

## Review

### Discovery of m<sup>7</sup>G-cap in eukaryotic mRNAs

By Yasuhiro FURUICHI<sup>\*1,†</sup>

(Communicated by Takao SEKIYA, M.J.A.)

**Abstract:** Terminal structure analysis of an insect cytoplasmic polyhedrosis virus (CPV) genome RNA in the early 1970s at the National Institute of Genetics in Japan yielded a 2'-O-methylated nucleotide in the 5' end of double-stranded RNA genome. This finding prompted me to add S-adenosyl-L-methionine, a natural methylation donor, to the *in vitro* transcription reaction of viruses that contain RNA polymerase. This effort resulted in unprecedented mRNA synthesis that generates a unique blocked and methylated 5' terminal structure (referred later to as "cap" or "m<sup>7</sup>G-cap") in the transcription of silkworm CPV and human reovirus and vaccinia viruses that contain RNA polymerase in virus particles. Initial studies with viruses paved the way to discover the 5'-cap m<sup>7</sup>GpppNm structure present generally in cellular mRNAs of eukaryotes. I participated in those studies and was able to explain the pathway of cap synthesis and the significance of the 5' cap (and capping) in gene expression processes, including transcription and protein synthesis. In this review article I concentrate on the description of these initial studies that eventually led us to a new paradigm of mRNA capping.

**Keywords:** mRNA capping, methylation-coupled transcription, 7-methyl-guanosine, mechanism of cap formation, cap dependent protein synthesis, cap snatching by influenza virus

#### I. Introduction

Cap structures of the type m<sup>7</sup>GpppNmpNm are at the 5' ends of nearly all eukaryotic cellular and viral mRNAs (Fig. 1).<sup>1)</sup>

The cap is added to cellular mRNA precursors made by RNA polymerase II and to transcripts of viruses that replicate in the nucleus. It is added during the initial phases of transcription and before other RNA processing events, including internal N<sup>6</sup>A methylation, 3'-poly(A) addition and exon splicing.<sup>2),3)</sup> Most capped mRNAs have a single methyl group on the terminal G residue at the N-7 position, whereas adjacent nucleotides can be 2'-O-methylated to different extents, providing a basis for cap nomenclature: m<sup>7</sup>GpppN (Cap 0), m<sup>7</sup>GpppNm (Cap 1), and m<sup>7</sup>GpppNmpNm (Cap 2).<sup>4)</sup> Methyl groups in Cap 0 and Cap 1 structures are added in the nucleus; additional 2'-O-methylation in Cap 2 is a cytoplasmic event. In many lower eukaryotes,

including yeast, mRNAs contain mainly Cap 0, and higher organisms usually have more extensively methylated caps.<sup>1)-3)</sup> Despite variations in the extent of methylation important biological consequences of a cap structure appear to correlate with the N<sup>7</sup>-methyl on the 5'-terminal G and the two pyrophosphoryl bonds that connect m<sup>7</sup>G in a 5'-5' linkage to the first nucleotide of mRNA. Caps increase mRNA stability by protecting against 5'→3' exonucleolytic degradation.<sup>5)-7)</sup> Splicing accuracy and efficiency both increase by the presence of 5'-terminal m<sup>7</sup>GpppN, possibly because of the participation of nuclear cap-binding protein(s) analogous to the cytoplasmic cap-binding initiation factor required for eukaryotic protein synthesis.<sup>8)-10)</sup> The 5'-terminal m<sup>7</sup>GpppN may also participate in other critical steps in gene expression, including transport from nucleus to cytoplasm.

The 5' cap and 3'-poly (A) are hallmark structures of eukaryotic mRNAs and are not present in rRNA, tRNA or prokaryotic mRNAs. These two structures have been the subjects of many review articles on their biological significance. Since their discovery in the early 1970s, the 5'- and 3'-terminal

<sup>\*1</sup> GeneCare Research Institute Co., Ltd., Kanagawa, Japan.

<sup>†</sup> Correspondence should be addressed: Y. Furuichi, GeneCare Research Institute Co., Ltd., 19-2 Kajiwara, Kamakura, Kanagawa 247-0063, Japan (e-mail: furuichi@genecare.co.jp).

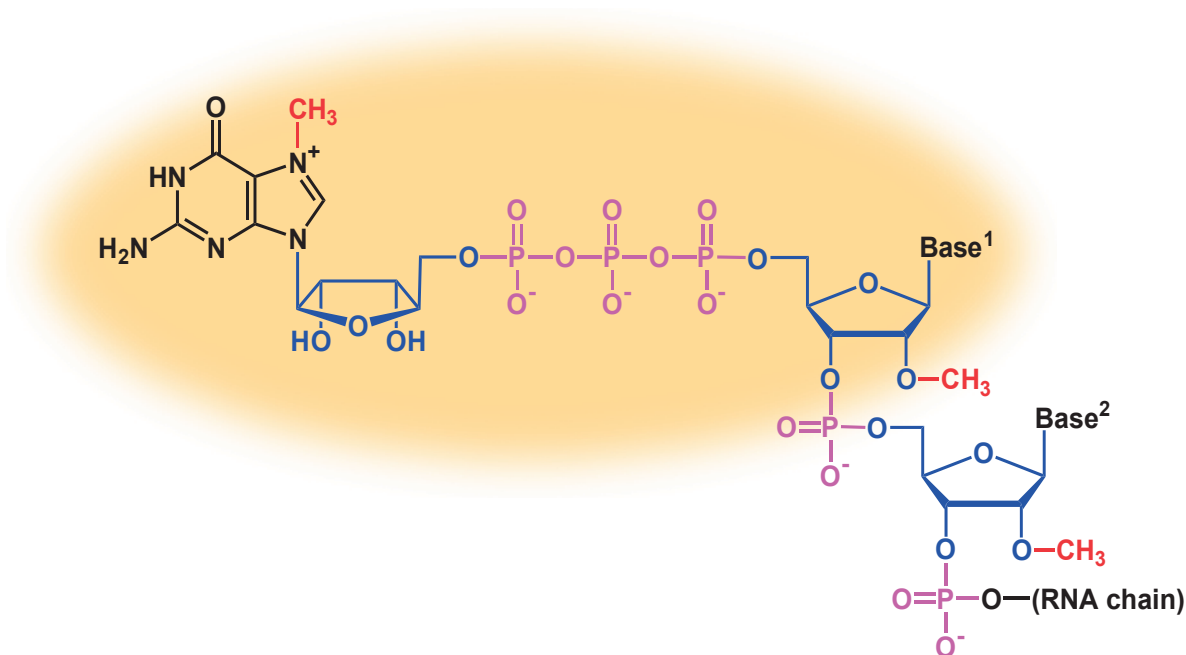


Fig. 1. Cap structure. m<sup>7</sup>GpppNm pN<sup>m</sup>- (Cap 2) representing the m<sup>7</sup>G linked to the 5'-end of the primary transcript via a 5'-5' triphosphate.

elements have contributed to the progress of molecular biology in many important ways including their utilization to distinguish and isolate various eukaryotic mRNAs. This article describes how the cap structure m<sup>7</sup>GpppNm was discovered, how the cap is synthesized, and what kind of biological functions the cap has in eukaryotic cellular events.

## II. Historical view of the discovery of cap structure

**A. Prologue.** Before the identification of the cap in several viral mRNAs in 1975, the 5' ends of all eukaryotic mRNAs, like those of *Escherichia coli* and bacteriophage mRNA, were generally believed to be triphosphorylated pppN forms, although no eukaryotic mRNA had been reported as 5' end-labeled by the standard, polynucleotide kinase method. Large-scale isolation of mRNA from eukaryotic cells was technically difficult, and the prevailing biochemical studies on *E. coli* and its phage had clearly shown that the mRNAs were triphosphorylated. *E. coli* mRNA contained pppA and pppG 5' ends, and all RNA phages, such as MS2, Q-beta, and R17 contained a 5'-triphosphorylated purine nucleoside in the genomic positive-strand RNA (and thus mRNA) extracted from purified particles. In contrast, rRNAs and tRNAs, which are abundant in both eukaryotic and bacterial cells, had been shown

to contain monophosphorylated 5' ends.<sup>11)</sup> Modified methylated nucleotides were believed to be present only in tRNA and rRNA. Here, like other important biological findings made in the 1970s, including 3'-polyadenylation, splicing, and reverse transcription, viral systems had a pivotal role in defining the structure of the cap and the biological significance of the cap in the 5' terminus of mRNA.

**B. Stage 1: Identification 2'-O-methylated nucleotides at the 5' ends of CPV genomes.** At the National Institute of Genetics in Japan, I and Kin'ichiro Miura, a senior co-researcher at the institute, tried to define the 5'- and 3'-terminal sequences of dsRNA of cytoplasmic polyhedrosis virus (CPV) that infects silkworm *Bombyx mori*. Our initial interest was to understand how 10 genomic segments of CPV could be correctly arranged within the viral particle (Fig. 2A).<sup>12),13)</sup> We divided the experiments in half: 3' sequencing by myself and 5' sequencing by Miura. The 3' sequencing was done by using a classical method that contains a series of chemical reactions with a stepwise reductive labeling of 3'-terminal nucleotides by <sup>3</sup>H-borohydride after converting the *cis*-diol ribose moiety to two aldehyde residues by periodate-mediated oxidation (Fig. 2B). The <sup>3</sup>H-labeled 3'-terminal nucleoside was cleaved by ribonuclease and was analyzed by two dimensional paper chromatog-

