Kinetics of Template-Directed Pyrophosphate-Linked Dideoxyguanylate Synthesis as a Function of 2-MeImpdG and Poly(C) Concentration: Insights into the Mechanism

Anastassia Kanavarioti* and Sumana Gangopadhyay[1]

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

Received June 22, 1999

Aqueous solutions of deoxyguanosine 5′-monophosphate 2-methylimidazolide, 2-MeImpdG, yield primarily deoxyguanosine 5′-monophosphate, 5′dGMP, and pyrophosphate-linked dideoxyguanylate, $dG^{5'}ppdG$, abbreviated G_2^p (see Chart 1). The initial rate of G_2^p formation, $d[G_2^p]/dt$ in M h^{-1}, determined at 23 °C, pH 7.8, 1.0 M NaCl and 0.2 M Mg^{2+} by timed high-performance liquid chromatography (HPLC) analysis, exhibits a second-order dependence on 2-MeImpdG concentration, $[G]_o$, indicating a bimolecular mechanism of dimerization in the range $0.02\ M \leq [G]_o \leq 0.09\ M$. In the presence of polycytidylate, poly(C), G_2^p synthesis is accelerated and oligodeoxyguanylate products are formed by incorporation of 2-MeImpdG molecules. The kinetics of G_2^p formation as a function of both monomer and polymer concentration, expressed in C equivalents, were also determined under the above conditions and exhibited a complex behavior. Specifically, at a constant [poly(C)], values of $d[G_2^p]/dt$ typically increased with $[G]_o$ with a parabolic upward curvature. At a constant $[G]_o$, values of $d[G_2^p]/dt$ increase with [poly(C)], but level off at the higher poly(C) concentrations. As $[G]_o$ increases this saturation occurs at a higher poly(C) concentration, a result opposite to expectation for a simple complexation of two reacting monomers with the catalyst prior to reaction. Nevertheless, these results are shown to be *quantitatively* consistent with a template-directed (TD) mechanism of dimerization where poly(C) acts as the template to bind 2-MeImpdG in a cooperative manner and lead, for the first time, to the formulation of principles that govern template-directed chemistry. Analysis of the kinetic data via a proposed TD cooperative model provides association constants for the affinity between polymer and monomer and the intrinsic reactivity of 2-MeImpdG toward pyrophosphate synthesis. To the best of our knowledge, poly(C)/2-MeImpdG is the first system that could serve as a textbook example of a TD reaction under conditions such that the template is fully saturated by monomers and under conditions that it is not.

Introduction

Template-directed (TD) synthesis with phosphoimidazolide-activated mononucleotides has frequently been used as a model for the nonenzymatic construction of polynucleotides that carry information.[2] However, the principles which govern TD synthesis are not well understood and the presently known TD chemical systems are far from being efficient. Dimerization represents the formation of the primer that is necessary for further polymerization, and the dimer yield is one of the parameters that critically determine polymerization efficiency. Insights into the mechanism of dimerization would not only contribute to our understanding of TD synthesis, but also provide design principles toward more efficient polynucleotide synthesizing systems.

Polycytidylate (poly(C)) facilitates oligoguanylate synthesis with guanosine 5′-monophosphate 2-methylimidazolide, 2-MeImpG, as the activated monomer.[3,4] In the 2-MeImpG/poly(C) system the dimer (in this case the 3′-5′-linked one) does not accumulate because its formation is substantially slower than elongation.[4a] Dimer concentration could only be determined indirectly from the concentration of the oligomer products, thus making the kinetic analysis less straightforward. For this reason we were in search of a TD-oligomerizing system where the dimer accumulates.

The dimerization process was recently investigated in the reaction between guanosine 5′-monophosphate morpholinamide, mor-pG, and 2-MeImpdG.[5] Based on the observation that the yield of the 3′-5′-linked dimer increased linearly with poly(C) concentration, the conclusion was drawn that this dimerization is TD. It was determined that the polymer catalyzes dimerization of template-bound monomers by a factor of 4 to 5 over the reaction in solution and that it reverses the regioselectivity.[5] However, this system was not amenable to a determination of the kinetics as a function of monomer concentration because of low yields and low solubility of the guanosine derivatives.

(1) On leave of absence from the Department of Chemistry, Gurudas College, Calcutta-54, India.

(2) Joyce, G. F. *Cold Spring Harbor Symp Quant Biol* **1987**, *52*, 41−51. Orgel, L. E. *Nature* **1992**, *358*, 203−209. Joyce, G. F.; Orgel, L. E. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Lab. Press: Cold Spring Harbor, **1993**; pp 1−25. Kanavarioti, A. *Origins Life Evol. Biosph.* **1994**, *24*, 479−495. Bag, B. G.; von Kiedrowski, G. *Pure Appl. Chem.* **1996**, *68*, 2145−2152. Ferris, J. P.; Hill, A. R.; Liu, R.; Orgel, L. E. *Nature* **1996**, *381*, 59−61. Luther, A.; Brandsch, R.; von Kiedrowski, G. *Nature* **1998**, *396*, 245−248. Beier, M.; Reck, F.; Wagner, T.; Krishnamurthy, R.; Eschenmoser, A. *Science* **1999**, *283*, 699−703.

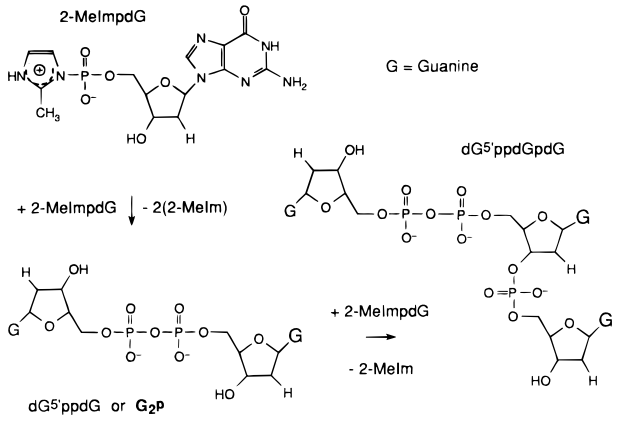
(3) Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201−218. Inoue, T.; Orgel, L. E. *Science* **1983**, *219*, 859−862. Fakhrai, H.; Inoue, T.; Orgel, L. E. *Tetrahedron* **1984**, *40*, 39−45.

(4) (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.

