ORGANIC LETTERS

2002
Vol. 4, No. 2
161−164

# Synthesis and Application of a Chain-Terminating Dinucleotide mRNA Cap Analog

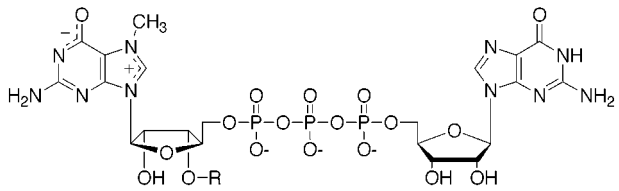
**Zheng-Hong Peng, Vivek Sharma, Scott F. Singleton,[†] and Paul D. Gershon***

*Institute of Biosciences and Technology, Texas A&M University System Health Science Center, Houston, Texas 77030, and Department of Chemistry and Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005*

*pgershon@tamu.edu*

**Received September 18, 2001**

## ABSTRACT




**We describe the synthesis of a chain-terminating mRNA cap dinucleotide and its use in the in vitro transcription of homogeneously capped RNA. Computer modeling strongly indicates that RNA capped with the new compound will be a substrate for cap-dependent translation.**

The mRNA cap structure, comprising an $N^7$-methylguanosine ($m^7G$) residue joined to the 5′ end of the mRNA via a 5′-5′ triphosphate linkage, serves a number of central roles in eukaryotic gene expression including the targeting of mRNA to the intracellular mRNA translation and degradation machineries. In vivo, the cap is added enzymatically. However, over the past 20 years or so, numerous studies have required the synthesis of proteins in an in vitro translation extract supplemented with in vitro synthesized mRNA. The prevailing method for the in vitro synthesis of capped mRNA employs a preformed dinucleotide of the form $m^7G(5')ppp(5')G$ (**6**) as an initiator of transcription. A disadvantage of using this pseudosymmetrical dinucleotide has always been the propensity of 3′OH of either the G or $m^7G$ moiety to serve as the initiating nucleophile for transcriptional elongation[1] leading to the synthesis of two isomeric RNAs of the form $m^7G(5')pppG(pN)_n$ and $G(5')ppp^7G(pN)_n$, in approximately equal proportions depending upon the ionic conditions of the transcription reaction.[2] This may be problematic for various downstream processes, such as in vitro translation or crystallization studies.[3,4] In the current study, we have synthesized a "chain-terminating" dinucleotide cap analogue, 3′-*O*-Me-$m^7G(5')$-pppG (**5**) in which the $m^7G$ moiety is doubly methylated: in addition to the characteristic guanosine $N^7$-methyl substituent, a methyl group is included at the 3′-OH of the $m^7G$ ribose. We demonstrate that this doubly methylated cap analogue supports efficient transcriptional initiation with only the desired forward-oriented cap.

[†] Rice University.

(1) A templating C residue is universally employed by phage RNA polymerase promoters at the site of transcription initiation. See: Dunn, J. J.; Studier, F. W. *J. Mol. Biol.* **1983**, *166*, 477−535. Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783−8798.

(2) Pasquinelli, A. E.; Dahlberg, J. E.; Lund, E. *RNA* **1995**, *1*, 957−967.

(3) In one notable instance, contaminating amounts of the $G(5')ppp^7G(pN)_n$-form isomer were inhibitory to cocrystallization of vaccina protein VP39 with purified $m^7G(5')pppG(pN)_n$-form RNA (Gershon, P. D., unpublished observation). Rigorous purification of the $m^7G(5')pppG(pN)_n$-form using RP-HPLC is possible only for RNAs ≤ ∼6 nt in length). See ref 4.

(4) Lockless, S. W.; Cheng, H.-T.; Hodel, A. E.; Quiocho, F. A.; Gershon, P. D. *Biochemistry* **1998**, *37*, 8564−8574.