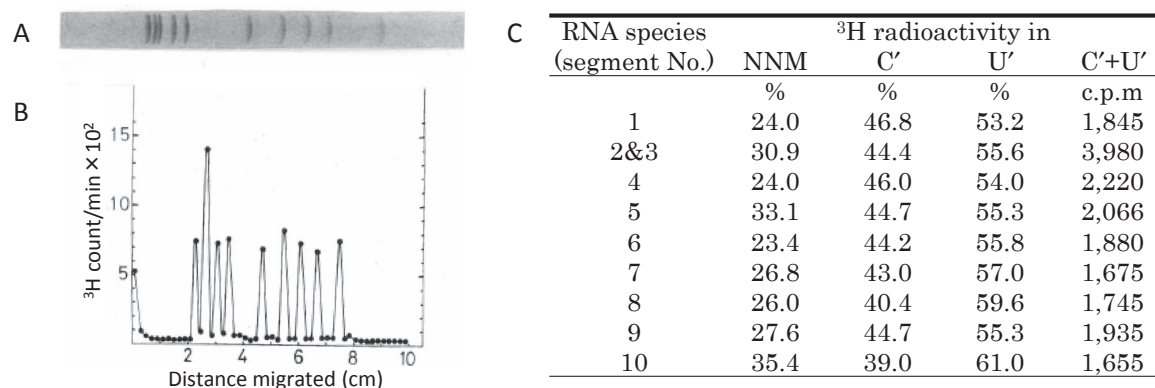
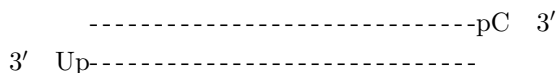


Fig. 2. CPV dsRNA genome segments and the 3'-terminal analysis. A: Polyacrylamide gel electrophoresis of CPV genome dsRNA segments. B: 3' <sup>3</sup>H-labeling of CPV genome dsRNAs by periodate oxidation followed by [<sup>3</sup>H]-NaBH<sub>4</sub>. C: Distribution of <sup>3</sup>H-radioactivity in 3'-terminal nucleosides and NNM (non-nucleoside material). Parts of the data are reprinted from the paper by Furuichi and Miura (1973).<sup>13)</sup>

raphy that can identify nucleoside derivatives (A', C', G' and U') with modified ribose residues. The method also can label the penultimate nucleotide after removal of the 3' terminal nucleotide by  $\beta$ -elimination reaction with aniline and subsequent removal of 3'-phosphate by alkaline phosphatase. These efforts yielded identical terminal sequences of



for each dsRNA segment that was separated by polyacrylamide gel electrophoresis (Fig. 2A).<sup>12)</sup> However, unexpected was that nearly 30% of the <sup>3</sup>H radioactivity was always incorporated into an undefined component of dsRNA, referred to as non-nucleoside material (NNM), while the remaining <sup>3</sup>H counts were evenly distributed in the two terminal nucleosides C' and U' that were identified after digestion of labeled RNA by ribonucleases (Fig. 2B and C).<sup>13)</sup>

The NNM could be ignored as rubbish, but I analyzed intensively the property of NNM. The NNM (which we later proved to be Cap 1) was resistant to treatments with alkali, protease, DNase, and ribonucleases, but was sensitive to snake venom phosphodiesterase as characterized by paper chromatography. It had a net negative charge of -4.5 at neutral pH and a size of 5.4 Å as determined by DEAE-cellulose column chromatography in the presence of 7M urea.<sup>13)</sup>

Miura, however, was faced with the difficulty of labeling the 5' terminal nucleotides, possibly G and A, based on the predicted end-to-end duplex structure of CPV dsRNA, by treatment with polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP.<sup>14)</sup> However, this was achieved by dsRNA, whose 3'-terminal nucleotides

C and U were removed by sequential oxidation (by sodium periodate),  $\beta$ -elimination (by aniline), and dephosphorylation (by alkaline phosphatase). The processed RNA yielded a clear-cut result of 3'-terminal C, for both strands, in the penultimate positions to the terminal C and U, indicating that the 3'-terminal sequences of CPV RNA are —pCpC and —pCpU. Curiously, the NNM that was labeled by <sup>3</sup>H in the first round 3' terminal nucleoside analysis was eliminated after the second <sup>3</sup>H-labeling to the penultimate C. The dsRNA that was shortened by the removal of terminal nucleoside C and U provided a far better substrate for 5'-<sup>32</sup>P-labeling by polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]-labeled ATP than the unprocessed dsRNA.

*The results have remained unclear until the end-to-end base paired CPV dsRNA genome structure was clarified later and one of the 5' termini was found to be protected by m<sup>7</sup>G-cap. The capped 5' end provides an additional 2'-3' cis-diol moiety of ribose, besides two 3' terminal riboses, all of which are susceptible to periodate oxidation and  $\beta$ -elimination reaction by aniline.*

*It was indeed enigmatic incidents at the time, however, as it turned out, the <sup>3</sup>H-labeling by NaBH<sub>4</sub> took place on the ribose moiety of m<sup>7</sup>G-cap to yield <sup>3</sup>H-NNM and the removal of cap and 3' terminal nucleotides converted the 5' ends of CPV dsRNA to an one-base-overhanged structure, which facilitated polynucleotide kinase to phosphorylate 5' pA<sup>m</sup> and pG (Fig. 3).*

*The Fig. 3 depicts the reaction scheme employed for analysis of terminal structure of CPV RNA and retrospective explanation for the outcome of individual reactions.*

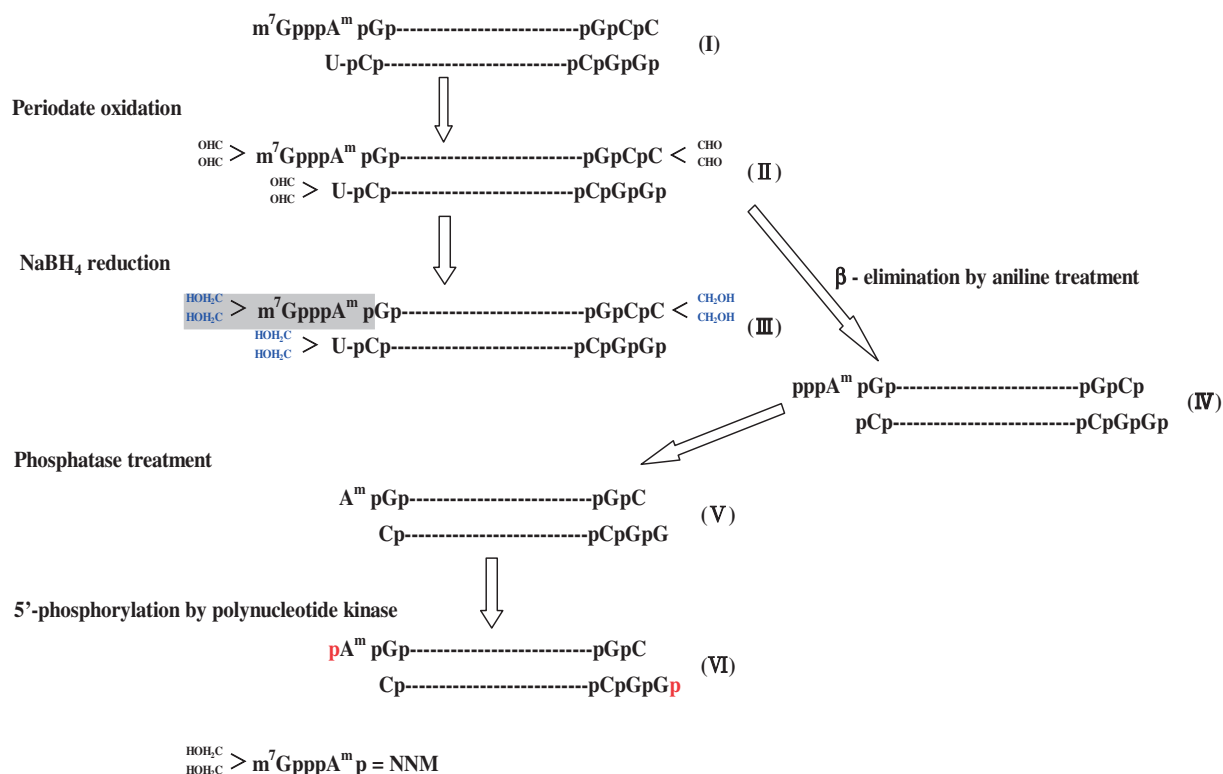
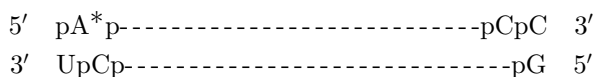


Fig. 3. Reaction scheme employed for analysis of 5' and 3' terminal structures of CPV genome RNA and retrospective identification of NNM and cap structure. (I): CPV genome RNA. (II): Periodate oxidized RNA, 2' and 3' OH were oxidized to aldehydes. (III): [<sup>3</sup>H]-NaBH<sub>4</sub> reduced RNA which contains <sup>3</sup>H labeled-CH<sub>2</sub>OH (shown by blue). The shadowed structure was turned out to be NNM later. (IV): CPV RNA produced from (III) by aniline treatment which removes the ribose-oxidized nucleoside. (V): Alkaline phosphatase-treated CPV RNA (IV), which is devoid of sensitive phosphates. (VI): 5'-<sup>32</sup>P-labeled CPV RNA which was phosphorylated by polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP (the radioactive phosphates are shown by red).

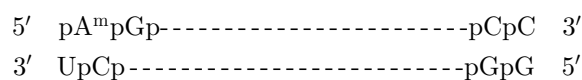
Subsequently, the 5'-nucleotides of CPV RNA were found to have 5' pG and 5' pA that made a perfect match with the complementary strand in the form:



However, Miura was puzzled by the slight difference in mobility of <sup>32</sup>P-labelled pA\* compared with authentic pA in the two-dimensional paper chromatography system used to identify the <sup>32</sup>P-labelled 5'-nucleotides: the pA\* obtained by P1 nuclease digestion migrated slightly faster than the authentic marker pA. Further analyses of 5' oligonucleotides resulting from the 5'-<sup>32</sup>P-labelled RNA after digestion by guanine-specific RNase T<sub>1</sub>, pyrimidine-specific pancreatic ribonuclease A or nonspecific ribonuclease T<sub>2</sub> indicated that pG was the penultimate nucleotide to both 5' pA\* and 5' pG, establishing pA\*pG and pGpG as the 5' sequences of CPV dsRNA segment. More importantly, the results showed that

the 2'OH of the ribose in pA\* of pA\*pG was modified because the phosphodiester linkage between pA\* and pG could not be digested by RNase T<sub>2</sub>, which requires a free OH for 2', 3'-cyclic phosphate formation as an RNA hydrolysis intermediate.