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

10.1021/jo990998k CCC: $18.00
Published on Web 09/25/1999

Chart 1



Recently, Orgel and co-workers reported the poly(C)-directed oligomerization of the *deoxy* derivative, 2-Me-ImpdG.[6] They established that the products include dG$^{5'}$ppdG (G_2^p) and a series of pyrophosphate-capped oligomers, dG$^{5'}$pp(dG)$_n$ with $n \leq 8$ (see Chart 1). Our own observations of this oligomerization indicated that oligomer formation lags behind G_2^p synthesis, allowing the conclusion to be drawn that oligomers longer than the dimer are elongation products of G_2^p. It is known that pyrophosphate synthesis is faster than internucleotide bond formation[7] and that TD elongation by incorporation of monodeoxynucleotides is slower than elongation by incorporation of ribonucleotides.[8] On the basis of these notions, we anticipated that the poly(C)/2-MeImpdG system should exhibit an enhanced dimerization accompanied by a diminished elongation that would result in the accumulation of higher dimer yields compared to the poly(C)/2-MeImpG system. Indeed, this is the case. Here we report the kinetics of G_2^p formation as a function of 2-MeImpdG as well as poly(C) concentration at 23 °C in the presence of 1.0 M NaCl, 0.2 M $MgCl_2$ and 0.5 M HEPES (pH 7.80 ± 0.05). The kinetics of dimerization in the absence of poly(C) were also determined.

Experimental Section

Materials, preparation of samples, pH measurements, and product identification follow already developed methods.[5] The sodium salt of 2-MeImpdG was 95.5% pure. Analysis by HPLC shows one unidentified contaminant at 4% eluting ahead of 5′dGMP which is only present at 0.5%. The contaminant stays intact during incubation. HPLC analysis was performed with a 1090 LC from Hewlett-Packard equipped with a diode array detector. Samples were incubated at 23 °C in the thermostated autosampler of the HPLC instrument, and the analysis was run automatically. Absorbance was monitored at 254 nm. The analysis of the samples containing only 2-MeImpdG was performed with a C18 Alltima[9] (3.2 × 250 mm, 5 μm by Alltech) solvent minimizer column run at 0.5 mL/min in conjuction with C18 chromatography (see Figure 1, top): solvent A: 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 45% B in 12 min. Analysis of samples containing poly(C) was done by HEMA chromatography[10] (see Figure 1, bottom). A HEMA-BIO 1000Q 10 μm analytical column from TESSEK Ltd, Czechoslovakia, was purchased from Melcor Technologies. Solvent A: 0.01 M NaOH, solvent B: 0.4 M $NaClO_4$ in solvent A. Gradient: 0 to 60% B in 30 min, 3 min wash with 80% B and 12 min reequilibration with solvent A. For the faster reactions, we used a somewhat steeper gradient: 0 to 21% B in 7 min and 21% to 50% in 7 min; 80% B for 2 min elutes poly(C). The $NaClO_4$ used was 99% ACS pure from Aldrich.

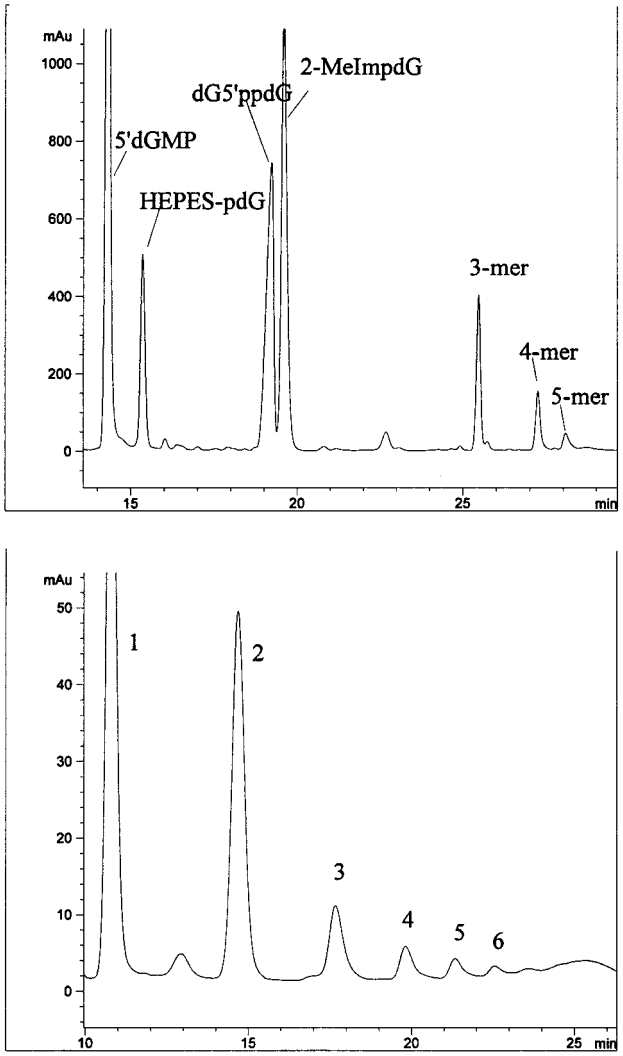


Figure 1. Self-condensation of 0.092 M 2-MeImpdG in the presence of 0.02 M poly(C) at 23 °C; *y*-axis, mAU stands for milliabsorbance units, *x*-axis, time in min. (Top) 5 days of incubation analyzed with C18 chromatography; (bottom) 19 days of incubation analyzed with HEMA chromatography. Identification of peaks: 1, 5′dGMP, HEPES-pdG and 2-Me-ImpdG coelute; 2 stands for dG$^{5'}$ppdG, abbreviated G_2^p in text, 3 stands for 3-mer, 4 stands for 4-mer, etc. (see Experimental Section).

Product distribution was obtained directly from HPLC reports as the percent of the total HPLC area corresponding

(6) Kozlov, I. A.; Politis, P. K.; Van Aerschot, A.; Busson, R.; Herdewijn, P.; Orgel, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 2653−2656. Kozlov, I. A.; Orgel, L. E. *Origins Life Evol. Biosph.*, **1999**, in press.

(7) Dolinnaya, N. G.; Tsytovich, A. V.; Sergeev, V. N.; Oretskaya, T. S.; Shabarova, Z. A. *Nucl. Acids Res.* **1991**, *19*, 3073−3080.

(8) Lohrmann, R.; Orgel, L. E. *J. Mol. Biol.* **1977**, *113*, 193−198. Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 317−322.

(9) C18 columns, including Alltima, retain poly(C) and oligoguanylates longer than five nucleotides. We also noticed that in the presence of poly(C), shorter oligoguanylates, including dimers, do not elute quantitatively probably because they are complexed by poly(C). Thus C18 chromatography, most likely, underestimates yields of oligoguanylates in samples containing poly(C). The advantage of C18 compared to HEMA chromatography is that it allowed the determination of substrate disappearance as well as the appearance of hydrolysis products that are eluted quantitatively even in the presence of poly-(C).

