Unique Catalysis and Regioselectivity Observed in the Poly(C)-Directed RNA Dimer Formation from 2-MeImpG: Kinetic Analysis as a Function of Monomer and Polymer Concentration

Anastassia Kanavarioti,* Eldon E. Baird, T. Brian Hurley, Julie A. Carruthers, and Sumana Gangopadhyay[1]

Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064

Received August 3, 1999

Polycytidylate, poly(C), serves as a scaffold or template to direct and catalyze the synthesis of long oligoguanylates from guanosine 5′-phosphate 2-methylimidazolide, 2-MeImpG. In the absence of poly(C), small amounts of three isomeric dimers, i.e., the 2′-5′-, the 3′-5′-, and the pyrophosphate-linked, are formed slowly. In the presence of poly(C) oligomers that are primarily 3′-5′-linked are formed quickly and in high yield. Product analysis suggests that the oligomers are elongation products of the 3′-5′-linked dimer, abbreviated D. Assuming that D is formed slowly from two molecules of 2-MeImpG (Scheme 1) and elongates relatively fast, the initial rate of dimerization, d[D]/dt in M h^{-1}, was determined using two independent methods. The first method is based on the approximation that at the onset of the reaction the substrate is consumed only via hydrolysis and dimerization, and thus elongation can be neglected. The second, more accurate, method exploits the assertion that every oligomer was once a 3′-5′-linked dimer. Hence the concentration of D was obtained indirectly from the concentration of the oligomer products. These two methods gave comparable results. Experiments were run in aqueous solution in the presence of 1.0 M NaCl, 0.2 M $MgCl_2$ at pH 7.9 ± 0.1 and 23 °C. Controls were run in the absence of poly(C) and in the presence of other polynucleotides. The kinetics were determined as a function of both monomer and polymer concentration the latter expressed in C equivalents. The kinetic data obtained in the presence of poly(C) confirmed an earlier conclusion regarding the remarkable effect of poly(C) on the formation of the 3′-5′-linked diguanylate. Initial dimerization rates were quantitatively correlated using a simple template-directed (TD) model that presumes cooperative binding (two association constants) of 2-MeImpG on poly(C) and reaction between adjacent template-bound molecules. The model allows for the estimation of the association constants and the intrinsic rate constant of dimerization, k_2*. Insights into the detailed mechanism are also gained from this analysis. The fact that the proposed model can successfully correlate kinetic data that vary by more than 5000-fold between the slowest and the fastest reaction adds confidence and suggests the suitability of this model for describing TD reactions in general. It is anticipated that similar analysis of other known TD reactions may lead to clues that will facilitate the design of more efficient polynucleotide-synthesizing systems.

Introduction

Template-directed (TD) enzyme-assisted synthesis of polynucleotides is the method chosen by Nature to replicate its gene pool. In the process of evolution, Nature has developed a rather complex procedure for transmitting information accurately and in a regulated fashion. Recently TD *nonenzymatic* synthesis has found many applications in the pharmaceutical industry for the large-scale production of antisense polynucleotides.[2] Such synthesis is performed in an elegant and efficient way by TD ligation[3] of two oligomers, but requires oligomer synthesis by another method. TD ligations of unprotected monomers may have been the method of choice, albeit they are inefficient.[4a] In the context of prebiotic chemistry and in an attempt to unravel simple self-replicating systems, the nonenzymatic TD synthesis of polynucleotides and analogs thereof has received considerable attention.[5] TD polymerizations are initiated by a relatively slow two-monomer ligation step (dimerization), followed by, most likely, faster and perhaps more regi-

(1) On leave of absence from the Depart. of Chemistry, Gurudas College, Calcutta-700054, India.

(2) Bennett, M. R. *J. Drug Dev. Clin. Practice* **1995**, *7*, 225−235. Field, A. K. *Antiviral Res.* **1998**, *37*, 67−81.

(3) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3332−3339. Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3340−3344. Boli, M.; Micura, R.; Eschenmoser, A. *Chem. Biol.* **1997**, *4*, 309−320. Kuznetsova, S. A.; Merenkova, I. N.; Kanevsky, I. E.; Shabarova, Z. A.; Blumenfeld, M. *Antisense Nucleic Acid Dev.* **1999**, *9*, 95−100. Selvasekaran, J.; Turnbull, K. D. *Nucleic Acid Res.* **1999**, *27*, 624−627. Xu, Y.; Kool, E. T. *Nucleic Acid Res.* **1999**, *27*, 875−881. Koppitz, M.; Nielsen, P. E.; Orgel, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 4563−69.

(4) (a) Hill, A. R., Jr.; Orgel, L. E.; Wu, T. *Origins Life Evol. Biosphere* **1993**, *23*, 285−290. (b) Inoue, T. and Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201−218. (c) Fakhrai, H.; Inoue, T.; Orgel, L. E. *Tetrahedron* **1984**, *40*, 39−45. (d) Inoue, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1981**, *103*, 7666−7667.

(5) Inoue, T.; Orgel, L. E. *Science* **1983**, *219*, 859−862. Joyce, G. F. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 41−51. Orgel, L. E. *Nature* **1992**, *358*, 203−209. Kanavarioti, A. *Origins Life Evol. Biosphere* **1994**, *24*, 479−495. Orgel, L. E. *Acc. Chem. Res.* **1995**, *28*, 109−118. Bag, B. G.; von Kiedrowski, G. *Pure Appl. Chem.* **1996**, *68*, 2145−2152. Ertem, G.; Ferris, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 7197−7201. Joyce, G. F.; Orgel, L. E. *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Lab. Press: Cold Spring Harbor, 1993; pp 1−25. Luther, A.; Brandsche, R.; von Kiedrowski, G. *Nature* **1998**, *396*, 245−248. Zhan, Z.-Y. J.; Lynn, D. G. *J. Am. Chem. Soc.* **1997**, *119*, 12420−12421. Prakash, T. P.; Roberts, C.; Switzer, C. *Angew Chem., Int. Ed. Engl.* **1997**, *36*, 1522−1523.

10.1021/jo991216q CCC: $18.00
Published on Web 10/14/1999

oselective, primer extension (elongation) steps. Hence the overall efficiency of a TD polymerization depends critically on the yield and regioselectivity of the dimerization step. Insights in TD mononucleotide dimerization may facilitate the design of more efficient TD polynucleotide-synthesizing systems and find application in the large-scale synthesis of polynucleotide-based pharmaceuticals.

To the best of our knowledge the most efficient nonenzymatic TD polymerization so far is the poly(C)-directed oligoguanylate synthesis from 2-MeImpG or G for simplicity.[4bc,6] In the presence of poly(C) this reaction yields quantitatively oligoriboguanylates,[4b] whereas in the absence of poly(C) only a small percentage of dimers is formed (Figure 1a).[7] Enzymatic degradation of the oligomers indicated that the majority of the linkages is 3′-5′,[4b] leading to the conclusion that, depending on the temperature, the primer is an all 3′-5′-linked-dimer or -trimer.[4c] Kinetic determinations at 23 °C in the range 0.005 M ≤ [G] ≤ 0.045 M at a constant 0.05 M poly(C) concentration suggested substantial catalysis by poly(C) and indicated that d[D]/d*t* exhibits a third-order dependence on monomer concentration.[6a] To explain this strong dependence, a model was proposed in which dimerization occurs in long stacks, six or longer, of template-bound monomers. However, the effect of poly(C) concentration on dimerization was not investigated and thus the proposed model was not fully tested.[6a]

Here we have extended the earlier kinetic studies by including a dependence on poly(C) concentration, a larger range of monomer concentration, and several controls. The product distribution and the kinetics of dimerization were determined in the presence/absence of poly(C) as well as in the presence of other polymers, such as polyinosinate, poly(I), polyuridylate, poly(U), and double-stranded polycytidylate·polyguanylate, poly(C)·poly(G). Perhaps not surprising it was established that these polymers exhibit practically no effect on 2-MeImpG dimerization, whereas the effect of poly(C) is dramatic in comparison. Moreover, the extended set of data was found to be quantitatively consistent with a TD mechanism of dimerization and allowed an important refinement of the earlier proposed cooperative model describing such a mechanism.[6a]

Experimental Section

Materials, Methods, and HPLC Analysis. Acquisition of materials, preparation of samples, pH measurements, and product identification were done following already developed methods.[4b,6,7] 2-MeImpG with $\epsilon = 12\,000$ at 253 nm was better than 97% pure as tested by C18 chromatography. The polymers poly(C), poly(I), poly(U), and poly(C)·poly(G) were purchased from Sigma. The potassium salt of poly(C) is about 100 to 300 units long. Analysis of samples was performed with high performance liquid chromatography (HPLC) using a 1090 LC from Hewlett Packard equipped with a diode array detector set at 254 nm. Samples were incubated at 23 °C in HPLC vials in the thermostated autosampler of the HPLC instrument, and the analysis was run with C18 chromatography (see below). Alternatively, samples were incubated in a Lauda bath at 23 ± 0.1 °C, quenched in regular intervals by dilution and addition of acidic EDTA. Then these samples were hydrolyzed at pH 3 and 50 °C overnight in order to remove the 2-MeIm groups and analyzed with RPC5 chromatography (see below). RPC5 chromatography resolves oligoguanylates according to length and isomerism.[4b] Evaluation of dimer concentration based on oligomer yields has been described earlier.[6a] In contrast to RPC5, C18 chromatography resolves 2-MeImpG from 5′GMP, i.e., the hydrolysis product as well as from the dimerization products. Product distribution was obtained directly from HPLC reports as the percent of the total HPLC area corresponding to the initial substrate (G_o). For example (% 5′GMP) = 100 (HPLC area of 5′GMP peak)/(total HPLC area). It follows that [5′GMP] = $[G]_o$ (% 5′GMP)/100 where $[G]_o$ is the initial or formal concentration of substrate obtained by weight. Percent total dimer yields (see Scheme 1), % D_{all}, are reported in monomer equivalents and are uncorrected for hypochromicity h.[8] Hence $[D_{all}] = h\,[G]_o$ (% D_{all})/200.