10.1021/ol0167715 CCC: $22.00 
**Published on Web 12/20/2001**

Prior to synthesizing the dinucleotide, we addressed the possibility that the 3′-OMe group might compromise interaction with the m$^7$G cap-binding proteins VP39 (a cap-specific mRNA 2′-*O*-methyltransferase from vaccinia virus) and eukaryotic translation initiation factor eIF4E. In the cocrystal structures of VP39 with capped RNA[5] and eIF4E with m$^7$GDP,[6] the m$^7$G moiety of each ligand was 3′-*O*-methylated in silico, and the resulting structures were subjected to energy minimization (Figure 1). For the VP39-capped RNA complex, the 3′-OMe group experienced only a single, minor van der Waals steric clash with the protein—with Ala27 (Figure 1B, inset). For the eIF4E·m$^7$GDP complex, the 3′-OMe group experienced no significant close contact with any residue of the protein (Figure 1D, inset). As a result of these experiments we expected no significant energetic barrier to the interaction of **5**-capped RNA with either of the two proteins, and we therefore proceeded with the synthesis of **5**.

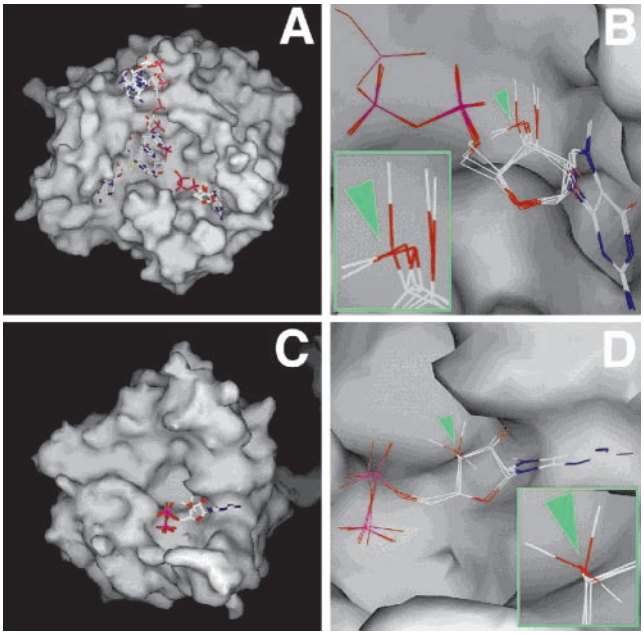
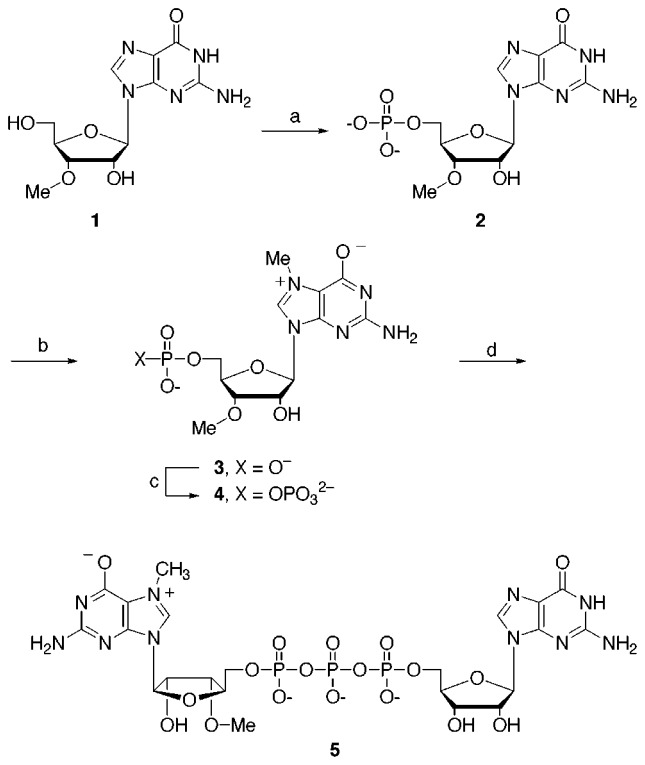



