Kinetic Dissection of Individual Steps in the Poly(C)-Directed Oligoguanylate Synthesis from Guanosine 5′-Monophosphate 2-Methylimidazolide

Anastassia Kanavarioti,* Claude F. Bernasconi, Diann J. Alberas, and Eldon E. Baird

Contribution from the Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064

*Received May 26, 1992. Revised Manuscript Received March 22, 1993**

Abstract: A kinetic study of oligoguanylate synthesis on a polycytidylate template, poly(C), as a function of the concentration of the activated monomer, guanosine 5′-monophosphate 2-methylimidazolide, 2-MeImpG, is reported. Reactions were run with 0.005–0.045 M 2-MeImpG in the presence of 0.05 M poly(C) at 23 °C. The kinetic results are consistent with a reaction scheme (eq 1) that consists of a series of consecutive steps, each step representing the addition of one molecule of 2-MeImpG to the growing oligomer. This scheme allows the calculation of second-order rate constants for every step by analyzing the time-dependent growth of each oligomer. Computer simulations of the course of reaction based on the determined rate constants and eq 1 are in excellent agreement with the product distributions seen in the HPLC profiles. In accord with an earlier study (Fakhrai, H.; Inoue, T.; Orgel, L. E. *Tetrahedron* **1984**, *40*, 39), rate constants, k_i, for the formation of the tetramer and longer oligomers up to the 16-mer were found to be independent of length and somewhat higher than k_3 (formation of trimer), which in turn is much higher than k_2 (formation of dimer). The k_i ($i \geq 4$), k_3, and k_2 values are not true second-order rate constants but vary with monomer concentration. Mechanistic models for the dimerization (Scheme I) and elongation reactions (Scheme II) are proposed that are consistent with our results. These models take into account that the monomer associates with the template in a cooperative manner. Our kinetic analysis allowed the determination of rate constants for the elementary processes of covalent bond formation between two monomers (dimerization) and between an oligomer and a monomer (elongation) on the template. A major conclusion from our study is that bond formation between two monomer units or between a primer and a monomer is assisted by the presence of additional next-neighbor monomer units. This is consistent with recent findings with hairpin oligonucleotides (Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 317). Our study is the first of its kind that shows the feasibility of a thorough kinetic analysis of a template-directed oligomerization and provides a detailed mechanistic model of these reactions.

Introduction

An experiment done a quarter of a century ago demonstrated that polyadenylate catalyzes the condensation of two molecules of hexathymidylic acid[1] and marked the beginning of template-directed chemistry.[2] Although new to organic chemists, template-directed reactions are used by nature in the reproduction of cells. These syntheses use a poly- or oligonucleotide, acting as the template, to direct the polymerization of nucleoside triphosphates, leading to the formation of the complementary oligonucleotide chain. The complementarity is achieved via Watson–Crick base-pairing between derivatives of guanine (G) and cytosine (C) on one hand, and of adenine (A) with thymine (T) or uracil (U) on the other. On the basis of this notion, the catalysis seen in the synthesis of dodecathymidylic acid from two molecules of hexathymidylic acid was attributed to an increased association of the two oligomers brought about by the action of the polyadenylate template. Although the catalysis detected was much more modest than that observed in the presence of the appropriate enzyme, the effect was large enough to make this reaction interesting in the context of chemical evolution and the RNA world hypothesis.[3]

Since the above original experiment, Orgel and others have repeatedly shown that template-directed reactions are feasible in the absence of enzymes and that RNA and DNA homo- and heteropolymers or analogs thereof catalyze and direct the synthesis of polymers with some degree of information transfer.[4,5] This work has, by and large, been qualitative in nature, although recently several efforts have been directed at a kinetic characterization of these reactions.[6-9] Such kinetic studies can clearly advance our understanding of template-directed reactions and help in designing more efficient systems in a more systematic way.

Oligoguanylate synthesis on a C-template is the most efficient template-directed reaction known to date. A few attempts to determine rates of this reaction have thus far resulted in conflicting reports (see Discussion). In the present study, we investigated the oligomerization kinetics of guanosine 5′-monophosphate 2-methylimidazolide, 2-MeImpG, in the presence of polycytidylate, poly(C), acting as the template at 23 °C.[10] It has been generally assumed that template-directed reactions work because synthesis of the complementary strand is facilitated by the presence of a stable double helix between the template and stacks of activated monomer which hold the monomer in the right conformation for condensation. Oligomerization then proceeds as a zipping-up process within this helix.[2] An alternate scheme is eq 1, where oligomerization is defined as a stepwise process (M stands for the monomer 2-MeImpG; G_2, G_3,...G_i are oligomers of length 2, 3,...i). For simplicity we used the symbols G_i to represent the sum of the 3′–5′, 2′–5′, and pyrophosphate capped

* Abstract published in *Advance ACS Abstracts*, September 1, 1993.

(1) Naylor, R.; Gilham, T. *Biochemistry* **1966**, *5*, 2722.

(2) For reviews, see: Orgel, L. E. *J. Theor. Biol.* **1986**, *123*, 127. Joyce, G. F. *Cold Spring Harbor Symp. quant. Biol.* **1987**, *52*, 41. Orgel, L. E. *Nature* **1992**, *358*, 203.

(3) Gilbert, W. *Nature* **1986**, *319*, 619. Cech, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4360. Cech, T. R. *Sci. Am.* **1986**, *255*, 64.

(4) Inoue, T.; Orgel, L. E. *Science* **1983**, *219*, 859. Schwartz, A. W.; Orgel, L. E. *Science* **1985**, *228*, 585.

(5) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497.

(6) Kanavarioti, A.; White, D. H. *Origins Life* **1987**, *17*, 333.

(7) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 317.

(8) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 5496.

(9) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 7963.

(10) Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201.

0002-7863/93/1515-8537$04.00/0

$$M \xrightarrow{k_2[M]} G_2 \xrightarrow{k_3[M]} G_3 \xrightarrow{k_4[M]} G_4 \longrightarrow G_{i-1} \xrightarrow{k_i[M]} G_i \longrightarrow \xrightarrow{k_n[M]} G_n \quad (1)$$

isomers. $k_2, k_3, \ldots k_i$ are the individual second-order rate constants for dimerization, addition of a monomer to the dimer, or, generalized, for the addition of a monomer to an oligomer of length $i-1$, respectively.

The results to be reported in this paper are, in fact, amenable to an analysis according to eq 1 which allows the determination of all apparent second-order rate constants k_2, $k_3, \ldots k_i$ up to k_{15}. We call them "apparent" second-order rate constants because they are not true constants but depend on monomer concentration. From their dependence on monomer concentration, details about the reaction mechanism will be deduced.

Experimental Section

1. Materials. Solvents used were of HPLC quality. Buffers and other reagents were purchased from Sigma or Aldrich and are abbreviated as follows: 2-methylimidazole, 2-MeIm; guanosine 5′-phosphate, 5′GMP; diguanosine 5′,5′-pyrophosphate, GppG; polycytidylate potassium salt, poly(C), about 100 units long; *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid, HEPES; tris(hydroxymethyl)aminomethane, TRIZMA; (ethylenedinitrilo)tetraacetic acid disodium salt, EDTA. The sodium salt of 2-MeImpG was synthesized according to a known procedure.[11] The preparations were 96% pure, the impurities being 5′GMP and GppG as shown with authentic samples. The preparations contained 0.8–0.9% GppG, measured in guanosine equivalents. The amount of GppG turned out to be very important because experiments performed with 2-MeImpG preparation that included higher amounts of GppG exhibit decreased reactivity.[12] Pancreatic ribonuclease A (RNase A) was purchased from Boeringer Mannheim or Sigma, and phosphodiesterase from *Crotalus durissus* (PDE) from Boehringer Mannheim.

2. Oligomerization of 2-MeImpG in the presence of Poly(C)/Sample Preparation. Samples containing appropriate amounts of materials to make 0.05 M poly(C) in monomer equivalents, 0.2 M $MgCl_2$ and 1.2 M NaCl, were lyophilized in disposable polystyrene culture tubes. A 60-μL solution containing the desired amount of 2-MeImpG in 0.5 M HEPES buffer (pH 7.95 ± 0.05) was then added to initiate the reaction. Temperature was maintained at 23 ± 0.1 °C with a Lauda K-2/R bath by Brinkmann. Samples were quenched at appropriate times with EDTA (175 μL of 0.11 M EDTA, pH 4.2) to chelate Mg^{2+} and diluted with 1100 μL of water. This solution was incubated at 72 °C overnight to hydrolyze the unreacted 2-MeImpG and prevent further oligomerization. The solution was neutralized (155 μL of 1 M TRIZMA, pH 7.6), poly(C) was digested by adding 10 μL of RNase A (3.0 mg/mL), and the samples at an approximately final 25-fold dilution were incubated at 37 °C for 1–2 h. Control experiments with undigested poly(C) showed that no additional reaction takes place after quenching or during overnight storage at −40 °C and that the procedures used for the degradation of unreacted 2-MeImpG as well as poly(C) leave the oligoguanylates intact. This additional workup to digest poly(C) to cytidine 3′-monophosphate, 3′CMP, is necessary because poly(C) elutes in the same region as the longer oligoguanylates, prohibiting the quantitative analysis of oligomers longer than the 8-mer.

