J|A|C|S
ARTICLES
Published on Web 06/13/2003

Sequence- and Regioselectivity in the Montmorillonite-Catalyzed Synthesis of RNA

Shin Miyakawa† and James P. Ferris*

Contribution from the Department of Chemistry and New York Center for Studies on the Origins of Life, Rensselaer Polytechnic Institute, Troy, New York 12180

Received January 24, 2003; E-mail: ferrij@rpi.edu

Abstract: The possible role of catalysis in forming a limited number of RNAs from activated monomers is investigated by examining the sequence- and regioselectivity in the montmorillonite-catalyzed formation of RNA dimers and trimers. The reactivity of A was similar to that of G, and C was comparable in reactivity to U. Yet the reactivity of the purine nucleotides differed from that of the pyrimidines. In the reaction of nucleotides (pN) with activated monomers (ImpN), the sequence- and regioselectivity was $Pu^{3'}Py > Pu^{3'}Pu = Pu^{2'}Py > Pu^{2'}Pu$. The 5′-pyrimidine initiated dimers formed less efficiently than the 5′-purine initiated dimers. Trimer formation was investigated by the synthesis of 8 dimers (pNpN) and measuring the yields of trimers formed in the reaction of each dimer with a mixture of equal molar amounts of four activated monomers. The reactivity of the dimers depended on the nucleotide attached to the 3′-end of the RNA and the regiochemistry of the phosphodiester bond. Rules based on these studies are proposed to predict the sequence- and regioselectivity of the RNAs formed in montmorillonite-catalyzed reactions. These rules are consistent with the structures of the 2−5-mers formed in the reaction of equimolar amounts of ImpA and ImpC. This research establishes that the montmorillonite catalyst limits the number of RNA oligomer isomers formed. The potential significance of these findings to the origins of life is discussed.

Introduction

Polymers such as proteins and nucleic acids have a central role in life on Earth. The prebiotic synthesis of ordered polymers by random synthetic pathways was recognized early on as a major problem in the origins of life on Earth.[1] It was calculated that it would be necessary to synthesize 10^{130} isomers to generate a protein of defined sequence that contained the 20 biological amino acids. Recent studies on the origins of life have focused on the origins of the postulated RNA world where RNA was the most important polymer in this early life form. The same problem of the formation of too many isomers has been discussed for the prebiotic synthesis of RNA. An RNA of sufficient length to initiate catalysis and replicate with high fidelity may have contained only 40 monomer units.[2,3] A minimal requirement for the origins of life is two RNAs each catalyzing the synthesis of the other. The random formation of two RNAs containing 40 nucleotides would require the formation of 10^{48} isomers weighing 10^{28} g, an amount comparable to the mass of the Earth.[3]

We focus on the role of catalysis in prebiotic synthesis as a possible solution to the selective formation of RNAs.[4] This approach originated from the discovery that some montmorillonite clays catalyze the condensation of activated monomers (5′-phosphorimidazolides of nucleosides, ImpN, N = A, G, U, and C, Figure 1) to form oligomers of RNA containing 6−14 monomers. These oligomers contain 2′,5′-, 3′,5′- phosphodiester bonds as well as pyrophosphate linkages. More recent studies demonstrated that a 10-mer of a nucleic acid could be elongated to a 40−50-mer by the montmorillonite-catalyzed addition of activated monomers to the 10-mer.[5,6] Our proposal is that the montmorillonite is effective in generating 40−50-mers because it limits the number of isomeric RNAs formed. The purpose of this study was to examine the sequences and the types of phosphodiester bonds present (2′,5′ or 3′,5′) in the dimers and trimers formed from the reactions of mixtures of ImpN and 5′-nucleotides to determine if sequence- and regioselectivity is actually present in these oligomers.

Initial studies of the sequence- and regioselectivity in the reaction of a mixture of equal amounts of the four activated monomers of RNA to form dimers revealed that the predominant reaction products contained a 5′-purine nucleotide and that 84% of the products were the following dimers in decreasing order of formation pApC, pGpC, pApU, pApA, pGpA, pGpU, pGpG, and pApG.[7] A different approach to determining the reactivity

† Present address: Department of Basic Medical Science, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan.