Analysis with C18 chromatography was performed on a C18 Alltima (3.2 × 250 mm, 5 μm by Alltech) solvent minimizer column run at 0.5 mL/min.[9] Solvent A is 0.02 M KH_2PO_4 with 0.2% w/v trifluoroacetic acid (TFA) pH 2.5; solvent B: 30% CH_3CN in water v/v with 0.2% w/v TFA. 0 to 20% B in 10 min; isocratic at 20% B for 4 min and then 20% to 32% B in 6 min; 2 min wash with 100% B. This chromatography was used for the "faster" samples. A slightly modified gradient that exhibited better resolution was used for the "slower" samples; 0 to 15% B in 10 min; isocratic at 15% B for 4 min and then 15% to 45% B in 16 min; 2 min wash with 100% B. The order of eluting guanosine derivatives with both of these gradients is 5′GMP, HEPES-pG, guanosine cyclic 3′-5′ monophosphate, $pG^{2'}pG$, $G^{5'}ppG$, 2-MeImp$G^{2'}$pG, 2-MeImpG, $pG^{3'}pG$, and 2-MeImp$G^{3'}$pG coelute, whereas oligomers longer than the dimers elute later.

Some of the experiments with samples devoid of polymer were analyzed with a different C18 chromatography (pH 6.5 in contrast to the one described above at pH 2.5). This analysis was performed with Hypersil C18 column 4.6 × 250 mm 5 μm from Phenomenex used at 1.0 mL/min flow. Solvent A is 0.02 M KH_2PO_4 at pH 6.5, and solvent B is 30% CH_3CN in water (v/v). Gradient is 0 to 13% B in 10 min; isocratic at 13% B for 8 min. Order of elution and typical, retention times were 5′GMP, 4.2 min; pGpG (both isomers), 4.9 min; $G^{5'}ppG$, 9.5 min; HEPES-pG, 9.8 min; 2-MeImpGpG (both isomers), 10.5 min; 2-MeImpG, 13.5 min. pGpG does not always resolve well from 5′GMP. Both chromatographies, the one at pH 2.5 and the one at pH 6.5, gave comparable results.

Results

Dimer Formation with or without Polymer, Excluding Poly(C). All experiments were performed with 0.0005 M ≤ [G] ≤ 0.083 M in the presence of 1.2 M or 1.0 M NaCl, 0.2 M $MgCl_2$, and 0.5 M HEPES buffer at pH 7.90 ± 0.1 at 23 °C. Samples of 2-MeImpG in the presence and in the absence of non-C polymers were analyzed by C18 chromatography (see Experimental Section) which allowed monitoring of 2-MeImpG disappearance and product formation. In accord with the reaction pathways proposed for 2-MeImpC and 2-MeImpU investigated under identical conditions with the ones reported here,[7] we found that 2-MeImpG hydrolyzes to 5′GMP, reacts with the buffer HEPES to form HEPES-pG, and yields three isomeric dimers. These are the two internucleotide linked dimers, $pG^{2'}pG$ and $pG^{3'}pG$, and

(6) (a) Kanavarioti, A.; Bernasconi, C. F.; Alberas, D. J.; Baird, E. E. *J. Am. Chem. Soc.* **1993**, *115*, 8537−8546. (b) Kanavarioti, A.; Baird, E. E. *J. Mol. Evol.* **1995**, *41*, 169−173. (c) Kanavarioti, A.; Bernasconi, C. F.; Baird, E. E. *J. Am. Chem. Soc.* **1998**, *120*, 8575−8581.

(7) Kanavarioti, A. *Origins Life Evol. Biosphere* **1997**, *27*, 357−376.

(8) Hypochromicity $h = 1.16$ (in neutral and acidic solutions) was determined by enzymatic degradation of $G^{5'}ppG$ to 5′GMP with PDE (snake venom phosphodiesterase from *Crotalus durrisus* from Boeringer Mannheim) from the ratio of the areas between produced 5′GMP and consumed $G^{5'}ppG$. An internal standard, 5′AMP, which was not degraded by the enzyme was included. It was presumed that $h = 1.16$ is the same for all guanosine dimers.

(9) C18 packing elutes short oligoguanylates and oligocytidylates (not shown here) but retains poly(C) and oligoguanylates longer than the tetramer. Shorter than tetramer oligoguanylates are, most likely, not quantitatively eluted either.

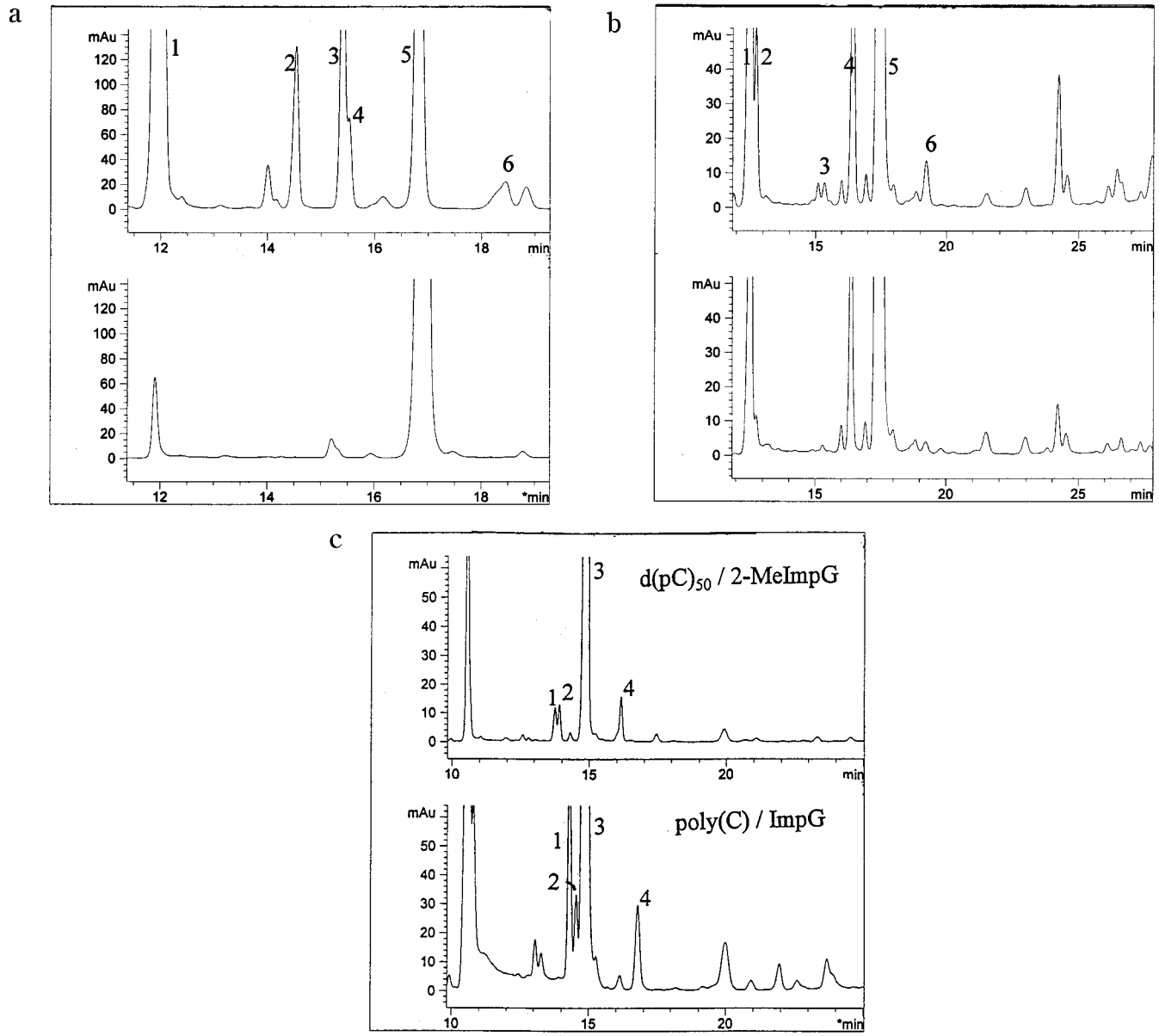
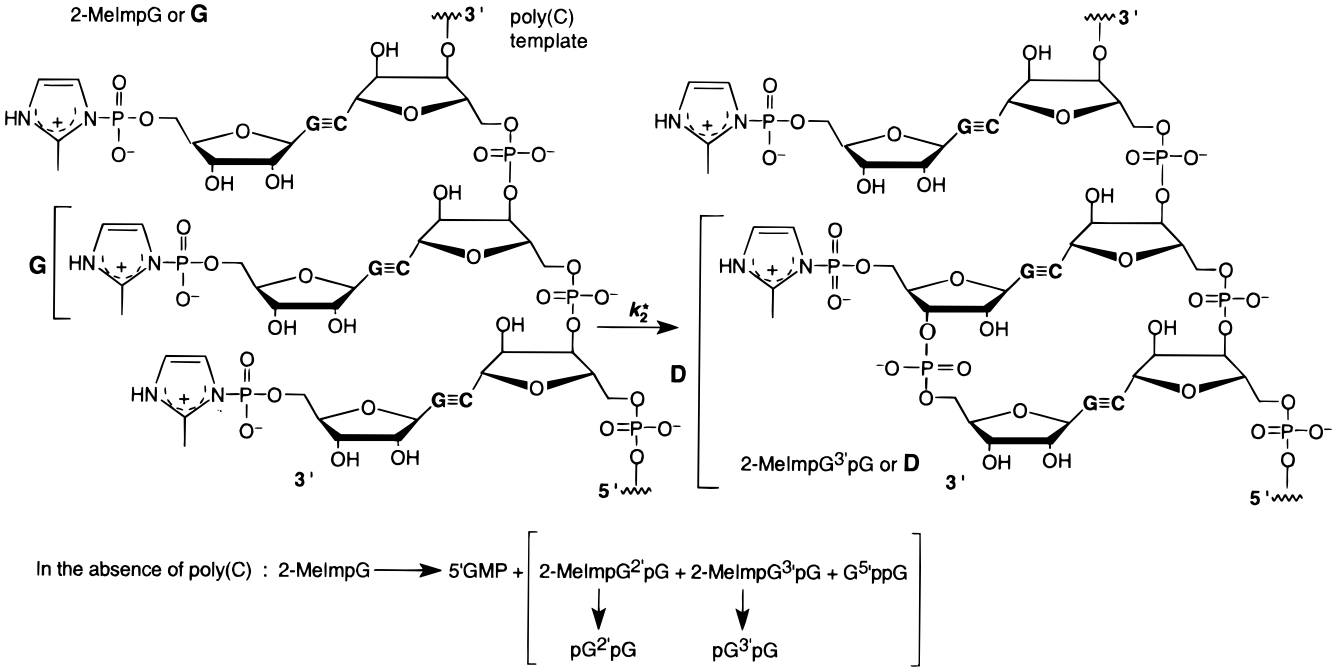