3. HPLC Analysis of Samples and Conversion of HPLC Areas to Concentrations. Analysis was performed on a 1090M Hewlett-Packard HPLC, using RPC-5 chromatography at 254 nm.[10,13] The RPC-5 column (25 cm) was packed by us using a slurry packer from Micrometrics Inc. and a Spectra Physics pump working in the purge mode at 4000 psi. Packing material was kindly provided by Dr. L. E. Orgel. Mobile phase: solvent A, 0.01 M NaOH; solvent B, 0.1 M $NaClO_4$ in 0.01 M NaOH; 100% A for 4 min, 0 to 21% B in 16 min and 21 to 40% B in 23 min. The high pH prevents oligo(G) aggregation.[14]

Our quantitative analysis required that both large and small peaks could be integrated accurately. To assess the linearity of our HPLC diode array detector, we constructed calibration curves with 5′GMP sodium salt supplied from Sigma. This compound is of highest purity, stable toward hydrolysis, exhibits a UV–vis spectrum almost identical to that of 2-MeImpG, and hence is suited as a calibration standard. Solutions of 5′GMP in the range of 5×10^{-7} up to 1×10^{-3} M were tested at 254 nm at pH 6 and 12. Analysis was done with isocratic elution: at pH 6 using a C18 5-μm 4.6 × 200-mm column supplied by Hewlett-Packard and a 0.02 M potassium phosphate buffer; at pH 12 using a 4.6- × 200-mm HEMA Bio Q column[15] supplied by Alltech Associates and 0.1 M sodium perchlorate in 0.01 M NaOH buffer as the eluent. In most cases, reproducibility was better than 1%. At pH 6, a linear response was observed, with HPLC areas ranging from 17 to 57 000 HPLC units, whereas at pH 12 the linear range was limited to areas from 4 to 20 000. Note that 4 HPLC units corresponds to 12.7 pmol of 5′GMP. At 254 nm and pH 6, we obtained a conversion factor of 2.54 pmol, and at pH 12 a factor of 3.08 pmol of 5′GMP per HPLC unit. Using the conversion factor at pH 12, we calculated the total HPLC area of a sample attributed to guanosine derivatives, area M_o, from eq 2 and the concentration of an oligoguanylate from eq 3:

$$\text{area } M_o = [M]_o(\text{inj vol in liters})/(3.08 \times 10^{-12})\text{dilution factor} \quad (2)$$

$$[G_i] = h[M]_o(\text{area } G_i)/i(\text{area } M_o) \quad (3)$$

$[M]_o$ is the molarity of the 2-MeImpG solution. The dilution factor corresponds to the ratio final volume/volume of reacting sample. The hypochromicity correction factor, h, was determined at pH 12 and 254 nm.[16] For G_2, $h = 1.34$ was obtained by enzymatic (PDE) hydrolysis of GppG to 5′GMP; for G_i, $i \geq 3$, $h = 1$ was obtained by hydrolysis of poly(G) with KOH and analysis on a HEMA Bio Q column. It was assumed that h for all G_i, $i \geq 3$, is approximately the same.

We also discovered that with RPC-5 chromatography, 5′GMP partially coelutes with the 3′CMP that is formed from poly(C) digestion. This was confirmed by analysis of samples with undigested poly(C) and also by analysis performed on the HEMA Bio Q column that resolves 5′GMP from 3′CMP. However, oligoguanylate analysis on the HEMA column is less satisfactory than that on the RPC-5, so it was necessary to analyze the samples with the latter. Since there was no simple way to correlate the amount present in the sample with the observed HPLC area under the 5′GMP peak, we calculated it for each sample by subtracting the areas of the observed oligomers corrected for hypochromicity from the total HPLC area of the sample (eq 4).

$$\text{area 5′GMP} = \text{area } M_o - \{[(\text{area 2-mer}) \times 1.34] + \sum_{i=3}^{n} \text{area } i\text{-mer}\} \quad (4)$$

4. Computer Simulation of Kinetics (KINSIM). The simulation was performed with KINSIM[17] on a VAX computer. KINSIM can simulate the time course of up to 33 coupled reactions. In our simulation of eq 1, we included the hydrolysis of 2-MeImpG with $k_h = 6.4 \times 10^{-3}$ h^{-1}.[18] Simulations were performed using a Δt of 0.01 h, one iteration per point, and flux tolerance and integral tolerance settings of 1×10^{-7}. These settings were chosen to make the calculation time reasonably short but keep the accuracy better than 99%, as checked by doing calculations with different settings. As an additional check, the calculated concentrations were subjected to the same analysis as the experimental ones and yielded rate constants that were indistinguishable from the ones given to the KINSIM program originally.

5. Computer Simulation of Monomer Distribution on the Template. The monomer distribution on the template for a given fraction of occupied template was approximated by a Monte Carlo procedure in which monomers were accepted at random positions on a template containing 100 binding sites (average length of poly(C)). A statistical distribution

(11) Lohrman, R.; Orgel, L. E. *Tetrahedron* **1978**, *34*, 853.
(12) Kanavarioti, A.; Alberas, D. J.; Baird, E. E., unpublished results.
(13) RPC-5 packing is made by adsorbing Adogen, a commercially available tetraalkylammonium ion, on Kel F or Plaskon of 5–35-μm particle size. RPC-5 chromatography combines both ion-exchange and reverse-phase modes of adsorption (Pearson, R. L.; Weiss, J. F.; Kelmers, A. D. *Biochim. Biophys. Acta* **1971**, *228*, 770).
(14) Selsing, E.; Larson, J. E.; Wells, R. D. *Anal. Biochem.* **1979**, *99*, 213.
(15) Stribling, R. J. *Chromatogr.* **1991**, *538*, 474.
(16) Hurley, T. B. Chemistry Senior Thesis, UCSC, 1993. The observation of $h = 1$ for poly(G) at pH 12 is consistent with the notion that single-stranded poly(G) at 92 °C ($\epsilon_{255} = 11\,800$) has practically the same absorption as 5′-GMP ($\epsilon_{256} = 12\,200$) from ref 31.
(17) Barshop, B. A.; Wrenn, R. F.; Frieden, C. *Anal. Biochem.* **1983**, *130*, 134.
(18) This value refers to the hydrolysis of the free monomer. It was determined by interprolation from an Arrhenius plot at [2-MeImpG] = 1 mM under conditions otherwise identical to the ones used in this study from: Kanavarioti, A.; Chang, S.; Alberas, D. J. *J. Mol. Biol.* **1990**, *31*, 462. The observation that the biomolecular rate constants exhibit no detectable trend as a function of time indicates that [M] is well estimated and suggests that hydrolysis of template-bound or free monomers may not differ markedly.

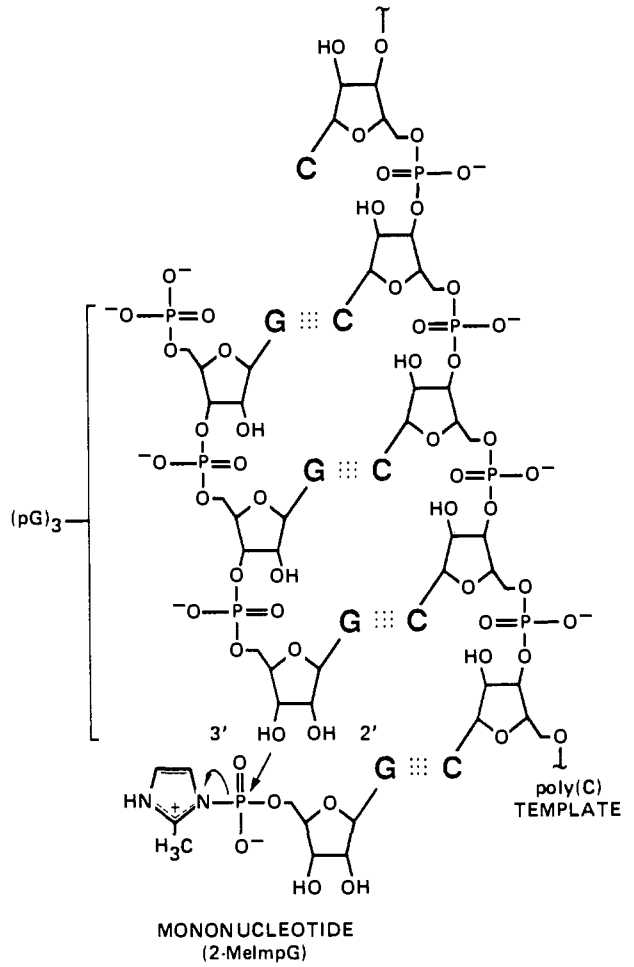


Figure 1. Schematic representation of the elongation of a template-bound trimer (G_3) by reaction of the 3′-OH group with 2-MeImpG.

of monomers on a template was generated with a probability which was selected to give the correct preference of binding adjacent to an already bound monomer (see below). This procedure was repeated 1000 times. The number of stacks of a given size was then summed for all these runs to obtain the relative abundance of the various stacks.

Results

General Features. The best conditions for oligomerization are pH 8.0, 1.2 M NaCl, and 0.2 M $Mg(Cl)_2$.[10] Figure 1 depicts the template effect of poly(C) and the reactive sites for formation of the internucleotide bonds between an oligoguanylate (G_3) and 2-MeImpG. Besides the 3′–5′-linked oligoguanylates, minor products consist of the 2′–5′-linked oligomers formed by reaction with the 2′OH (not detectable in our runs) and the pyrophosphate-capped isomers from reaction at the 5′ end of the elongating strand. Elongation of an oligomer could, in principle, occur at the 3′ and 5′ ends simultaneously, but there is convincing evidence[10] that elongation in the poly(C)·2-MeImpG systems occurs only at the 3′ position.

At concentrations ≤0.001 M, 2-MeImpG undergoes mainly hydrolysis of the P–N bond (eq 5).[19] At higher concentrations,

$$\text{2-MeImpG} \xrightarrow{k_h} \text{5′GMP} + \text{2-MeIm} \qquad (5)$$

dimerization starts to compete with hydrolysis, even in the absence of poly(C), but yields mainly the 2′–5′-linked and pyrophosphate-capped dimers.[20] In the presence of poly(C) and at 2-MeImpG concentrations of ≥0.0025 M, the formation of dimer (mainly 3′–5′-linked) and longer oligomers becomes significant within hours.

