA capping: progress and prospects

STEWART SHUMAN

Molecular Biology Program, Sloan Kettering Institute, New York, New York 10065, USA

The 5' m⁷GpppN cap, co-discovered by Shatkin, Furuichi, and Moss in 1975, is the signature feature of eukaryal cellular and viral messenger RNA that confers mRNA stability and efficient translation. Cap formation entails three sequential enzymatic modifications targeted to nascent pre-mRNAs synthesized by cellular or viral RNA polymerases. First, the 5' triphosphate end of the pre-mRNA is hydrolyzed to a diphosphate by RNA 5' triphosphatase (RTPase). Second, the diphosphate RNA is capped with GMP by RNA guanylyl-transferase (GTase) via a two-step mechanism: (i) reaction of GTase with GTP to form a covalent enzyme-(lysyl-N ζ)-GMP intermediate and PP_i; and (ii) transfer of GMP from GTase to the ppRNA end to form GpppRNA. Third, the GpppRNA cap is converted to m⁷GpppRNA by AdoMet:RNA(guanine-N7)-methyltransferase (MTase). This pathway was elucidated between 1975 and 1984 via the analysis of the purified vaccinia virus capping enzyme, a heterodimer of 97 kDa and 33 kDa subunits that catalyzes all three steps in cap formation. The same biochemical pathway (though not the same organization of the capping apparatus) is conserved in all eukaryal taxa.

Studies of capping enzymes have transited many stages, especially as new enabling methods emerged and new investigators joined the fray. Quanta of progress include: (i) purification of vaccinia, mammalian, and yeast capping enzymes from their native sources and characterization of their physical and functional properties; (ii) delineation of the domain organization of recombinant vaccinia capping enzyme and location of the RTPase, GTase and MTase active sites via mutagenesis and crosslinking to substrates; (iii) using these functional data to identify, clone and characterize cap-forming enzymes from DNA virus (baculovirus, *Chlorella* virus, mimivirus) and cellular (metazoan, fungal, and protozoan) sources spanning a wide evolutionary spectrum; (iv) obtaining atomic structures of exemplary RTPase, GTase, and MTase enzymes via X-ray crystallography; and (v) dissecting the physical and functional interactions of the poxvirus capping enzyme with the viral transcription machinery and of the cellular capping enzymes with RNA polymerase II. The great strides made on these fronts owes to the efforts of many PIs,

whom I cite here, in alphabetical order: David Bentley, Stephen Buratowski, Stephen Cusack, Chris Lima, Kiyohisa Mizumoto, Alfonso Mondragon, Bernard Moss, Edward Niles, Beate Schwer, Aaron Shatkin, Dale Wigley.

Henceforth, I will briefly review, from a purely personal and slanted perspective, one facet of the mRNA capping story that unfolded in the RNA era: the structural biology of the capping apparatus and what it tells us about the evolution of an uniquely eukaryal RNA modification.

If asked 20 years ago which of the capping enzymes would be most interesting from a structural standpoint, I would have replied instantly that the GTase is the one, what with its ping-pong mechanism and covalent enzyme-GMP phosphoramidate intermediate. I had described these features of the GTase mechanism when I was a graduate student (in 1981) and noted then the similarity of capping to the chemistry of covalent catalysis by DNA ligases, via enzyme-(lysyl)-AMP and AppDNA intermediates. In the early 1990s, my lab and others mapped the lysine sites of covalent GMP attachment to the vaccinia and fungal GTases; the lysine nucleophile resided within a peptide motif (KxDG) also found at the sites of covalent AMP attachment to DNA ligases. By visual inspection of the primary structures of then-known GTases and polynucleotide ligases, I discerned a collinear ensemble of conserved peptide motifs, hypothesized that these motifs comprised the active site of nucleotidyl transfer of capping enzymes and DNA ligases, and posited a shared structural basis and evolutionary ancestry for capping and ligation. The landmark crystal structures from the Wigley lab—of a T7 DNA ligase•ATP complex (1996) and *Chlorella* virus capping enzyme•GTP and covalent enzyme–GMP complexes (1997)—showed that DNA ligase and capping enzyme share a tertiary structure, composed of an N-terminal nucleotidyl-transferase domain and a C-terminal OB-fold domain, and that the active site is indeed composed of the conserved peptide motifs. Comprehensive mutational analyses, together with crystal structures of yeast, mammalian, and poxvirus GTases and bacterial, viral and human DNA ligases, have confirmed their conserved NTase-OB domain organization and active site architectures. Our hypothesis that RNA ligases also share a structural and evolutionary history with capping

Corresponding author: s-shuman@ski.mskcc.org

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.049973.115>. Freely available online through the RNA Open Access option.

