Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 18 (2008) 880–884
Bioorganic & Medicinal Chemistry Letters

Synthesis and biological evaluation of trimethyl-substituted cap analogs

Anilkumar R. Kore* and Muthian Shanmugasundaram

Ambion, Inc., An Applied Biosystems Business, Bioorganic Chemistry Division, 2130 Woodward Street, Austin, TX 78744-1832, USA

Received 3 December 2007; revised 17 December 2007; accepted 19 December 2007
Available online 4 January 2008

Abstract—The N^7-methyl guanosine cap located on the 5′-terminus of mRNAs is important for a number of biochemical processes. A new dinucleoside triphosphate cap analog was synthesized with methyl groups on the N^7 of both guanine moieties, as well as the 3′-OH of one of the ribose moieties ($m_2^{7,3'O}G[5']ppp[5']m^7G$). The function of this trimethylated cap analog was compared with those of three other, less-methylated cap analogs: one omitting the ribose methylation ($m^7G[5']ppp[5']m^7G$), one omitting the N^7 methylation linked to the unmodified ribose ($m_2^{7,3'O}G[5']ppp[5']G$), and the standard cap analog, $m^7G[5']ppp[5']G$. These cap modifications were assayed with respect to their effects on capping efficiency, yield of RNAs during in vitro transcription, and the translational activity of these RNAs upon transfection into HeLa cells. The translational activity was monitored by measuring the luciferase activity of a luciferase-fusion protein produced from the in vitro synthesized RNAs. The RNA capped with the trimethylated analog ($m_2^{7,3'O}G[5']ppp[5']m^7G$) was translated the most efficiently, with ~2.6-fold more activity than the conventional cap ($m^7G[5']ppp[5']G$). The other two variants were also more efficient, generating, ~2.2 times (for the $m_2^{7,3'O}G[5']ppp[5']G$ analog) and, ~1.6 times (for the $m^7G[5']ppp[5']m^7G$ analog) more luciferase function than the conventional cap.
© 2007 Elsevier Ltd. All rights reserved.

The nucleotide structure found at the 5′ terminus of many eukaryotic messenger RNAs, consisting of a 7-methylguanosine, connected, via a triphosphate bridge, to the 5′-end of the first transcribed nucleotide, resulting in $m^7G(5')$ ppp(5′)N, where N is any nucleotide. The 5′-terminal 'cap' structure, $m^7G(5')ppp(5')N$, that is present in most eukaryotic mRNAs, serves several functions. It facilitates their function in protein synthesis at the level of initiation.[1] The presence of the N^7-methyl group on the terminal guanosine promotes the formation of stable initiation complexes between mRNA and ribosomal 40S subunits.[2,3] The mRNA cap also plays an important role in processing, it protects the mRNAs from degradation by exonucleases and enables transport of RNAs from the nucleus to the cytoplasm. The prevailing method for generating capped mRNA using in vitro transcription employs a preformed dinucleotide of the form $m^7G[5']ppp[5']G$ (mCAP) in excess over regular G in the transcription reaction, so that it is incorporated as the initial base in preference to regular G.[4–9] However, the drawback of this mCAP analog is that the 3′ OH of either the G or m^7G can serve as the initiating nucleophile for transcriptional elongation leading to the synthesis of two isomeric RNAs of either forward or reverse form in approximately equal proportions depending upon the ionic conditions of the transcription reaction. The reverse form of capped mRNAs, i.e., $G[5']pppm^7G[pN]^n$ will not be recognized during the translation process, with only forward oriented sequences i.e. $m^7G[5']pppG[pN]^n$ being translated.[10] The recent finding that the synthesis of two 'anti-reverse cap analogs' (ARCAs) such as $m_2^{7,3'-O}G[5']ppp[5']G$ and $m^7,3'dG[5']ppp[5']G$ is also exclusively incorporated only in the forward orientation because of modifications at the 3′ position of the N^7-methylguanosine ribose.[11,12] In a subsequent study, it was also reported that chemical modification at either the 2′ or 3′ position of N^7-methylguanosine, the cap incorporated solely in the forward orientation, even though the 2′ OH group does not participate in the phosphodiester linkage.[13] As the N^7-methylguanosine cap is involved in diverse molecular events as stated above, another natural cap structure, the $m_3^{2,2,7}G[5']ppp[5']G$ trimethylated cap, found on small nuclear RNAs is required for shuttling these RNAs between nucleus and cytoplasm, indicating the potentially powerful roles additional methyl groups