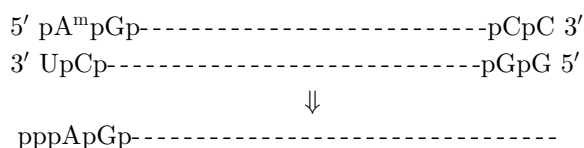
As in the days when chemistry was used to identify modified bases often found in tRNA, we readily obtained authentic 2'-O-methyl-pA to compare its chromatographic mobility with that of the <sup>32</sup>P-labelled pA\*. Both migrated to the identical position, showing clearly that A\* was 2'-O-methyladenosine. This finding indicated for the first time that virus RNAs, such as genomic CPV RNA, contains methylated residues.<sup>14)</sup> Accordingly, CPV genome RNA was considered to be in an end-to-end structure:



A similar study by Miura *et al.*<sup>15)</sup> showed that human reovirus dsRNA genomes also contained a 2'-

O-methylated nucleotide in the 5' terminal sequence of pG<sup>m</sup>pCp in one of the strands. Despite the exciting findings of 2'-O-methylated nucleotides in the 5' sequence of virus RNAs, the NNM in CPV dsRNA and its structural relationship to the 5' pA<sup>m</sup>pGp remained unclear, although apparently NNM somehow protected the strand containing pA<sup>m</sup>pGp from phosphorylation by polynucleotide kinase and from end-labelling.

**C. Stage 2: Identification of mRNA strands formed *in vitro* by CPV transcriptase.** Viruses in the *Reoviridae* family containing dsRNA, including the prototype human reovirus, contain RNA polymerase, which conservatively copy the plus strands in duplex genomic RNAs to produce viral mRNAs. Thus, one of the two strands comprising the dsRNA has the same polarity of mRNA. Synthesis of viral mRNAs by *in vitro* transcription using purified reovirus had already been demonstrated.<sup>16)</sup> Shimotohno *et al.*<sup>17,18)</sup> examined which of the two strands of CPV dsRNA had the same polarity as mRNA, and they found that CPV RNAs synthesized *in vitro* by virus-associated RNA polymerase had the same polarity as the pA<sup>m</sup>pGp-strand. Those CPV transcripts made *in vitro* were labeled by [ $\gamma$ -<sup>32</sup>P]ATP in the 5'-triphosphorylated form, <sup>32</sup>p-ppApGp-, suggesting that the strand with 5'-pA<sup>m</sup>pG in the genomic dsRNA has the same polarity as virus mRNA.



As a result, questions were raised: 1) when the methyl group in 5'-pA<sup>m</sup>pG was added or 2) is the methyl group found in the genomic RNA added during the mRNA synthesis or during the genome RNA synthesis, for example, to assemble segmented genomes?

**D. Stage 3: "Methylation-coupled transcription" in CPV mRNA synthesis that uses S-adenosyl-L-methionine.** In the fall of 1973, I tested an idea that CPV transcriptase may require the methyl donor S-adenosyl-L-methionine (AdoMet), considering that methylation of mRNA might be a prerequisite for natural viral mRNA synthesis in infected silkworm cells. The results were indeed surprising, far beyond expectation, because the CPV transcriptase activity was stimulated nearly 100-fold, and mRNA synthesis appeared to proceed in an AdoMet dependent manner (Fig. 4).<sup>19)</sup>

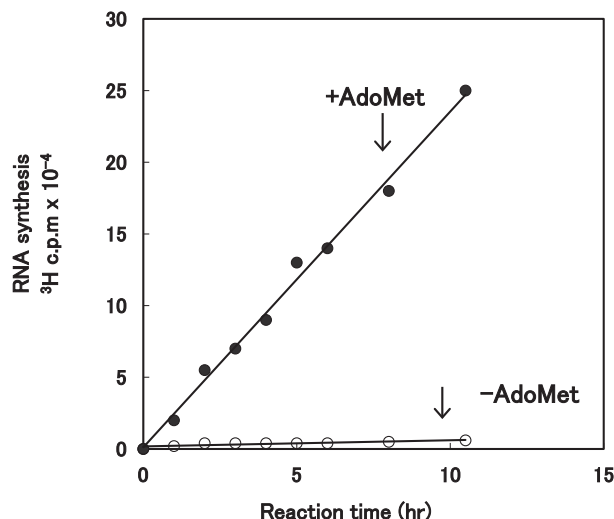


Fig. 4. Stimulation of CPV mRNA synthesis by S-adenosyl-L-methionine. Data with slight modifications are obtained from the paper by Furuichi (1974).<sup>19)</sup>

In addition, an experiment that used [<sup>3</sup>H-methyl]-labeled AdoMet clearly showed for the first time that virus mRNAs were methylated. The methylation of mRNA appeared to occur at the initiation of transcription and a limited number of methyl groups from [<sup>3</sup>H] methyl-labeled AdoMet were incorporated into each molecule of CPV RNA.<sup>19)</sup> One methyl group each was incorporated into the 2'-OH residue of 5' terminal pA and into NNM that links to the 5'-terminal pA<sup>m</sup> in a structure of NNM~pA<sup>m</sup> (representing the pA<sup>m</sup>-containing NNM). When [ $\gamma$ -<sup>32</sup>P]ATP was included in the CPV transcription in the presence of AdoMet, the product mRNA did not contain the <sup>32</sup>P-labeled phosphate, in contrast to the 5'-pppA containing mRNA synthesized in the absence of AdoMet, indicating that the  $\gamma$  phosphate of 5'-terminal <sup>32</sup>pppA was somehow eliminated when AdoMet was included in the transcription reaction mixture.<sup>19)</sup>

By contrast, the radioactive  $\beta$  phosphate in [ $\beta$ -<sup>32</sup>P]ATP remained in the newly synthesized mRNA and was found in the 5' NNM~pA<sup>m</sup>. This finding helped me to decide the overall mechanism of cap synthesis later. In fact, the <sup>32</sup>ppA moiety of p-<sup>32</sup>ppA was incorporated into the CPV mRNA in the form of NNM~<sup>32</sup>ppAm in which <sup>32</sup>P-labeled phosphate was resistant to alkaline phosphatase. I reported some of these unique observations as "Methylation-coupled transcription" in volume 1 of *Nucleic Acids Research* in 1974.<sup>19)</sup> Since this report, methylation in mRNA synthesis has quickly become

popular, and AdoMet has been added in many laboratories as a regular component to the *in vitro* transcription system of cells and viruses that contain RNA polymerase.<sup>20)–22)</sup>

At the same time, Perry and Kelley<sup>23)</sup> reported the presence of 2.2 residues of methylated nucleotides in every 1000 residues in mRNA containing poly(A) prepared *in vivo* from mouse L cells cultured in the presence of [<sup>3</sup>H-methyl]methionine. Here, earlier findings of 3'-poly (A) in vaccinia virus and cellular mRNAs, and an invention of a purification method for mRNAs containing poly (A) by oligo(dT)-cellulose column chromatography greatly facilitated isolation and analysis of cellular mRNAs. Desrosiers *et al.*<sup>24)</sup> reported that mRNAs from Novikoff hepatoma cells contain m<sup>7</sup>G, 2'-O-methylated nucleotides and m<sup>6</sup>Ap, suggesting that the presence of methylated nucleosides in mRNA may reflect a cellular mechanism to process certain mRNA sequences. After discussions of methylation on cellular mRNAs and reovirus RNAs at the 1974 Gordon Research Conference, Rottman, Shatkin and Perry<sup>25)</sup> jointly predicted m<sup>7</sup>GppNm as the 5' structure of eukaryotic mRNA in the newly started journal *Cell*. This structure (a freehand sketch) was close to m<sup>7</sup>GpppNm, but was incorrect with respect to the number of phosphates: two phosphates, instead of the correct three phosphates, between m<sup>7</sup>G and the first nucleotide of RNA. Those authors regrettably did not have solid chemical data supporting the 5' blocked structure, nor did they have knowledge about the mechanism of how the structure was synthesized. The report was just too early to conclude the unprecedented blocked mRNA structure consisting of three phosphates, two methyl groups and one plus charge at the 5' terminus. Thus, 1974 was the dawn of mRNA methylation; it was very competitive in the race to reach the truth, but it was certainly close to a complete explanation of the cap structure.

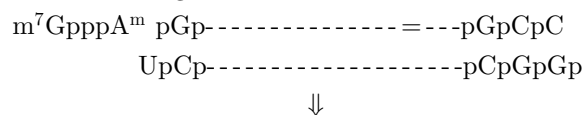
**E. Stage 4: Characterization of NNM'~ppA<sup>m</sup> and NNM'-ppG<sup>m</sup>, the 5' terminal ends of CPV and reovirus mRNAs.** In June 1974, I joined Shatkin at the Roche Institute Molecular Biology to continue research on mRNA methylation using reovirus and CPV. Soon, similar NNM-ppG<sup>m</sup> (representing the pG<sup>m</sup>-containing NNM) was found in the reovirus mRNA synthesized by reovirus transcriptase *in vitro* in the presence of AdoMet. The <sup>32</sup>P-labeled β-phosphate of [β,γ-<sup>32</sup>P]-labeled guanosine triphosphate (GTP) was incorporated into reovirus mRNA suggesting that CPV and reovirus mRNAs share a

similar 5'-structure, although the first nucleotide of virus mRNA was different. Meanwhile, the NNM of CPV and reovirus remained to be characterized. The 5' m<sup>7</sup>GppNm structure proposed by Rottman *et al.*<sup>25)</sup> failed to account for the 5' NNM~ppA<sup>m</sup> and NNM~ppG<sup>m</sup> in CPV and reovirus mRNAs, respectively, because the net charge of the postulated m<sup>7</sup>GppNm was -1 short owing to the + charge in m<sup>7</sup>G while NNM had -2. The NNM of CPV and reovirus mRNAs synthesized in the presence of [<sup>3</sup>H-methyl] AdoMet was found to contain a <sup>3</sup>H-methyl-labeled material that might be m<sup>7</sup>G or m<sup>7</sup>A, or an yet unidentified AdoMet-derivative (AdoMet has a +1 charge). An experiment that used [2-<sup>14</sup>C-methionyl]-AdoMet (containing non-radioactive methyl group) failed to label the mRNAs of either CPV or reovirus, nor it was incorporated into mRNA, showing that the NNM did not contain AdoMet itself. These results clearly limited the candidates for NNM to methylated nucleotide m<sup>7</sup>G- or m<sup>7</sup>A-related compound.