© 2015 Shuman This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

enzymes and DNA ligases was affirmed in due course by crystallography and mutational analysis, the salient theme being that RNA ligases have an N-terminal nucleotidyltransferase domain, but they lack the OB domain characteristic of capping enzymes and DNA ligases.

Notwithstanding the happy outcomes of GTase research, my answer to the “most interesting structure” question was badly off the mark. The prize goes to the RNA triphosphatase enzyme, about which there was no molecular information known 20 years ago. The hydrolysis of a β - γ phosphoanhydride in a 5' NTP is such a ubiquitous chemical reaction that it was hard for me to imagine that Nature would re-invent the wheel to accomplish this seemingly pedestrian step in cap synthesis. But there was reason to think that the reaction was not so simple, insofar as there were two biochemically distinct classes of RTPase that differed in their reliance on a metal cofactor. Whereas vaccinia and yeast RTPase required a metal, the metazoan RTPases did not. Although the vaccinia RTPase was known by then to reside within a 60 kDa polypeptide fragment, there were no instructive sequence similarities. The key advance was the cloning of the essential *S. cerevisiae* gene encoding the RTPase Cet1 in 1997 by the Mizumoto lab. Armed with this information, we quickly identified three conserved collinear motifs, containing amino acids essential for catalysis, in the metal-dependent RTPases of yeast and DNA viruses; we proposed that these RTPases defined a novel family of metal-dependent phosphohydrolases with a shared active site. After lots of slicing and dicing to delineate a minimized biologically active domain, we (i.e., my lab colleague LiKai Wang) succeeded in growing crystals of Cet1. Despite LiKai's protestations all the while that the crystals were too small and therefore “no good,” we schlepped them across York Avenue to Chris Lima, then at Cornell Medical College, who proceeded to collect a 2.1 Å data set and to solve the structure of this truly slick protein.

Cet1 is a homodimer with a fold that, at first glance, looks like a pair of binoculars. The homodimer consists of two parallel topologically closed 8-strand antiparallel β -barrels (the triphosphatase tunnel) that rest on a predominantly α -helical globular pedestal at the homodimer interface. The active site in the center of the tunnel comprises a large ensemble of essential basic and acidic amino acids that emanate from the strands to coordinate the γ phosphate and the manganese cofactor and to activate a nucleophilic water. Successive structures of mimivirus and vaccinia RTPases and fission yeast *S. pombe* RTPase revealed that they too are members of the “triphosphate tunnel metalloenzyme” (TTM) family, defined originally by Cet1. Indeed, all fungal and protozoan RTPases (except for *Trichomonas*) belong to the TTM family.

By contrast, as first reported by Buratowski, the metazoan RNA triphosphatase domain of a bifunctional triphosphatase-guanylyltransferase enzyme belongs to the cysteinyl-phosphatase superfamily that includes familiar tyrosine-specific and dual-specificity protein phosphatases and lipid phosphatases. Biochemical studies and crystal structures of

the mammalian RNA triphosphatase and a baculovirus homolog (via the Mondragon lab in collaboration with my lab) revealed distinctive properties of the RNA 5'-phosphatase branch of the cysteinyl-phosphatase superfamily.

It initially appeared that Nature chose to re-purpose the cysteinyl-phosphatase fold for RNA 5' phosphate hydrolysis in metazoan (and plant) mRNA capping, but invented the TTM fold as an entirely new structural and mechanistic solution to the same chemical reaction in viruses, fungi and protozoa. This highlighted the TTM-family RNA triphosphatases as promising targets for anti-infective drug discovery. As it turns out, TTMs embrace more than just capping RTPases. They include a wide range of metal-dependent phosphohydrolase/transferase enzymes in bacterial, archaeal, and eukaryal taxa that act on such substrates as NTPs, thiamine triphosphate, and inorganic polyphosphates. Just goes to prove that Nature doesn't let a good idea languish narrowly.