(1) Kaplan, R. W. In *Molecular Evolution I. Chemical Evolution and the Origin of Life*; Buvet, R., Ponnamperuma, C., Eds.; North-Holland Publishing Company: Amsterdam-London, 1971; pp 319−329.
(2) Szostak, J. W.; Ellington, A. D. In *The RNA World. The Nature of Modern RNA Suggests a Prebiotic RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1993; Vol. Monograph 24, pp 511−533.
(3) Joyce, G. F.; Orgel, L. E. In *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA*, 2nd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1999; pp 49−77.
(4) Ferris, J. P. *Origins Life Evol. Biosphere* **1993**, *23*, 307−315.
(5) Ferris, J. P.; Hill, A. R., Jr.; Liu, R.; Orgel, L. E. *Nature* **1996**, *381*, 59−61.
(6) Ferris, J. P. *Origins Life Evol. Biosphere* **2002**, *32*, 311−332.
(7) Ertem, G.; Ferris, J. P. *Origins Life Evol. Biosphere* **2000**, *30*, 411−422.

10.1021/ja034328e CCC: $25.00

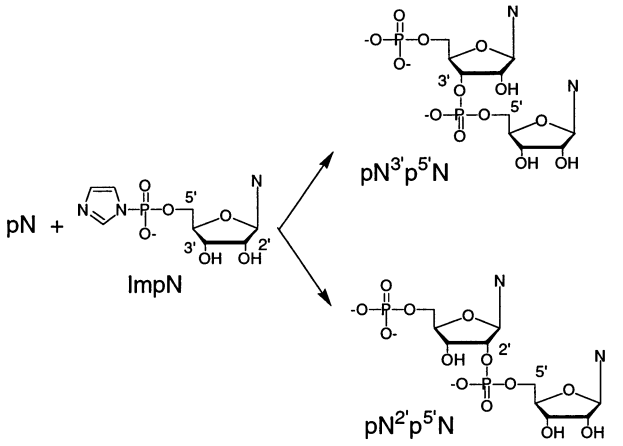

Figure 1. Linear dimers formed by the reaction of ImpN (N = A, G, C, and U).

Table 1. Yields of pNpN Synthesized from pN + ImpN

PyPy	yields (%)[a]		PyPu	yields (%)[a]	
reactions	3′,5′	2′,5′	reactions	3′,5′	2′,5
pA + ImpU → pApU	26	6.7	pA + ImpA → pApA	11	2.9
pA + ImpC → pApC	25	7.5	pA + ImpG → pApG	9.2	2.1
pG + ImpU → pGpU	22	9.3	pG + ImpA → pGpA	7.2	2.6
pG + ImpC → pGpC	23	14	pG + ImpG → pGpG	6.3	4.3
PyPy	yields (%)[a]		PyPu	yields (%)[a]	
reactions	3′,5′	2′,5′	reactions	3′,5′	2′,5′
pC + ImpC → pCpC	2.0	5.2	pC + ImpA → pCpA	0.28	0.66
pC + ImpU → pCpU	2.1	3.5	pC + ImpG → pCpG	0.78	0.73
pU + ImpC → pUpC	1.9	3.6	pU + ImpA → pUpA	0.45	0.45
pU + ImpU → pUpU	2.8	5.2	pU + ImpG → pUpG	0.69	0.69

[a] Dimer yields corrected for the molar extinction coefficients and hypochromicity.[8]

of the monomers was undertaken in this study, which was then extended to the formation of trimers.

Results and Discussion

Dimer Selectivity. The selectivity and reactivity of monomers was investigated by the reactions of equimolar amounts of 5′-nucleotides (pN) with each ImpN to obtain the corresponding dimers (Figure 1). In general, it was observed that purine nucleotides (Pu) react at the same rate and pyrimidine nucleotides (Py) react at the same rate but the rates of reaction of the two groups differ. The dimer yields were used as a measure of the reactivity of the 5′-nucleotides and the ImpNs used (Table 1). The results obtained were comparable to those obtained in the previous study.[7] One important difference is that the yields of pGpG are higher in the present study because it was observed that a 0.1 M NaCl−30% acetonitrile mixture is more effective in eluting the G dimers from montmorillonite than is the 0.1 M ammonium acetate used in the previous study.