(10) Stribling, R. *J. Chromatogr.* **1991**, *538*, 474−477.

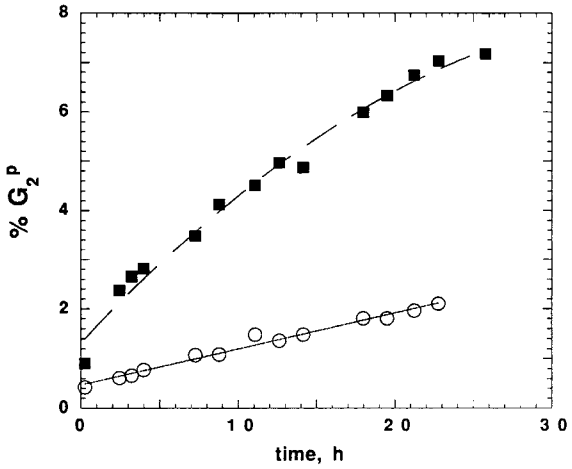


Figure 2. Percent yield of G_2^p formed as a function of time during the reaction of 0.09 M 2-MeImpdG (open circles) and the reaction of 0.08 M 2-MeImpdG with 0.005 M poly(C) (filled squares). The percent yield is expressed in monomer equivalents and is not corrected for hypochromicity. The downward curvature indicates that the rate of G_2^p formation is slowed at later times. Initial slope is equal to d(% G_2^p)/dt in h^{-1}.

to the initial substrate. This could be done because all guanosine monomers exhibit similar extinction coefficients. Percent G_2^p yields are reported in monomer equivalents and are uncorrected for hypochromicity. G_2^p concentration was calculated from $[G_2^p] = h[G]_o(\%\ G_2^p)/200$ where h is a hypochromicity correction factor with $h = 1.16$ (acidic solutions, C18 chromatography) and $h = 1.34$ (basic solutions, HEMA chromatography); the value of h was assumed to be the same as the one determined for $G^{5'}$ppG.[11] The hypochromicity of deoxyoligoguanylates longer than the dimer was assumed to be the same as the one determined for the ribooligoguanylates $h = 1$ (basic solutions).[4a]

Results

General Features. Under our conditions and in the absence of poly(C), there is still approximately 65% of 2-MeImpdG left intact after 1 day and 20% to 30% left after 5 days. Practically all substrate is consumed after about 20 days of incubation after which a solution of 0.09 M 2-MeImpdG yields 58.5% of 5′dGMP, 17.2% of HEPES-pdG (formed by reaction of the buffer), 16.2% of G_2^p, and less than 1% of longer oligomers. In the presence of poly(C), besides the above products, oligomers longer than the dimer are formed. After 20 days of incubation, a mixture of 0.02 M poly(C) and 0.092 M 2-MeImpdG produces 29% G_2^p and approximately 12% of longer oligomers up to the 6-mer, i.e., substantially higher yields than the 16.2% G_2^p and 1% oligomers observed in the absence of the polymer. These observations imply that poly(C) has, at least, doubled dimerization and has dramatically catalyzed the polymerization. A better way to quantify the effect of poly(C) on the dimerization is to monitor the percent yield of G_2^p formed as a function of time, as shown in Figure 2 for two reactions, one with 0.005 M poly(C) and the other without it. Plots such as these were either linear or slightly curved mainly because the reaction rate slows down as the substrate is being consumed.

(11) Diguanylates exhibit hypochromicity correction factor of $h = 1.16$ in neutral and acidic solutions (from T. Brian Hurley, Senior Thesis for the Degree of B. S. in Chemistry, 1993, University of California at Santa Cruz) and $h = 1.34$ in basic solutions.[4a]

Table 1. Detailed Kinetic Analysis of the Self-Condensation of 2-MeImpdG at 23 °C and pH 7.85 ± 0.05 in the Presence of 0.5 M HEPES, 0.2 M $MgCl_2$, and 1.0 M NaCl Performed with C18 Chromatography (see Experimental Section)

$[G]_o$, M[a]	d(% H)/dt,[b] h^{-1}	-d(% G)/dt,[c] h^{-1}	d(% G_2^p)/dt,[d] h^{-1}	d[G_2^p]/dt,[e] M h^{-1}
0.020	0.79	0.86	0.025	2.90×10^{-6}
0.040	0.78	0.77	0.046	1.07×10^{-5}
0.060	0.79	0.75	0.072	2.51×10^{-5}
0.080	0.70	1.00	0.086	4.00×10^{-5}
0.085	0.93	1.05	0.105	5.18×10^{-5}
0.090	0.83	0.87	0.094	4.91×10^{-5}
0.104	1.36	1.20	0.085	5.13×10^{-5}
0.092[f]	0.92	1.16	*g*	*g*

[a] Initial 2-MeImpdG concentration. [b] Slope of a plot of percent hydrolysis product, H, with time where H includes both 5′dGMP and HEPES-pdG. [c] Slope of a plot of percent 2-MeImpdG as a function of time. [d] Slope of a plot of percent G_2^p formed as a function of time; experimental error estimated at 15−20%. [e] d[G_2^p]/dt calculated based on eq 1 with $h = 1.16$. [f] Experiment run in the presence of 0.02 M poly(C) and analyzed with C18 chromatography. [g] Not determined.

G_2^p is one of two possible dimers anticipated from the dimerization of 2-MeImpdG.[12a] The other dimer is the internucleotide-linked pdGpdG that was not detected. In principle, the dimers are formed as phosphoimidazolide-activated derivatives, $dG^{5'}$p(2-MeIm)pdG and 2-MeImpdGpdG, which slowly hydrolyze. Nevertheless, the absence of any other peak in the solutions that could be attributed to $dG^{5'}$p(2-MeIm)pdG is interpreted to indicate that this derivative hydrolyzes much faster than what can be observed with our detection methods. Reaction of 5′dGMP with 2-MeImpdG provides an additional pathway for the formation of G_2^p, in analogy to the reaction of the corresponding ribo-derivatives, 5′GMP with 2-MeImpG leading to $G^{5'}$ppG.[12a] This reaction becomes increasingly important for long incubation times as the hydrolysis product accumulates and should result in an increased G_2^p formation. Enhanced G_2^p formation at longer times was, apparently, counterbalanced by substrate consumption, and this is why many plots of % G_2^p vs time showed good linearity. Experiments (not listed) indicate that the reaction between 5′dGMP and 2-MeImpdG, to produce G_2^p, is also catalyzed by poly(C). We believe that it is this additional pathway that produces the substantial yields of G_2^p that are observed both in the absence and in the presence of the polymer after long incubation times. To minimize this problem, samples were analyzed immediately after mixing and initial rates were determined from the first few hours of incubation.