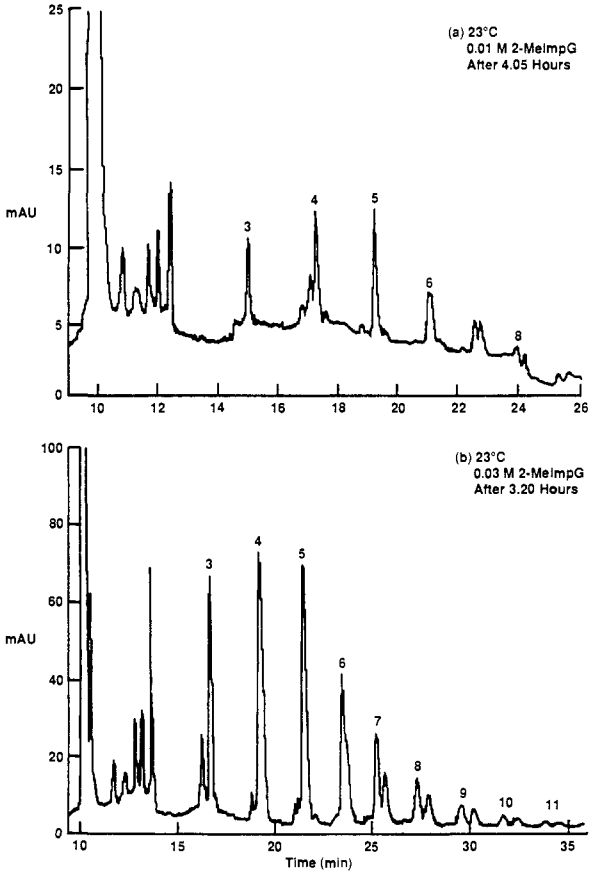


Figure 2. Representative HPLC profiles of the oligomerization of 2-MeImpG for different initial conditions. Numbers refer to length of oligomers.

In order to monitor the progress of these reactions, 6–10 identical samples for each specific set of conditions were prepared as described in the Experimental Section and quenched at regular intervals. Some representative HPLC profiles are shown in Figure 2. They illustrate the increased yield obtained with higher initial monomer concentrations for samples quenched at similar times. Note that in Figure 2b (0.03 M monomer) the y-scale is 0–100 mAU (AU = absorbance units), whereas in Figure 2a (0.01 M monomer) the y-scale is 0–25 mAU. Our observations are in agreement with earlier reports by Orgel and co-workers.[10,21]

The HPLC profiles also show that the 3′–5′ isomers are seen to be dominant for short oligomers, but with increasing length the ratio of the two detectable isomers gradually shifts in favor of the pyrophosphate-capped isomers. This change may be attributed to the fact that an oligomer that contains only 3′–5′ linkages can elongate by making either another 3′–5′ linkage or a pyrophosphate linkage at the 5′ end while pyrophosphate-capped oligomers can elongate only at their 3′ end. A kinetic scheme which accounts for this reactivity pattern has been discussed by Kanavarioti and White.[6]

In this paper we focus on the dependence of the oligomerization on monomer concentration. The reaction was studied with 0.005, 0.008, 0.01, 0.015, 0.02, 0.03, 0.04, and 0.045 M 2-MeImpG, pH 8.0, at 23 °C in the presence of 0.05 M poly(C), 1.2 M NaCl, and 0.2 M $MgCl_2$. The raw data of the HPLC reports are summarized in Table S1.[22] Analysis of the HPLC profiles, as described below, indicates that the system exhibits a surprisingly

(19) Kanavarioti, A. *Origins Life* **1986**, *17*, 85. Kanavarioti, A.; Bernasconi, C. F.; Doodokyan, L. D.; Alberas, D. J. *J. Am. Chem. Soc.* **1989**, *111*, 7247.
(20) Kanavarioti, A.; Hurley, T. B., unpublished results; consistent with ref 11.
(21) Fakhrai, H.; Inoue, T.; Orgel, L. E. *Tetrahedron* **1984**, *40*, 39.
(22) See paragraph concerning supplementary materials at the end of this paper.

Table I. Rate Constants[a] Determined from Experiment with $[M]_0 = 0.04$ M

$t_n - t_{n-1}$[b]	$10^2[M]_{av}$[c]	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}
1.1–0.6	3.67	0.29	26.5	38.2	48.1	53.1	49.5	38.5	43.6					
1.6–1.1	3.58	0.31	25.6	51.3	44.6	49.8	45.7	43.7	47.9					
2.1–1.6	3.49	0.31	23.5	33.0	41.2	43.4	44.4	47.5	49.0	52.9	57.7	38.2		
2.6–2.1	3.39	0.37	23.6	44.6	44.3	42.8	42.3	43.5	43.6	42.1	39.6	44.4	44.3	
3.2–2.6	3.28	0.38	24.0	39.5	40.0	39.7	36.5	36.2	40.7	40.5	37.4	34.2	27.0	30.8
						Average Rate Constants								
		0.33	24.6	41.3	43.7	45.8	43.7	41.9	45.0	45.2	44.9	38.9	35.7	30.8

[a] k_i in units of M^{-1} h^{-1}. [b] Time interval in hours of consecutive quenches. [c] $[M]_{av}$ = average [2-MeImpG] in M during time interval, see text.

Table II. Rate Constants of Individual Steps of 2-MeImpG Oligomerization at pH 7.95 ± 0.05 in the presence of 0.05 M poly(C) in 0.5 M HEPES with 0.2 M $MgCl_2$ and 1.2 M NaCl at 23 °C

	$[2\text{-MeImpG}]_0$,[a] M													
k_i, M^{-1} h^{-1}	0.005	0.005[b]	0.008	0.008[b]	0.01	0.015	0.015[b]	0.02	0.02[b]	0.02[b]	0.03	0.03[c]	0.04	0.045
k_2	0.053[d]	0.038	0.036	0.046	0.11		0.11	0.16	0.19	0.22	0.24	0.26	0.33	0.41[e]
k_3	9.0[d]	32.0	18.0	30.0	29.0	37.6	37.8	24.2	34.6	33.1	26.2	24.8	24.7	
k_i ($i \geq 4$)[f]	49 ± 10	54 ± 10	48 ± 15	64 ± 10	65 ± 7	57 ± 9	63 ± 10	63 ± 7	57 ± 4	58 ± 4	50 ± 5	49 ± 5	44 ± 5	42 ± 5
last oligomer	*11*	*10*	*14*	*16*	*11*	*10*	*12*	*11*	*13*	*15*	*12*	*11*	*14*	*17*

[a] Concentration of substrate preparation given as weighted out, preparation was 96% pure and contained 0.8% GppG, calculated as guanosine equivalents. [b] Experiment repeated. [c] With 1.2 M LiCl instead of NaCl. [d] Calculated assuming $k_4 = 53$. [e] Calculated assuming $k_3 = 30$. [f] Error limits given as standard deviations.

simple kinetic behavior and that each specific experiment that monitors a relatively short time in the oligomerization process can be described satisfactorily by eq 1.

Kinetic Analysis. The kinetic analysis is based on eq 1, for which we can write the rate eqs 6 and 7. Note that we define

$$d[G_i]/dt = k_i[M][G_{i-1}] - k_{i+1}[M][G_i] \approx \Delta[G_i]/\Delta t \quad (6)$$

$$d[G_n]/dt = k_n[M][G_{n-1}] \approx \Delta[G_n]/\Delta t \quad (7)$$

$i = 2, 3, 4, 5...(n-1)$; G_n is the longest oligomer that can be detected in a given experiment, and hence the rate term for its conversion into G_{n+1} is negligible and omitted from eq 7. $[G_i]$, which refers to the sum of the 3′–5′ and pyrophosphate-capped isomers, can be determined from the respective HPLC areas via eq 3.[23] [M] is calculated from eq 8, with the area 5′GMP obtained

$$[M] = [M]_0\{e^{-k_h t}\}(\text{area 5'GMP})\text{purity}/(\text{area } M_0) \quad (8)$$

from eq 4 by adjusting for the purity of the preparation and for the fact that 2-MeImpG slowly hydrolyzes during incubation. The concentrations of M and G_i used in eqs 6 and 7 were taken as average values between the initial and the endpoint of the time interval Δt. With k_n the only unknown in eq 7, it is easily obtained. Using this k_n in eq 6 for $i = n-1$ allows calculation of $k_i = k_{n-1}$. This procedure is repeated for $i = n-2$, $n-3$, etc., all the way back to $i = 2$. Depending on the rates, Δt was between 0.5 and 5 h; this gave the best compromise between a small enough Δt to justify the approximation of $d[G_i]/dt$ with $\Delta[G_i]/\Delta t$ and large enough concentration changes to be evaluated accurately by HPLC. All rate constants except for k_n and k_{n-1} were obtained as an average of 5–6 individual determinations that corresponded to solving eq 6 for a specific k_i at different times during a given run. These calculations were performed using Microsoft Excel on a Macintosh IIci computer.