**Figure 1.** Screening for potential steric clashes of added 3′-*O*-methyl group by free energy minimization. Panels A, B: Solvent-accessible surface of VP39 with bound *S*-adenosyl methionine and capped RNA. Panels C, D: Solvent-accessible surface of yeast eIF4E with bound m$^7$GDP. Panels B and D are close-ups taken from A and C, respectively, showing m$^7$GDP moieties. Insets in B, D: Extreme close-ups showing calculated rotamers of the 3′-OMe group. Bound ligands (stick form): *N*, blue; *O*, red; *P*, mauve. Equal spacing of rotamers indicates an absence of significant energetic barriers to free rotation (panel D, inset); nonpenetration of rotameters into a rotational sector (panel B, inset) indicates the presence of an energetic barrier.

**Scheme 1**[a]



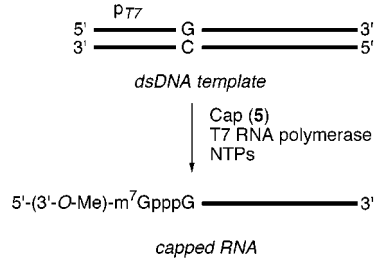

[a] (a) $POCl_3$, $PO(CH_3)_3$, 0 °C, 50%; (b) MeI, MPD, 99%; (c) (i) carbonyldiimidazole, DMF; (ii) $Bu_3NH{\cdot}H_2PO_4$, DMF, 30% for two steps; (d) GMP-morpholidate, tetrazole, DMSO, 72 h, 60%.

Compound **5** was prepared according to Scheme 1. Following chemical phosphorylation of 3′-*O*-methylguanosine with $POCl_3$,[7] the pyridinium salt of 3′-*O*-methylguanosine-5′-monophosphate was regioselectively methylated at $N^7$ using MeI.[8] Next, the tributylammonium salt of 3′-*O*-Me-m$^7$GMP was further phosphorylated to dimethylated GDP with tributylammonium orthophosphate in the presence of carbonyldiimidazole.[9] Finally, the triethylammonium salt of the resulting 3′-*O*-Me-m$^7$GDP was condensed[10] with activated[11] GMP-morpholidate. When carried out essentially as described, the coupling did not proceed to completion in 1 week. However, the addition of 1*H*-tetrazole[12] allowed the reaction to reach the same point after only 2.5 days. After DEAE-Sephadex A25 chromatography and counterion exchange, 17 mg of purified **5** (sodium salt) was obtained.

The newly synthesized dinucleotide cap analogue (sodium salt)[13] was used to prime the in vitro synthesis of two species of short, G-less RNA (of the form **5**-$(pA)_3$ and **5**-$(pA)_5$) by T7 transcription of the corresponding templates (Scheme 2).

(5) Hodel, A. E.; Gershon, P. D.; Quiocho, F. A. *Mol. Cell* **1998**, *1*, 443−447.
(6) (a) Marcotrigiano, J.; Gingras, A.-C.; Sonenberg, N.; Burley, S. K. *Cell* **1997**, *89*, 951−961. (b) Matsuo, H.; Li, H.; McGuire, A. M.; Fletcher, C. M.; Gingras, A. C.; Sonenberg, N.; Wagner, G. *Nat. Struct. Biol.* **1997**, *4*, 717−724.
(7) Yoshikawa, M.; Kato, T.; Takenishi, T. *Tetrahron Lett.* **1967**, *50*, 5065−5068.
(8) Fukuoka, K.; Suda, F.; Suzuki, R.; Ishikawa, M.; Takaku, H.; Hata, T. *Nucleosides Nucleotides* **1994**, *13*, 1557−1567.
(9) Hoard, D. E.; Ott, D. G. *J. Am. Chem. Soc.* **1965**, *87*, 1785−1788.
(10) Adam, A.; Moffatt, J. G. *J. Am. Chem. Soc.* **1966**, *88*, 838−842.
(11) Moffatt, J. G.; Khorana, H. G. *J. Am. Chem. Soc.* **1961**, *83*, 649−658.
(12) Wittmann, V.; Wong, C.-H. *J. Org. Chem.* **1997**, *62*, 2144−2147.