I will end by opining that the canonical mRNA capping pathway is but one example of Nature's capacity to regulate nucleic acid function and fate via the installation of an end-blocking covalent modification. Recent studies highlight the existence of non-canonical RNA 5' cap structures and/or non-canonical cap-forming enzymes, the upshot of which is that end-blocking caps are a deeply rooted and widespread phenomenon about which we need to know much more. A few bullet points on this theme convey what I think are exciting areas for exploration and food for thought.

- *Recapping of 5'-monophosphate ends with m⁷G*. Schoenberg has shown that endonucleolytically cleaved RNAs with 5'-monophosphate ends can be re-capped with m⁷G by an enzyme system in the mammalian cytoplasm. This requires a novel RNA kinase activity, encoded by a gene yet to be defined, that converts pRNA to ppRNA, which can then be capped by a cytoplasmic form of the mammalian GTase. The structure and mechanism of the pRNA kinase are of acute interest, as is the prospect that similar or distinctive re-capping systems exist in other eukarya. Re-capping implies a vast new potential for expanding the information content of genes, by post-transcriptional re-sets of stable, translatable mRNA ends.
- *AMP capping of 5'-monophosphate ends*. RNA 3'-phosphate cyclase (Rtc) is a widely distributed end-modifying enzyme, originally studied by Filipowicz, that catalyses ATP-dependent conversion of RNA 3'-monophosphate ends to 2',3'-cyclic phosphate ends via covalent Rtc-(histidinyl)-AMP and RNA_(3')pp_(5')A intermediates. We showed that RtcA is also adept at transferring AMP to RNA 5'-monophosphate termini to form “A-capped” AppRNA products. Which is the real physiological reaction of Rtc enzymes? Or are both biologically relevant, perhaps to different extents in different contexts? (We find that many classic ATP-dependent RNA ligases are also quite adept at forming AppRNA caps as end-products.) AMP capping would protect the

RNA from 5'-monophosphate triggered decay. Are there proteins that recognize A-caps and target the AppRNA for specific transactions, akin to recognition of m⁷G caps by eIF4E? Are there enzymes that further modify A-caps, e.g., by adenine N6-methylation or 2'-O-methylation?

- *Methylphosphate capping*. The γ -mono-methyl-phosphate cap structure, discovered in human U6 sRNA by Reddy in 1989, is formed by methyl transfer from AdoMet to a γ -phosphate oxygen of the initiating nucleotide of the transcript. In 2007 Coulombe identified human BCDIN3 as the relevant methyltransferase enzyme, which also modifies 7SK RNA. In 2012, Kouzarides identified a paralogous human enzyme BCDIN3D that catalyzes two methyl transfers from AdoMet to the 5'-monophosphate oxygens of pre-miRNA-145; this dimethyl-phosphate modification negatively regulates miRNA maturation. What dictates the distinctive RNA end specificities and reaction outcomes of these phosphate methylation enzymes? Is this just the tip

of the iceberg on the variety of RNA phosphate methylation reactions?

- *5' capping with NAD⁺*. Mass spec analysis of *E. coli* RNA by the Liu lab in 2009 revealed the existence of 5' NAD⁺ capped RNAs, at an estimated level of 3000 copies per cell. The Jäschke lab recently isolated and sequenced the *E. coli* NAD⁺ capped RNAs, which were enriched for certain regulatory small RNAs (sRNAs) and 5' fragments of mRNAs. The biochemical pathway of NAD⁺ capping is presently *tabula rasa*. Liu weighed in against a mechanism of NAD⁺ priming of transcription initiation by bacterial RNA polymerase. My bet is that NAD⁺ caps are formed by attack of the phosphate of nicotinamide mononucleotide on the α -phosphorus of de novo initiated 5' adenosine triphosphate-terminated RNA, yielding a 5'-5' pyrophosphate bridged nicotinamide adenine dinucleotide capped RNA and expelling inorganic pyrophosphate: $(\text{Nic})\text{p} + \text{pppA}(\text{pX})_n \Rightarrow (\text{Nic})\text{ppA}(\text{pX})_n + \text{PP}_i$.