Figure 1. (a) HPLC profiles for the self-condensation of 0.066 M 2-MeImpG in the presence of 0.0073 M poly(U) after 6 days (top) and after 0.35 h (bottom) of incubation at 23 °C. Similar HPLC profiles were obtained in the absence of a control polymer. *Y*-axis: milliabsorbance units at 254 nm. Identification of peaks: 1, 5′GMP and HEPES-pG; 2, pG$^{2'}$pG; 3, G$^{5'}$ppG; 4, 2-MeImpG$^{2'}$-pG; 5, 2-MeImpG; 6, pG$^{3'}$pG and 2-MeImpG$^{3'}$pG. C18 Chromatography at pH 2.5 long run as described in the Experimental Section. No other peaks are detected at later elution times. (b) HPLC profiles for the self-condensation of 0.060 M 2-MeImpG in the presence of 0.010 M poly(C) after 7.3 h (top) and after 0.2 h (bottom) of incubation at 23 °C. *Y*-axis: milliabsorbance units at 254 nm. Identification of peaks: 1, 5′GMP; 2, HEPES-pG; 3, pG$^{2'}$pG; 4, G$^{5'}$ppG; 5, 2-MeImpG; 6, pG$^{3'}$pG and 2-MeImpG$^{3'}$pG. The peaks seen eluting at later times correspond to oligoguanylates and are presumably products of pG$^{3'}$pG. Chromatography as described under part a. Retention times between the samples in Figure 1a and 1b vary because of a guard column that was used in conjuction with the analytical column but only for the samples under part a. (c) HPLC profiles for the self-condensation of guanosine derivatives in the presence of a C-template. *Y*-axis: milliabsorbance units at 254 nm. Top: 0.010 M d(pC)$_{50}$ with 0.045 M 2-MeImpG after 2.5 h incubation. Bottom: 0.012 M poly(C) with 0.045 M ImpG after 10.5 h incubation. Identification of peaks: 1, G$^{5'}$ppG; 2, Activated pG$^{2'}$pG; 3, substrate; 4, activated pG$^{3'}$pG. 2, 3, and 4, 2-MeIm derivatives for the top and Im derivatives for the bottom profile. Experimental conditions and HPLC analysis the same as in Figure la. Retention times and yields of these two profiles are not comparable.

the pyrophosphate-linked dimer, G$^{5'}$ppG (Scheme 1 and Figure 1a). The first two appear initially as the imidazolide-activated derivatives, 2-MeImpG$^{2'}$pG and 2-MeImpG$^{3'}$-pG, which hydrolyze slowly to pG$^{2'}$pG and pG$^{3'}$pG. In contrast, G$^{5'}$ppG has been detected only in the form of the deactivated dimer, most likely because the corresponding activated derivative, G$^{5'}$p(2-MeIm)pG, hydrolyzes rapidly compared to our detection methods. Dimerization products were detected with [G] > 0.005 M.

None of the tested control polymers had a substantial effect on the rate or product distribution of the 2-MeImpG reaction. Figure 2 shows percent substrate, percent 5′GMP, and total percent dimers, % D_{all}, as a function of time as described in the Experimental Section. From such plots, initial rates were obtained from the slope of the line at 0 time using a second-order fit of the data (Kaleidagraph, Abelbeck software). Initial rates were determined for the formation of the hydrolysis product,

Scheme 1. Representation of the Proposed Poly(C)-Directed 2-MeImpG Dimerization with k_2* the TD Intrinsic Rate constant for Dimerization (see Discussion). The Third Monomer is Mechanistically Important But Its Position Is Arbitrarily Shown Here Upstream



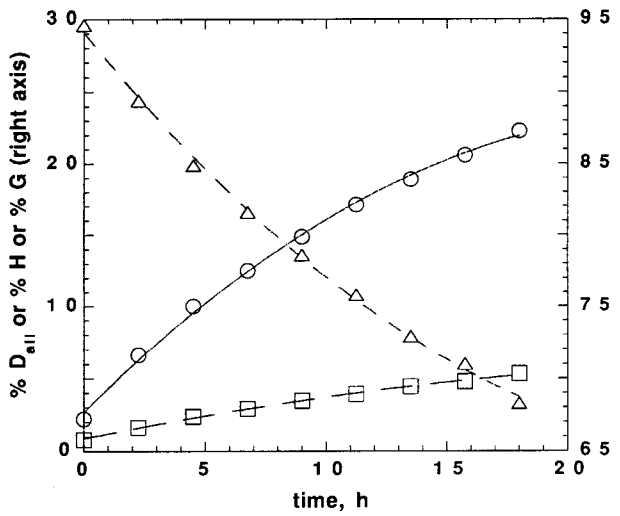


Figure 2. Reaction of 0.066 M 2-MeImpG in the presence of 0.0073 M poly(U). Percent product distribution as a function of time: H = hydrolysis products, (circles, left axis); D_{all} = sum of all dimers (squares, left axis); G = unreacted 2-MeImpG (triangles, right axis). The percent dimer yield is expressed in monomer equivalents and is not corrected for hypochromicity (see Experimental Section). The downward curvature indicates that the rate of hydrolysis as well as dimer formation is slowed down at later times. The lines are second-order fits to the data; d(% H)/dt, d(% G)/dt, d(% D_{all})/dt in h^{-1} are obtained from such plots from the slopes at 0 time (see Results).

d(%H)/dt (H includes 5′GMP and HEPES-pG), the disappearance of the substrate, d(% G)/dt, and the appearance of the sum of all dimers, d(% D_{all})/dt. These values are listed in Table 1. The experimental error associated with the determination of the slopes is estimated at 20% or less.

Product ratios, 2′/3′, were obtained from the yields of the two internucleotide-linked dimers 2′-5′-linked vs 3′-5′-linked. Product ratios, p/(2′+3′), were obtained from the yield of G5′ppG and the total yield of the internucleotide-linked dimers. These ratios stayed constant during the first two days of reaction, and the averages are listed in Table 1. They indicate that in solution the synthesis of the 2′-5′-linked dimer is about four times more favored than the synthesis of any of the other two dimers which are formed about equally. At later times the product distribution favors the pyrophosphate dimer, because of an additional pathway. This pathway is reaction of the substrate with the hydrolysis product 5′GMP that accumulates with time and yields G[5′]ppG as the major product.[7]

Dimer Formation in the Presence of Poly(C). Product distribution in samples of 2-MeImpG was conveniently monitored by C18 chromatography every 37 to 45 min depending on the gradient. Samples without poly(C) exhibit a decrease from 95% to 75% substrate during the first 15 h of incubation, whereas samples with poly(C) exhibit a similar decrease in 2 to 4 h. After 2 h of incubation and in the presence of poly(C) the product distribution includes 0.5% of the RNA dimer, about 6% of oligomers, 0.1% of the 2′-5′-linked dimer, and no detectable change in the pyrophosphate-linked dimer which is an impurity (≈2%) of the substrate preparation (Figure 1b). Longer incubation than 2 h shows that the RNA dimer stays constant at 0.5% (plateau, see later), whereas small amounts of the other two dimers are being formed and large amounts of longer oligomers accumulate. On the basis of the notion that the oligoguanylate products are more than 90% 3′-5′-linked,[4b] our observations can be most easily explained as follows. A poly(C)-catalyzed process strongly accelerates the synthesis of the RNA dimer, which does not accumulate beyond a certain plateau value because it serves as a primer for further extension. The small fraction of 2′-5′-linked and pyrophosphate-linked dimers forms in solution rather than on the template; their yield increases steadily because no further reaction occurs with these dimers. These observations show that the poly(C)/2-MeImpG

Table 1. Initial Rates in the Reaction of 2-MeImpG (G) in the Absence/Presence of Nontemplating Polymers at 23 °C and pH 7.90 ± 0.05, 0.5 M HEPES, 1.0 M NaCl, and 0.2 M $MgCl_2$

$[G]_o$, M[a]	with	d(% H)/dt,[b] h^{-1}	d(% G)/dt,[c] h^{-1}	d(% D_{all})/dt,[d] h^{-1}	p/(2′+3′)[e]	2′/3′[f]
0.0005		0.49	0.53	g		
0.0007		0.50	0.58	g		
0.0010		0.49	0.59	g		
0.0053		1.01	1.12	0.053		
0.0060	0.05 M KCl	1.04	1.02	0.091		
0.0101		1.15	1.14	0.094		
0.0219	0.05 M KCl	1.13	1.29	0.150		
0.0310	0.033 M[h]	1.47	1.67	0.251		
0.0333		1.37	1.68	0.198		
0.0363		1.26	1.57	0.200		
0.0370		1.36	1.51	0.220		
0.0450		1.35	1.66	0.272		4.3
0.0530		1.80	2.17	0.378	0.26	4.8
0.0690		1.67	2.00	0.414	0.30	4.6
0.0830		1.55	1.94	0.472	0.33	4.8
0.0660	0.0073 M[h]	1.67	2.08	0.379		3.6
0.0710	0.0148 M[h]	1.59	2.08	0.409	0.24	4.0
0.0720	0.0076 M[i]	1.49	1.98	0.424	0.34	3.6
0.0680	0.0137 M[i]	1.43	1.90	0.333	0.24	3.7
0.0690	0.0061 M[j]	1.46	1.86	0.390	0.42	3.7
0.0680	0.0147 M[j]	1.35	1.69	0.342	0.26	4.6

[a] Formal 2-MeImpG concentration determined by weight. [b] Initial slope of a plot of percent hydrolysis product, H, with time where H includes both 5′GMP and HEPES-pG. [c] Initial slope of a plot of percent 2-MeImpG as a function of time. [d] Initial slope of a plot of percent total dimer, D_{all}, formed as a function of time. Experimental error of the slopes estimated at less than 20%. [e] Average ratio (±10%) of pyrophosphate-linked vs internucleotide-linked dimers. [f] Average ratio (±10%) of 2′-5′ vs 3′-5′-linked dimers. [g] Not detected. [h] Poly(U). [i] Poly(C)·poly(G). [j] Poly(I).