Marked sequence selectivity was observed in dimer formation. The reactivity of the purine nucleotides was greater than that of the pyrimidine nucleotides in initiating dimer formation. For example, the yields of pApU and pUpU were 33% and 8.0%, respectively. The RNAs with a 5′-purine will be emphasized in the discussion because the longer RNAs formed will all have purines at the 5′-ends of their chains. The reactivity of the phosphorimidazolides of the pyrimidine nucleosides is greater than that of the corresponding purine nucleosides. For example, in the reaction of ImpU with pA and in the reaction of ImpA with pA, the yields of pApU and pApA were 33% and 14%, respectively. The sequence- and regioselectivity for dimers with a 5′-purine nucleotide is in the order $Pu^{3'}Py > Pu^{3'}Pu = Pu^{2'}Py > Pu^{2'}Pu$. The 3′,5′-phosphodiester bond predominates over the 2′,5′-linkage, when the pN is a purine nucleotide. For example, the yields of $pA^{3'}pU$ and $pA^{2'}pU$ were 26% and 6.7%, respectively. In contrast, the 2′,5′-linkage predominates over the 3′,5′-linkage, when the pN is a pyrimidine. For example, the yields of $pC^{2'}pU$ and $pC^{3'}pU$ were 3.5% and 2.1%, respectively.

In dimer formation, there are two reaction pathways possible, the reaction of ImpN with pN and the reaction of ImpN with ImpN. In the reaction of pA and ImpC, the pApC was 33%, while the yield of pCpC was less than 1%. This shows that the ImpC reacts more efficiently with pA than with ImpC. In contrast, in the reaction of pC with ImpA, the yield of pApA was greater than that of pCpA, indicating that the ImpA reacts with the ImpA more efficiently than with the pC. The same reactivities were observed in the other purine−pyrimidine reactions. In a mixture of ImpA and pG, pGpA was more abundant than pApA, indicating that the ImpA reacts with the pG more efficiently than with the ImpA. The same order of reactivity was observed for the reaction of other activated purine nucleotides with a 5′- purine nucleotide and with an activate pyrimidine nucleotide with a 5′-pyrimidine nucleotide.

Trimer Selectivity. Trimer selectivity was determined by the reaction of dimers (pNpN) with equimolar amounts of ImpN. Eight of the 32 possible dimers were synthesized for the investigation of their reactivity for elongation. The dimers used were selected on the basis of their yields in dimer formation,[7] their ease of synthesis, and their utility in understanding the selectivity observed in the elongation reactions (Table 1). The 8 dimers and the number of each prepared belong to the following general classes: $Pu^{3'}Py$ (2), $Pu^{3'}Pu$ (3), $Pu^{2'}Pu$ (1), $Pu^{2'}Py$ (1), and $Py^{3'}Pu$ (1). The first 5 are major products in dimer synthesis, $Pu^{2'}Pu$ and $Pu^{2'}Py$ were used to determine the effect of the adjacent 2′,5′-phosphodiester bond on elongation, and $Py^{3'}Pu$ was prepared for use in an investigation the effect of an adjacent 5′-Py on the elongation of a 3′-Pu. Two of the dimers that were not investigated, $Py^{3'}Py$ and $Py^{2'}Pu$, were formed on low yields (Table 1), so they would not have a major role in RNA elongation, while the third, $Py^{2'}Py$, would not be expected to elongate efficiently because it has a $-^{2'}Py$ at its 3′-terminus (see below).

Four of the dimers used were prepared by the reaction of an equimolar amount of the phosphoramidate of the nucleoside (NH_2pN) with an equimolar amount of the phosphorimidazolides of the nucleosides in the presence of montmorillonite (eq 1).

$$NH_2pN + ImpN' \rightarrow NH_2pNpN' \rightarrow pNpN' \qquad (1)$$

The NH_2 group was removed from NH_2pNpN by acid hydrolysis. The dimers prepared by this route included $pA^{3'}pC$, $pA^{3'}pU$, $pA^{2'}pC$, and $pA^{3'}pA$. $pG^{3'}pA$ and $pA^{3'}pG$ were synthesized by the reaction of pG with ImpA and pA with ImpG, respectively. $pU^{3'}pA$ is formed in very low yields by these synthetic procedures (Table 1), so it was prepared by the phosphorylation of a commercial sample of $U^{3'}pA$ using a mixture of T4 polynucleotide kinase and ATP. $pA^{2'}pA$ was prepared by the uranyl acetate-catalyzed reaction of ImpA.[9] All

(8) Warshaw, M. M.; Tinoco, I., Jr. *J. Mol. Biol.* **1966**, *20*, 29−38.