Keywords: Capping efficiency; In vitro transcription; Translation efficiency; Trimethylated cap analog; Luciferase activity; HeLa cells.
* Corresponding author. Tel.: +1 512 721 3589; fax: +1 512 651 0201; e-mail: anil.kore@appliedbiosystems.com

0960-894X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2007.12.049

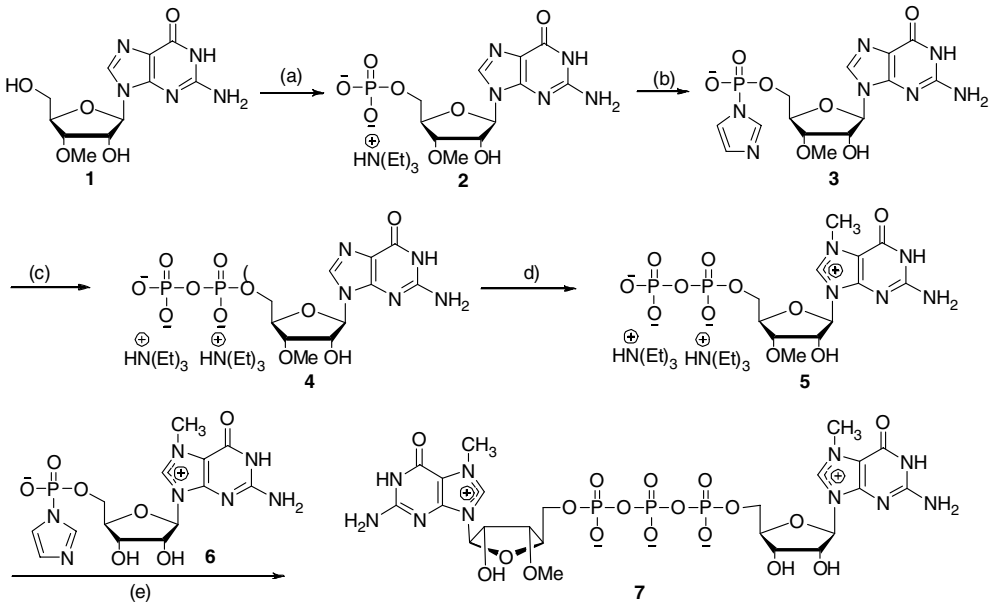
could play.[14,15] In continuation of our efforts[16] to test the potential beneficial effects additional chemical modifications could play, we have synthesized a trimethylated cap analog, $m_2^{7,3'O}G[5']ppp[5']m^7G$, and studied it with other methylated cap variants incorporated into in vitro synthesized RNA. The variants we created were $m_2^{7,3'O}G[5']ppp[5']m^7G$ **7**, $m_2^{7,3'O}G[5']ppp[5']G$ **9**, and $m^7G[5']ppp[5']m^7G$ **10**. The biological activity of RNAs capped with these and $m^7G[5']ppp[5']G$ was studied with respect to incorporation efficiency, in vitro transcription yield, and translational activity. These three variants all hold an advantage over the conventional cap analog. Compound **10**, $m^7G[5']ppp[5']m^7G$, presents an identical cap structure to be recognized during translation regardless of its orientation. The cap variant $m_2^{7,3'O}G[5']ppp[5']G$ **9** is the anti-reverse cap analog (ARCA) currently supplied by Applied Biosystems. The 'anti-reverse' designation in ARCA refers to a modification at the 3′ position of the ribose attached to the methylated guanine where the OH group is replaced with an OCH_3 group. Because of this substitution, the RNA polymerase can only initiate transcription with the remaining 3′-hydroxyl group, thus forcing ARCA incorporation in the forward orientation. The compound **7**, $m_2^{7,3'O}G[5']ppp[5']m^7G$, is a new version of ARCA, in which one more methyl group is added at the N^7 position of the second guanine. As a result, unlike transcripts synthesized with conventional cap analog, 100% of the transcripts synthesized with our anti-reverse compounds **7** and **9** at the 5′ end should be translatable leading to a strong increase in translational yield.