Table 2. Rate Data for Dimer Formation in the Presence of Poly(C) Determined by C18 Chromatography

[T],[a] M	$[G]_o$,[b] M	r[c]	d(% H)/dt[d] h^{-1}	d(% G)/dt[e] h^{-1}	d(% D)/dt[f] h^{-1}	d[D]/dt (C),[g] M h^{-1}	TD d[D]/dt,[h] M h^{-1}
0.0050	0.0600	0.85	1.92	3.94	2.02	7.03×10^{-4}	6.87×10^{-4}
0.0050	0.0550	0.86	1.71	3.77	2.06	6.57×10^{-4}	6.43×10^{-4}
0.0080	0.0600	0.85	1.81	5.23	3.42	1.19×10^{-3}	1.17×10^{-3}
0.0080	0.0550	0.85	2.19	5.69	3.50	1.12×10^{-3}	1.10×10^{-3}
0.0080	0.0500	0.88	2.21	5.15	2.94	8.53×10^{-4}	8.40×10^{-4}
0.0080	0.0450	0.86	1.38	3.35	1.97	5.14×10^{-4}	5.05×10^{-4}
0.0140	0.0510	0.87	1.77	5.55[i]	3.78	1.12×10^{-3}	1.11×10^{-3}

[a] Poly(C) concentration in cytidine equivalents. [b] Formal 2-MeImpG concentration determined by weight. [c] Factor r corrects for substrate purity (see ref 14). [d] Initial slope of a plot of percent hydrolysis product, H, with time where H includes both 5′GMP and HEPES-pG. [e] Initial slope of a plot of percent 2-MeImpG as a function of time. [f] Calculated difference between d(% H)/dt and d(% G)/dt. Experimental error of the slopes estimated at 15 to 20%. [g] Initial rate of 3′-5′-linked dimer formation in the presence of poly(C) (see Results, eq 2′). [h] Calculated rate of TD 3′-5′-linked dimer formation (see Results). [i] Experiments with higher rates of dimerization could not be analyzed with this method.

system exhibits a substantially higher regioselectivity for 3′-5′-linkage formation compared to the other two linkages. A lower limit of 15−20 for preference of the formation of the 3′-5′-linkage over the 2′-5′-linkage can be estimated from our data, i.e., by comparing the sum of the HPLC areas under peak 6 and the areas of later eluting peaks with the area under peak 3 (Figure 1b).

Additional observations in similar systems suggest that the remarkable catalysis and regioselectivity exhibited in the poly(C)/2-MeImpG is unique for this system. For example, product distributions determined under our analytical conditions with systems $d(pC)_{50}$/2-MeImpG[10,11] and poly(C)/ImpG[4d] indicate that both internucleotide-linked dimers are detected and formed in comparable yields especially with the former system (Figure 1c: peak 2 represents the 2′-5′-linked and peak 4 the 3′-5′-linked). More importantly, the small yield of the RNA dimer detected with $d(pC)_{50}$/2-MeImpG cannot be attributed to consumption by oligomerization, because the latter reaction is slow (Figure 1c, top). Interestingly, neither of the above three TD reactions catalyzes the formation of the pyrophosphate-linked dimer (product distribution as a function of time, data not shown), which is a major product in the reaction occurring in solution (Figure 1a and Scheme 1, bottom).

Two methods were employed for the determination of the rate of RNA dimer formation. The first method, abbreviated C18 because samples were analyzed by C18 chromatography (see Experimental Section), was used here for the first time and it is based on the approximation that at early reaction times the substrate is consumed by the hydrolysis and the dimerization pathways and that the elongation pathway can be neglected, i.e., $-$d(% G)/dt = d(% H)/dt + d(% D_{all})/dt. Since with poly(C) the dimer formed is practically only the RNA dimer (D_{all} = D), then d(% D)/dt can be determined from the difference between d(% G)/dt and d(% H)/dt. Values for d(% G)/dt and d(% H)/dt were determined as described for samples without poly(C), and they are listed in Table 2 together with the calculated d(% D)/dt values.

The second method, abbreviated RPC5 because samples were analyzed by RPC5 chromatography (see Experimental Section), had been used earlier.[6a] This method is based on the hypothesis that the observed concentration

(10) Chen, C. B.; Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1985**, *181*, 271−279. Gangopadhyay, S.; Kanavarioti, A. Unpublished results.

(11) Kozlov, I. A.; Politis, P. K.; Van Aerschot, A.; Busson, R.; Herdewijn, P.; Orgel, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 2653−2656, and references therein.

Table 3. Rate Data for Dimer Formation in the Presence of Poly(C) Determined by RPC5 Chromatography

entry no.	$[G]_o$, M[a]	[T], M[b]	r[c]	d[D]/dt (C),[d] M h^{-1}	TD d[D]/dt,[e] M h^{-1}	ref
1	0.005	0.050	0.88	1.0×10^{-6}	8.8×10^{-7}	*f,g*
2	0.008	0.050	0.88	2.9×10^{-6}	2.6×10^{-6}	*f*
3	0.010	0.050	0.91	9.0×10^{-6}	8.5×10^{-6}	*f*
4	0.015	0.050	0.92	2.9×10^{-5}	2.8×10^{-5}	*f*
5	0.020	0.050	0.91	7.8×10^{-5}	7.6×10^{-5}	*f,g*
6	0.030	0.050	0.90	2.4×10^{-4}	2.4×10^{-4}	*f*
7	0.040	0.050	0.87	5.3×10^{-4}	5.2×10^{-4}	*f,g*
8	0.045	0.050	0.88	8.1×10^{-4}	8.0×10^{-4}	*f*
9	0.010	0.030	0.92	8.7×10^{-6}	8.2×10^{-6}	
10	0.015	0.030	0.91	3.2×10^{-5}	3.1×10^{-5}	
11	0.020	0.030	0.92	6.2×10^{-5}	6.0×10^{-5}	
12	0.030	0.030	0.92	2.2×10^{-4}	2.2×10^{-4}	*g*
13	0.040	0.030	0.90	4.6×10^{-4}	4.5×10^{-4}	*g*
14	0.020	0.020	0.92	5.7×10^{-5}	5.5×10^{-5}	*g*
15	0.005	0.010	0.89	3.3×10^{-7}	2.1×10^{-7}	*g*
16	0.020	0.010	0.93	5.2×10^{-5}	5.0×10^{-5}	
17	0.040	0.010	0.92	3.4×10^{-4}	3.3×10^{-4}	
18	0.020	0.035	0.91	6.2×10^{-5}	6.0×10^{-5}	
19	0.008	0.025	0.87	3.3×10^{-6}	3.0×10^{-6}	
20	0.045	0.005	0.93	4.0×10^{-4}	3.9×10^{-4}	
21	0.020	0.002	0.93	2.0×10^{-5}	1.8×10^{-5}	*g*

[a] Formal 2-MeImpG concentration. [b] Poly(C) concentration in cytidine equivalents. [c] Factor r corrects for substrate purity (see ref 14). [d] Initial rate of 3′-5′-linked dimer formation in the presence of poly(C) (see Results and ref 6a). [e] Calculated rate of TD 3′-5′-linked dimer formation (see Results). [f] Kinetic data from ref 6a. [g] From ref 6b; unidentified, this work.

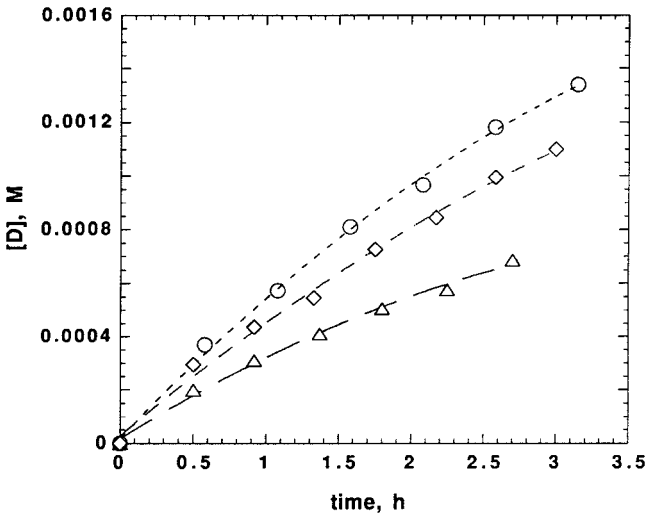


Figure 3. Plot of the concentration of 3′-5′-linked dimer, [D], as a function of time obtained from a reaction with 0.040 M 2-MeImpG and 0.010 M (triangles), 0.030 M (diamonds), and 0.050 M (circles) poly(C). The lines are second-order fits to the data, and d[D]/dt (C) was obtained from the initial slopes of these plots.

of oligomers is equal to the concentration of D, if D was the only product of the reaction. Hence $[D] = [pG^{3'}pG] + \Sigma[(pG)i]$, $i \geq 3$ where $(pG)_i$ are the 3′-5′-linked oligomers of length i. This hypothesis is valid at early reaction times, when the concentration of the oligomers is much less than the concentration of the template, and hence synthesis of oligomers by ligation is negligible. The yield of oligoguanylates as a function of time was determined by RPC5 chromatography. Figure 3 illustrates representative plots of [D] as a function of time. The slight downward curvature is attributed to a combination of two factors: the depletion of monomer and to the fact that, with time, substantial amounts of oligomers are formed which form stable complexes with poly(C) and therefore inhibit the dimerization process. The slope of the lines at zero time provides initial rates, henceforth referred to as d[D]/dt(C) to indicate the presence of poly(C); such slopes are listed in Table 3.