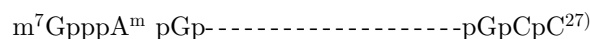
Regarding the origin of phosphate in 5' NNM~ppA<sup>m</sup> of CPV mRNA, the β phosphate of adenosine triphosphate (ATP) was shown to be included in the structure by the experiment with [β-<sup>32</sup>P]ATP prepared by a newly invented procedure (the details were published later in 1977),<sup>26)</sup> while γ phosphate of ATP was eliminated. Finally, the experiment that used [α-<sup>32</sup>P-labeled]GTP and <sup>3</sup>H-methyl-labeled AdoMet showed that both <sup>32</sup>P and <sup>3</sup>H radioactivity were incorporated together into CPV NNM~ppA<sup>m</sup> and reovirus NNM~ppG<sup>m</sup>, indicating that NNM is the structure containing 7-methyl guanosine 5' monophosphate, m<sup>7</sup>Gp. These results also clearly concluded that the 5' terminal structure of CPV mRNA is m<sup>7</sup>GpppAmp-, which fitted in all aspects with the mysterious <sup>3</sup>H-borohydride-labeled NNM found before in the analysis of CPV genome RNA. Indeed, CPV m<sup>7</sup>GpppAmp- had a net negative charge of -4.5 and about 5.4 Å in size and is susceptible to digestion by venom phosphodiesterase.<sup>13)</sup>

All these studies explained the initial questions regarding the genome structure of CPV and reovirus dsRNAs, and the mRNA structures synthesized in the presence of AdoMet.<sup>27),28)</sup>

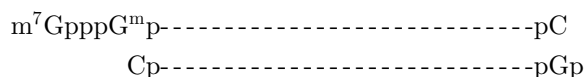
Silkworm CPV genome:



CPV mRNA:

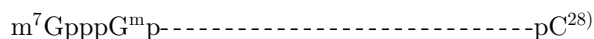


Human reovirus genome:



↓

Reovirus mRNA:



At the same time, in Japan and USA, similar structures of  $\text{m}^7\text{GpppA}^{\text{m}}$  and  $\text{m}^7\text{GpppG}^{\text{m}}$  were found in vaccinia virus mRNA synthesized *in vitro* in the presence of AdoMet by virion-associated RNA polymerase.<sup>(21),(22)</sup> Adding AdoMet into the *in vitro* viral transcription system initiated by me was so simple that the method encouraged many researchers. This method has indeed led a new insight into eukaryotic virus transcription, resulting in new mRNAs with methylation and a blocked 5'-terminal structure.<sup>(19)</sup> The next question was "Does this bizarre structure found in viral mRNAs also exist in eukaryotic cellular mRNAs?"

**F. Stage 5: Employment of nickname "cap" for the "blocked and methylated structure".** The  $\text{m}^7\text{GpppA}^{\text{m}}$  and  $\text{m}^7\text{GpppG}^{\text{m}}$  found in viral mRNAs were referred as to "blocked and methylated structures" in the initial publications. However, they were soon given the short nickname of "cap" in a study to confirm the precursor-product relationship between heterogeneous nuclear RNA (hnRNA) and cytoplasmic mRNA. Poly(A)-containing mRNA of HeLa cells and precursor hnRNA were examined for their 5'-terminal structure after <sup>32</sup>P-labeling and oligo (dT)-cellulose selection. A collaboration between the Shatkin and Darnell groups at the Rockefeller University clearly showed that both the mRNA in the cytoplasm and its precursor hnRNA in the nucleus contained the same general "blocked and methylated structures",  $\text{m}^7\text{GpppN}^{\text{m}}$ .<sup>(29)</sup> In one manuscript preparation meeting, Darnell suggested that we needed a nickname to replace the complicated "blocked and methylated 5' structure" and said "Let's call it 'cap' for short": cap was initially introduced by Rottman *et al.*<sup>(25)</sup> in their prediction of  $\text{m}^7\text{GppNm}$ , but it was soon popularized by Shatkin and Darnell groups to mean  $\text{m}^7\text{GpppN}^{\text{m}}$ .<sup>(30)</sup> Subsequent experiments done by us with <sup>32</sup>P pulse-labeled hnRNA suggested that caps were made at the initiation of transcription. The m7G caps were added to the products of RNA polymerase II shortly after transcriptional initiation within a narrow window of nascent oligonucleotides (+20 to +40), indicating that the process is rapid and efficient.<sup>(30),(31)</sup> The name

of cap became widespread quickly in the field of molecular biology and virology, and has further been expanded to wording for mRNA activity, such as "capping mRNA" or "decapping enzymes".<sup>(32)</sup>

### III. Caps in eukaryotic mRNAs

**A. Cellular mRNAs.** After the discovery of the cap structure in several virus mRNAs by us and others, studies were made worldwide to examine if the cap structure  $\text{m}^7\text{GpppN}$  is present in mRNAs of various organisms. For example, I identified caps in mRNAs and precursor hnRNA of HeLa cells.<sup>(29),(30)</sup>

Also, Adams and Cory<sup>(33),(34)</sup> in Australia reported the bizarre 5' terminal structure (namely the caps) in mRNA of mouse myeloma cells. Soon, the cap structure of  $\text{m}^7\text{GpppN}_{(\text{m})}$  consisting of 7-methylguanosine linked to the 5' end of the transcript by a 5'-5' triphosphate bridge was found in all eukaryotic cells examined (Table 1). But prokaryotic cells, such as *E. coli*, do not contain a cap in their mRNAs, suggesting that caps in mRNA are restricted to eukaryotic cells that contain the nucleus. An interesting trend is in the degree of methylation with respect to the levels of evolution of eukaryotes. Most lower eukaryotes, including yeast, fungi and amebas, were found to contain the less methylated Cap 0 ( $\text{m}^7\text{GpppN}$ ); higher eukaryotes, including humans, have Cap 1 ( $\text{m}^7\text{GpppN}_{(\text{m})}$ ) and 2 ( $\text{m}^7\text{GpppN}_{(\text{m})}\text{pN}_{(\text{m})}$ ) in which the first and second nucleotides of mRNA are 2'-O-methylated. In addition to methylation in caps, one or two N-6-methyl adenosine ( $\text{m}^6\text{A}$ ) were found in every one thousand nucleotides, as mentioned before by Perry and Kelley<sup>(23)</sup> and by Desrosier *et al.*<sup>(24)</sup> Because methylation on the cap generally became more complex as organisms evolved, the existence of a cap on mRNA and its structural complexity are among the characteristics that distinguish eukaryotes from prokaryotes and higher eukaryotes from lower eukaryotes.

**B. Virus mRNAs.** Viruses adopt various types of strategies for their parasitic replication and proliferation in infected cells. With a few exceptions, notably the picornavirus group,<sup>(35)</sup> eukaryotic viral mRNAs were found to contain the same cap structure as cellular mRNAs (Table 2), irrespective of their genomic structures, for example DNA or RNA, single-stranded or double-stranded, or negative-strand or positive-strand RNA or (DNA), and their replication strategies. This is probably because capped mRNAs are the fundamental, as well as functional, in host cells. Accordingly, the degree of methylation on viral mRNA caps correlates with host

Table 1. Caps in eukaryotic cellular mRNAs

Source	Types of Cap <sup>a</sup>			Other methylations
	Cap 0	Cap 1	Cap 2	
Human HeLa cells	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
HeLa cell histone	—	N <sub>1</sub> =A <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup> ,G <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	n.d.
Mouse myeloma	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
Mouse erythroid	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
Mouse fibroblasts	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
Mouse kidney	—	N <sub>1</sub> =N <sup>m</sup>	N <sub>2</sub> =N <sup>m</sup>	m <sup>6</sup> A
Rat hepatoma	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
Hamster kidney (BHK-21)	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A,m <sup>5</sup> C
Monkey kidney (BSC-1)	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
Mouse immunoglobulin	—	N <sub>1</sub> =G <sup>m</sup> ,m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =A <sup>m</sup>	m <sup>6</sup> A
Human globin	—	N <sub>1</sub> =m <sup>6</sup> A <sup>m</sup> ,A <sup>m</sup>	N <sub>2</sub> =C <sup>m</sup>	n.d.
Human gp 130	—	N <sub>1</sub> =N <sup>m</sup>	n.d.	n.d.
Human WS <sup>b</sup>	—	N <sub>1</sub> =A <sup>m</sup> ,G <sup>m</sup> ,U <sup>m</sup>	n.d.	n.d.
Human RTS <sup>c</sup>	—	N <sub>1</sub> =A <sup>m</sup> ,G <sup>m</sup> ,U <sup>m</sup>	n.d.	n.d.
Mouse globin	—	N <sub>1</sub> =m <sup>6</sup> A <sup>m</sup> ,A <sup>m</sup>	N <sub>2</sub> =C <sup>m</sup>	n.d.
Rabbit globin	—	N <sub>1</sub> =m <sup>6</sup> A <sup>m</sup>	N <sub>2</sub> =C <sup>m</sup>	n.d.
Duck globin	—	N <sub>1</sub> =N <sup>m</sup>	n.d.	n.d.
Chick ovalbumin	N <sub>1</sub> =A	N <sub>1</sub> =m <sup>6</sup> A <sup>m</sup> ,A <sup>m</sup>	N <sub>2</sub> =Py <sup>m</sup>	n.d.
Trout protamine	—	N <sub>1</sub> =N <sup>m</sup>	—	n.d.
Drosophila	N <sub>1</sub> =C,A,G	N <sub>1</sub> =Py <sup>m</sup> >Pu <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	n.d.
<i>Bombyx mori</i> silk fibroin	—	N <sub>1</sub> =A <sup>m</sup>	N <sub>2</sub> =U <sup>m</sup>	m <sup>6</sup> A
<i>Aedes albopictus</i>	—	N <sub>1</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	N <sub>2</sub> =Pu <sup>m</sup> ,Py <sup>m</sup>	m <sup>6</sup> A
<i>C. elegans</i>	TMG-capped mRNA <sup>2,2,7</sup> m <sub>3</sub> GpppN(m)-			—
Tobacco hornworm oocyte	GpppN	—	—	n.d.
Brine shrimp	—	N <sub>1</sub> =Am,Gm	n.d.	m <sup>6</sup> A
Sea urchin embryo	—	N <sub>1</sub> =Pu <sup>m</sup> >Py <sup>m</sup>	n.d.	m <sup>6</sup> A
Slime mold	N <sub>1</sub> =A>G	N <sub>1</sub> =A <sup>m</sup> (10%)	n.d.	n.d.
Neurospora	N <sub>1</sub> =A>G	n.d.	n.d.	n.d.
Yeast <i>S. cerevisiae</i>	N <sub>1</sub> =A>G	n.d.	n.d.	n.d.
Wheat embryo	N <sub>1</sub>	n.d.	n.d.	n.d.
Maize	N <sub>1</sub>	n.d.	n.d.	n.d.
Soybean seeds	N <sub>1</sub>	N <sub>1</sub> =N <sup>m</sup>	n.d.	n.d.