The reaction pathway leading to the formation of desired $m_2^{7,3'O}G[5']ppp[5']m^7G$ **7** is shown in Scheme 1. The starting compound 3′-*O*-methylguanosine **1** was prepared as previously described.[17] Treatment of 3′-*O*-methylguanosine with $POCl_3$ in the presence of trialkyl phosphate afforded the corresponding 3′-*O*-methyl GMP **2**.[18] The imidazolide reaction of 3′-*O*-methyl GMP **2** with imidazole, triphenyl phosphine, and aldrithiol afforded the corresponding imidazolide salt **3** with 76% yield.[19] The phosphorylation reaction of 3′-*O*-methyl ImGMP with $(Et_3NH)_3PO_4$ in the presence of $ZnCl_2$ as the catalyst afforded the corresponding 3′-*O*-methyl GDP **4** with 65% yield.[19] The desired $m_2^{7,3'O}GDP$ **5** to make $m_2^{7,3'O}G[5']ppp[5']m^7G$ **7** was obtained by the methylation of 3′-*O*-methyl GDP **4** using dimethyl sulfate as the methylating agent with 78% yield. It is noteworthy that the methylation reaction is highly regioselective affording single m^7 methylated product, $m_2^{7,3'O}GDP$ **5**.[20] Treatment of $m_2^{7,3'O}GDP$ **5** with m^7ImGMP **6** in the presence of $ZnCl_2$ as the catalyst furnished the final $m_2^{7,3'O}G[5']ppp[5']m^7G$ **7** with 61% yield.[21,22]

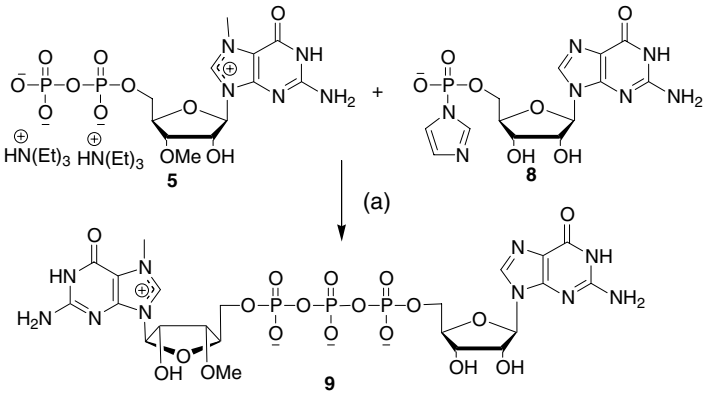
The preparation of $m_2^{7,3'O}G[5']ppp[5']G$ **9** is depicted in Scheme 2. The coupling reaction of $m_2^{7,3'O}GDP$ **5** with ImGMP **8** took place in the presence of $ZnCl_2$ as the catalyst to form compound **9** with 70% yield.[21] The structure of compounds **7** and **9** was confirmed by ^1H and ^{31}P NMR and mass data.

In order to determine the capping efficiency of the methylated cap variants, compounds **7**, **9**, and **10** and conventional mCAP, were next tested in an in vitro transcription system. The template was a pTri β actin vector from the MAXIscript® Kit (Ambion, Inc.) and the products were analyzed by a gel shift assay.[23] Under the reaction conditions, of the four NTPs, only ATP and GTP were used, while CTP and UTP were omitted from the transcription reaction. Due to this omission only the six nucleotides at the 5′ end were transcribed by T7 RNA polymerase, producing a transcript of a short enough length to distinguish whether the cap or regular G had been incorporated. During the transcription reaction, GTP was included along with compounds **7**, **9**, **10** or conventional mCAP in presence of (α-^{32}P) ATP, and the control reaction was performed without any cap analog. The resulting transcription products (6-mer RNAs) were analyzed by 20% denaturing polyacrylamide/8 M urea gel electrophoresis. The outcome of the gel shift assay, shown in Figure 1, indicates that the capped RNAs migrate slower than uncapped RNA, allowing determination of the relative incorporation of cap versus unmodified G. All reactions were performed in triplicate and the capping efficiency was determined by quantitating the intensities of capped versus uncapped RNA using phosphorimagery. From these gel shift assays we calculate that conventional cap $m^7G[5']ppp[5']G$ has a capping efficiency of 61%, while two-way cap variant, compound **10**, has a 56% capping efficiency, the ARCA, compound **9**, has 52% capping efficiency, and compound **7**, which has a N^7 methyl on both guanine of the cap structure and a 3′-OCH_3 substitute, has a 51% capping efficiency. The capping efficiency experiment clearly indicates that the methylated cap variants were substrate for T7 RNA polymerase, although the capping efficiency of compounds **7** and **9** were slightly lower compared to the conventional mCAP. This could be because of the modification and the polymerase can still fit the dinucleotide in either orientation, but since only one led to a productive complex, the aborted complexes decrease yield or at least the speed of production.