Table I reports the rate constants determined from an experiment with 0.04 M monomer. The rate constants were determined for each step of eq 1 by comparing consecutive samples and applying eq 7 for the last observed oligomer and eq 6 for all the others in each sample. In view of the approximations that underlie eqs 6 and 7 and the relatively small HPLC areas from which oligomer concentrations had to be calculated, the consistency in the rate constants obtained for a given oligomer at different times is quite remarkable and demonstrates the soundness of the method. Taking averages, we obtain $k_2 = 0.33$ and $k_3 = 25$ M^{-1} h^{-1}; the averages for the k_is ($4 \leq i \leq 11$) are indistinguishable and yield a "global average" of 44 M^{-1} h^{-1} with a standard deviation of less than 5%. For k_{12}, k_{13}, and k_{14}, somewhat lower averages are obtained. This is probably an artifact resulting from the heavy weighting of rate constants calculated from the truncated eq 7. The truncation error is largest for k_{14} but decreases rapidly for k_{13} and k_{12}.

The rate constants obtained from the experiments conducted at other 2-MeImpG concentrations are summarized in Table II (for k_i ($i \geq 4$) only the average values are reported). The following points are noteworthy. (1) Irrespective of the initial monomer concentration, the results invariably show that $k_2 \ll k_3 < k_i$ ($i \geq 4$) and that k_i ($i \geq 4$) for a given $[M]_0$ is essentially independent of i, except when i approaches n, and truncation errors may distort k_i. When this was the case, these k_i values were not included in calculating the global average of k_i ($i \geq 4$). (2) More than one experiment was performed with each $[M]_0$ (not all shown) in order to test the reproducibility and for the effects of varyiations in the total run time and preparation of 2-MeImpG and slight variations in the procedure of sample preparation and analysis. The reproducibility was very good, as seen from a comparison of three runs at 0.02 M monomer: k_i ($i \geq 4$) = 63 ± 7 (up to 11-mer), 57 ± 4 (up to 13-mer), and 58 ± 4 M^{-1} h^{-1} (up to 15-mer). With $[M]_0 = 0.03$ M, we also performed an experiment in the presence of 1.2 M LiCl instead of NaCl. The results show no measurable change in any of the determined rate constants with changing metal ion.

Verification of Rate Constants with Computer Simulation. The reliability of our kinetic analysis was further tested by computer simulations using KINSIM.[17] KINSIM calculates product distributions as a function of time, given initial concentrations and rate constants for each reaction. With few exceptions the discrepancy between observed and simulated concentrations is below 10%. That such low discrepancies strongly support the soundness of our kinetic analysis becomes even more apparent in a simulation based on deliberately altered rate constants. A

(23) A more rigorous kinetic treatment would have to take the changing ratio of 3′–5′ to pyrophosphate-capped isomers into account and base the analysis on Scheme II of ref 6. However, the change in the isomer ratios per elongation step is quite small, indicating that the rate constants for capping at the 5′ end are only about one-tenth of the rate constants for elongation at the 3′ end. Hence, this more rigorous treatment is not warranted at this time, especially in view of the fact that the resolution of the two isomers was only of good quality when a freshly packed RPC-5 column was used. KINSIM simulations with and without the reactions at the 5′ end of the oligomers confirmed that the effect of neglecting those reactions on the calculated rate constants is minimal.

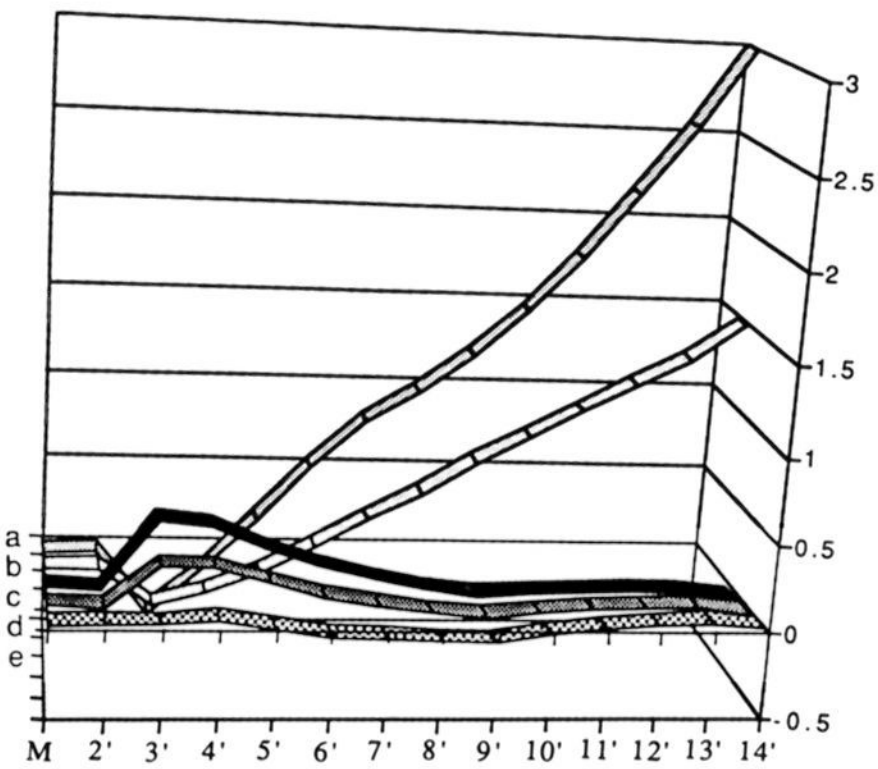


Figure 3. Plot of log $[(pG)_i]_{obs}/[(pG)_i]_{calc}$ as a function of oligomer length for experiment with $[M]_o = 0.04$ M. Line e is based on the reported, the other lines on reduced rate constants. For all lines, $k_3 = 25$ M^{-1} h^{-1}; Lines a, b, $k_2 = 0.33$ M^{-1} h^{-1}; line a, $k_i = 11$ M^{-1} h^{-1}; line b, $k_i = 22$ M^{-1} h^{-1}. Lines c–e, $k_i = 44$ M^{-1} h^{-1}; line c, $k_2 = 0.09$ M^{-1} h^{-1}; line d, $k_2 = 0.18$ M^{-1} h^{-1}; line e, $k_2 = 0.33$ M^{-1} h^{-1}.

comparison between observed and simulated concentrations for every oligomer length up to the 14-mer is shown in Figure 3. Lines a–d are based on reduced values of k_2 and/or k_i, whereas line e is based on the observed rate constants, i.e., $k_h = 6.4 \times 10^{-3}$, $k_2 = 0.33$, $k_3 = 25$, and k_i $(i \geq 4) = 44$ M^{-1} h^{-1}. Line e illustrates the excellent agreement between simulation and experiment. Line b corresponds to a k_i reduced by a factor of 2 and line a a factor of 4. The disparity is seen to become quite severe as the length of the oligomers increases: for the 14-mer this disparity amounts to a factor of about 30 when $k_i = 22$ M^{-1} h^{-1} (line b) and a factor of about 1000 when $k_i = 11$ M^{-1} h^{-1} (line a).

Simulations with either k_2 (Figure 3, see lines c and d) or k_3 (not shown) reduced by a factor of 2 or 4 indicate a much smaller sensitivity of the product distribution to variations in these two rate constants. This means that here the computer simulation is not as helful as for k_i $(i \geq 4)$ in establishing the confidence levels for these rate constants.

Independent Determination of Dimer Formation. Because of the difficulty in obtaining reliable k_2 values via eq 6 at low monomer concentrations, an attempt at evaluating k_2 by a different method was made. It is based on expressing the rate of dimer formation by eq 9 in

$$d[\text{“D”}]/dt = d\sum_{i=2}^{n}[G_i]/dt = k_2[M]^2 \qquad (9)$$

which $[\text{“D”}] = \sum_{i=2}^{n}[G_i]$ is equivalent to the concentration of dimer that would prevail if the reaction were to stop at the dimer stage. Equation 9 is valid because every oligomer was once a dimer; the possibility that some oligomers are formed by ligation of smaller units is remote. This is because the concentration of oligomers was very low under all experimental conditions, thus disfavoring reaction of oligomers with each other. We have calculated [“D”] for every sample from [“D”] = $[G_2]$ (only 3′–5′) + $[G_i]$ $(i \geq 3)$. Amounts of [“D”] as a function of time and monomer concentration are listed in Table S2.[22] Two representative plots of [“D”] vs time are shown in Figure 4. The slight downward curvature reflects a decrease in [M] with time. From the initial slopes of these plots the d[“D”]/dt values were obtained, and these are reported in Table III. The comparison between d[“D”]/dt values and the equivalent $k_2[M]^2_{t=0}$ also reported in the same table indicate that the agreement between the two sets is quite satisfactory considering the approximations involved in the two methods. In further analysis we will use the average of the d[“D”]/dt and $k_2[M]^2_{t=0}$ values, called d[D]/dt for simplicity.

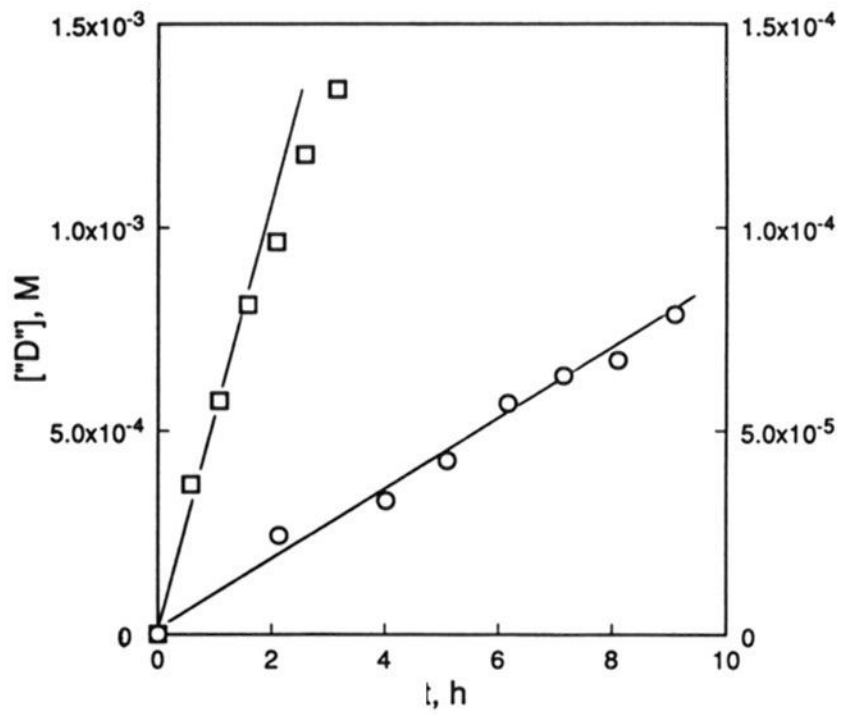


Figure 4. Formation of “D” as a function of time for two representative runs: □, $[M]_o = 0.04$ M, left y-axis; ○, $[M]_o = 0.01$ M, right y-axis.