Kinetics of 2-MeImpdG Dimerization in the Absence of Poly(C). Experiments in the absence of poly(C) were run with $0.02\ M \le [G]_o \le 0.104\ M$ for about 15 to 20 h and produced 1−4% of G_2^p, up to 18% of 5′dGMP, and about 4% of HEPES-pdG. Plots of % G_2^p as a function of time were generally linear and provided slopes, d(% G_2^p)/dt, listed in Table 1. The experimental error of these slopes is estimated at 15−20%. This table includes also slopes of lines obtained by plotting the percent of hydrolysis product H (H = 5′dGMP + HEPES-pdG) formed as well as the percent of substrate left unreacted, d(% H)/dt and d(% G)/dt, respectively. Rates of G_2^p formation, d[G_2^p]/dt, were calculated via eq 1 and are included in Table 1. In eq 1, $h = 1.16$ is the hypochro-

(12) (a) Kanavarioti, A. *Origins Life Evol. Biosph.* **1997**, *27*, 357−376. (b) Kanavarioti, A. *J. Mol. Evol.* **1998**, *46*, 622−632.

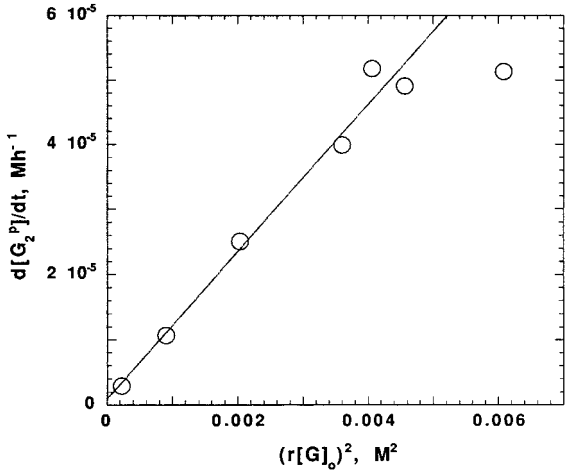


Figure 3. Rate of G_2^p formation, $d[G_2^p]/dt$ in M h^{-1}, in the absence of poly(C) as a function of the square of monomer concentration, $(r[G]_o)^2$ in the range 0.02 M $\leq [G]_o \leq$ 0.09 M. $d[G_2^p]/dt$ values from Table 1, $[G]_o$ the initial monomer concentration and r corrects for substrate purity (see Results). The line is a linear fit of the data, excluding 0.104 M G_o, with slope 0.0108 M^{-1} h^{-1}.

micity correction for the G_2^p under acidic conditions.[11] The last experiment listed in Table 1 was performed in the presence of 0.02 M poly(C) and was analyzed with C18 chromatography but for reasons explained in the Experimental Section d(% G_2^p)/d*t* was not determined from this experiment.[9]

$$d[G_2^p]/dt = \frac{h}{2 \times 100}[G]_o[d(\%\ G_2^p)/dt] \qquad (1)$$

Table 1 indicates that the rate of G_2^p formation increases substantially with monomer concentration, but levels off at $[G]_o \geq 0.09$ M. The effect of the enhanced dimerization on the disappearance of the substrate, d(% G)/d*t*, is negligible due to the fact that hydrolysis is by far the most important reaction in this system. In contrast to d(% G_2^p)/d*t*, d(% H)/d*t* values (excluding 0.104 M) in Table 1 indicate that hydrolysis is independent of monomer concentration and allow the calculation of an average d(% H)/d*t* = 0.80. From d[H]/d*t* = 0.01[d(% H)/d*t*] one calculates $k_H = 8 \times 10^{-3}$ h^{-1}, in excellent agreement with the k_H value determined earlier.[5] Moreover, the constancy of d(% H)/d*t* adds confidence to the increases seen with d(% G_2^p)/d*t*.

To correct for the fact that the substrate was only about 95% pure and that it was further consumed during incubation, we included a correction factor, *r*, with *r* = 0.75. This value is approximately the average value of the fraction of 2-MeImpdG observed at the beginning ([2-MeImpG]/$[G]_o$ = 0.85) and at the end (([2-MeImpG]/$[G]_o$ = 0.65) of the analyses. Hence, the $d[G_2^p]/dt$ values reported in Table 1 are plotted vs $(r[G]_o)^2$ and not vs $[G]_o^2$ (see Figure 3). Figure 3 shows that $d[G_2^p]/dt$ values, calculated via eq 1, exhibit a linear relationship with the square of substrate concentration up to 0.09 M. The linearity confirms the expected bimolecular mechanism of dimerization, and the slope of the straight line in Figure 3 provides the rate constant for the formation of the pyrophosphate bond, k_p. Based on $d[G_2^p]/dt = k_p(r[G]_o)^2$ we find $k_p = 1.08 \times 10^{-2}$ M^{-1} h^{-1}. The plateau seen in Figure 3 will be dealt with in the Discussion.

Table 2. Kinetics of G_2^p Formation in the Poly(C)-Catalyzed Self-Condensation of 2-MeImpdG at 23 °C and pH 7.85 ± 0.05 in the Presence of 0.5 M HEPES, 0.2 M $MgCl_2$, and 1.0 M NaCl as a Function of 2-MeImpdG and Poly(C) Concentration

entry no.	$[G]_o$, M[a]	poly(C), M[b]	d(% G_2^p)/d*t*, h^{-1} [c]	$d[G_2^p]/dt$ (C), M h^{-1} [d]	TD $d[G_2^p]/dt$, M h^{-1} [e]
1	0.060	0.016	0.34	1.37×10^{-4}	1.12×10^{-4}
2	0.057	0.008	0.30	1.15×10^{-4}	9.21×10^{-5}
3	0.060	0.005	0.22	8.84×10^{-5}	6.36×10^{-5}
4	0.080	0.016	0.50	2.68×10^{-4}	2.24×10^{-4}
5	0.080	0.005	0.27[f]	1.45×10^{-4}	1.00×10^{-4}
6	0.100	0.030	0.96	6.43×10^{-4}	6.09×10^{-4}
7	0.110	0.024	0.81	5.97×10^{-4}	5.46×10^{-4}
8	0.110	0.019	0.61	4.50×10^{-4}	3.92×10^{-4}
9	0.100	0.008	0.31[f]	2.08×10^{-4}	1.49×10^{-4}
10	0.110	0.004	0.19	1.41×10^{-4}	6.22×10^{-5}
11	0.10	0.005	0.36	2.41×10^{-4}	1.79×10^{-4}
12	0.02	0.005	0.09	1.21×10^{-5}	9.30×10^{-6}
13	0.12	0.016	0.74	5.95×10^{-4}	5.20×10^{-4}
14	0.10	0.016	0.64	4.29×10^{-4}	3.80×10^{-4}
15	0.05	0.016	0.28	9.38×10^{-5}	7.65×10^{-5}
16	0.04	0.016	0.18	4.82×10^{-5}	3.72×10^{-5}

[a] Initial 2-MeImpdG concentration. [b] Poly(C) concentration expressed in C equivalents. [c] Slope of a plot of percent G_2^p formed as a function of time; experimental error of the slopes estimated at 15−20%. [d] $d[G_2^p]/dt$ (C) calculated based on eqs 1 and 2 with *h* = 1.34. [e] TD $d[G_2^p]/dt$ is the contribution to the rate by the TD process only; calculated from $d[G_2^p]/dt$ (C) by subtracting the contribution to the rate from the reaction of the monomer in solution (see eq 3). [f] Average slope of duplicate experiments.