<sup>a</sup>Cap 0: m<sup>7</sup>GpppN<sub>1</sub>pN<sub>2</sub>p-, Cap 1: m<sup>7</sup>GpppN<sub>1</sub><sup>m</sup>pN<sub>2</sub>p-, Cap 2: m<sup>7</sup>GpppN<sub>1</sub><sup>m</sup>pN<sub>2</sub><sup>m</sup>p-. n.d., Not done; —, absent.

<sup>b</sup>WS: Human Werner syndrome gene transcript.

<sup>c</sup>RTS: Human Rothmund-Thomson syndrome gene transcript.

<sup>d</sup>TMG, Trimethylguanosine (<sup>2,2,7</sup>m<sub>3</sub>G).

mRNA cap methylation, *i.e.*, viruses that infect higher eukaryotic cells generally contain Cap 1 and Cap 2 structures, whereas mRNAs of viruses that infect unicellular hosts and plant cells contain Cap 0. After finding caps in mRNAs of cytoplasmic viruses, such as insect CPV, human reovirus and vaccinia virus, I investigated oncogenic viruses, such as avian Rous sarcoma virus and adenovirus, which have a replication phase in the nucleus. The studies showed these two viruses contained capped mRNAs synthesized by RNA polymerase II of host cells.<sup>36),37)</sup>

#### IV. Mechanism of cap synthesis

Subsequent to the discovery of the cap structure in mRNAs of many viruses, normal cells and tumor cells, I began to explain the mechanism of mRNA capping by studying the process associated with the *in vitro* synthesis of reovirus mRNA by viral cores. Because the components in the capping reaction and the origin of three phosphates that constitute two pyrophosphate linkages had been identified previously during the characterization steps of cap



Table 2. 5'-structures of virus mRNAs

Viruses	Genome	Transcription	5' Structures		Other methylations
			Genome 5' end	mRNA 5' end	
Mammals/Birds (RNA type)					
Reo	dsRNA(±)	cytoplasm	m <sup>7</sup> GpppG <sup>m</sup> (+)/ppG(−)	m <sup>7</sup> GpppG <sup>m</sup> pC <sup>m</sup> p	n.d.
Rous sarcoma	ssRNA(+)	nucleus	m <sup>7</sup> GpppG <sup>m</sup> pCp	n.d.	10–12 m <sup>6</sup> A
Avian sarcoma	ssRNA(+)	nucleus	m <sup>7</sup> GpppG <sup>m</sup> pCp	n.d.	10 m <sup>6</sup> A
Moloney murine leukemia	ssRNA(+)	nucleus	m <sup>7</sup> GpppG <sup>m</sup> p	n.d.	15–23 m <sup>6</sup> A
Feline leukemia	ssRNA(+)	nucleus	m <sup>7</sup> GpppG <sup>m</sup> p	m <sup>7</sup> GpppG <sup>m</sup> pAp	10 m <sup>6</sup> A
Sindbis	ssRNA(+)	cytoplasm	m <sup>7</sup> GpppA <sub>p</sub> U <sub>p</sub> Y <sub>p</sub> G <sub>p</sub> <sup>b</sup>	m <sup>7</sup> GpppApUpG	m <sup>5</sup> C
Calci	ssRNA(+)	cytoplasm	protein combined	base unspecified	
Dengue	ssRNA(+)	cytoplasm	m <sup>7</sup> GpppA <sup>m</sup> pNp	m <sup>7</sup> Gpppm <sup>6</sup> A <sup>m</sup> p	n.d.
VSV	ssRNA(−)	cytoplasm	(p)ppA	m <sup>7</sup> Gpppm <sup>6</sup> A <sup>m</sup> pA <sup>m</sup> p	n.d.
Influenza	ssRNA(−)	nucleus	pppA	m <sup>7</sup> GpppN <sup>m</sup> p	3 m <sup>6</sup> A
Newcastle disease	ssRNA(−)	nucleus	n.d.	m <sup>7</sup> GpppN <sup>m</sup> p	n.d.
Polio	ssRNA(+)	cytoplasm	protein-pUp	pUp	n.d.
EMC	ssRNA(+)	cytoplasm	protein-pUp	pNp	n.d.
Mammals (DNA type)					
Vaccinia	DNA	cytoplasm	—	m <sup>7</sup> Gpppm <sup>6</sup> A <sup>m</sup> p/G <sup>m</sup> p	n.d.
Adeno	DNA	nucleus	—	m <sup>7</sup> Gpppm <sup>6</sup> A <sup>m</sup> p/A <sup>m</sup> pN <sub>2</sub> <sup>m</sup> p	m <sup>6</sup> A, m <sup>5</sup> C
Simian virus 40	DNA	nucleus	—	m <sup>7</sup> Gpppm <sup>6</sup> A <sup>m</sup> p/G <sup>m</sup> p/C <sup>m</sup> p	m <sup>6</sup> A
Herpes simplex type 1	DNA	nucleus	—	m <sup>7</sup> GpppN <sub>1</sub> <sup>m</sup> pN <sub>2</sub> <sup>m</sup> p	m <sup>6</sup> A
Polyoma	DNA	nucleus	—	m <sup>7</sup> Gpppm <sup>6</sup> A <sup>m</sup> p/A <sup>m</sup> pN <sub>2</sub> <sup>m</sup> p	m <sup>6</sup> A

structure, a series of five reactions catalyzed by distinct enzymatic activity, were readily found to be required to form the Cap 1 structure (Fig. 5).<sup>38)</sup> The same general mechanism was also shown for CPV by Shimotohno *et al.*<sup>39)</sup> and for vaccinia virus by Moss *et al.*<sup>40)</sup> and was confirmed with purified enzymes from yeast and mammalian cells.<sup>41)</sup> This mechanism was confirmed for various cellular mRNAs using purified enzymes isolated from mammalian cells and yeast. For most cellular and nuclear viral mRNAs, transcriptional initiation sites are the capping sites, and capping and subsequent methylations occur by the scheme illustrated in Fig. 5. The 5' end of the nascent RNA is first modified by removal of the  $\gamma$ -phosphate by RNA triphosphatase (RTase) to yield a diphosphorylated end (ppNpNp-). It is then capped by adding guanosine monophosphate (GMP) transferred from GTP by mRNA guanylyltransferase (GTase) to form GpppNpNp-. In mammalian cells, the GTase, which is present together with RTase in a bifunctional protein, reacts with GTP to make a covalent GMP-enzyme complex by phosphoamide linkage to the  $\epsilon$ -amino group of the lysine residue in the signature KXDG sequence. Subsequent methylations catalysed by methyltransferases using AdoMet

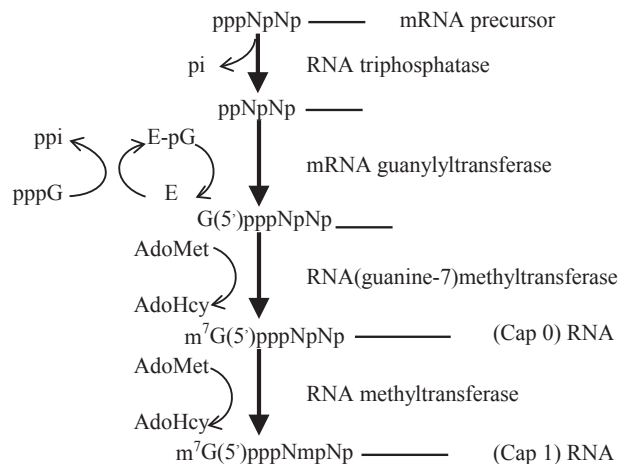


Fig. 5. Mechanism of cap formation.

produces the cap structure m<sup>7</sup>GpppN(m)pN(m)p-. The N7-methylation on guanine occurs cotranscriptionally, catalysed by the RNA (guanine-7-) methyltransferase (MTase) that binds to GTase-Pol II complexes. The first 2'-O-ribose methylation is m<sup>7</sup>G-cap dependent and occurs in the nucleus, while the second 2'-O-ribose methylation occurs in the cytoplasm. Accordingly, capping is completed at an

early stage of transcription when the nascent RNA is close to or is less than ~30 nucleotides away; co-transcriptional rather than post-transcriptional capping is likely for cellular mRNA synthesis. Consistent with selective RNA capping on the Pol II transcripts, capping enzymes were found to bind to the Pol II C-terminal domain heptad (7 amino acids) repeat sequences in the largest subunit after they are specifically phosphorylated on serine residues and during the switch from initiation to processive elongation in mRNA synthesis.<sup>42),43)</sup>