Table III. Rate Data for Dimerization

$[M]_o$,[a] M	$10^3[M]$,[b] M	$r = [M]/[M]_o$	$k_2[M]^2$ at $t=0$,[c] M h^{-1}	d[“D”]/dt,[d] M h^{-1}	θ[e]	$10^3[M]_T$,[f] M
0.005	4.3	0.86	1.2 ± 10^{-6}	9.0×10^{-7}	0.026	1.31
0.005	4.5	0.89	8.8×10^{-7}	1.1×10^{-6}	0.026	1.31
0.008	6.9	0.86	2.1×10^{-6}	2.8×10^{-6}	0.065	3.25
0.008	7.2	0.90	2.7×10^{-6}	4.1×10^{-6}	0.065	3.25
0.010	9.1	0.91	1.0×10^{-5}	8.0×10^{-6}	0.097	4.85
0.015	13.7	0.91		3.0×10^{-5}	0.181	9.05
0.015	13.9	0.93	2.3×10^{-5}	3.3×10^{-5}	0.181	9.05
0.020	18.1	0.91	5.9×10^{-5}	7.5×10^{-5}	0.271	13.6
0.020	18.1	0.91	7.0×10^{-5}	8.5×10^{-5}	0.271	13.6
0.020	18.1	0.91	8.1×10^{-5}	1.0×10^{-4}	0.271	13.6
0.030	26.8	0.89	2.0×10^{-4}	2.6×10^{-4}	0.461	23.1
0.030 (Li^+)	27.2	0.91	2.2×10^{-4}	2.7×10^{-4}	0.461	23.1
0.040	34.8	0.87	4.9×10^{-4}	5.7×10^{-4}	0.647	32.4
0.045	38.8	0.86	7.6×10^{-4}	8.5×10^{-4}	0.739	37.0

[a] Input concentration of 2-MeImpG. [b] Concentration of 2-MeImpG midway through a kinetic run, corrected for the purity of 2-MeImpG. [c] Equivalent to d[“D”]/dt according to eq 9, see text; k_2 calculated from the analysis according to eq 6, k_2 values reported in Table II; $[M]_{t=0} = 0.96[M]_o$, with 0.96 correcting for the purity of 2-MeImpG. [d] Rate of dimer formation according to eq 9. [e] Fraction of occupied template sites. [f] Concentration of monomer associated with the template.

Discussion

A most remarkable result of our study is that the kinetic behavior of eq 1 is so simple and amenable to analysis in terms of only three second-order rate constants, i.e., k_2, k_3, and k_i $(i \geq 4)$. This apparent simplicity does not, however, mean that the mechanistic details are necessarily simple. Just the fact that the reactions occur on a template rather than in solution implies a complex mechanism. An indication of such complexity is our finding that k_2, k_3, and k_i $(i \geq 4)$ are not true second-order rate constants but depend on initial monomer concentration (Table II).[24]

Mechanism of Dimer Formation. Dimerization off the template leads primarily to 2′–5′-linked and pyrophosphate-capped dimers.[20]

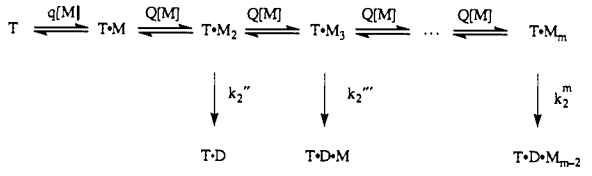
(24) This concentration dependence of the second-order rate constant does not invalidate our kinetic treatment as (pseudo)-second-order processes (eqs 6 and 7). This is because the reactions were only followed over a relatively short period of time, during which [M] did not decrease by more than 15% at high $[M]_o$ and by less than 5% at low $[M]_o$.

Table IV. Analysis of the Dimerization in Terms of the M_2, M_6, and M_{10} Mechanisms[a]

		F_D			k_2^*, h^{-1} [c]		
$[M]_0$, M	d[D]/dt,[b] M h^{-1}	M_2	M_6	M_{10}	M_2	M_6	M_{10}
0.005	1.0×10^{-6}	6.7×10^{-4}	4.8×10^{-5}	3.5×10^{-6}	1.7×10^{-3}	2.4×10^{-2}	(3.3×10^{-1})
0.008	2.9×10^{-6}	2.2×10^{-3}	4.2×10^{-4}	8.2×10^{-5}	1.5×10^{-3}	7.9×10^{-3}	4.0×10^{-2}
0.010	9.0×10^{-6}	3.5×10^{-3}	9.5×10^{-4}	2.6×10^{-4}	2.8×10^{-3}	1.0×10^{-2}	3.8×10^{-2}
0.015	2.9×10^{-5}	7.4×10^{-3}	3.1×10^{-3}	1.3×10^{-3}	4.3×10^{-3}	1.0×10^{-2}	2.4×10^{-2}
0.020	7.8×10^{-5}	1.2×10^{-2}	6.1×10^{-3}	3.1×10^{-3}	7.1×10^{-3}	1.4×10^{-2}	2.8×10^{-2}
0.030	2.4×10^{-4}	2.1×10^{-2}	1.4×10^{-2}	9.0×10^{-3}	1.3×10^{-2}	1.9×10^{-2}	3.0×10^{-2}
0.040	5.3×10^{-4}	3.1×10^{-2}	2.4×10^{-2}	1.8×10^{-2}	2.0×10^{-2}	2.5×10^{-2}	3.4×10^{-2}
0.045	8.1×10^{-4}	3.6×10^{-2}	2.9×10^{-2}	2.4×10^{-2}	2.6×10^{-2}	3.3×10^{-2}	3.9×10^{-2}

[a] Analysis by eq 11. [b] Reported values are averages from duplicate experiments and from the two methods used (see Table III). [c] Average values of k_2^* are $(9.5 \pm 9) \times 10^{-3}$, $(1.8 \pm 0.9) \times 10^{-2}$, and $(3.3 \pm 0.6) \times 10^{-2}$ h^{-1} for the M_2, M_6, and M_{10} mechanisms, respectively. Value in parentheses is not included in the average, and errors are given as standard deviations.

Scheme I



The potential competition by off-template formation of the 3′–5′ linked dimer is further reduced by the fact that, in the presence of the template, the amount of free monomer in solution is only a few millimolar (see below). Therefore, dimerization in solution would contribute little to the observed yields of the 3′–5′-linked dimer formed during the short reaction times of our experiments. This is contrary to an earlier notion, according to which at room temperature most of the 3′–5′ G_2 is formed in solution.[21]

Rate constants for dimer formation were determined by two different methods, eqs 6 and 9. The average from the two methods, reported as d[D]/dt in Table IV, is probably more reliable and will be used thereafter. The simplest mechanism one can propose is shown in Scheme I. The symbols used are not meant to imply that dimer D is necessarily located at one end of a monomer stack, i.e., T·D·M could also be T·M·D, or T·D·M_2 could also be T·M·D·M or T·M_2·D, etc. q represents the equilibrium constant for association of M to an isolated site of poly(C), while Q is the equilibrium constant for association to a site adjacent to an already associated monomer; Q is presumed to be independent of the length of the stack. Inasmuch as stacking interactions render association adjacent to another monomer more favorable, we have $Q > (\gg) q$, i.e., the association process is a cooperative process.[25]

The second assumption with regard to Scheme I is that the various template-bound stacks have the same *intrinsic* reactivity, i.e., the rate for bond formation between two adjacent units is independent of stack length. We shall use the symbol k_2^* for the rate constant of this process. However, since in stacks containing more than two monomers bond formation can occur between any two neighbors, statistical corrections as described by eq 10 are needed. For reasons to become apparent below, we call Scheme

$$k_2'' = k_2^*;\quad k_2''' = 2k_2^*;\quad k_2'''' = 3k_2^*;\quad \ldots\quad k_2^m = (m-1)k_2^* \qquad (10)$$

I coupled with eq 10 the "M_2 mechanism". According to this mechanism, the rate of dimer formation is given by eq 11, with F_D given by eq 12:

$$\mathrm{d[D]}/dt = k_2^* r F_D \qquad (11)$$

$$F_D = [T{\cdot}M_2] + 2[T{\cdot}M_3] + 3[T{\cdot}M_4] + \ldots + (m-1)[T{\cdot}M_m] \qquad (12)$$

r, which is equal to $[M]/[M]_o$[26] (Table III), corrects for the fact that the concentration of the reactive monomer is slightly less than $[M]_o$. $[T{\cdot}M_2]$, $[T{\cdot}M_3]$,...$[T{\cdot}M_m]$ are the concentrations of the various stacks. Mass balance requires that $[M]_T = \sum_{j=1}^{m} j[TM_j]$, where $[M]_T$ is the concentration of monomers associated with the template.