Kinetics of 2-MeImpdG Dimerization in the Presence of Poly(C). Experiments were conducted with 0.02 M $\leq [G]_o \leq$ 0.12 M and 0.004 M $\leq$ [poly(C)] $\leq$ 0.03 M; the upper range was dictated by solubility constraints, the lower range by detectable G_2^p yields. Samples were incubated for up to 7 h and analyzed automatically by HEMA chromatography (see Experimental Section). Lines of % G_2^p as a function of time were either linear or concave down (for a representative example see Figure 2); in the latter case the slope, d(% G_2^p)/d*t*, was determined as the zero time tangent by a second-order fit with a polynomial function. Values of d(% G_2^p)/d*t* have an estimated experimental error of 15−20%. Due to the relatively short reaction times, most experiments did not produce products longer than the G_2^p. However with high concentrations of monomer and polymer, oligomerization became evident and the corrected percent G_2^p (hypochromicity included, see Experimental Section), $h(\%\ G_2^p)_{cor}$, was calculated from eq 2. $h(\%\ G_2^p)_{cor}$ was then used instead of $h(\%\ G_2^p)$ in eq 1 in order to calculate $d[G_2^p]/dt$ (C), where C indicates that these measurements were done in the presence of poly(C).

$$h(\%\ G_2^p)_{cor} = 1.34(\%\ G_2^p) + 2/3(\%\ dG^{5'}p(pdG)_2) + 2/4(\%\ dG^{5'}p(pdG)_3) + \text{etc.} \qquad (2)$$

We conducted experiments with 0.058 ± 0.002 M (entries 1−3 in Table 2), 0.08 M (entries 4, 5), and 0.105 ± 0.005 M 2-MeImpdG (entries 6−11) at various poly(C) concentrations and additional experiments with 0.005 M (entry 12) and 0.016 M poly(C) (entries 13−16) at different monomer concentrations. The results summarized in Table 2 include calculated values of $d[G_2^p]/dt$ (C), using eq 1 with *h* = 1.34 (see Experimental Section). Table 2 also lists values TD $d[G_2^p]/dt$ that represent the contribution to the rate by the TD pathway only. TD $d[G_2^p]/dt$ were calculated from $d[G_2^p]/dt$ (C) by subtracting

out the rate contributed by the reaction in solution as shown in eq 3.

$$\text{TD } d[G_2^p]/dt = d[G_2^p]/dt\ (C) - k_p(r[G])^2 \qquad (3)$$

In eq 3 $k_p = 1.08 \times 10^{-2}$ M^{-1} h^{-1} and r, discussed earlier, was estimated from the C18 analysis of a 0.092 M 2-MeImpdG/0.02 M poly(C) sample (see last entry in Table 1). The value $r = 0.8$ was used for all the experiments done with poly(C). This value is somewhat higher than $r = 0.75$ used for the experiments in the absence of poly(C), but was dictated by the shorter incubation times used in the presence of the polymer. Also in eq 3 we used $[G] = [G]_o$ for experiments that have low template occupancy and $[G] = [G]_o - [T]_o$ for experiments that have high template occupancy (see Discussion). Specifically, $[G] = [G]_o - [T]_o$ was used only with experiments done at 0.11 ± 0.01 M 2-MeImpG. This approximation is an oversimplification, but it is justified because the template concentration is much lower than the monomer concentration and the contribution of the template-bound monomer to the total amount minimal. It is seen that TD $d[G_2^p]/dt$ are 10−25% less than $d[G_2^p]/dt$ (C); the above-mentioned approximations do not affect the conclusions.

Discussion

Synthesis of G_2^p in Solution. Figure 3 indicates that G_2^p formation in solution is second-order in the range $0.02\ M \le [G]_o < 0.09$ M and most likely proceeds by a bimolecular mechanism. The slope of the straight line in Figure 3 provides $k_p = 1.08 \times 10^{-2}$ M^{-1} h^{-1}. This value can be compared with $k_p = 1.7 \times 10^{-2}$ M^{-1} h^{-1} determined under otherwise identical conditions but in the presence of 0.2 M Mn^{2+} instead of Mg^{2+}.[5] The 60% lower k_p value observed with Mg^{2+} compared to Mn^{2+} is consistent with earlier observations indicating that Mn^{2+} is a better catalyst for dimerization than Mg^{2+}.[12b]

The low solubility of 2-MeImpdG precluded measurements at $[G]_o > 0.105$ M. The plateau observed in the range $0.09\ M \le [G]_o \le 0.104$ M supports the idea of a switch from a second-order to a pseudo-first-order mechanism. The latter could be attributed to intermolecular stacking interactions[13] becoming important at [2-MeImpdG]>0.09 M and, consequently, to a dimerization occurring within a stack of monomers. On the basis of this scenario, one calculates a pseudo-first-order rate constant, $k_p(\text{stack}) = 4.9 \times 10^{-4}$ h^{-1} (from $0.085 \times 1.16/200$ where 0.085 h^{-1} is the rate at the plateau in Figure 3), for dimerization within a stack and in the absence of a template.

Synthesis of G_2^p on the Template. Values of slopes, $d(\%\ G_2^p)/dt$, obtained in the presence of poly(C) are listed in Table 2, together with calculated $d[G_2^p]/dt$ (C) values (see Results, eqs 1 and 2). From the latter the rate of G_2^p formation that occurs exclusively on the template, TD $d[G_2^p]/dt$, was obtained by subtracting out a small contribution coming from the dimerization that occurs in solution (see eq 3, Results). These values, also listed in Table 2, depend both on monomer and polymer concentration, but in a complex way. For this reason, the data were grouped in families and plotted accordingly.

(13) Saenger, W. *Principles of nucleic acid structure*; Springer-Verlag: New York, 1984; p 134.