Discussion

Bimolecular Mechanism of 2-MeImpG Dimerization in Solution. Control experiments were performed with 2-MeImpG alone as well as with 2-MeImpG in the presence of polymers, such as poly(U), poly(I), and the double-stranded poly(C)·poly(G). The experiments with poly(U) and poly(C)·poly(G) were designed to test whether or not 2-MeImpG dimerization is facilitated based on the notion that guanosine forms a non-Watson−Crick base-pair with U and two types of hydrogen-bonded triplets with C·G.[12] The effect on the rate was tested at two different concentrations of polymer. An additional reaction was performed with 0.05 M KCl (instead of 0.05 M polymer that is supplied as the potassium salt) at two different monomer concentrations. None of these controls exhibited noticeable differences compared to the results obtained in the reaction of G alone. Therefore all these controls were pooled together, listed in Table 1, and plotted in Figures 4 and 5. It is likely that the scatter seen in these figures is due to slight inhibition exhibited by some of the polymers.

Table 1 indicates that at very low concentration of monomer the rate of substrate disappearance is equal to the rate of 5′GMP appearance, i.e. d(% H)/dt = d(% G)/dt = 0.53×10^{-2} h^{-1} (average from the first three entries), whereas at the higher concentrations d(% H)/dt < d(% G)/dt = 2.04×10^{-2} h^{-1} (average from the values at 0.053, 0.069, and 0.083 M monomer). The observed increase in the rate of substrate disappearance with initial monomer concentration, $[G]_o$, could, in principle, be attributed to the onset of dimerization that becomes important with the higher concentration of monomer. However, Figure 4 indicates hydrolysis is also accelerated and therefore partially responsible for the increased substrate disappearance. The increase in the rate of hydrolysis is surprising. It is interpreted to indicate a change in the thermodynamics of the system, perhaps by increasing stacking interactions[12] in the range 0.005 M < [2-Me-

(12) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; for hydrogen-bonding see p 120, for stacking interactions see p 134.

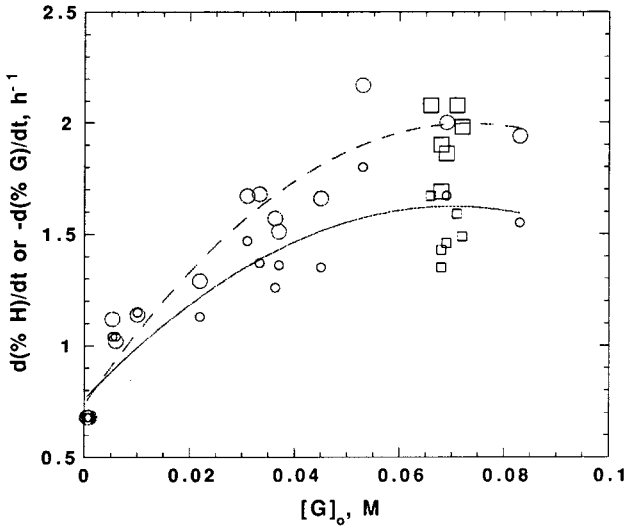


Figure 4. Rate of disappearance of 2-MeImpG (G) and appearance of the hydrolysis product (H) as a function of total monomer concentration, $[G]_o$. d(% G)/d*t*, large markers; d(% H)/d*t*, small markers; without polymer, circles; with polymer, squares. The lines are second-order fits through the data; no parameters are obtained from these fitting curves.

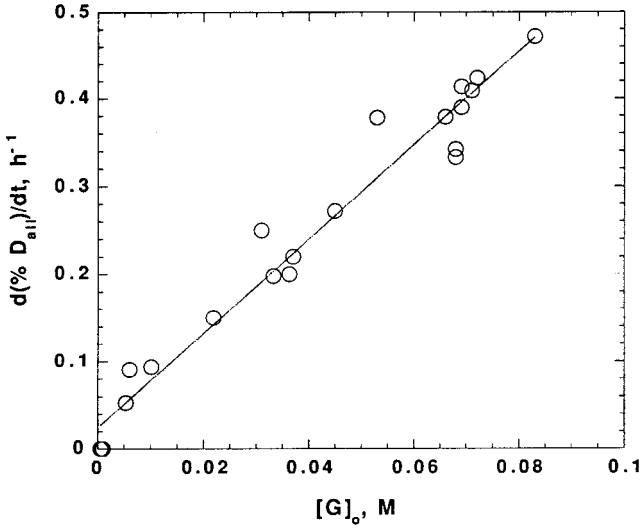


Figure 5. Rate of total dimer formation, d(% D_{all})/d*t* in h^{-1}, in the absence of poly(C) as a function of total monomer concentration, $[G]_o$, in the range 0.005 M ≤ $[G]_o$ ≤ 0.083 M; all values from Table 1 are included.

ImpG] < 0.05 M. This proposition relies on the notion that guanosine derivatives favor stacking interactions.[12,13]

The constancy of the product ratios p/(2′+3′) and 2′/3′ (see Experimental Section and Table 1) confirms the simultaneous formation of these three dimers from two molecules of 2-MeImpG.[7] The experimentally determined values of d(% D_{all})/d*t* (see Results) are reported in Table 1. They are in good agreement with the calculated difference d(% G)/d*t* − d(% H)/d*t* (not shown). A bimolecular mechanism of dimerization is described by the rate law given in eq 1 where *r* corrects for substrate purity[14] and k_d is the bimolecular rate constant for the formation of all three dimers. Equation 2 describes the relationship between the total dimer concentration and the percent yield obtained directly from the HPLC reports (see Experimental Section, *h* = 1.16 is the hypochromicity factor[8]). From eqs 1 and 2 it follows that d(% D_{all})/d*t* = 200/h $k_d r^2[G]_o$. The linear relationship between the experimentally determined d(% D_{all})/d*t* and $[G]_o$ (shown in Figure 5) is consistent with the proposed bimolecular mechanism of dimerization. The straight line drawn to fit the data in Figure 5 has a slope equal to 5.36 M^{-1} h^{-1}.[15] From this slope and *r* = 0.86,[14] one calculates k_d = 4.2×10^{-2} M^{-1} h^{-1}. In eq 3 k_d^2, k_d^3, k_d^p are the bimolecular rate constants for formation of the 2′-5′-, the 3′-5′-, and the pyrophosphate bond, respectively. From $k_d = 4.2 \times 10^{-2}$ M^{-1} h^{-1} and the average values *p*/(2′+3′) = 0.28 and 2′/3′ = 4.17 of the product ratios reported in Table 1, one obtains $k_d^3 = 6.3 \times 10^{-3}$ M^{-1} h^{-1}.

$$d[D_{all}]/dt = k_d(r[G]_o)^2 \quad (1)$$

$$d[D_{all}]/dt = h/200[G]_o d(\%\ D_{all})/dt \quad (2)$$

$$k_d = k_d^2 + k_d^3 + k_d^p \quad (3)$$

Synthesis of the RNA Dimer on the Template. Several lines of evidence summarized in Results lead to the conjecture that the synthesis of RNA dimer is, for all practical purposes, exclusively catalyzed by poly(C); estimated lower limit of TD 3′-5′:2′-5′ ≈ 15−20 (see Results). Values of d(% D)/d*t*, obtained as the difference d(% G)/d*t* − d(% H)/d*t* (see Results, C18 method), are listed in Table 2. From d(% D)/d*t* the rate of dimer formation in the presence of poly(C), d[D]/d*t* (C), was calculated via eq 2′ which is another version of eq 2 and is also listed in Table 2.

$$d[D]/dt\ (C) = h/200[G]_o d(\%\ D)/dt \quad (2')$$

Values d[D]/d*t* (C) determined with the RPC5 method (see Results, RPC5 method) were pooled together with values obtained from earlier measurements conducted under identical conditions with the ones used here; they are all reported in Table 3. In both Tables, 2 and 3, the rate of dimer formation that occurs exclusively on the template, TD d[D]/d*t*, was obtained by subtracting out a small contribution coming from the dimerization that occurs in solution ($d[D]/dt = k_d^3(r[G]_o)^2$ see eqs 1 and 3). Taken together the two tables suggest a complex dependence of the TD rate on both monomer and polymer concentration. What is more surprising is that the reactivity of this TD reaction varies by more than 5000-fold with a 12-fold change in monomer concentration. This is seen by comparing the lowest rate obtained in this study (2.1×10^{-7} M^{-1} h^{-1} with $[G]_o$ = 0.005 M and [T] = 0.01 M, see entry 15 in Table 3 where [T] stands for poly(C) concentration expressed in monomer equivalents) with the highest rate (1.1×10^{-3} M^{-1} h^{-1}, see entries 3, 4, and 7 in Table 2 with 0.051 M ≤ $[G]_o$ ≤ 0.060 M and 0.008 M ≤ [T] ≤ 0.014 M). The strong dependence of the rate on monomer concentration confirms earlier

(13) Zimmerman, S. B. *J. Mol. Biol.* **1976**, *106*, 663−672.

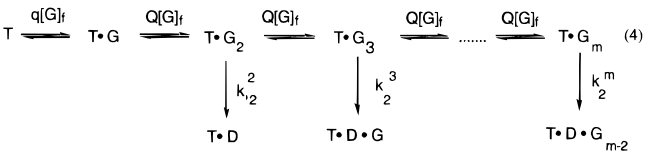
(14) Factor *r* is the ratio of [G]/$[G]_o$ where [G] is the average value for the monomer concentration between the initial and the endpoint of the interval Δ*t*, and $[G]_o$ is the formal concentration of the substrate determined by weight. Substrate purity was better than 97%. In actuality *r* corrects for both the purity of the substrate as well as the fact that the activated monomer is consumed during incubation. For the reactions in the absence of poly(C) *r* = 0.86 ± 0.04.