To test whether eq 11 for the M_2 mechanism accounts for the experimental results, a method for calculating $[M]_T$ and F_D in eq 11 as a function of q, Q, and total monomer concentration had to be devised. Two complementary approaches were used and gave indistinguishable results. In the first, eq 13,[27] was applied,

$$\theta = \frac{Q^{\alpha_H}[M]_f^{\alpha_H}}{1 + Q^{\alpha_H}[M]_f^{\alpha_H}} \qquad (13)$$

in which θ is the fraction of occupied template sites, $[M]_f$ the concentration of free monomer in solution, and α_H, the Hill constant,[28,29] is a measure of the cooperativity of the association, with $\alpha_H = \sqrt{Q/q}$.

Using eq 13 in conjunction with the mass balance equation 14,

$$[M]_o = [M]_T + [M]_f \qquad (14)$$

where $[M]_o$ is the total monomer concentration ($M_T = 0.05\theta$, with 0.05 being the total concentration of template sites), $[M]_T$ and $[M]_f$ could be calculated for a given $[M]_o$, Q, and q. Q = 140 M^{-1} and q = 1.73 M^{-1}, determined recently,[30] were used for these calculations.

The relative abundances of the various monomer stacks on the template, i.e., the $f_{T\cdot M_j}$ values defined in eq 15, for a given θ were then obtained by the Monte Carlo simulation as described earlier. This simulation was done with a built-in preference of $Q/q = 81$ for association adjacent to an already bound monomer. $[T{\cdot}M_j]$ values were obtained from eq 16:

(25) (a) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Addison-Wesley: Reading, MA, 1960; p 235. (b) Hill, T. L. *Cooperativity Theory in Biochemistry*; Springer-Verlag: New York, 1985; Chapter 8.

(26) [M] is the concentration of 2-MeImpG midway through a kinetic run.

(27) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; Freeman and Co.: San Francisco, CA, 1971; Part III, p 864. Equation 13 is considered most accurate for θ = 0.25–0.75 and more approximate for θ outside this range. It was therefore applied primarily to the experiments of $[M]_o \geq 0.015$ M where θ is ≥ 0.18.

(28) Hill, A. V. *J. Physiol. (London)* **1910**, *40*, iv.

(29) The equality $\alpha_H = \sqrt{Q/q}$ follows from a comparison of the slope at the midpoint of the function θ of eq 13 $(d\theta/dQ[M]_f|_{\theta=0.5} = d\theta/d\ln Q[M]_f|_{\theta=0.5} = \alpha_H/4)$, with the same expression resulting from the more accurate treatment by Hill[25] $(d\theta/d\ln Q[M]_f|_{\theta=0.5} = {}^1/_4\sqrt{Q/q}$.

(30) (a) Kanavarioti, A.; Hurley, T. B.; Baird, E. E., unpublished results. (b) It is surprising that Miles and Frazier (Miles, H. T.; Frazier, J. *J. Mol. Biol.* **1982**, *162*, 219) concluded that association between 2-MeImpG and poly(C) is undetectable under conditions that were virtually identical with ours. On the other hand, our q and Q values are consistent with several reports of cooperative interactions of guanosine derivatives with poly(C). For example, see: Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. T. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 93. Davies, R. J. H.; Davidson, N. *Biopolymers* **1971**, *10*, 1455.

(31) Howard, F. B.; Frazier, J.; Miles, H. T. *Biopolymers* **1977**, *16*, 791.

$$f_{T\cdot M_j} = [T\cdot M_j]/\sum_{j=1}^{m}[T\cdot M_j] \quad (15)$$

$$[T\cdot M_j] = [M]_T \frac{f_{T\cdot M_j}}{\sum_{j=1}^{m} j f_{T\cdot M_j}} \quad (16)$$

The second method for calculating $[T\cdot M_j]$ was based on the mass law and mass balance eqs 17 and 18, where $[T]_f$ is the

$$[T\cdot M] = q[T]_f[M]_f \quad (17)$$

$$[T\cdot M_j] = q[T]_f[M]_f Q^{j-1}[M]_f^{j-1} \qquad (j \geq 2) \quad (18)$$

concentration of template sites that are free and whose next-neighbor sites on both sides are also unoccupied.[32] In conjunction with eq 14, eq 19 provides a relationship between $[M]_T$ and $[M]_o$ which can be used in place of eq 13:

$$[M]_T = [T\cdot M] + 2[T\cdot M_2] + 3[T\cdot M_3] + ... + m[T\cdot M_m]$$
$$= q[T]_f[M]_f(1 + 2Q[M]_f + 3Q^2[M]_f^2 + ... + mQ^{m-1}[M]_f^{m-1}) \quad (19)$$

The validity of the M_2 mechanism may now be tested by evaluating F_D via eq 12 for each experiment and checking whether a constant k_2^* value is obtained by applying eq 11. The calculated F_D and k_2^* values are reported in Table IV. Far from being constant, k_2^* is seen to increase from 1.5×10^{-3} to 2.6×10^{-2} h^{-1} with increasing $[M]_o$, indicating that the M_2 mechanism is deficient.

A better fit between the experimental and calculated $d[D]/dt$ values is obtained if one assumes that bond formation between two monomers is much more efficient when the monomer pair that undergoes reaction has one or several next neighbors. For example, if a monomer pair needs at least two next neighbors on each side to be reactive, the statistical corrections take on the form of eq 20. We shall call this the "M_6 mechanism", since $T\cdot M_6$ is the shortest stack that has significant reactivity. F_D in eq 12 for the M_6 mechanism thus becomes eq 21. Table IV

$$k_2'' = k_2''' = k_2'''' = k_2^{v} \approx 0; \quad k_2^{vi} = k_2^*; \quad k_2^{vii} = 2k_2^*; \quad ... \quad k_2^m = (m-5)k_2^* \quad (20)$$

$$F_D = [T\cdot M_6] + 2[T\cdot M_7] + 3[T\cdot M_8] + ... + (m-5)[T\cdot M_m] \quad (21)$$

shows that the values of k_2^* obtained for the M_6 mechanism scatter around the average $(1.8 \pm 0.9) \times 10^{-2}$ h^{-1}, demonstrating that the M_6 mechanism is much more realistic than the M_2 mechanism. By the same criterion, the M_6 mechanism is superior to the M_3, M_4, and M_5 mechanisms (F_D and k_2^* not included in Table IV).

Mechanisms that require an even larger number of next neighbors improve the fit between experimental and calculated dimerization rates at high $[M]_o$ but make it worse at very low $[M]_o$. This is illustrated with the M_{10} mechanism, for which F_D is given by eq 22:

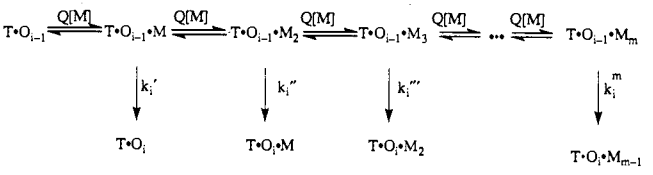
$$F_D = [T\cdot M_{10}] + 2[T\cdot M_{11}] + 3[T\cdot M_{12}] + ... + (m-9)[T\cdot M_m] \quad (22)$$

In this case the calculated k_2^* values for the experiments with $[M]_o \geq 0.008$ M vary by a factor of <2 (average $k_2^* = (3.3 \pm 0.6) \times 10^{-2}$ h^{-1}), but the calculated k_2^* at $[M]_o = 0.005$ M is 10-fold higher than the above average. Similar problems exist with the M_7, M_8, and M_9 mechanisms.

What conclusions can be drawn from the above analysis? The poor fit with the M_2 mechanism clearly shows that bond formation between two template-bound monomer units is substantially more favorable in stacks that contain more than two units. The fact that no completely satisfactory fit can be found for any of the "higher" mechanisms suggests that these higher mechanisms represent oversimplifications. For example, it is unlikely that in the M_{10} mechanism there is essentially zero intrinsic reactivity for all stacks from TM_2 to TM_9, as implied by eq 22. Allowance for non-zero intrinsic reactivity of the shorter stacks has the effect of decreasing the value of k_2^* calculated from eq 11 at low $[M]_o$ without significantly affecting it at high $[M]_o$. This would, e.g., greatly improve the fit of the $[M]_o = 0.005$ M experiment in the M_{10} mechanism. However, the absence of a good model for predicting the relative intrinsic reactivity of the various stacks precludes further quantitative analysis.

Mechanism of Elongation. Oligoguanylate products are known to be tightly held on the template.[33] The process of elongating a template-bound oligomer or primer, $T\cdot O_{i-1}$, by one unit requires a fast equilibrium association of at least one monomer with $T\cdot O_{i-1}$ to form a reaction complex, $T\cdot O_{i-1}\cdot M_j$, (Scheme II). Since association of a monomer adjacent to a primer involves the same kind of hydrogen bonding to the template and stacking interaction with its neighbor as in the association of a monomer adjacent to another monomer, the association equilibrium constant is assumed to be Q, as in Scheme I. Note that the symbols used ($T\cdot O_{i-1}\cdot M_j$) are not meant to imply that all Ms are piled up at the same end of the primer. As schematically shown in Figure 5, $T\cdot O_{i-1}\cdot M$ represent two, $T\cdot O_{i-1}\cdot M_2$ three, $T\cdot O_{i-1}\cdot M_3$ four, etc., different configurations. For simplicity, $T\cdot O\cdot M_j$ will henceforth be used instead of $T\cdot O_{i-1}\cdot M_j$.