**Scheme 2.** Preparation of Capped mRNA




Control reactions were primed with the standard dinucleotide cap analogue, **6**. The products of each of the four reactions were fractionated by $C_{18}$ RP-HPLC (Figure 2), and their

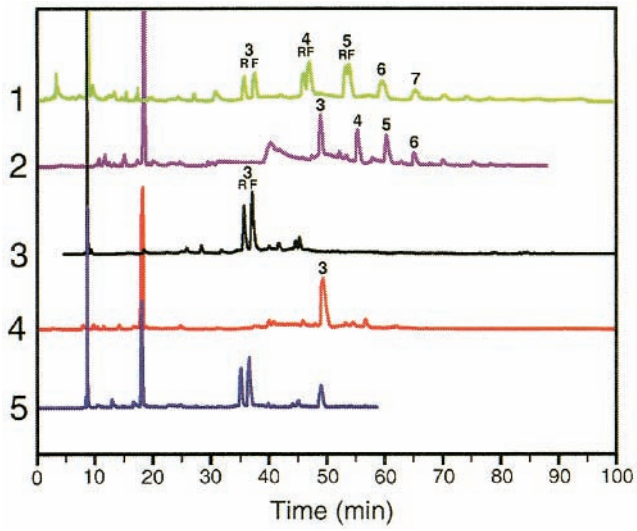



**Figure 2.** $C_{18}$ RP-HPLC fractionation of short G-less in vitro transcripts. Traces 1 and 2: Transcripts generated from a template encoding the RNA $G(pA)_5$ in the presence of cap analogues **6** and **5**, respectively. Traces 3 and 4: Transcripts generated from a template encoding the RNA $G(pA)_3$ in the presence of cap analogues **6** and **5**, respectively. Trace 5: As in traces 3 and 4, except that the cap analogues were premixed in equal concentration. Numbers above peak clusters: Value of *n* in cap-$(pA)_n$. For each value of *n*, F and R denote forward- and reverse-capped RNAs, respectively, after ref 4. At each value of *n*, **5**-$(pA)_n$ RNAs eluted later in chromatograms than the corresponding **6**-$(pA)_n$ ones, presumably as a result of the hydrophobicity of the additional methyl group. For visual clarity, the vertical axes of some traces have been rescaled slightly so that peaks are comparable in height from one trace to another.

identities were established by ESI-MS. Within each cluster[14] of products in the **6**-primed reactions (Figure 2, traces 1 and 3), the earlier- and later-eluting peaks comprised the reverse- and forward-capped RNA isomers $G(5')pppm^7G(pA)_n$ and $m^7G(5')pppG(pA)_n$, respectively.[4] Notably, for the **5**-primed products (Figure 2, traces 2 and 4), each cluster was reduced to a single peak consistent with our blockage of the $N^7$-methylguanylate ribose by 3′-*O*-methylation, permitting the generation of only forward-capped products of the form **5**-$(pA)_n$.

To determine whether **5** and **6** were incorporated with equivalent efficiency during transcription, a reaction was performed in which the two cap dinucleotides were present at equal concentration (Figure 2, trace 5). Quantitation of areas under the peaks revealed that **5** was incorporated approximately 6-fold less efficiently than **6**. Importantly, there was no apparent overall inhibition of transcription, indicating that the **5** preparation did not possess a dominant transcriptional inhibitor. Moreover, tobacco acid pyrophosphatase treatment of both **6** and **5** followed by MALDI-TOF MS confirmed that identical guanosine moieties (the only portion that is essential to transcription initiation) were present in the two caps. These data suggest that that the lower activity of **5** likely arises from a nontransferable inhibitor such as an inappropriate counterion or the absence of an appropriate one.