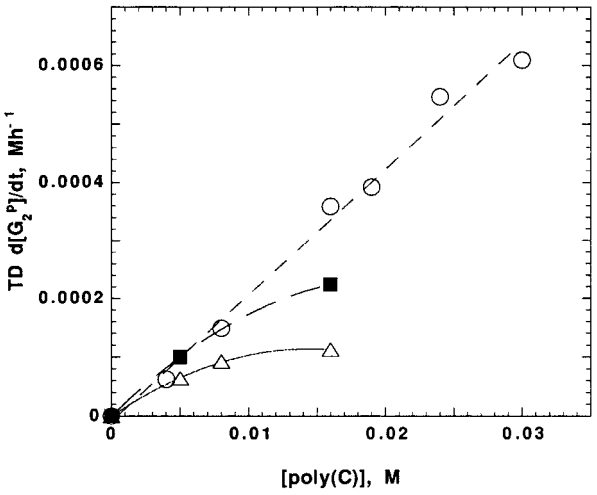


Figure 4. Rate of dimer formation on the template, TD $d[G_2^p]/dt$ in M h^{-1}, as a function of poly(C) at different initial 2-MeImpdG concentrations: circles, 0.105 ± 0.005 M; filled squares, 0.08 M; and triangles, 0.058 ± 0.002 M 2-MeImpdG. Poly(C) concentration expressed in C equivalents. TD $d[G_2^p]/dt$ values from Table 2. The lines are linear for the data at 0.105 M, and second-order fits for the lower concentrations.

Figure 4 illustrates the effect of increasing poly(C) concentration on the rates. In this figure TD $d[G_2^p]/dt$ are plotted vs [poly(C)] for three families of data (0.058 ± 0.002 M, 0.08 M and 0.105 ± 0.005 M 2-MeImpdG). Values of TD $d[G_2^p]/dt$ increase both with monomer and polymer concentration. Qualitatively speaking, the catalytic effect of the polymer is small at low $[G]_o$, and large at high $[G]_o$. The effect of the polymer at 0.06 M monomer is to catalyze the dimerization at most by a factor of 4.5, whereas the effect of the polymer at 0.10 M monomer becomes 12-fold at 0.03 M poly(C) compared to the reaction in solution. There is a clear saturation effect with the 0.058 and 0.08 M families and, perhaps, a beginning saturation with the 0.105 M family. It is worth noting that as monomer concentration decreases the saturation occurs at lower poly(C) concentration. This observation is inconsistent with the hypothesis that poly(C) acts as a "simple" catalyst facilitating dimerization by a three-molecule complexation prior to reaction. If this were the mechanism, then at low monomer concentration complete complexation would require relatively high polymer concentration (late plateau), and at high monomer concentration complete complexation would occur with lower poly(C) concentration (early plateau). Both predictions are opposite to observation.

Figure 5 illustrates the effect of increasing $[G]_o$ on the rates. In this figure values of TD $d[G_2^p]/dt$ are plotted vs monomer concentration for the 0.016 and 0.005 M poly(C) families. The rates were plotted vs $r[G]_o$, instead of $[G]_o$, to account for substrate purity (see Results). With both families, the rate exhibits approximately a second-order dependence on $r[G]_o$, but with the 0.016 M family the rates become increasingly faster with monomer concentration compared to the 0.005 M family. How can we understand these observations?

Template-Directed Mechanism of Dimerization. Because of the known affinity of guanosine monomers for poly(C),[3,13] it is presumed that any catalysis observed in di- or oligoguanylate synthesis is due to a TD mechanism. Such a mechanism was proposed for the first time in the poly(C)-directed 3′-5′-linked dimer formation from

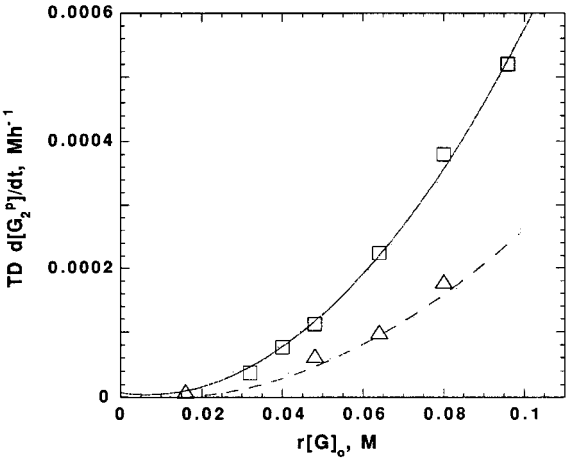
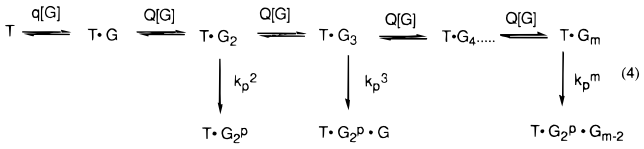


Figure 5. Rate of dimer formation on the template, TD d[G_2^p]/dt in M h^{-1}, as a function of 2-MeImpdG concentration, r[G]$_o$, at a constant poly(C) concentration, expressed in C equivalents; squares at 0.016 M and triangles at 0.005 M. TD d[G_2^p]/dt values from Table 2. [G]$_o$ is the initial monomer concentration and r corrects for substrate purity (see Results). The lines are computer simulations of a cooperative association model with $q = 7$ M^{-1}, $Q = 9.5$ M^{-1} and an intrinsic rate of TD dimerization $k_p^* = 0.06$ h^{-1} (line at 0.016 M) and $k_p^* = 0.07$ h^{-1} (line at 0.005 M, see Discussion).

2-MeImpG which was investigated as a function of monomer at a constant polymer concentration.[4a] This model is summarized below. It presumes that 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 $Q > q$. Q is assumed to be independent of the length of the stack. T·G_m is a template-bound stack of m monomers. T·G_2 stands for a stack of two monomers in contrast to T·G_2^p that identifies a template-bound dG$^{5'}$ppdG dimer standing alone or within a stack of monomers, T·G_2^p·G_{m-2} (eq 4). As long as the relative positioning of two molecules is right for pyrophosphate bond formation, reaction can occur between any two adjacent monomers within a stack of monomers hydrogen-bonded to the template. For simplicity it is assumed here that the rate of G_2^p formation is independent of the length of the stack, so that $k_p^2 = k_p^*$; k_p^* is the intrinsic rate constant for TD pyrophosphate synthesis from of 2-MeImpdG. For statistical reasons $k_p^3 = 2k_p^*$ and $k_p^m = (m - 1)k_p^*$.



(4)

Equations 5 and 6 describe the relationships between total monomer concentration, [G]$_o$, and the concentration of template-bound monomer, [G]$_{tem}$ (see later on how to calculate [G]$_{tem}$ from the mass balance and mass law of eq 4). [G]$_f$ is the concentration of the monomer present in solution, [T]$_o$ stands for the total poly(C) concentration, expressed in C equivalents, and θ is the occupancy.