To study the ability of these methylated cap variants to stimulate translation, all four compounds were used in a transcription reaction to generate a luciferase message.[24] The supercoiled plasmid, AmbLuc Poly(A) (Ambion, Inc.), used as directed, creates a luciferase mRNA with a 60-base poly(A) tail. This in vitro transcription was performed with compounds **7**, **9**, **10**, and conventional mCAP, along with a control reaction which contained no cap. The transcript containing 5′-capped and -uncapped mRNAs was purified by using the MEGAclear™ Kit (Ambion, Inc.). Transcripts produced with T7 RNA polymerase using compounds **7**, **9**, and **10** were of the predicted length (1.85 kb) and indistinguishable in size and homogeneity from those produced with either conventional mCAP or no cap analog. Analysis of each transcribed mRNA with poly(A) tail was performed on RNA chips in the Agilent 2100 Bioanalyzer, which revealed that no mRNAs were degraded and all were of similar size (data not shown). The Yield of the T7 RNA polymerase transcription reaction with compounds **7**, **9**, and **10**, conventional mCAP, and no cap analog is shown in Figure 2. All methylated



Scheme 1. Reagents and conditions: (a) $POCl_3$, $(OMe)_3P$, 0 °C, 2 h; (b) imidazole, aldrithiol, PPh_3, DMF, rt, 5 h; (c) $(Et_3NH)_3PO_4$, DMF, rt, 3 h; (d) dimethyl sulfate, water, pH 4.0, rt, 4 h; (e) $ZnCl_2$, DMF, rt, 6 h.



Scheme 2. Reagents and conditions: (a) $ZnCl_2$, DMF, rt, 6 h.

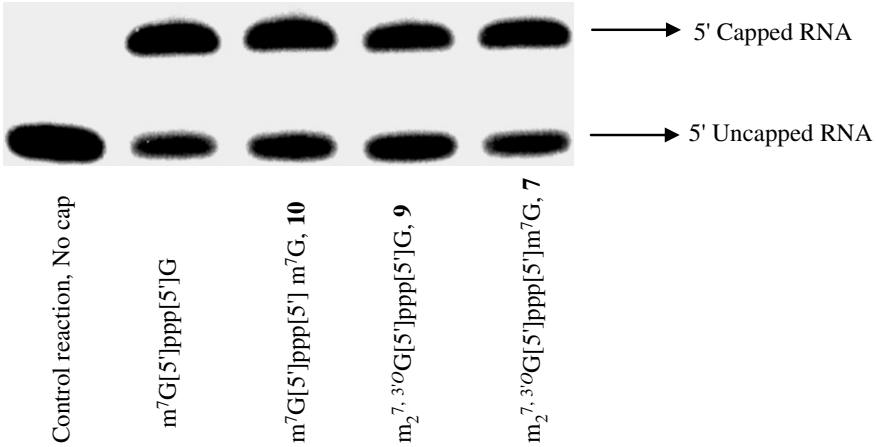


Figure 1. dPAGE gel (20%) showing capping efficiency of compounds **7**, **9**, and **10** along with conventional cap analog. The control reaction was normal in vitro transcription reaction, in which no cap analog was added. Radiation in the gel bands of interest was quantified by a phosphorimager (GE Healthcare).[23]

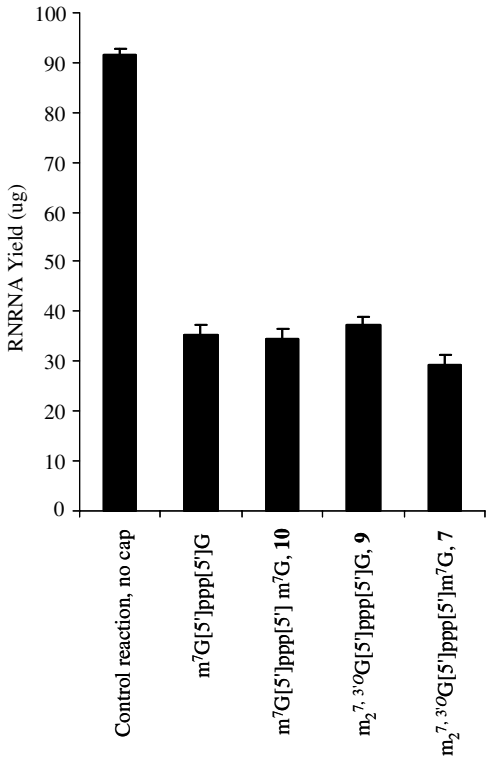


Figure 2. Yield of T7 RNA polymerase transcription reaction with methyl-substituted cap variants.[24]