Table 2. Yields (%) of Trimers Synthesized by the Reaction of pNpN with 4 ImpNs[a]

pA$^{3'}$pC + ImpA + ImpG + ImpC + ImpU		pA$^{2'}$pC + ImpA + ImpG + ImpC + ImpU	
pA$^{3'}$pC$^{3'}$pU	0.26	pA$^{2'}$pC$^{3'}$pU	0.21
pA$^{3'}$pC$^{2'}$pU	0.68	pA$^{2'}$pC$^{2'}$pU	<0.21
pA$^{3'}$pC$^{3'}$pC	0.25	pA$^{2'}$pC$^{3'}$pC	<0.25
pA$^{3'}$pC$^{2'}$pC	0.62	pA$^{2'}$pC$^{2'}$pC	<0.25
pA$^{3'}$pC$^{3'}$pG	0.19	pA$^{2'}$pC$^{3'}$pG	<0.21
pA$^{3'}$pC$^{2'}$pG	0.28	pA$^{2'}$pC$^{2'}$pG	<0.21
pA$^{3'}$pC$^{3'}$pA	<0.16	pA$^{2'}$pC$^{3'}$pA	<0.19
pA$^{3'}$pC$^{2'}$pA	0.23	pA$^{2'}$pC$^{2'}$pA	<0.19
pA$^{3'}$pU	0.42	pA$^{3'}$pU	0.62
pG$^{3'}$pU	0.34	pG$^{3'}$pU	0.43
		pG$^{3'}$pC	0.34
pA$^{3'}$pU + ImpA + ImpG + ImpC + ImpU		pA$^{3'}$pA + ImpA + ImpG + ImpC + ImpU	
pA$^{3'}$pU$^{3'}$pU	0.37	pA$^{3'}$pA$^{3'}$pU	2.4
pA$^{3'}$pU$^{2'}$pU	0.90	pA$^{3'}$pA$^{2'}$pU	0.98
pA$^{3'}$pU$^{3'}$pC	0.33	pA$^{3'}$pA$^{3'}$pC	1.8
pA$^{3'}$pU$^{2'}$pC	0.94	pA$^{3'}$pA$^{2'}$pC	0.93
pA$^{3'}$pU$^{3'}$pG	0.24	pA$^{3'}$pA$^{3'}$pA	1.2
pA$^{3'}$pU$^{2'}$pG	0.67	pA$^{3'}$pA$^{2'}$pA	0.71
pA$^{3'}$pU$^{3'}$pA	0.29	pA$^{3'}$pA$^{3'}$pG	1.3
pA$^{3'}$pU$^{2'}$pA	0.41	pA$^{3'}$pA$^{2'}$pG	0.60
pG$^{3'}$pU	0.27		
pA$^{2'}$pA + ImpA + ImpG + ImpC + ImpU		pG$^{3'}$pA + ImpA + ImpG + ImpC + ImpU	
pA$^{2'}$pA$^{3'}$pU	1.4	pG$^{3'}$pA$^{3'}$pU	4.0
pA$^{2'}$pA$^{2'}$pU	0.25	pG$^{3'}$pA$^{2'}$pU	0.58
pA$^{2'}$pA$^{3'}$pC	1.3	pG$^{3'}$pA$^{3'}$pC	3.8
pA$^{2'}$pA$^{2'}$pC	<0.20	pG$^{3'}$pA$^{2'}$pC	0.77
pA$^{2'}$pA$^{3'}$pA	0.68	pG$^{3'}$pA$^{3'}$pA	2.5
pA$^{2'}$pA$^{2'}$pA	<0.16	pG$^{3'}$pA$^{2'}$pA	0.52
pA$^{2'}$pA$^{3'}$pG	0.68	pG$^{3'}$pA$^{3'}$pG	2.5
pA$^{2'}$pA$^{2'}$pG	<0.17	pG$^{3'}$pA$^{2'}$pG	0.48
pU$^{3'}$pA + ImpA + ImpG + ImpC + ImpU		pA$^{3'}$pG + ImpA + ImpG + ImpC + ImpU	
pU$^{3'}$pA$^{3'}$pU	3.7	pA$^{3'}$pG$^{3'}$pU	3.6
pU$^{3'}$pA$^{2'}$pU	0.83	pA$^{3'}$pG$^{2'}$pU	0.4
pU$^{3'}$pA$^{3'}$pC	3.3	pA$^{3'}$pG$^{3'}$pC	3.3
pU$^{3'}$pA$^{2'}$pC	0.85	pA$^{3'}$pG$^{2'}$pC	<0.67
pU$^{3'}$pA$^{3'}$pG	2.1	pA$^{3'}$pG$^{3'}$pA	2.4
pU$^{3'}$pA$^{2'}$pG	0.57	pA$^{3'}$pG$^{2'}$pA	<0.53
pU$^{3'}$pA$^{3'}$pA	1.9	pA$^{3'}$pG$^{3'}$pG	2.2
pU$^{3'}$pA$^{2'}$pA	0.61	pA$^{3'}$pG$^{2'}$pG	0.42