(15) Another interpretation of the data in Figure 5 is a curvilinear behavior with a plateau at $[G]_o$ > 0.05 M that would support the stacking interactions proposed earlier. Both interpretations coincide at $[G]_o$ < 0.05 M.

results at a constant 0.05 M poly(C).[6a] In order to rationalize this dependence, it is helpful to review the conclusions drawn from another TD dimerization.

Recently we completed a kinetic study of the dimerization of the deoxyguanosine 5′monophosphate 2-methylimidazolide, 2-MeImpdG, on poly(C) leading exclusively to the synthesis of the pyrophosphate-linked dimer, dG$^{5'}$ppdG.[16] In the presence of poly(C), dG$^{5'}$ppdG elongates by incorporation of additional monomers and formation of 3′-5′-internucleotide linkages to producedG$^{5'}$ppdG-(pdG)$_n$.[17] Accumulation of dG$^{5'}$ppdG at the onset of reaction was attributed to a slow elongation, most likely due to the poorer nucleophilicity of deoxyribose compared to the ribose. In contrast to the small yield of RNA dimer detected here, the accumulation of the pyrophosphate dimer allowed the kinetics to be determined directly from the dimer yield.[16] The highlights of that study were as follows: First, the dependence of the rate of dG$^{5'}$ppdG formation as a function of 2-MeImpdG and poly(C) concentration suggested a TD mechanism of dimerization. Secondly, the kinetic behavior of the 2-MeImpdG/poly(C) system as a function of monomer and polymer concentration led to the formulation of a number of principles governing TD chemistry. Thirdly, using a simple cooperative model of template−monomer association, the kinetic results were quantitatively correlated and yielded binding constants for 2-MeImpdG on poly(C) and an intrinsic rate constant of TD dG$^{5'}$ppdG formation.[16]

Cooperative Model for Template-Directed Dimerization. The same model used for the above TD pyrophosphate synthesis is exploited here for the synthesis of the RNA dimer. This model is summarized below. It presumes that 2-MeImpG, G, binds cooperatively at C sites of the template, T. Binding of G to an isolated site on the template occurs with an association constant, q, and binding of G adjacent to an occupied site occurs with an association Q with $Q > q$ and Q independent of the length of G-bound molecules (eq 4). $Q >> q$ implies that the process is cooperative.[18] As long as the relative positioning of two molecules is right for internucleotide bond formation, reaction can occur between any two adjacent monomers within a stack of monomers hydrogen-bonded to the template. The rate constants k_2^2, k_2^3, k_2^m, were defined as rate constants for formation of the dimer from a stack containing two, three, or m monomers, respectively. Initially, it is assumed that the rate of dimer formation is independent of the length of the stack, so that $k_2^2 = k_2$*, where k_2* is the intrinsic rate constant for TD 3′-5′-internucleotide bond synthesis from two molecules of 2-MeImpG. For statistical reasons $k_2^3 = 2k_2$*, $k_2^m = (m - 1)k_2$*, etc.



Equations 5−7 describe relationships between total or formal monomer concentration, $[G]_o$, the concentration of template-bound monomer, $[G]_{tem}$, and total concentration of template sites, [T]. $[G]_f$ is the concentration of the free monomer present in solution and θ is template occupancy. The system is close to saturation when $\theta \approx 1$ and far from saturation when $\theta \ll 1$. Another useful parameter is the free monomer concentration at half occupancy, $[G]_{0.5\theta}$, and it can be shown that $[G]_{0.5\theta} \approx 1/Q$.[18,19]

$$[G]_o = [G]_{tem} + [G]_f \quad (5)$$

$$\theta = [G]_{tem}/[T] \quad (6)$$

When $[G]_f = [G]_{0.50}$ then

$$[G]_{tem} = 0.5[T] \text{ and } [G]_o = [G]_{0.5\theta} + [G]_{tem} \quad (7)$$

Some of the principles first established with the 2-MeImpdG/poly(C) system were exploited here to facilitate the *experimental* estimation of k_2* and Q. (i) At a constant [T] and as long as the system is close to saturation, an increase in monomer concentration leaves the rate of dimer formation unchanged. This is because the additional monomer goes into solution. (ii) At a constant $[G]_o$ and as long as the system stays close to saturation, an increase in polymer concentration yields a proportional increase in the rate of dimer formation. This is because the additional template sites will be occupied by G. In such case the rate of dimer formation for a first-order TD reaction will be given by eq 8 where k_2* is the intrinsic rate constant of the process. (iii) When $[G]_o \gg [T]$, departure from the anticipated proportionality between TD d[D]/dt and [T] (eq 8) induced by a *small* decrease in $[G]_o$ implies that $\theta < 1$ and that $[G]_f \approx [G]_{0.5\theta}$. Under such conditions Q is estimated from eq 9 which is derived from eq 7.

$$\text{TD } d[D]/dt = k_2*[G]_{tem} \text{ and}$$

$$\text{when } \theta = 1, \text{ TD } d[D]/dt = k_2*[T] \quad (8)$$

$$Q = 1/[G]_{0.5\theta} \approx 1/([G]_o - 0.5[T]) \quad (9)$$

Based on the above concepts, experiments were designed in order to estimate k_2* and Q values (Table 2). For example, the first two entries in Table 2 at 0.055 M and 0.060 M monomer with 0.005 M template exhibit almost identical values of TD d[D]/dt (see principle i) and provide an average TD d[D]/dt $= 6.65 \times 10^{-4}$ h^{-1} which yields k_2* $= 0.13$ h^{-1} from [T] $= 0.005$ M and eq 8. Similarly, the 3rd and 4th entries (at 0.008 M poly(C)) in Table 2 yield k_2* $= 0.14$ h^{-1}, in excellent agreement with the value of k_2* obtained with 0.005 M poly(C) (see principle ii). Furthermore, a small decrease in $[G]_o$ from 0.055 M to 0.050 M (5th entry in Table 2) slows down the rate, indicating that the system is moving away from saturation. The observation that with $[G]_o = 0.045$ M the rate is half of the maximum obtained at saturation leads to the conclusion that under these conditions the template is approximately half occupied. These conditions (6th entry in Table 1) yield an estimate for $Q \approx 1/0.041 = 24.4$ M^{-1} from [T] $= 0.008$ M and eq 9. In addition, it is predicted from eq 8 and k_2* $= 0.14$ h^{-1} that a "saturated" system with [T] $= 0.014$ M should exhibit TD d[D]/dt $= 1.96 \times 10^{-3}$ h^{-1}. On the basis of this calculated rate, the conditions described by the 7th entry in Table 2 must be close to half-occupancy (from $1.96 \times 10^{-3}/1.11$

(16) Kanavarioti, A.; Gangopadhyay S. *J. Org. Chem.* **1999**, *21*, 7957−7964.

(17) Kozlov, I. A.; Orgel, L. E. *Origins Life Evol. Biosphere* **1999**, in press.

(18) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; Freeman and Co.: San Francisco, CA; Part III, 1980; p 864.

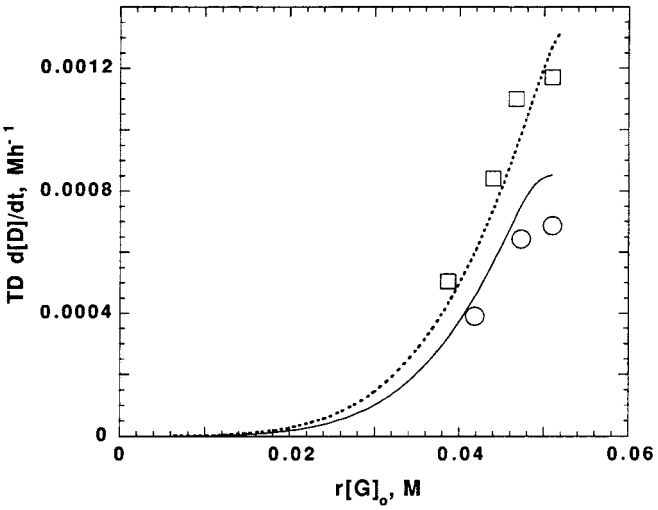


Figure 6. Rate of dimer formation on the template, TD d[D]/dt in M h^{-1}, as a function of 2-MeImpG concentration, r[G]$_o$, at a constant poly(C) concentration, expressed in C equivalents; squares at 0.008 M and circles at 0.005 M. TD d[D]/dt values from Table 2. [G]$_o$ is the formal monomer concentration, and r corrects for substrate purity.[14] The lines are computer simulations of a cooperative association model with q = 2.2 M^{-1}, Q = 22.5 M^{-1} (eq 4) and an intrinsic rate constant for TD dimerization k_2* = 0.18 h^{-1} assuming a TM$_3$ mechanism (see Discussion).

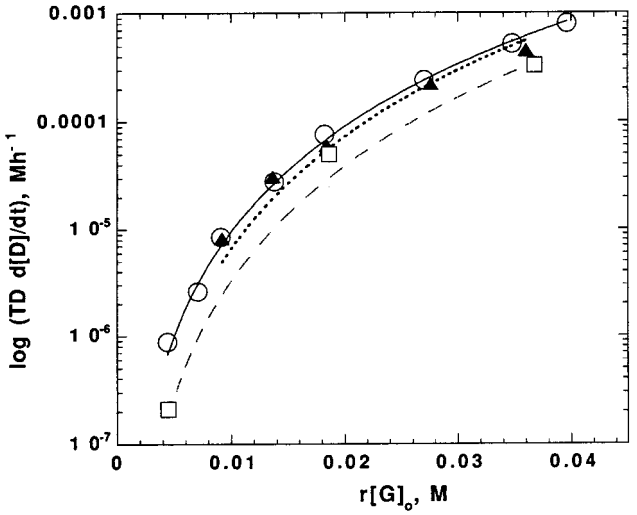


Figure 7. Rate of dimer formation on the template, TD d[D]/dt in M h^{-1}, as a function of 2-MeImpG concentration, r[G]$_o$, at a constant poly(C) concentration, expressed in C equivalents; circles at 0.05 M, filled triangles at 0.03 M, and squares at 0.01 M. TD d[D]/dt values from Table 3. [G]$_o$ is the formal monomer concentration, and r corrects for substrate purity.[14] The lines are computer simulations of a cooperative association model with q = 2.2 M^{-1}, Q = 22.5 M^{-1} (eq 4), and an intrinsic rate constant for TD dimerization k_2* = 0.18 h^{-1} assuming a TM$_3$ mechanism (see Discussion).