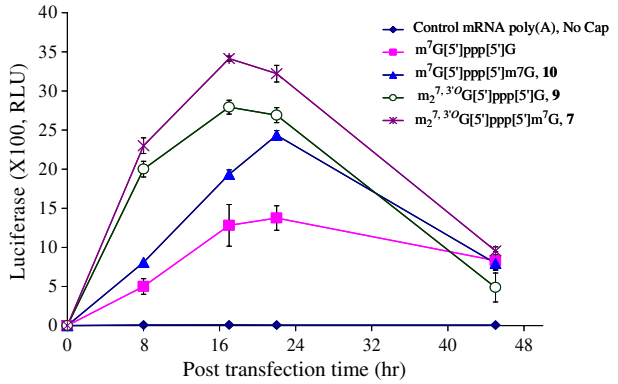


Figure 3. Luciferase activity of methylated cap variants from compounds **7**, **9**, and **10**, and conventional mCAP. Translation efficiency of compound **7**, **9**, and **10**, and conventional mCAP-capped luciferase RNA poly(A). Comparison of protein expression from conventional mCAP and methylated cap variants capped luciferase RNAs poly(A). The conventional mCAP, and modified methyl-substituted cap variants capped in vitro transcribed poly(A) tailed luciferase RNA (400 ng) were transfected into HeLa cells. The fold induction of luciferase protein data was normalized to the control reaction, that is, no cap, mRNA poly(A) transfection results.[25]

cap variants produced optimum yields as compared to conventional mCAP.

The firefly luciferase mRNA expressed by the AmbLuc Poly(A) vector will direct synthesis of a functional luciferase protein, which can be quantitated with a luminescence assay. The RNA product generated from the above transcription reactions was next transfected into HeLa cells[25] to determine the functionality, whether the presence of methyl groups in cap variants **7**, **9**, and **10**, and conventional mCAP influenced the amount of protein expressed, cells were harvested and lysed at 8, 17, 22, 45 h post-transfection. Luciferase activity was measured in each extract using the luciferase assay reagent (Promega) and the recommended procedure. Luciferase activity data were normalized to the control reaction—that is, no cap, poly(A) transfection results. Comparison of capped luciferase RNAs with poly(A) tail luciferase expression at different time points after post transfection from conventional mCAP, and methyl-substituted cap variants **7**, **9**, and **10** is illustrated in Figure 3. All the modified caps outperformed the regular cap in terms of serving as a translation template.

In conclusion, we report biological evaluation of a trimethyl-substituted cap variant with respect to its capping efficiency and impact on translation activity. The methyl-substituted cap variants, when used in appropriate in vitro transcription protocols, allow the generation of 5′-capped poly(A) mRNA. The yield from the transcription reactions in the presence of methyl-substituted cap variants **7**, **9**, and **10** was in agreement with conventional mCAP, except that compounds **7** and **9** was slightly lower, presumably due to the forced incorporation direction. The methyl-substituted cap analogs **7**, **9**, and **10** were compatible with luciferase enzyme production from mRNAs incorporating them and therefore useful for intracellular or in vitro translation experiments. The high translation activity could be attributed from the chemical replacement of the 3′-hydroxyl group of N^7-methylguanosine with a 3′-*O*-methyl moiety, which produces only forward capped mRNA poly(A) transcript with cap variants **9** and **10**. Compound **7** has shown even better translation efficiency, presumably because of its trimethyl substitution, which also produces homogeneously capped transcript and also the extra N^7-methyl group. These modifications at ribose and N^7-methyl group might be increasing its stability of transcribed mRNAs in cell culture. The methyl-substituted cap variants **7**, **9**, and **10** were all readily available substrates for cap-dependent translation, and their translation efficiency was ~2.6-, 2.2-, and 1.6-fold more than mRNA poly(A) capped with conventional $m^7G[5']ppp[5']G$, respectively. The high-translation activity for $m^7G[5']ppp[5']m^7G$ **10** could be due to the fact that it is a symmetrical 'two-way' cap variant, which has two methyl groups at N^7 position that provides an identical cap structure and will be both incorporated by RNA polymerase during transcription and recognized by the initiating machinery during translation regardless of its orientation. The compound **10** has a modification at both N^7positions, but lacks the modification at ribose, as a result it might be producing a transcript which is less stable as compared to compounds **7** and **9**.