**Genes and enzymes participating in capping reaction.** Genes encoding enzymes participating in individual capping reactions have been cloned and characterized from several sources. In yeast *Saccharomyces cerevisiae* and fungus *Candida albicans*, the capping enzymes RTase, GTase and MTase are coded for by three separate genes. However, human RTase and GTase are on the same polypeptide of two-domain proteins, and MTase is coded for by a separated gene. Chu and Shatkin<sup>44)</sup> showed by gene silencing studies that each reaction participating in cap formation is essential for proliferation of cells and viruses. Later, I, who had left the research field of capping since 1984 and returned to Japan, collaborated with Shatkin again to identify the location of capping genes: human GTase and RTase genes were mapped on chromosome 6q16, and MTase gene on chromosome 18p11.22–p11.23.<sup>45)</sup>

The strange behavior of AdoMet-directed methylation-coupled transcription that triggered an expansion of mRNA methylation research was investigated thoroughly by me.<sup>46)</sup> The stimulatory effect of AdoMet was then found to be due to a lowering of K<sub>m</sub> for the initiating ATP so that virus RNA polymerase can more readily form the first dinucleotide pppApG. Kinetic studies indicated that AdoMet stimulated RNA polymerase through an allosteric effect by interacting with methyltransferase contained in the CPV transcription apparatus.<sup>47)</sup> This strange transcription stimulatory effect by AdoMet was also observed with the *Spring Viremia Carp Virus* transcription system *in vitro*.<sup>48)</sup>

**When and where does the capping reaction occur in cellular mRNA synthesis?** Earlier, I found that HeLa cell mRNAs contained a cap structure at the 5'-terminus.<sup>29)</sup> Subsequently, Salditt-Georgieff *et al.*<sup>30)</sup> found that adding 5'-cap structures occurs early in HeLa hnRNA synthesis. Similarly, Sommer *et al.*<sup>37)</sup> reported that adenovirus RNAs in the nucleus and cytoplasm of Hella cells were both capped. Because capping appears to be completed in

the nucleus at a very early stage of transcription, co-transcriptional capping, rather than post-transcriptional capping, is postulated for cellular mRNA synthesis *in vivo*.<sup>49)</sup> Biochemically, post-transcriptional capping can occur *in vitro* on 5' triphosphorylated RNA, oligonucleotides, or even on ppNpN dinucleotide structures, by the action of purified 5'-RTase and GTase.<sup>50)</sup> However, this may not occur *in vivo* because a strong 5'-exonuclease activity that hydrolyzes uncapped RNA exists in the nucleus, and nascent RNAs would be digested readily before the completion of transcription unless their 5' ends are protected by m<sup>7</sup>G-cap.<sup>5),6)</sup> As the action of 5'-exonuclease yields 5' mono-phosphorylated RNAs, which are no longer substrates for re-capping, capping of nascent RNAs must be completed when they are short enough to be protected by RNA polymerase II complexes.

Historically, a prototype concept "Methylation-coupled transcription" was first proposed by me for CPV mRNA synthesis before identification of the cap structure.<sup>19)</sup> Next, "Post-transcriptional capping" was shown by Ensinger *et al.*<sup>50)</sup> using a purified vaccinia virus capping enzyme. Finally, the concept of "Co-transcriptional capping in the initiation of transcription *in vivo*" has been supported on the basis that a direct interaction occurs between RNA polymerase II and capping enzymes in the initiation of transcription.<sup>42),43)</sup>

## V. Biological functions of m<sup>7</sup>G-cap in mRNA

**Cap stabilizes mRNA and stimulates mRNA translation.** To investigate the potential biological function of 5' cap in mRNA, Furuichi and Shatkin<sup>51)</sup> invented a method to prepare reovirus mRNAs having three different types of 5' termini, *i.e.*, m<sup>7</sup>GpppG(m)-, GpppG- and ppG-, based on the synthetic pathway shown in Fig. 5. Here, inclusion of AdoHcy, a methylation inhibitor, and pyrophosphatase in the reaction mixture produced predominantly GpppG-mRNA, while the presence of AdoHcy and a high concentration of pyrophosphate induced a reverse reaction of pyrophosphorolysis and produced ppG-mRNA. When we microinjected <sup>32</sup>P-labelled reovirus mRNAs having the three different types of 5' structures, into a *Xenopus* oocyte, or incubated in the wheat protein synthesizing system, m<sup>7</sup>GpppGm-mRNA was most stable in the cells and was incorporated into polysomes for protein synthesis, while GpppG-mRNA remained stable but was unable to be incorporated into polysomes (Fig. 6).<sup>5)</sup> By contrast, ppG-mRNA was unstable and was quickly

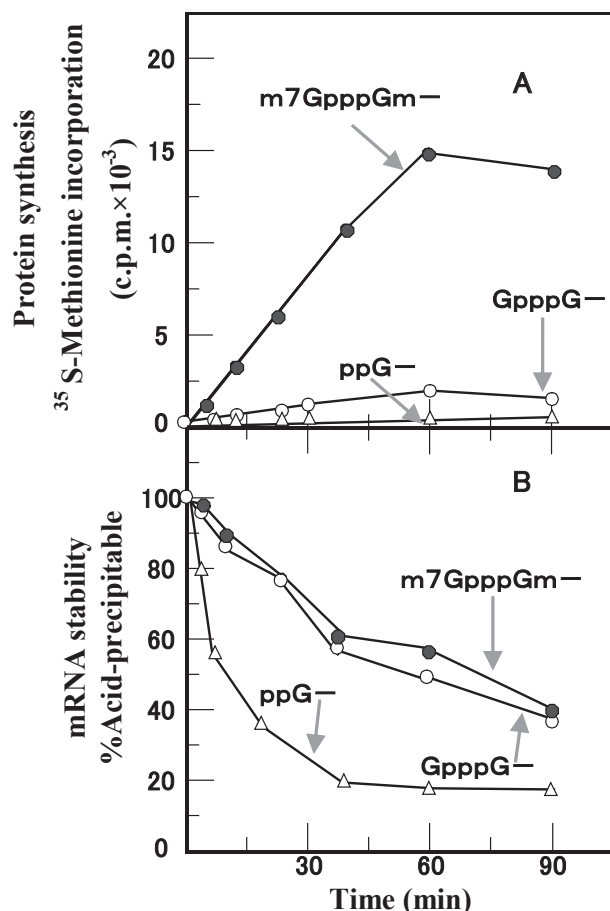


Fig. 6. Two important biological functions of m<sup>7</sup>G-cap: mRNA stabilization and stimulation of protein synthesis at the initiation. The <sup>32</sup>P-labelled reovirus mRNAs having the three different types of 5' structures, namely m<sup>7</sup>GpppG<sup>m</sup>-, GpppG- and ppG-, were prepared by the method of Furuichi and Shatkin (1976)<sup>51</sup> and were either microinjected to frog oocytes or incubated in the wheat germ protein synthesizing extract and examined mRNA stability and translational activity. Data are obtained from the paper by Furuichi *et al.* (1977).<sup>5)</sup>

degraded to 5'-monophosphorylated nucleotides, indicating clearly two important conclusions:

- (i) 5' Blocked structure is important for mRNA stability,
- (ii) m<sup>7</sup>G in cap structure is required for initiation of translation.

Indeed, Hickey *et al.*<sup>52)</sup> showed that m<sup>7</sup>GMP, a cap analogue, is a strong and specific inhibitor of eukaryotic translation, suggesting a cap is important for mRNA translation. A number of other studies used the *in vitro* translation systems of wheat germ extracts or rabbit reticulocyte lysates and they all showed the importance of a cap structure in protein synthesis. I participated in a study by

Muthukrishnan *et al.*<sup>53)</sup> that showed 5'-terminal 7-methylguanosine of a cap is required for translation in the wheat germ system, and in another study by Both *et al.*<sup>54)</sup> that showed ribosome binding to mRNA requires 5' terminal 7-methylguanosine. However, I evaluated highly the microinjection experiment using frog eggs with reovirus mRNA having the three different types of 5' structures, because the experiments were done in a way close to the *de novo* natural cell condition, and the results most explicitly showed the fundamental role of mRNA cap structure in the cytoplasm.

The study also predicted the presence in the nucleus and cytoplasm of a 5'-to-3' exonuclease that preferentially digests uncapped RNA to 5' mononucleotides exonucleolytically. Indeed, such an exonuclease, XRN1, having the expected enzymatic properties and the responsibility for digestion of decapped mRNA, was later found in *S. cerevisiae* and mammalian cells.<sup>55)</sup> Studies with different reovirus mRNAs indicated that transcripts with 5'-terminal m<sup>7</sup>GpppGm were preferentially translated in extracts of both plant and animal cells. More direct evidence came from findings that the 5' cap of mRNA is part of the ribosome-binding site in 40S initiation complexes, because caps were retained in the sequences protected against RNase digestion. Furthermore, the protected capped mRNA fragments rebound efficiently, while only low levels of ribosome rebinding were observed using the corresponding uncapped fragments that were collected from RNase-treated 80S complexes. These findings on the sizes, sequences and initiation properties of 40S versus 80S ribosome-protected mRNAs provided insights that led to the "scanning model" proposed by Kozak to explain how eukaryotic ribosomes select initiation regions in mRNAs, usually the first AUG codon downstream of the cap.<sup>56),57)</sup>

**Cap function as a hallmark of regulation of eukaryotic protein synthesis.** The effects of the 5' cap in mRNA translation was studied extensively and soon it was found that most eukaryotic translation is initiated with a cap-binding protein (named later as eIF4E) that was identified by Sonenberg *et al.*<sup>58)</sup> This finding opened a new stage of cap research with regard to studying the regulation of protein synthesis and for understanding how viruses manage to replicate by making use of the cap-dependent living system of host cells. Selection of capped over uncapped transcripts occurred at the level of mRNA binding to 40S ribosomal subunits. Several other observations provided support for the

idea that caps are recognized during the early stages of protein synthesis and have a pivotal role in regulation of translation.<sup>10)</sup>