Scheme II



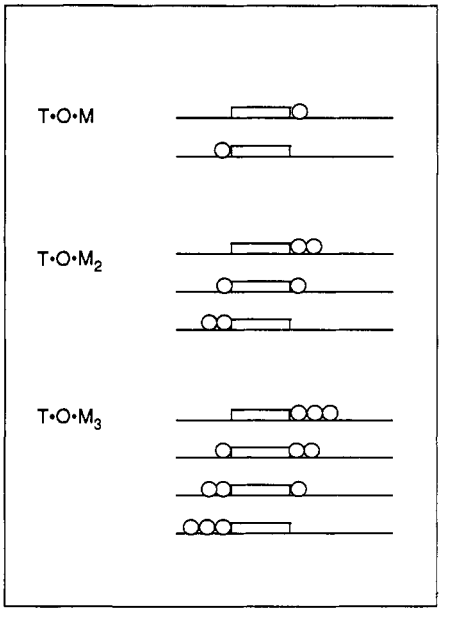


Figure 5. Schematic representation of template-bound oligomer with various possible arrangements of one, two, or three next-neighbor monomers.

(32) $[T]_f$ was found by multiple iterations carried out with the Microsoft Excell Solver.

(33) Lipsett, M. N. *J. Biol. Chem.* **1964**, *239*, 1256.

Regarding the rate constants, the simplest assumption is that there is an inherent rate constant for bond formation (k_i^*) that is independent of the number of monomers associated with the primer. However, a statistical correction is required because elongation can occur only at the 3′ end of the primer. As easily seen from Figure 5, the statistical factor for T·O·M is $^1/_2$, for T·O·M$_2$ $^2/_3$, for T·O·M$_3$ $^3/_4$, etc. These considerations lead to eq 23.

$$k_i' = {}^1/_2k_i^*;\quad k_i'' = {}^2/_3k_i^*;\quad k_i''' = {}^3/_4k_i^*;\quad \dots\quad k_i^m = \frac{m}{m+1}k_i^* \tag{23}$$

The rate of elongation is then given by eq 24,

$$-\frac{d[O]}{dt} = k_i^*\left\{{}^1/_2[T\cdot O\cdot M] + {}^2/_3[T\cdot O\cdot M_2] + {}^3/_4[T\cdot O\cdot M_3] + \dots + \frac{m}{m+1}[T\cdot O\cdot M_m]\right\} \tag{24}$$

T·O·M$_j$ values[34] and is given by eq 25. Dividing both sides of eq 24 by [O] affords eq 26:

$$[O] = \sum_{j=0}^{m}[T\cdot O\cdot M_j] \tag{25}$$

$$-\frac{d[O]}{[O]dt} = k_i^* r F_E \tag{26}$$

with

$$F_E = {}^1/_2f_{T\cdot O\cdot M} + {}^2/_3f_{T\cdot O\cdot M_2} + {}^3/_4f_{T\cdot O\cdot M_3} + \dots + \frac{m}{m+1}f_{T\cdot O\cdot M_m} \tag{27}$$

and r being again $[M]/[M]_0$ as in eq 11. The $f_{T\cdot O\cdot M_j}$ terms are the fraction of [O] in the form of T·O·M$_j$. Note that -(d[O]/dt)/[O] corresponds to k_i[M] ($i \geq 4$) or k_3[M], with k_i ($i \geq 4$) and k_3 being the experimentally determined second-order rate constants.[26]

The calculation of the $f_{T\cdot O\cdot M_j}$ values can be based on the distribution of the T·M$_j$ species described earlier.[35] Since the affinity of an oligomer for monomers is the same as that of an M molecule on the template, the distribution of the various T·O·M$_j$ species should parallel that of the corresponding T·M$_{j+1}$ species. Thus the fraction of T·O should be the same as the fraction of T·M, the fraction of T·O·M the same as that of T·M$_2$, etc. A plot of -(d[O]/dt)/[O] vs rF_E (F_E from eq 27) calculated by this method should yield a straight line with a slope = k_i^*. Curves labeled O·M in Figures 6 and 7 show such plots for k_i[M] ($i \geq 4$) and k_3[M], respectively. They are seen to deviate substantially from a straight line.

The reasons for the negative deviations may be understood by assuming that bond formation between the primer and the monomer is not very efficient unless there is one or several additional adjacent monomers present in the reaction complex. This reduced reactivity is analogous to the low rates of dimerization of very short stacks such as T·M$_2$ and T·M$_3$. For example, if the reactivity of an oligomer is assumed to be negligible unless there are at least two monomers adjacent to the 3′ end (first isomer to T·O·M$_2$ and first two isomers of T·O·M$_3$ in Figure 5), then eq 23 reduces to eq 28 and F_E is given by eq 29.

$$k_i' \approx 0;\quad k_i'' = {}^1/_3k_i^*;\quad k_i''' = {}^2/_4k_i^*;\quad \dots\quad k_i^m = \frac{m-1}{m+1}k_i^* \tag{28}$$

$$F_E = {}^1/_3f_{T\cdot O\cdot M_2} + {}^2/_4f_{T\cdot O\cdot M_3} + {}^3/_5f_{T\cdot O\cdot M_4} + \dots + \frac{m-1}{m+1}f_{T\cdot O\cdot M_m} \tag{29}$$

An even better fit (O·M$_3$ mechanism) is obtained assuming that efficient elongation requires at least three monomers adjacent

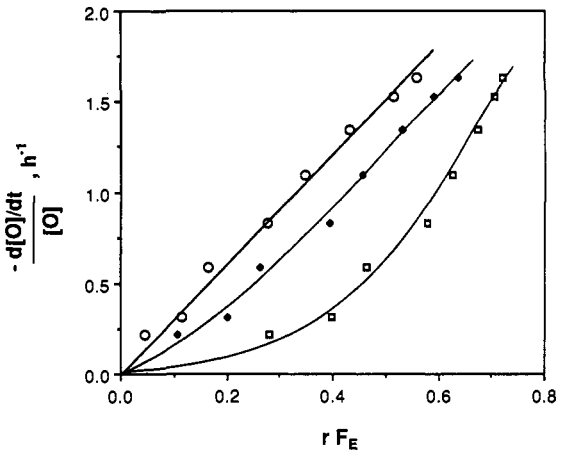


Figure 6. Plots of -(d[O]/dt)/[O] = k_i[M] according to eq 26 for the k_i ($i \geq 4$) elongation process: □, O·M mechanism; ◆, O·M$_2$ mechanism; ○, O·M$_3$ mechanism.

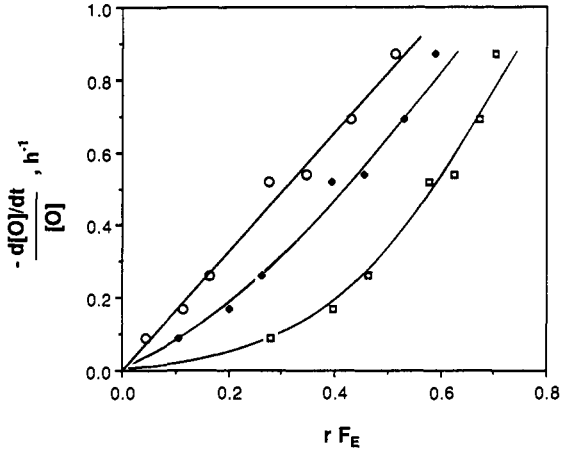


Figure 7. Plots of -(d[O]/dt)/[O] = k_3[M] according to eq 26 for the k_3 elongation process: □, O·M mechanism; ◆, O·M$_2$ mechanism; ○, O·M$_3$ mechanism.

Table V. Analysis of the Elongation in Terms of the O·M, O·M$_2$, and O·M$_3$ Mechanisms[a]

			F_E		
[M$_o$], M	k_3[M],[b] h^{-1}	k_i[M][b] ($i \geq 4$), h^{-1}	O·M	O·M$_2$	O·M$_3$
0.005	0.09[c]	0.22[c]	0.32	0.12	0.05
0.008	0.17[c]	0.31[c]	0.45	0.23	0.13
0.010	0.26	0.59	0.51	0.29	0.18
0.015	0.52	0.83[c]	0.63	0.43	0.30
0.020	0.54[c]	1.09[c]	0.69	0.50	0.38
0.030	0.69[c]	1.34[c]	0.75	0.59	0.48
0.040	0.87	1.53	0.81	0.68	0.59
0.045		1.63	0.84	0.74	0.65

[a] Analysis by eq 26. [b] Equivalent to -(d[O]/dt)/[O]; [M] is the concentration of 2-MeImpG midway through a kinetic run. [c] Reported values are averages from two experiments (see Table II).

(34) Since the rate is expressed as disappearance of an oligomer, such as O$_{i-1}$, T·O·M$_j$ refers to T·O$_{i-1}$·M$_j$ in Scheme II.

(35) If there was no significant association of monomers with the template at sites other than those adjacent to oligomers, the various f terms in eq 27 could easily be calculated as

$$f_{T\cdot O\cdot M_j} = \frac{Q^j[M]_f^j}{\sum_{j=0}^{m}Q^j[M]_f^j}$$

This equation is indeed a good approximation at very low θ. However, when θ is substantial this equation breaks down, because as the fraction of template sites occupied by monomer grows, the likelihood that a primer happens to have a monomer as an immediate neighbor irrespective of [M]$_f$ increases.

to the 3′ end of the primer (eqs 30 and 31):

$$k_i' \approx k_i'' \approx 0;\quad k_i''' = {}^1/_4 k_i^*;\quad k_i'''' = {}^2/_5 k_i^*;\quad \dots\quad k_i^m = \frac{m-2}{m+1} k_i^* \tag{30}$$

$$F_E = {}^1/_4 f_{T\cdot O\cdot M_3} + {}^2/_5 f_{T\cdot O\cdot M_4} + {}^3/_6 f_{T\cdot O\cdot M_5} + \dots + \frac{m-2}{m+1} f_{T\cdot O\cdot M_m} \tag{31}$$

From the slope of the straight line in Figures 6 and 7 ($O\cdot M_3$) we obtain $k_i^* = 1.7\ h^{-1}$ for the k_3 process and $k_i^* = 2.9\ h^{-1}$ for the $k_i (i \geq 4)$ process. The lower rate constant for the k_3 process may reflect the weaker association of the dimer on the template as compared with the longer oligomers[33] and/or an intrinsically slower process.