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

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

It is convenient to differentiate between systems that are close to saturation, abbreviated "saturated" and systems that are far from saturation, abbreviated "unsaturated". Assuming a one to one C:G binding, "saturated" systems are such that [G]$_{tem}$ = [T]$_o$ and unsaturated when [G]$_{tem}$ < [T]$_o$. Another useful parameter in TD systems is the free monomer concentration at half occupancy, [G]$_{0.5\theta}$, where [G]$_{0.5\theta} \approx 1/Q$.[15]

Under saturated conditions (S) and at a constant poly(C) concentration an increase in the monomer concentration should have no effect on template occupancy because $\theta = 1$. Therefore any experimental parameter, such as the rate of a TD process, that is proportional to [G]$_{tem}$ should remain unchanged upon further increase in monomer concentration. Similarly, an increase in template concentration at constant monomer concentration will have no effect on θ as long as saturation applies. On the other hand, an experimental parameter that depends on [G]$_{tem}$ will increase proportionally to template concentration.

Under unsaturated conditions (U) and at constant poly(C) concentration, an increase in monomer concentration will increase occupancy. As a consequence of the enhanced θ, the experimental parameter will increase. At constant monomer concentration, an increase in template concentration will result in a lower occupancy. In this case the experimental parameter will be subject to two effects. One is an increase because of the enhanced [T]$_o$, the other a decrease because of the lower occupancy. These two effects will counteract each other and, depending on the specific case, the experimental parameter will either exhibit an increase or remain unchanged. If, e.g., the measured parameter exhibits an increase, this increase will be less than the increase in template concentration, i.e., parameter increase will not be proportional to template concentration (see different scenario above for a saturated system).

The features that distinguish saturated from unsaturated conditions are summarized in Table 3. They are applicable to other TD reactions, not only dimerizations. For example, TD oligoguanylate elongation was found to exhibit a decreased rate with poly(C) concentration increasing from 0.002 to 0.05 M at a constant 0.02 M 2-MeImpG (see condition [T]$_o \uparrow \Rightarrow \theta \downarrow$, at constant [G]$_o$).[4b] Moreover, these conditions or principles can be exploited as diagnostic tools to identify whether a system is close or far from saturation. Alternatively, if Q is known or can be estimated, the relationship between [G]$_o$, [G]$_{0.5\theta}$, and [T]$_o$ is useful in predicting whether a system is close to saturation or not. Knowing if a system is close or far from saturation is critical when the efficiency of polynucleotide synthesizing systems is concerned. This is because dimerization and elongation become kinetically optimal processes in a saturated system.[4] In this context, comparison of oligomerization yields between two TD systems, one being close to saturation and the other far, could lead to erroneous conclusions regarding which one is more efficient under optimal conditions.

Below we will show how the relationships in Table 3 rationalize the data in Figures 4 and 5. As the concentration of polymer increases and sites for binding become available, catalysis is evident, as seen by the increased

(14) Kanavarioti, A.; Hurley, T. B.; Baird, E. E. *J. Mol. Evol.* **1995**, *41*, 161−168.
(15) For an explanation see ref 14 and references therein.

Table 3. Relationships between Monomer, Polymer, and Occupancy in a Cooperative System under "Saturated" and "Unsaturated" Conditions

S	$\theta = 1$	$[G]_o > [G]_{0.5\theta} + [T]_o$[16]	$[G]_o \uparrow \Rightarrow \theta$ constant at constant $[G]_o$	$[T]_o \uparrow \Rightarrow \theta$ constant at constant $[T]_o$
U	$0 < \theta < 1$	$[G]_o \leq [G]_{0.5\theta} + [T]_o$[16]	$[G]_o \uparrow \Rightarrow \theta \uparrow$ at constant $[T]_o$	$[T]_o \uparrow \Rightarrow \theta \downarrow$ at constant $[G]_o$

$d[G_2^p]/dt$ in the presence of the template compared to the reaction in solution (Figure 4). TD $d[G_2^p]/dt$ is the measured experimental parameter that in this system is presumed to be directly proportional to the saturated template/monomer complex. In the range $0 \leq$ [poly(C)] ≤ 0.004 M and with 0.08 and 0.105 M monomer the values of TD $d[G_2^p]/dt$ are the same, suggesting that under these conditions, i.e., $[T]_o \leq 0.004$ M and $[G]_o > 0.08$ M, the system is close to saturation (see $[G]_o \uparrow \Rightarrow \theta$ const., at constant $[T]_o$). The fact that the two lines digress at $[T]_o > 0.004$ M implies that the system switches from saturated to unsaturated allowing an estimate for $[G]_{0.5\theta} = 0.10 \pm 0.01$ M from $[G]_o \approx [G]_{0.5\theta} + [T]_o$ and $[T]_o \ll [G]_o$.

Furthermore, with the family of 0.105 M monomer an increase in template concentration results in a proportional increase in TD $d[G_2^p]/dt$ (see above discussion for $[T]_o \uparrow \Rightarrow \theta$ constant, at constant $[G]_o$) up to about 0.024 M poly(C) where there is, perhaps, an indication for a plateau. This plateau is clearly evident with the 0.08 and 0.058 M families and occurs at about 0.013 and 0.008 M poly(C), respectively. The curvarture seen with the 0.08 and 0.058 M families is a manifestation of the condition $[T]_o \uparrow \Rightarrow \theta \downarrow$, at constant $[G]_o$, indicating unsaturated conditions. The plateaus that occur earlier with lower monomer concentration can be attributed to the fact that it takes much less polymer before the system's occupancy is lowered.

Figure 5 depicts the effect of increasing monomer concentration on the rate with two familes of data, one with 0.005 M and the other with 0.016 M poly(C). Here there is an approximately 3-fold increase in $[T]_o$ between the two families, but less than a 3-fold increase in the rate in the range of measurements up to 0.1 M 2-MeImpdG. For example, at 0.1 M 2-MeImpdG TD $d[G_2^p]/dt$ measures 1.79×10^{-4} M h^{-1} with 0.005 M and 3.8×10^{-4} M h^{-1} with 0.016 M. This amounts to a 2.1-fold increase in the rate for a 3.2-fold increase in $[T]_o$, suggesting that one or both tested systems are unsaturated. Moreover, the conclusion can be drawn that with decreasing $[G]_o$ these two systems are led farther from saturation. Thus, Figure 5 can be seen as the experimental manifestation of the condition $[T]_o \uparrow \Rightarrow \theta \downarrow$, at constant $[G]_o$.