Acknowledgments

We thank the NIH (SBIR Phase II: R44GM070156-02) for supporting this research. We are thankful to Rick

Conrad and Shirley A. Recipon (Applied Biosystems) for commenting and critical reading of the manuscript. We also thank Timothy Barta and Angie Cheng for their valuable assistance in gel electrophoresis.

References and notes

1. Shatkin, A. J. *Cell* **1976**, *9*, 645.
2. Both, G. W.; Furuichi, Y.; Muthukrishnan, S.; Shatkin, A. J. *Cell* **1975**, *6*, 185.
3. Both, G. W.; Furuichi, Y.; Muthukrishnan, S.; Shatkin, A. J. *J. Mol. Biol.* **1976**, *104*, 637.
4. Banerjee, A. K. *Microbiol. Rev.* **1980**, 175.
5. Furuicht, Y.; Shatkin, A. J. *Adv. Virus Res.* **2000**, *55*, 135.
6. Reddy, R.; Singh, R.; Shimba, S. *Pharm. Therap.* **1992**, *54*, 249.
7. Wei, C. M.; Gershowitz, A.; Moss, B. *Cell* **1975**, *4*, 379.
8. Rottman, F. M.; Shatkin, A. J.; Perry, R. P. *Cell* **1974**, *3*, 197.
9. Sachs, A. B.; Varani, G. *Nat. Struct. Biol.* **2000**, *7*, 356.
10. Pasquinelli, A. E.; Dahlberg, J. E.; Lund, E. *RNA* **1995**, *1*, 957.
11. Stepinski, J.; Waddell, C.; Stolarski, R.; Darzynkiewicz, E.; Rhoads, R. E. *RNA* **2001**, *7*, 1486.
12. Peng, Z. H.; Sharma, V.; Singleton, S. F.; Gershon, P. D. *Org. Lett.* **2002**, *4*, 161.
13. Jemielity, J.; Fowler, T.; Zuberek, J.; Stepinski, J.; Lewdorowicz, M.; Niedzwiecka, A.; Stolarski, R.; Darzynkiewicz, E.; Rhoads, R. E. *RNA* **2003**, *9*, 1108.
14. Fischer, U.; Darzynkiewicz, E.; Tahara, S. M.; Dathan, N. A.; Luhrmann, R.; Mattaj, I. W. *J. Cell Biol.* **1991**, *113*, 705.
15. Hamm, J.; Darżynkiewicz, E.; Tahara, S. M.; Mattaj, I. W. *Cell* **1990**, *62*, 569.
16. Kore, A. R.; Shanmugasundaram, M.; Charles, I.; Cheng, A. M.; Barta, T. J. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5295.
17. Kore, A. R.; Parmar, G.; Reddy, S. *Nucleosides, Nucleotides, Nucleic Acids* **2006**, *25*, 307.
18. Yoshikawa, M.; Kato, T.; Takenishi, T. *Tetrahedron Lett.* **1967**, *50*, 5065.
19. Kore, A. R.; Parmar, G. *Synth. Commun.* **2006**, *36*, 3393.
20. Kore, A. R.; Parmar, G. *Nucleoside, Nucleotides, Nucleic Acids* **2006**, *25*, 337.
21. Kadokura, M.; Wada, T.; Urashima, C.; Sekine, M. *Tetrahedron Lett.* **1997**, *38*, 8359.
22. *Typical procedure to make 7*. To a stirred solution of $m_2^{7,3'O}$GDP **5** (0.08 g, 0.12 mmol) and m^7ImGMP **6** (0.08 g, 0.18 mmol) in 10.0 mL dry DMF, zinc chloride (0.033 g, 0.24 mmol) was added under nitrogen atmosphere and the reaction mixture was stirred at rt for 6 h. After 6 h, the reaction mixture was added to a solution of EDTA disodium (0.18 g, 0.48 mmol) in 100.0 mL of water at 0 °C. The resulting aqueous solution was adjusted to pH 5.5 and loaded on a DEAE Sephadex column. The desired product was eluted using a linear gradient of 0–1 M TEAB and the fractions containing the product were pooled, evaporated, and concentrated to 10.0 mL TEA salt of **7**. The resulting 10.0 mL was passed through a Strata-X-AW column and washed with 10.0 mL water followed by 10.0 mL MeOH. Then, the desired compound was eluted with 15.0 mL of $NH_4OH/MeOH/H_2O$ (2/25/73) and the collected solution was evaporated and dried to give a fine white powder **7**. (Yield: 0.065 g, 61%). *Data for* **7**. ^{1}H NMR (D_2O, 400 MHz) δ 6.10 (m, 2H), 4.65 (m, 1H), 4.52–4.17 (m, 9H), 4.13 (s, 3H), 4.12 (s, 3H), 3.51 (s, 3H); ^{31}P NMR (D_2O, 162 MHz) δ −10.12 (d, J = 19.6 Hz), −10.18 (d, J = 19.3 Hz), −21.42 (t, J = 18.5 Hz); MS (*m/z*): 831 $[M]^+$.