The above data were obtained using a simple model transcription system in which the transcripts were short and contained no G residues at internal positions. For in vitro translation and other studies, much longer, G-containing capped transcripts would typically be required. The synthesis of RNA containing internal G residues requires the presence of GTP in the transcription reaction, which in turn leads to an unavoidable competition at the initiating position between GTP and cap dinucleotide. We dissected these "real world" in vitro transcription conditions by generating full-length (1762 nt) transcripts of the luciferase gene under conditions where **5** or **6** was permitted to compete for initiation with guanosine 5′-monophosphate (GMP), a non-cap initiator nucleotide. In each reaction, overall initiator concentration (GMP + **5** or GMP + **6**) remained constant. The concentration of GTP (included to permit transcript elongation) was kept ~160-fold lower than that of initiator to avoid competition for initiation[15] and remained constant in all reactions. The extent of capping of the internally labeled transcripts was then assayed by annealing the transcripts to complementary oligonucleotides, cleaving off the transcripts' extreme 5′ ends using RNaseH, and resolving capped from uncapped 5′ ends by PAGE (Figure 3). Although reactions with unannealable control oligonucleotides showed small amounts of degradation yielding, in some instances, an apparent band of mobility similar to that of the uncapped 5′ end fragments ("Anneal -" lanes, Figure 3), results with the

(13) Prior to its use in transcription, **6** was further purified by strong anion exchange chromatography followed by preparative $C_{18}$ RP-HPLC. The major peak from the latter column appeared homogeneous when a sample of it was rechromatographed over an analytical $C_{18}$ column.

(14) As described previously (ref 4), **6**-primed RNA products from templates for the transcript $G(pA)_5$ (Figure 2 trace 1) eluted from the column as a series of clusters of peaks in which each cluster represented a single RNA chain length. The presence of such a series indicated heterogeneity in the length of the oligo(A) portion of the transcript presumably as a result of polymerase "stuttering" during runoff transcription. **6**-primed products from this template (Figure 2, trace 2) also eluted as a series representing equivalent heterogeneity of the oligo(A) portion. Products from the shorter, $G(pA)_3$ template were homogeneous in length (Figure 2, traces 3 and 4), suggesting that this template was too short to support polymerase stuttering.

(15) Kreig, P. A.; Melton, D. A. *Methods Enzymol.* **1987**, *155*, 397−415.

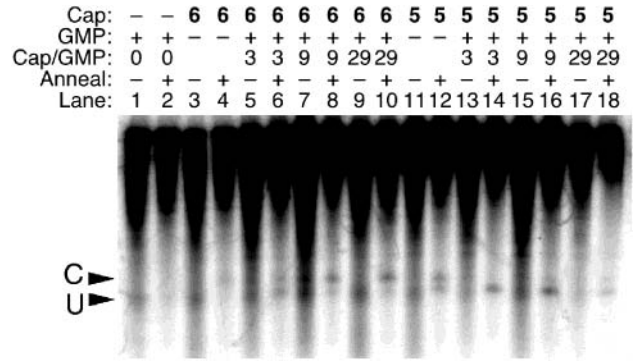



**Figure 3.** Extent of capping of luciferase transcripts generated in transcription reactions containing various mixtures of initiator nucleotide. C = capped, U = uncapped transcript 5′ ends. The cap analog/GMP concentration ratio is indicated by "Cap/GMP" over each lane. Lanes marked "Anneal +" and "Anneal −" indicate the use of transcript-sense (control, unannealable) and -antisense (annealable) oligonucleotides, respectively, in combination with RNase H to produce the 5′ end fragments.