Computer Simulation of Monomer Distribution on the Template. The lines shown in Figure 5 are the result of a computer simulation. This simulation was based on the mathematical equations derived from eq 4. The concentration of the various $T{\cdot}G_n$ species in eq 4 was calculated assuming a 100-unit long template and was based on the mass balance equations.[17] Optimization of q and Q values to fit the experimental data in Figure 5 provided values $q = 7 \pm 0.5$ M^{-1} and $Q = 9.5 \pm 0.5$ M^{-1}, the latter being in good agreement with $[G]_{0.5\theta} = 0.10 \pm 0.01$ M estimated from Figure 4. The relatively large value of q is a manifestation of the fact that the template exhibits a catalytic effect even at low concentrations. The best fit in Figure 5 was obtained with an intrinsic rate constant for TD pyrophosphate formation $k_p^* = 0.07$ h^{-1} (for the 0.005 M) and $k_p^* = 0.06$ h^{-1} (for the 0.016 M). Simulations were also performed for the rest of the experimental conditions listed in Table 2 using $q = 7$ M^{-1} and $Q = 9.5$ M^{-1}. The k_p^* values derived from these simulations and the experimentally determined TD $d[G_2^p]/dt$ ranged from 0.035 h^{-1} to 0.075 h^{-1}. Thus an average $k_p^* = 0.055 \pm 0.02$ h^{-1} is seen to satisfactorily fit all the data in Table 2. Models presuming more efficient dimerization in longer stacks, i.e., where $k_p^2 = 0$, $k_p^3 = k_p^*$, $k_p^m = (m - 2)k_p^*$, or where $k_p^2 = k_p^3 = 0$, $k_p^4 = k_p^*$, $k_p^m = (m - 3)k_p^*$, etc., did not result in better fits.

The $k_p^* = 0.055$ h^{-1} for TD pyrophosphate bond formation suggests a 110-fold catalysis by the template when compared to the value of k_p(stack) $= 4.9 \times 10^{-4}$ h^{-1} which assumes synthesis within a stack of monomers. The value $k_p^* = 0.055$ h^{-1} compared to $k_d^* = 4.1 \times 10^{-3}$ h^{-1}, the TD rate constant for 3′-5′-internucleotide bond formation between morpholine-pG and 2-MeImpdG,[5] indicates a 13-fold higher reactivity of the phosphate anion compared to the ribose 3′-OH at pH 7.8 and in the presence of Mg^{2+}. These two systems, i.e., poly(C)/2-MeImpdG and poly(C)/mor-pG/2-MeImpdG, are similar in the sense that the reacting molecules act either as a nucleophile or as an electrophile. This is in contrast to the poly(C)/2-MeImpG, system, where each monomer has dual reactivity. An even larger reactivity difference, 30-fold, has been observed between pyrophosphate synthesis and internucleotide bond formation with deoxyribose 3′-OH as the nucleophile and carbodiimide activation.[7] Such large reactivity differences, perhaps 30-fold or higher, can explain why in the present study no internucleotide-linked dimer was detected.

Conclusions

In nonenzymatic TD polymerizations, dimerization is slower than elongation so that the dimer does not

(16) These relationships are not mathematically exact, and they only serve as a guide to deduce whether a system is close or far from saturation or, alternatively, to estimate Q.

(17) Calculations were executed using Microsoft Excel on a Power Mac computer. The following is based on the description given in ref 4a. According to the proposed mechanism (eq 4), the rate of G_2^p formation is given by eq 7 where r corrects for substrate purity and F_D is given by eq 8. $[G]_{tem}$ can be calculated from eq 9.

$$\text{TD } d[G_2^p]/dt = k_p r F_D \quad (7)$$

$$F_D = [T{\cdot}G_2] + 2[T{\cdot}G_3] + 3[T{\cdot}G_4] + 4[T{\cdot}G_5] + \ldots + (m-1)[T{\cdot}G_m] \quad (8)$$

$$[G]_{tem} = [T{\cdot}G] + 2[T{\cdot}G_2] + 3[T{\cdot}G_3] + 4[T{\cdot}G_4] + 5[T{\cdot}G_5] + \ldots + m[T{\cdot}G_m] \quad (9)$$

The concentration of $T{\cdot}G_j$ species can be calculated from eqs 10 and 11 where $[G]_f$ is the free monomer concentration in solution, $[T]_f$ is the free concentration of template sites that are free and whose next-neighbor sites on both sides are unoccupied (eq 12). $[T]_f$ is found by multiple iteration carried out with Microsoft Excel.

$$[T{\cdot}G] = q[T]_f[G]_f \quad (10)$$

$$[T{\cdot}G_j] = q[T]_f[G]_f Q^{j-1}[G]_f^{\,j-1} \quad (11)$$

$$[T]_f = [T]_o - [G]_{tem} - 2\sum_{j=1}^{j=m}[T{\cdot}G_j] \quad (12)$$

In practice, calculation was initiated by a chosen $[G]_f$, iteration provided the correct $[T]_f$ and then $[G]_o$ was obtained. To create the simulated lines, TD $d[G_2^p]/dt$ was obtained as TD $d[G_2^p]/dt = k_p F_D$, and it was plotted vs $r[G]_o$ with $r = 1$.

accumulate which makes it difficult to extract kinetic information about the dimerization step. However, the pyrophosphate-linked dimer, G_2^p, is formed relatively rapidly from two molecules of 2-MeImpdG and elongates slowly, so that the kinetics of its formation could be conveniently monitored by HPLC. Accumulation of G_2^p made it possible, for the first time, to determine the kinetics of a TD dimerization as a function of both monomer and polymer concentration. Rates of G_2^p formation obtained in the range 0.02 M ≤ [2-MeImpdG] ≤ 0.12 and 0.004 M ≤ [poly(C)] ≤ 0.03 M varied by 65-fold from the slowest to the fastest rate. This study allowed conditions to be outlined that could diagnose whether a *TD reaction* in general, not only a TD dimerization, is close or far from saturation based on a limited set of kinetic determinations. A TD cooperative model proposed earlier for another TD dimerization ((poly(C)/2-MeImpG that forms pG3′pG) was successful in fitting all the experimental results using one set of association constants, $q = 7\ M^{-1}$ and $Q = 9.5\ M^{-1}$, and one intrinsic rate constant $k_p^* = 0.055 \pm 0.020\ h^{-1}$ for TD pyrophosphate synthesis. Furthermore, it was concluded that dimerization occurs efficiently within a stack of two or more template-bound monomers.

Acknowledgment. This research was supported by the Exobiology Program of the National Aeronautics and Space Administration (Grant No NCC 2-5309). We thank Dr. L. E. Orgel from the Salk Institute for constructive criticism. 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. Bernasconi for helping us with the kinetic analysis.

JO990998K