23. *Gel shift assay*. Gel shift assay was performed by using MAXIscript kit (Ambion, Inc.) by following manufacturer's protocol. Typical 20 μL of T7 RNA polymerase transcriptions contained the following reagents at the final concentrations indicated: linearized pTri β actin vector template, 0.5 μg; ATP, 2 mM; GTP, 0.4 mM; methylated cap variants **7**, **9**, and **10**, and mCAP, 1.6 mM each in separate reaction; 10× reaction buffer, 4×; T7 RNA polymerase, 50 U/μL; (α-^{32}P) ATP 800 Ci/mmol, and DEPC water. The control reaction was normal in vitro transcription reaction, in which no cap analog was added. The transcription reactions were incubated at 37 °C for 2 h, after which the reaction mixtures (10 μL) were then applied to a 20% dPAGE gel. Radiation in the gel bands of interest were quantified by a phosphorimager (GE Healthcare).
24. *T7 RNA polymerase transcription*. T7 RNA polymerase transcription was performed by using MEGAscript kit (Ambion, Inc.) in 20 μL final volume and contains the following reagents at the final concentrations indicated: linearized AmbLuc Poly(A) DNA, 1.0 μg; 1× reaction buffer; ATP, UTP, and CTP, 7.5 mM each; while GTP was compensated with methylated cap variants **7**, **9**, and **10**, and along with conventional mCAP. GTP, 1.5 mM; methylated cap variants **7**, **9**, and **10** and along with conventional mCAP, at 6.0 mM; and 50 U/μl of T7 RNA polymerase. The transcription reaction mixtures were incubated at 37 °C for 2 h. In order to hydrolyze the remaining plasmid DNA, 1 μL of turbo DNAse was added to the reaction mixture and further incubated at 37 °C for 15 min. Purification of transcription reactions was done by using MEGAclear™ Kit (Ambion, Inc.) as per manufacturer's protocol.
25. *Luciferase assay*. HeLa cells (60,000/well in 24 well-plates) were seeded at least 12 h before transfection in growth medium without antibiotics. A complex of capped RNA was prepared by mixing 600 ng of RNA, 2.5 μL of TFX-20 (Promega), and 300 μL of serum-free DMEM in polystyrene tubes and incubated for 15 min at room temperature. After the incubation, media from the pre-plated HeLa cells were removed and 200 μL of the complex was added to each well. The plates were incubated for 1 h at 37 °C, and then 1 mL of pre-warmed media with serum was added. The transfected plates were incubated at 37 °C. Cells were harvested and lysed at 8, 17, 22, 45 h post-transfection. The cells are harvested by removing the media and adding 100 μL of 1× passive lysis buffer (Promega). The plate was mixed carefully to disrupt the cells and 10 μL of cell lysates from each transfections was mixed with 100 μL of luciferase substrate (Promega) and measured immediately on POLARstar OPTIMA Luminometer (BMG Labtech) in 96-well plates. The fold induction of luciferase protein data was normalized to the control reaction, that is, no cap, mRNA poly(A) transfection results.