In a similar manner as with the dimerization mechanism, the good fit with the $O\cdot M_3$ mechanism for elongation does not necessarily mean that $T\cdot O\cdot M$ and $T\cdot O\cdot M_2$ are completely unreactive. It is more likely that these complexes show some modest reactivity ($T\cdot O\cdot M_2$ more reactive than $T\cdot O\cdot M$) and that maximum reactivity is achieved with $T\cdot O\cdot M_j$ ($j \geq 3$). The main point, again, is that additional next-neighbor monomers are needed for optimal reactivity. Note, however, that with $O\cdot M_4$ or higher mechanisms, the plots of $k_i[M]$ ($i \geq 4$) and $k_3[M]$ vs the corresponding rF_E (not shown) display upward curvature.[36]

Catalytic Effect of Additional Associated Monomers. One of the most interesting conclusions emerging from our study is that bond formation between two monomer units or between a primer and a monomer is assisted by the presence of additional next-neighbor monomer units. This phenomenon is consistent with recent observations on hairpin oligonucleotides made in Orgel's laboratory.[7-9] These authors have synthesized oligodeoxynucleotide sequences that form unsymmetrical hairpin structures. The template sequences are the 5′-terminal single-strand portions of the hairpin. The 3′-terminal segments of the oligonucleotide, which are part of the double helical stems, act as intramolecular "primers" for template-directed incorporation of ribonucleotide monomers. Wu and Orgel[7] found that the addition of the penultimate and final G residues directed by a sequence of Cs is much slower than the addition of the first few 2-MeImpG monomers. This observation can also explain the low yield of complementary product obtained in reactions with other templates.[37] Evidence that these phenomena are not end effects but indicate the need for next neighbors comes from the observation that the addition of a C residue is efficient only if it is possible to stack another C or G residue downstream on the template.[8] Similarly, incorporation of single A or U residues is possible as long as 2-MeImpG, which evidently stacks well, is the downstream monomer and is present in the reaction mixture.[9]

Furthermore, Wu and Orgel's results[8] indicate that addition of 2-MeImpC (in the absence of 2-MeImpG) to hairpins that include consecutive G residues on the template followed by C residues is the most efficient when there are three or more G residues (approximate half-life 8 h), less efficient when there are only two (24 h), and inefficient when there is only one isolated G residue (longer than 10 days). This reactivity pattern for C incorporation fits remarkably well with the one observed for G incorporation in the present study.

Wu and Orgel[8] offered two possible explanations of these observations. The first is that occupancy of the site adjacent to the primer is low unless the next site "downstream" is occupied.[38] The second is that the presence of a neighboring base on the template catalyzes the elongation. Their findings, taken together with ours, suggest the presence of an additional effect, perhaps that the reaction becomes similar to a ligation process. This "ligation" is envisioned to be, e.g., between the primer and a "dimer" or "trimer" (i.e., an adjacent two- or three-monomer stack) in the elongation or between two "trimers" (i.e., two adjacent three-monomer stacks) in the dimerization. The advantage of a ligation vs a one-unit addition would be that in the ligation a more optimal conformation of the double helix conducive to bond formation is possible.

Comparison with Previous Studies. An earlier study of the elongation of preformed oligoguanylates with 2-MeImpG on poly(C) led to rate constants substantially lower than those reported here.[6] For example, $k_{10} = 8.4\ M^{-1}\ h^{-1}$[11] compares with $k_i \approx 60\ M^{-1}\ h^{-1}$ for similar [2-MeImpG]. At least two factors could contribute to this discrepancy. (1) Up to 6% GppG was present in the earlier investigation. GppG acts as a potent inhibitor[12] of the oligomerization reaction, which at the time of our previous study was not known. (2) Experimentally the previous study differed from the present one in that the reacting oligomer was not formed on the template but was added to the reaction mixture. It was noted in these experiments that the rate constants decreased with increasing amount of added oligoguanylate, which was attributed to oligoguanylate self-association.[31] Such self-association is prevented in the present study because the oligomers are synthesized directly on the template and strongly held there.

A system comparable with the one described here is the elongation of a G primer (part of a hairpin) with 2-MeImpG on a deoxy(C) sequence[7] with a half-life of 3 h at 4 °C, which should be compared with a half-life of 0.23 h in our study within an optimal reaction complex ($T\cdot O\cdot M_j, j \geq 3$) at 23 °C. The 12-fold shorter half-life observed in our reaction can be attributed, in part, to the higher temperature and, in part, to the use of a ribo-(C) that was shown elsewhere to be more efficient than the deoxy-(C) template.[39]

(36) A different analysis giving numerically identical results is based on the following reasoning. Let us assume that the small amount of oligomeric products leaves the distribution of the monomer stacks on the template practically unperturbed. In such a case, an oligomer, such as G_6, can be found in any stack or units equal or longer than 6, i.e., $T\cdot M_j, j \geq 6$. The distribution of the hexamer among all the possible stacks should be independent on their length because of the dynamic equilibrium that allows stacks to form and dissolve rapidly. Assuming an $O\cdot M_2$ mechanism, for example, the disappearance of G_6 is given by

$$-d[G_6]/dt = k_7^* r \sum [M_\lambda G_6 M_m]$$

with $\lambda \geq 0$ and $m \geq 2$ where k_7^* is the intrinsic first-order rate constant for elongation of G_6 and $r = [M]/[M]_o$, the correction for reactivity. Then,

$$\sum_{[M_\lambda G_6 M_m]} = [G_6] \frac{\text{sum of all stacks leading to elongation of } G_6}{\text{all stacks where } G_6 \text{ can be found}} = [G_6] F_E^6 / \sum_{j=6}^{n} f_{T\cdot M_j}$$

The sum of the stacks, expressed in fractions, leading to elongation of the hexamer, called F_E^6, can be obtained by considering the mechanistic requirements set above, i.e., $\lambda \geq 0$ and $m \geq 2$ and statistical corrections deduced from Figure 5. For the $O\cdot M_2$ mechanism,

$$F_E^6 = {}^1/_3 f_{T\cdot M_6} + {}^2/_4 f_{T\cdot M_7} + \dots + \frac{i-1}{i+1} f_{T\cdot M_{i+6}}$$

The above equations lead to

$$-\frac{d[G_6]/dt}{[G_6]} = k_7[M] = k_7^* r F_E^6 / \sum_{j=6}^{n} f_{T\cdot M_j}$$

Numerical manipulation shows that for a specific mechanism,

$$F_E^6 / \sum_{j=6}^{n} f_{T\cdot M_j} = F_E^m / \sum_{j=m}^{n} f_{T\cdot M_j} = F_E \qquad (F_E \text{ as defined in main text})$$

This alternative analysis confirmed the one presented in the main text and, in addition, led to the conclusion that $k_4^* = k_5^* = k_6^*$, etc. $= k_i^*$.

(37) Inoue, T.; Joyce, G. F.; Grzeskowiak, K.; Orgel, L. E.; Brown, J. M.; Reese, C. B. *J. Mol. Biol.* **1984**, *178*, 669. Haertle, T.; Orgel, L. E. *J. Mol. Biol.* **1986**, *188*, 77. Acevedo, O. L.; Orgel, L. E. *J. Mol. Biol.* **1987**, *197*, 187.

(38) This hypothesis requires that Q varies with the length of the stack and hence cannot be tested using our Schemes I and II.

(39) Chen, C. B.; Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1985**, *181*, 271.

Conclusions

This work constitutes the first comprehensive kinetic study of the poly(C)-directed oligomerization of 2-MeImpG, which provides rate constants for the elongation of oligoguanylates of varying length. What is most remarkable about our results is that despite the complexity of the mechanism the reaction can be described by just three experimental parameters: the second-order rate constants k_2, k_3, and k_i ($i \geq 4$). The complexity of the mechanism manifests itself in that the second-order rate constants are not true constants but show a complex dependence on monomer concentration.

The mechanistic models proposed for the dimerization (Scheme I) and elongation process (Scheme II) account satisfactorily for the kinetic results in qualitative and even quantitative terms. The dimerization is best understood in terms of a coupling between two monomer units that are part of a template-bound stack and is more efficient in stacks containing several monomer units. Likewise, the elongation appears to be most favorable when the primer–monomer complex contains at least two additional monomers as immediate neighbors on the template. This is an important conclusion that is consistent with Orgel's findings with hairpin nucleotides. The catalysis by the next neighbors, as well as the fact that $k_i^* (i \geq 4) > k_3^* (i = 3) \gg k_2^*$, may be visualized as arising from the reaction becoming more akin to a ligation process.

Acknowledgment. We thank NASA's Exobiology program for support of this work through Grant NCC 2-534, Dr. Ilan Benjamin of the University of California, Santa Cruz, for writing a program to do the Monte Carlo simulation, Dr. S. Chang of NASA/Ames Research Center for support and advice, and Dr. L. E. Orgel of the Salk Institute for criticism and for providing us with the RPC-5 material.

Supplementary Material Available: HPLC data, Tables S1 and S2 (9 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS. Ordering information is given on any current masthead page.