$\times 10^{-3} = 0.57$) and provide $Q \approx 1/0.044 = 22.7$ M^{-1} via eq 9 in good agreement with $Q \approx 24.4$ M^{-1} obtained above.

Computer Simulation of Monomer Distribution on the Template. Figure 6 illustrates the kinetic data obtained at 0.008 M and 0.005 M poly(C) (Table 2). The slowest rate with 0.005 M poly(C) corresponds to the rate listed as 20th entry in Table 3. The rates are plotted vs r[G]$_o$, to correct for substrate purity.[14] Similarly, Figure 7 illustrates the effect of changing monomer concentration (r[G]$_o$) at three different poly(C) concentrations, i.e., with 0.01 M (entries 15−17 in Table 3), 0.03 M (entries 9−13) and 0.05 M poly(C) (entries 1−8). It is worth noticing that under these conditions, highly "unsaturated" as will be shown later, the dependence of the rate on poly(C) concentration is negligible between 0.05 M and 0.03 M and weak between these two families of data and the family at 0.01 M poly(C). The rates here were plotted as the log of the rate reflecting the large range in reactivity. The lines in both figures are the result of a computer simulation of a specific version (see below) of the mechanism presented in eq 4. This simulation was initially based on the values of Q and k_2* obtained as described above which were then further optimized to fit all the data. The optimized values are Q = 22.5 ± 1.5 M^{-1} and k_2*= 0.18 ± 0.02 h^{-1}. Together with these two the value of q = 2.2 ± 0.2 M^{-1} was used.[19]

In its simplest version the model assumes that various template-bound monomer stacks have the same intrinsic reactivity k_2*. However, since in stacks containing more than two monomers bond formation can occur between any two neighbors, statistical corrections were applied.[6a] The relationship between each rate constant and k_2* for this mechanism, so called "M$_2$ mechanism", is given by eq 10. On the basis of this mechanism values TD d[D]/dt can be calculated from eqs 11 and 12. In eq 12 [T·G$_2$], [T·G$_3$], [T·G$_4$], and [T·G$_m$] are the concentrations of the various stacks.[6a,20]

$$k_2^{\,2} = k_2^*;\ k_2^{\,3} = 2k_2^*;\ k_2^{\,4} = 3k_2^*;$$
$$k_2^{\,m} = (m-1)k_2^* \quad (10)$$

$$\text{TD d[D]/d}t = k_2^* F_D \quad (11)$$

$$F_D = [\text{T·G}_2] + 2[\text{T·G}_3] + 3[\text{T·G}_4] + \ldots + (m-1)[\text{T·G}_m] \quad (12)$$

Assuming M$_2$ mechanism is operating and using eqs 11 and 12 rates, TD d[D]/dt were calculated for 0.005 M ≤ [G]$_o$ ≤ 0.060 M and 0.002 M ≤ [T] ≤ 0.050 M (not shown). While these calculated rates fit well data with relatively high monomer concentration, they were approximately one order of magnitude faster than the experimentally determined rates at low monomer concentration, i.e., most of the data in Table 3. Thus mechanism M$_2$ was discounted. The next mechanism that we attempted to fit the data with was the M$_3$ mechanism

(19) Kanavarioti, A.; Hurley, T. B.; Baird, E. E. *J. Mol. Evol.* **1995**, *41*, 161−168, and references therein.

(20) Calculations were executed using Microsoft Excel on a Power Mac computer. The following is based on the description given in ref 6a. The concentration of T·G$_j$ species can be calculated from eqs 16 and 17 where [G]$_f$ is the free monomer concentration in solution, [T]$_f$ is the concentration of template sites that are free and whose next-neighbor sites on both sides are unoccupied, and [T] is the total concentration of template sites (eq 18). [T]$_f$ is found by multiple iteration carried out with Microsoft Excel's Solver. In practice, calculation was initiated by a chosen [G]$_f$, iteration provided the correct [T]$_f$, and then [G]$_o$ was obtained. For the simulation TD d[D]/dt was obtained as TD d[D]/dt = k_2*F_D and was plotted vs [G]$_o$ with r = l.

$$[\text{T·G}] = q[\text{T}]_f[\text{G}]_f \quad (16)$$

$$[\text{T·G}_j] = q[\text{T}]_f[\text{G}]_f Q^{j-1}[\text{G}]_f^{\,j-1} \quad (17)$$

$$[\text{T}]_f = [\text{T}] - [\text{G}]_{tem} - 2\sum_{j=1}[\text{T·G}_j] \quad (18)$$

described by eqs 13 to 15. This mechanism implies that the reaction is assisted by at least one next-neighbor molecule. The fit with this mechanism is excellent, and it is the one with which the lines were created in both Figures 6 and 7.

$$k_2^{\,2} = 0;\ k_2^{\,3} = k_2^{\,*};\ k_2^{\,4} = 2k_2^{\,*};\ k_2^{\,m} = (m-2)k_2^{\,*} \quad (13)$$

$$\mathrm{TD}\ \mathrm{d[D]}/\mathrm{d}t = k_2^{\,*}F_{\mathrm{D}} \quad (14)$$

$$F_{\mathrm{D}} = [\mathrm{T}\cdot\mathrm{G}_3] + 2[\mathrm{T}\cdot\mathrm{G}_4] + 3[\mathrm{T}\cdot\mathrm{G}_5] + \ldots + (m-2)[\mathrm{T}\cdot\mathrm{G}_m] \quad (15)$$

The corresponding M_4 mechanism (with $k_2^{\,2} = k_2^{\,3} = 0$, $k_2^{\,4} = k_2^{\,*}$ and $k_2^{\,m} = (m-3)k_2^{\,*}$) was tested and found unsatisfactory because the calculated rates were about 1 order of magnitude lower than the experimental rates obtained at low $[G]_o$. The observation that all mechanisms fit the data at the high concentrations is due to the fact that at high $[G]_o$ the template is quite full and, consequently, the concentration of the small stacks, i.e., $[T\cdot G_2]$, $[T\cdot G_3]$, $[T\cdot G_4]$, is negligible. Therefore experiments at low occupancy, i.e., with predominantly short stacks, provide a good test for the elucidation of the specific mechanism. Other mechanistic possibilities, such as a mechanism with $k_2^{\,2} < k_2^{\,3} = k_2^{\,*}$ and $k_2^{\,m} = (m-2)k_2^{\,*}$, although probable, were not tested because of the increased number of unknown parameters involved. Thus the proposed M_3 mechanism is the simplest mechanism consistent with the data (see Scheme 1).

Insights in Template-Directed Chemistry. In addition to the principles mentioned earlier,[16] the present study led to following insights. (i) The best experimental conditions to obtain reliable kinetic data are the ones with $[G]_o \approx [T]$. This is because when $[G]_o \approx [T]$ it takes a lot of monomer to occupy the template. Thus small changes in $[G]_o$ do not alter the occupancy dramatically and with it the rates, in contrast to the situation when $[G]_o \gg [T]$. (ii) It is seen that an approximate doubling of poly(C) concentration leaves the rates of the process unchanged (0.05 M/0.03 M, see Figure 7). This negligible dependence of the dimerization rate on [T] can be mistakenly interpreted as absence of catalysis. This phenomenon occurs because with highly "unsaturated" systems, such as the ones with 0.05 M and 0.03 M template, the increase in template concentration is counterbalanced by a comparable decrease in occupancy leading to negligible rate effects. (iii) The dimerization of 2-MeImpG in the absence of poly(C) was shown to be consistent with a bimolecular mechanism. The low solubility of 2-MeImpG did not allow experimentation under higher concentrations, i.e., $[G]_o > 0.1$ M, where one may anticipate a switch from second-order to first-order kinetics consistent with extensive stacking.[12,13,15] In the presence of poly(C) the kinetic behavior is more complex. Specifically, the rate of dimer formation exhibited a third order dependence at very low monomer/high template concentrations, followed by a second order and then a first-order dependence (linear part in Figure 6) with increasing occupancy, followed by a zero order dependence at saturated conditions ($[G]_o > 0.05$ M, plateau in Figure 6 with 0.005 M poly(C)). This rich kinetic behavior is a composite of the kinetics exhibited under conditions close and far from saturation in conjuction with the mechanistic prerequisite that efficient dimerization in the poly(C)/2-MeImpG system takes place within a stack of three or more monomers.[21]

Comparisons with Earlier Conclusions and Other Systems. The value $Q = 22.5\ \mathrm{M}^{-1}$ for the association constant at a site adjacent to an occupied site is a factor of 9 lower than the value $Q = 180\ \mathrm{M}^{-1}$ obtained from the binding isotherm for poly(C)/2-MeImpG under the exact conditions used here.[19] The binding isotherm was determined via hypochromicity measurements of poly(C)/2-MeImpG mixtures. The disparity in Q determined from binding and from kinetic studies may indicate that there are a number of ways to construct a poly(C)/2-MeImpG duplex, but that not all duplexes lead to dimerization. Nevertheless, since $Q = 22.5\ \mathrm{M}^{-1}$ was obtained directly from the kinetic data of this system, its usage here for the evaluation of the model is justified. Interestingly, the value $q = 2.2\ \mathrm{M}^{-1}$ obtained from the binding isotherm[19] is the one that gives the best fit with the kinetic data in this study (see Figures 6 and 7). An attempt to fit the data in Tables 2 and 3 with values of $q = 2.2\ \mathrm{M}^{-1}$ and $Q = 180\ \mathrm{M}^{-1}$, instead of $q = 2.2\ \mathrm{M}^{-1}$ and $Q = 22.5\ \mathrm{M}^{-1}$, was unsuccessful. This is because the first set of constants predicts a factor of two lower rates for the 0.03 M family compared to the 0.05 M family which is not borne out by the data. The earlier study of the kinetics of dimerization in the range $0.005\ \mathrm{M} < [G] < 0.045$ M at 0.05 M poly(C), used values $q = 2.2\ \mathrm{M}^{-1}$ and $Q = 180\ \mathrm{M}^{-1}$ to determine product distribution of the various stacks.[6a] Based on this distribution, the conclusion was drawn that the process of dimerization is optimal in a stack of six or more template-bound molecules with $k_2^{\,*} = 0.018\ \mathrm{h}^{-1}$. Although the basic features of the model (eq 4) are intact, the revised $Q = 22.5\ \mathrm{M}^{-1}$ leads to the conclusion that the process of dimerization is optimal in a stack of only three or more template-bound molecules with a 10-fold faster intrinsic rate constant of dimerization ($k_2^{\,*} = 0.18\ \mathrm{h}^{-1}$). Consequently, the dependence of the rate on template concentration together with the evaluation of Q directly from the kinetics led to an important refinement of the mechanistic and kinetic interpretations of the original model.[22]