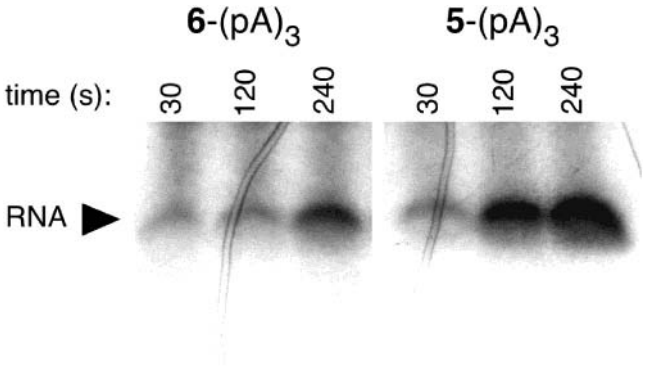



**Figure 4.** Methyltransferase timecourse assays for control RNA substrate **6**-$(pA)_3$ (purified forward-capped RNA) and **5**-$(pA)_3$. An autoradiograph of $^3$H-methylated RNA is shown after incubation of RNA substrates with VP39 and $^3$H-AdoMet under standard assay conditions.

"test" oligonucleotide were clear. The presence of **6** in 3-, 9-, and 29-fold molar excess over GMP yielded capping rates of 30%, 75%, and >90%, respectively (Figure 3). Thus, regular cap dinucleotide was a reasonable competitor of GMP for initiation. However, **5** was incorporated less efficiently. Thus, **5**-capped transcripts were detectable in reactions where **5** was present in at least 9-fold molar excess over GMP. In the reactions containing a 9- or 29-fold molar excess of **5** over GMP, capping rates of 5% and 25%, respectively, were observed (Figure 3). Comparing results at the 29:1 cap to GMP ratio, the competitive efficacy of our preparation of **5** for initiation of transcription appeared to be around 5- to 10-fold poorer than that of **6** in the initiation of a long G-containing transcript. This was comparable to the 6-fold difference observed in a direct competition between the two caps for the initiation of short $(pA)_n$ transcripts (Figure 2, trace 5).

To compare the interaction of the short G-less transcripts of the form **6**-$(pA)_3$ and **5**-$(pA)_3$ with VP39, they were investigated as substrates for the in vitro 2′-*O*-methylation assay (Figure 4). **5**-$(pA)_3$ appeared to be 2′-*O*-methylated by VP39 at least as efficiently as **6**-$(pA)_3$. Apparently, the 3′-OMe group had no deleterious effect on capped RNA-VP39 interaction, consistent with the modeling data (Figure 1). Moreover, since bona fide 7-methylation within the cap is an absolute prerequisite for VP39 substrate activity,[16] this experiment confirmed the presence of an $m^7G$ moiety in **5**.

To summarize, we report the synthesis of a novel dinucleotide cap analogue that, when used in in vitro transcription protocols, allows the generation of homogeneously forward-capped mRNA. Computer modeling strongly indicates that the novel cap is compatible with translation initiation factor eIF4E and therefore useful for in vitro translation experiments.

**Acknowledgment.** We thank Alec Hodel for discussions, Shi-Mei Wang for technical assistance, Ralph Orkizewski (Baylor College of Medicine) for ESI-MS, and Dr. Rowan Chang (University of Texas Institute of Molecular Medicine) for MALDI-TOF. This work was supported by grants from the National Science Foundation (MCB-9604188) to P.D.G. and from the Robert A. Welch Foundation (C-1374) to S.F.S.

**Supporting Information Available:** Synthetic procedures and analytical data, experimental details for transcription and methyltransferase assays, and procedures for computer modeling. This material is available free of charge via the Internet at http://pubs.acs.org.

OL0167715

(16) Hodel, A. E.; Gershon, P. D.; Shi, X.; Wang, S.-M.; Quiocho, F. A. *Nat. Struct. Biol.* **1997**, *4*, 350−354.