Many studies in the last 40 years discovered molecular mechanisms underlying the regulation of protein synthesis and yielded various important concepts, such as “cap-dependent translation” (or “m<sup>7</sup>G-dependent translation”), “cap-independent translation by virus strategy”, and “cap-snatching reaction by influenza virus”.<sup>58)</sup> The “5′ terminal cap as the 40S ribosome entry site”, “cap-dependent ribosome-scanning model”, “5′ cap-recognition for assembly of protein synthesis initiation complex”, and “internal ribosome entry site IRES in the uncapped mRNAs” are other titles given to some of the cap-related biochemical reactions. These concepts and inventions of new terminologies point out clearly the essence of many cap-mediated reactions and have now become popular in understanding various important regulations for protein synthesis and virus replication strategy.<sup>59)–61)</sup>

**Roles of 5′ caps in pre-mRNA splicing and nucleocytoplasmic transport.** Besides protein synthesis, caps were found to have important roles in the splicing of pre-mRNAs, cooperating with spliceosomes.<sup>62)</sup> The cap on pre-mRNA interacts with splicing components at the adjacent 5′ splice site to process the first exon and remove the first intron. This interaction is mediated by a nuclear cap-binding complex (CBC) that consists of two tightly associated proteins, CBP 20 and CBP 80. The m<sup>7</sup>G-cap on pre-mRNA is specifically bound by CBC, which facilitates association of the cap-proximal 5′ splice site with U1 snRNP. The CBC remains bound to the cap during RNA processing and has an active role in both splicing and RNA export. The CBC-mediated interaction between U1 snRNP and the 5′ splice site is one of the earliest steps in spliceosome assembly and is conserved in humans and yeast. CBC also participates in 3′ polyadenylation and nucleocytoplasmic transport of mRNA.<sup>63)</sup> One study indicated that the human mRNA export machinery complex, which mediates cap-first export of mRNA through nuclear pores, interacts with CBP80 and binds RNA close to the cap in a splicing-dependent interaction, facilitating the export of spliced mRNAs.<sup>64)</sup>

**Use of capping site in future diagnosis of human diseases; profiling of gene expression in patients by transcription start sites and promoter analysis with gene CAGE technology and next generation sequencers.** Nucleotide N in the 5′-cap m<sup>7</sup>GpppN(m) is the initiating nucleotide in

gene transcription by PolII RNA polymerase. Unique features of the cap structure that include 5′-5′ pyrophosphate linkages and a free 2′-3′ *cis*-hydroxyl group in m<sup>7</sup>G ribose permit isolation of RNA sequences adjacent to and upstream of 5′ terminal caps. For example, the capped 5′-terminal oligonucleotide of retrovirus RNA was isolated by affinity column chromatography by me with borate resin that binds specifically to 2′-3′ *cis*-OH groups in m<sup>7</sup>G ribose.<sup>35)</sup> Suzuki and Sugano<sup>65)</sup> developed cap-modifying technology that removes m<sup>7</sup>Gpp from mRNA by tobacco acid pyrophosphatase and recaps the monophosphorylated 5′-ends with an oligonucleotide adapter by RNA ligase. This method, termed “oligo-capping”, was used to determine the exact transcription initiation sites of individual genes and to prepare libraries of full-size cDNA clones.<sup>65)</sup> Kanamori-Katayama *et al.*<sup>66)</sup> developed another procedure to collect cDNA/RNA hybrid copies containing mRNA 5′ terminal sequences that used biotinylation of periodate-oxidized m<sup>7</sup>G residues. The cDNA fragments synthesized by reverse transcriptase are captured first by streptavidin, and then are hybridized to the oligo-dT on the flow cell surface. This cap-trapping method is referred to as HeliScopeCAGE (cap analysis of gene expression using HeliScope single-molecule sequencer). The isolated first-strand cDNAs are sequenced directly by an advanced generation sequencer. With this technology, a large short-tag CAGE library was made available to study transcription events on a genome-wide-scale.<sup>66)</sup> Analysis of cap-trapped sequences by high-throughput sequencers enables genome-wide quantification of expression, and diagnosis (or forecasting) of diseases. This innovative technology of the cap will surely open a new field to understand various levels of cell biology, including the nature of pluripotent and adult stem cells and their various differentiated progeny, as well as the nature of altered gene action in cancer initiation and progression.

## VI. Conclusion

**Discovery and impact of a cap on molecular biology past and future.** Methylation of genomic DNA and mRNA is now widely accepted as a major pathway to regulate both transcription and translation in eukaryotes. The m<sup>7</sup>G-cap formed on mRNA is fundamental for eukaryotic gene expression. Almost 40 years have passed since the discovery of the cap, during which much has been learned about the importance of biochemical mechanisms of capping

and the downstream effects on translation and its regulation. Genetic and RNAi studies have shown the lethal consequences of capping failure in yeast and higher organisms, confirming the importance of mRNA capping. In addition, comparative biological studies have shown evolutionary and functional conservation of capping from unicellular to multicellular organisms.

By contrast, viruses develop their specific strategies to inactivate or utilize the host capping system to proceed preferentially their replication cycle or to hijack the facility of host cells for their survival. The diverse strategies of cap formation among viruses that infect humans and domestic animals, such as influenza, corona and vesicular stomatitis viruses, have been reviewed extensively by me previously.<sup>1)</sup> Those information should be useful for drug screening in pharmaceutical companies to find efficacious antiviral compounds showing no or few adverse effects.

I was fortunate to be one of the discoverers of this important cap structure and was able to share with many colleagues and competitors the excitement of explaining how it is made and what it does. The year 1975 was the birth year of the m<sup>7</sup>G-cap appearing on the science bulletin board. It was also the starting year for The Current Content to identify the most-cited author, and I was honored to be selected as the first one.<sup>67)</sup> One day, there was a telephone call from the Current Content office asking "Why is this structure so important?" Although I missed the right answer then, now I will not disappoint the questioner by providing so much valuable information as described in this review article. In addition to the various cap effects on translational regulations that we learned in the last four decades, future studies of capping site and promoter-based analyses of gene expression by next-generation sequencers should lead to a deeper understanding of dynamic transcriptional regulation.

### Acknowledgement

I would like to dedicate this review to the late Dr. Kin'ichiro Miura and Dr. Aaron J. Shatkin, whom I respected as mentors, and whom I worked together on mRNA caps throughout the term of its discovery process and subsequent investigation of its biological significances. I thank Ms. Midori Kawaguchi and Ms. Aika Takahashi at the GeneCare Research Institute and Dr. Yoshihito Ueno and Mr. Kosuke Nakamoto at Gifu University for assistance in compiling this manuscript.

### References

- 1) Furuichi, Y. (2014) Caps on eukaryotic mRNAs. *In* eLS. John Wiley and Sons, Ltd., Chichester. doi: 10.1002/9780470015902.a0000891.pub3.
- 2) Weaver, R.F. (2008) Messenger RNA and processing II: Capping and Polyadenylation. *In* Molecular Biology. Fourth Edition, McGraw-Hill Companies, Inc., New York, Chapter 15.
- 3) Furuichi, Y. and Shatkin, A.J. (2000) Viral and cellular mRNA capping: past and prospects. *Adv. Virus Res.* **55**, 135–184.
- 4) Furuichi, Y. and Shatkin, A.J. (1989) Characterization of cap structure. *Methods Enzymol.* **180**, 164–177.
- 5) Furuichi, Y., LaFiandra, A. and Shatkin, A.J. (1977) 5'-Terminal structure and mRNA stability. *Nature* **266**, 235–239.
- 6) Shimotohno, K., Kodama, Y., Hashimoto, J. and Miura, K.I. (1977) Importance of 5' terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2734–2738.
- 7) Hsu, C. and Stevens, A. (1993) Yeast cells lacking 5'→3' exonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Mol. Cell. Biol.* **13**, 4826–4835.
- 8) Lewis, J.D. and Izaurralde, E. (1997) The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247**, 461–469.
- 9) Schwer, B. and Shuman, S. (1996) Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing *in vivo*. *RNA* **2**, 574–583.
- 10) Sonenberg, N. (1988) Cap-binding proteins of eukaryotic mRNA: Functions in initiation and control of translation. *Prog. Nucleic Acid Res. Mol. Biol.* **35**, 173–207.
- 11) Kominami, R. and Muramatsu, M. (1977) Heterogeneity of 5'-termini of nucleolar 45S, 32S and 28S RNA in mouse hepatoma. *Nucleic Acids Res.* **4**, 229–240.
- 12) Furuichi, Y. and Miura, K.-I. (1972) The 3'-termini of the genomic RNA segments of silkworm cytoplasmic polyhedrosis virus. *J. Mol. Biol.* **64**, 619–632.
- 13) Furuichi, Y. and Miura, K.-I. (1973) Identity of the 3'-terminal sequences in ten genome segments of silkworm cytoplasmic polyhedrosis virus. *Virology* **55**, 418–425.
- 14) Miura, K., Watanabe, K. and Sugiura, M. (1974) 5'-Terminal nucleotide sequences of the double-stranded RNA of silkworm cytoplasmic polyhedrosis virus. *J. Mol. Biol.* **86**, 31–48.
- 15) Miura, K.-I., Watanabe, K., Sugiura, M. and Shatkin, A.J. (1974) The 5'-terminal nucleotide sequences of the double-stranded RNA of human reovirus. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3979–3983.
- 16) Banerjee, A.K. and Shatkin, A.J. (1970) Transcription *in vitro* by reovirus-associated ribonucleic acid-dependent polymerase. *J. Virol.* **6**, 1–11.
- 17) Shimotohno, K., Kimura, K. and Miura, K. (1973)