Reaction of 2-MeImpdG with guanosine 5′-phosphorylmorpholinamide (mor-pG) in the presence of poly(C) leads to the formation of the two isomeric dimers mor-pG$^{3'}$pdG and mor-pG$^{2'}$pdG.[23] The dimerization kinetics in this system led to the conclusion that the template catalyzes dimerization as well as increases regioselectivity toward the RNA bond formation; the effect was small (about 4-fold) in both cases. The corresponding intrinsic rate for TD RNA bond formation was determined $k_d^{\,*} = 4.1 \times 10^{-3}$

(21) A critical evaluation of the numerical values of the constants obtained from the TD model is in order. First, it should be recognized that eq 4 represents the simplest possible model to describe a TD reactive system. The equilibria involved in this model are described by *only two* association constants, q and Q, and by a mechanism that limits the reactivity of two adjacent molecules to be either equal to zero or equal to the intrinsic rate constant k^*. On first inspection, the task is to fit a set of data with four "parameters". In actuality this is not so. It is seen that Q and k^* can be estimated directly from specifically designed experiments with systems close to saturation following certain rules.[16] Moreover, the value of q and the specific mechanism are both derived from the kinetics with systems at very low template occupancy. Therefore, if the experiments cover a large range of reactivity from highly unsaturated to saturated systems, then the obtained numerical values for q, Q, k^* and the choice of mechanism are, within the simple model, well evaluated.

(22) The argument can be made that the model for TD oligoguanylate elongation[6c] should be reevaluated by using the constants $q = 2.2\ \mathrm{M}^{-1}$ and $Q = 22.5\ \mathrm{M}^{-1}$, instead of $q = 2.2\ \mathrm{M}^{-1}$ and $Q = 180\ \mathrm{M}^{-1}$. However, we have no evidence, kinetic or otherwise, that $Q = 22.5\ \mathrm{M}^{-1}$ derived from the dimerization kinetics is a more reliable constant for the elongation reaction and hence a reevaluation will be postponed until a better understanding of the binding process is in place.

(23) Kanavarioti, A. *J. Org. Chem.* **1998**, *63*, 6830−6838.

h^{-1}. In view of $k_2^* = 0.18\ h^{-1}$ from 2-MeImpG/poly(C), the low k_d^* value determined with 2-MeImpG/mor-pG/poly(C) allows one of two conclusions to be drawn. Either the conditions under which $k_2^* = 4.1 \times 10^{-3}\ h^{-1}$ obtained in the 2-MeImpG/mor-pG/poly(C) were not "saturated" as presumed, or the configuration within the stack of mor-pG/2-MeImpdG is much less suited for reaction compared to the one in a stack of three 2-MeImpG molecules. The fact that 2-MeImpG has two reactive sites vs one each for 2-MeImpdG and mor-pG can only account for approximately 4-fold out of an approximately 40-fold different reactivity. Insights may come from molecular dynamics simulations (MD). Preliminary MD studies of a duplex between $(pC)_{10}$ and 10 2-MeImpG molecules suggest (data not shown) that 2-MeIm moieties of neighboring monomers are partially stacked on each other and form a third strand that wraps around the duplex formed by the polymer and the stacked guanosine monomers.[24] The third strand formed by the 2-MeIm moieties may then be responsible for a more rigid structure leading to a highly selective and efficient dimerization and oligomerization with poly(C)/2-MeImpG. The absence of such a stabilizing factor may lead to reduced catalysis and regioselectivity in the poly(C)/2-MeImpdG/mor-pG[23] and poly(C)/ImpG systems.[4d,10]

Association constants $q = 7\ M^{-1}$ and $Q = 9.5\ M^{-1}$ in the dimerization reaction with the 2-MeImpdG/poly(C) system were determined in a similar way as here. Comparison between these two sets of constants indicates that the deoxyribo derivative associates more strongly at an isolated site of the template and more weakly at a site adjacent to an occupied one compared to the ribo derivative. The conclusion that Q of 2-MeImpdG is smaller than Q of 2-MeImpG is reasonable, since it is known that the affinity decreases in the order RNA·RNA > RNA·DNA > DNA·DNA.[12] The reasons for the reversal in the corresponding values of q are unclear. Evidently, poly(C) catalyzes dimerization with high efficiency, regioselectivity, and specificity ($dG^{5'}ppdG$, from 2-MeImpdG and $pG^{3'}pG$ from 2-MeImpG). The value of $k_2^* = 0.18\ h^{-1}$ determined here for TD RNA dimer synthesis should be compared with the corresponding $k_p^* = 0.055\ h^{-1}$ for pyrophosphate formation in the 2-MeImpdG/poly(C) system. This comparison indicates that in the presence of the template the known higher reactivity[25] of pyrophosphate vs internucleotide bond synthesis is reversed, suggesting that the poly(C)/2-MeImpG complex is better suited for 3′-5′- than either 2′-5′- or pyrophosphate-bond formation. The recent discovery that hexitol nucleic acids, which also exhibit an A structure like RNA, are also efficient templates for polynucleotide synthesis with 2-MeIm-activated nucleotides points to the importance of the A structure among functional templates.[11]

To appreciate the catalytic effect of the template on the dimerization, it may be fair to compare the rate of dimerization under conditions that are optimal for dimerization in solution (0.1 M 2-MeImpG) and on the template (0.05 M template/0.1 M 2-MeImpG). These conditions are based on the upper solubility limits of monomer and polymer in water. One calculates that the rate of 3′-5′-linked dimer formation will be $d[D]/dt = 6.3 \times 10^{-5}\ M^{-1}\ h^{-1}$ in solution and $d[D]/dt\ (C) = 9 \times 10^{-3}\ M^{-1}\ h^{-1}$ on the template. These conditions yield a 150-fold catalysis and represent the minimum; the maximum is better than 5200 (Discussion and Table 3). In addition, the rate of hydrolysis is $5.3 \times 10^{-3}\ h^{-1}$ for both systems compared with the rate of dimerization on a saturated template ($k_2^* = 0.18\ h^{-1}$) renders the TD dimerization under optimal conditions 34 times more efficient than hydrolysis. With such a highly efficient 3′-5′ dimerization, it is not surprising that elongation to form oligoguanylates is also efficient.[4b,6] Hence the conjecture is made that dimerization represents a critical step in a TD polymerization and, perhaps, one that indicates whether or not the specific system under the conditions tested is well suited for TD chemistry.

Conclusions

The rate of 2-MeImpG dimerization to form the RNA-linked dimer has been obtained as a function of monomer and of poly(C) concentration. Rate comparisons suggest that poly(C) selectively catalyzes the synthesis of the 3′-5′-linked ribodiguanylate up to 5200-fold over the solution reaction. The dramatic effect of poly(C) on 2-MeImpG dimerization is contrasted with the undetectable effect of other polynucleotides that can not exploit Watson−Crick base-pairing. Dimerization in solution exhibits a simple second-order dependence on monomer concentration. In the presence of poly(C) and at very low template occupancy, the rate of RNA-dimer formation exhibits a third-order dependence on 2-MeImpG concentration followed by a second-order, then a first-order with increasing occupancy, and finally a zero-order dependence under saturated conditions. This behavior is quantitatively consistent with a TD mechanism of dimerization where efficient dimerization occurs within stacks consisting of three or more, but not two, monomers. The surprisingly small dependence of the rate of dimer formation on template concentration, when $[G] \approx [T]$, can be attributed to extremely low template occupancy. Rates of dimerization have been fitted to a simple cooperative model for poly(C)/2-MeImpG complexation and afforded an association constant at an isolated site $q = 2.2\ M^{-1}$ and adjacent to an occupied site $Q = 22.5\ M^{-1}$, and an intrinsic rate constant of TD dimerization $k_2^* = 0.18\ h^{-1}$. This surprisingly high rate constant suggests that strong catalysis and practically exclusive regioselectivity in the formation of the RNA dimer is exhibited within the duplex constructed from poly(C) and 2-MeImpG stacks. A better understanding of the factors involved in such catalysis could facilitate the design of efficient TD, nonenzymatic, polynucleotide syntheses with unprotected reactive monomers in water.

Acknowledgment. This research was supported by the Exobiology Program of the National Aeronautics and Space Administration (Grant No. NCC 2-534). We thank Dr. L. E. Orgel from the Salk Institute for discussions and for providing the RPC5 packing. We also thank Dr. Mark Fonda from the Planetary Biology Branch of NASA/Ames Research Center and Prof. C. F. Bernasconi from the Chemistry and Biochemistry Department of the University of California at Santa Cruz for providing the facilities and Prof. C. F. Bernasconi for helping us with the kinetic analysis and for reviewing the manuscript.

JO991216Q

(24) Cieplak, P.; Kanavarioti, A. Unpublished results.

(25) Rodriguez, L.; Orgel L. E. *J. Mol. Evol.* **1991,** *32,* 101−104. Kanavarioti, A.; Rosenbach, M. T. *J. Org. Chem.* **1991**, *56,* 1513−1521. Dolinnaya, N. G.; Tsytovich, A. V.; Sergeev, V. N.; Oretskaya, T. S.; Shabarova, Z. A. *Nucl. Acids Res.* **1991**, *19,* 3073−3080.