- Transcription of double-stranded RNA in cytoplasmic polyhedrosis virus *in vitro*. *Virology* **53**, 283–286.
- 18) Shimotohno, K. and Miura, K. (1974) 5'-Terminal structure of messenger RNA transcribed by RNA polymerase of silkworm cytoplasmic polyhedrosis virus containing double-stranded RNA. *J. Mol. Biol.* **15**, 21–30.
  - 19) Furuichi, Y. (1974) "Methylation-coupled" transcription by virus-associated transcriptase of cytoplasmic polyhedrosis virus containing double-stranded RNA. *Nucleic Acids Res.* **1**, 809–822.
  - 20) Shatkin, A.J. (1974) Methylated messenger RNA synthesis *in vitro* by purified reovirus. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3204–3207.
  - 21) Urushibara, T., Furuichi, Y., Nishimura, C. and Miura, K. (1975) A modified structure at the 5'-terminus of mRNA of vaccinia virus. *FEBS Lett.* **49**, 385–389.
  - 22) Wei, C.M. and Moss, B. (1975) Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 318–322.
  - 23) Perry, R.P. and Kelley, D.E. (1974) Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42.
  - 24) Desrosiers, R., Friderici, K. and Rottman, F. (1974) Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3971–3975.
  - 25) Rottman, F., Shatkin, A.J. and Perry, R.P. (1974) Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cell* **3**, 197–199.
  - 26) Furuichi, Y. and Shatkin, A.J. (1977) A simple method for the preparation of [ $\beta$ -<sup>32</sup>P]-purine nucleoside triphosphates. *Nucleic Acids Res.* **4**, 3341–3355.
  - 27) Furuichi, Y. and Miura, K. (1975) A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253**, 374–375.
  - 28) Furuichi, Y., Morgan, M., Muthukrishnan, S. and Shatkin, A.J. (1975) Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: m<sup>7</sup>G(5')ppp(5')G<sup>m</sup>pCp-. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 362–366.
  - 29) Furuichi, Y., Morgan, M., Shatkin, A.J., Jelinek, W., Salditt-Georgieff, M. and Darnell, J.E. (1975) Methylated, blocked 5' termini in HeLa cell mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1904–1908.
  - 30) Salditt-Georgieff, M., Jelinek, W., Darnell, J.E., Furuichi, Y., Morgan, M. and Shatkin, A.J. (1976) Methyl labeling of HeLa cell hnRNA: a comparison with mRNA. *Cell* **7**, 227–237.
  - 31) Shatkin, A.J. (1976) Capping of eucaryotic mRNAs. *Cell* **9**, 645–653.
  - 32) Nuss, D.L., Furuichi, Y., Koch, G. and Shatkin, A.J. (1975) Detection in HeLa cell extracts of a 7-methyl guanosine specific enzymatic activity that cleaves m<sup>7</sup>GpppN<sup>m</sup>. *Cell* **6**, 21–27.
  - 33) Adams, J.M. and Cory, S. (1975) Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature* **255**, 28–33.
  - 34) Cory, S. and Adams, J.M. (1975) The modified 5'-terminal sequences in messenger RNA of mouse myeloma cells. *J. Mol. Biol.* **99**, 519–547.
  - 35) Nomoto, A., Lee, Y.F. and Wimmer, E. (1976) The 5' end of poliovirus mRNA is not capped with m<sup>7</sup>G(5')ppp(5')Np. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 375–380.
  - 36) Furuichi, Y., Shatkin, A.J., Stavnezer, E. and Bishop, M. (1975) Blocked, methylated 5'-terminal sequence in avian sarcoma virus RNA. *Nature* **257**, 618–620.
  - 37) Sommer, S., Salditt-Georgieff, M., Bachenheimer, S., Darnell, J.E., Furuichi, Y., Morgan, M. and Shatkin, A.J. (1976) The methylation of adenovirus-specific nuclear and cytoplasmic RNA. *Nucleic Acids Res.* **3**, 749–765.
  - 38) Furuichi, Y., Muthukrishnan, S., Tomasz, J. and Shatkin, A.J. (1976) Mechanism of formation of reovirus mRNA 5'-terminal blocked and methylated sequence, m<sup>7</sup>GpppG<sup>m</sup>pC. *J. Biol. Chem.* **251**, 5043–5053.
  - 39) Shimotohno, K. and Miura, K. (1976) The process of formation of the 5'-terminal modified structure in messenger RNA of cytoplasmic polyhedrosis virus. *FEBS Lett.* **64**, 204–208.
  - 40) Moss, B., Gershowitz, A., Wei, C.M. and Boone, R. (1976) Formation of the guanylated and methylated 5'-terminus of vaccinia virus mRNA. *Virology* **72**, 341–351.
  - 41) Shuman, S. (1995) Capping enzymes in eukaryotic mRNA synthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **50**, 101–129.
  - 42) McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D.L. (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 3306–3318.
  - 43) Cho, E.J., Takag, T., Moore, C.R. and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **11**, 3319–3326.
  - 44) Chu, C. and Shatkin, A.J. (2008) Apoptosis and autophagy induction in mammalian cells by small interfering RNA knockdown of mRNA capping enzymes. *Mol. Cell. Biol.* **19**, 5829–5836.
  - 45) Pillutla, R.C., Shimamoto, A., Furuichi, Y. and Shatkin, A.J. (1998) Human mRNA capping enzyme (RNGTT) and cap methyltransferase (RNMT) map to 6q16 and 18p11.22–p11.23, respectively. *Genomics* **54**, 351–353.
  - 46) Furuichi, Y. (1978) "Pretranscriptional capping" in the biosynthesis of cytoplasmic polyhedrosis virus mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 488–492.
  - 47) Furuichi, Y. (1981) Allosteric stimulatory effect of S-adenosylmethionine on the RNA polymerase in cytoplasmic polyhedrosis virus: a model for the positive control of eukaryotic transcription. *J. Biol. Chem.* **256**, 483–493.

- 48) Gupta, K.C. and Roy, P. (1981) Synthesis of capped and uncapped methylated oligonucleotides by the virion transcriptase of spring viremia of carp virus, a rhabdovirus. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4758–4762.
- 49) Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S. and Darnell, J.E. Jr. (1980) The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* **19**, 69–78.
- 50) Ensinger, M.J., Martin, S.A., Paoletti, E. and Moss, B. (1975) Modification of the 5'-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2525–2529.
- 51) Furuichi, Y. and Shatkin, A.J. (1976) Differential synthesis of blocked and unblocked 5'-termini in reovirus mRNA: effect of pyrophosphate and pyrophosphatase. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3448–3452.
- 52) Hickey, E.D., Weber, L.A. and Baglioni, C. (1976) Inhibition of initiation of protein synthesis by 7-methylguanosine-5'-monophosphate. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 19–23.
- 53) Muthukrishnan, S., Both, G.W., Furuichi, Y. and Shatkin, A.J. (1975) 5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation. *Nature* **255**, 33–37.
- 54) Both, G.W., Furuichi, Y., Muthukrishnan, S. and Shatkin, A.J. (1975) Ribosome binding to reovirus mRNA in protein synthesis requires 5' terminal 7-methylguanosine. *Cell* **6**, 185–195.
- 55) Nagarajan, V.K., Jones, C.I., Newbury, S.F. and Green, P.J. (2013) XRN 5'→3' exoribonucleases: structure, mechanisms and functions. *Biochim. Biophys. Acta* **1829**, 590–603.
- 56) Kozak, M. (1978) How do eukaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123.
- 57) Kozak, M. (1989) The scanning model for translation: an update. *Cell Biol.* **108**, 229–241.
- 58) Sonenberg, N., Morgan, M.A., Merrick, W.C. and Shatkin, A.J. (1978) A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4843–4847.
- 59) Gingras, A.-C., Raught, B. and Sonenberg, N. (1999) eIF4 Initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963.
- 60) Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010) Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79**, 351–379.
- 61) Topisirovic, I., Svitkin, Y.V., Sonenberg, N. and Shatkin, A.J. (2010) Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA*. **2**, 277–298. doi: 10.1002/wrna.52. Epub 2010 Oct 28.
- 62) Lewis, J.D. and Izaurralde, E. (1997) The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247**, 461–469.
- 63) Gonatopoulos-Pournatzis, T. and Cowling, V.H. (2014) Cap-binding complex (CBC). *Biochem. J.* **457**, 231–242. doi:10.1042/BJ20131214.
- 64) Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A. and Reed, R. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389–1400.
- 65) Suzuki, Y. and Sugano, S. (2003) Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods Mol. Biol.* **221**, 73–91.
- 66) Kanamori-Katayama, M., Itoh, M., Kawaji, H., Lassmann, T., Katayama, S., Kojima, M., Bertin, N., Kaiho, A., Ninomiya, N., Daub, C.O., Carninci, P., Forrest, A.R. and Hayashizaki, Y. (2011) Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome. Res.* **21**, 1150–1159. doi: 10.1101/gr.115469.110. Epub 2011 May 19.
- 67) Garfield, E. (1976) "Current Contents": 1975 Life-Sciences Articles Highly Cited in 1975. *CURRENT CONTENTS*, Number 15 (April 12), 5–9.

(Received May 1, 2015; accepted June 22, 2015)

## Profile

Yasuhiro Furuichi was born in 1940 at Joshin in the present North Korea, which was part of Japan before 1945. He graduated from the University of Toyama in 1964 and entered the graduate school of the University of Tokyo, because he wished to study the chemistry of nucleic acids in the laboratory of Prof. Tyunosin Ukita. In 1969, he received a Ph.D., and he obtained a position at the National Institute of Genetics, where he worked with Dr. Kin'ichiro Miura to characterize the structures of CPV dsRNA genome. In 1974, he moved to Dr. Aaron J. Shatkin's laboratory at the Roche Institute of Molecular Biology in U.S.A. to extend his findings of mRNA methylation with CPV to other virus systems. He was cited by the Current Content as "most-cited author in 1975", and he received an award from the Japanese Biochemical Society in 1976 for his finding of mRNA methylation and discovery of the 5' cap structure present in eukaryotic mRNAs. In 1980–1985, he served as an associate editor of the Journal of Virology in the field of RNA viruses.



In 1985, he returned to Japan to direct the drug discovery program of Nippon Roche Research Center, where his team was successful in finding the endotheline antagonist Bosentan, which has long been used since as a specific medicine for pulmonary hypertension. From 1993 for seven years, he directed the national project AGENE Research Institute to elucidate the molecular basis behind Werner syndrome, a premature aging genetic disease prevalent in Japan. In 2000, he started the company GeneCare Research Institute to develop the findings made at AGENE into anticancer drugs free from adverse effects. From 2008 for five years, he directed the national project Hokuriku Innovation Cluster for Health Sciences. For the last 14 years since 2001, he has been an adviser to the Sakigake front-running programs of the Japanese government, in which he keeps encouraging young elite scientists to conduct innovation research to discover something new of value.