[a] Yields were obtained from peak areas of HPLC chromatograms based on all of the compounds detected. The peak areas were divided by the length of the oligomers, 3 or 2, and were corrected with the molar extinction coefficients. It was assumed that the molar extinction coefficients of trimers and dimers identified were obtained by adding those of the corresponding monomers, and those of the monomers and oligomers unidentified were the average values of pN and pNpN, respectively.

of the dimers were purified by HPLC on a semipreparative reversed-phase column to give samples that were >98% pure.

Reaction of solutions containing 7.5 mM dimer and 1.9 mM each of the four phosphorimidazolides resulted in the formation of trimers (Table 2). The reactivities of the trimers with 5′-purine nucleotides were determined from the relative yields of the trimeric reaction products (Table 2). As shown with dimer formation, the reactivity of ImpA is similar to that of ImpG, and the reactivity of a 3′-A on the dimer is comparable to that of a 3′-G. The reactivities of the corresponding U and C derivatives were also comparable. The sequence selectivity and the product ratios observed for the trimers with 5′-purine nucleotides are PuPuPy > PuPuPu > PuPyPy > PuPyPu.

Dimers with a 3′-Pu reacted with activated nucleotides faster than those with a 3′-Py. For example, pA$^{3'}$pA reacted 2.4 times faster than pA$^{3'}$pU. In contrast, the activated monomers of pyrimidine nucleotides [Imp(Py)] added to dimers faster than the corresponding purine nucleotides [Imp(Pu)]. For example, Imp(Py) reacted 1.6 times faster than Imp(Pu) in the reaction of pA$^{3'}$pA with a mixture of four activated monomers. The 3′,5′-phosphodiester bond predominates over the 2′,5′-linkage when the 3′-nucleotide in dimers is a purine nucleotide, but the 2′,5′-linkage predominates over the 3′,5′-linkage when the 3′-nucleotide is a pyrimidine nucleotide. For example, pA$^{3'}$pA$^{3'}$pN:pA$^{3'}$pA$^{2'}$pN was 2.1:1.0 in the reaction of pA$^{3'}$pA with 4 ImpNs, and the pA$^{3'}$pU$^{3'}$pN:pA$^{3'}$pU$^{2'}$pN ratio was 1.0:2.4 in the reaction of pA$^{3'}$pU with the 4 ImpNs.

The most reactive dimers were those with two purine nucleotides linked by a 3′,5′-phosphodiester bond (pA$^{3'}$pG, pG$^{3'}$pA, and pA$^{3'}$pA) and the pyrimidine-purine dimer pU$^{3'}$pA. Lower reactivity was found for pA$^{2'}$pA, pA$^{3'}$pC, and pA$^{3'}$pU, and the lowest reactivity was found for pA$^{2'}$pC. These data establish that a 3′-terminal Pu linked by a 3′,5′-phosphodiester bond reacts much faster than a dimer with the corresponding 3′-pyrimidine nucleotide. It did not matter whether there is a 5′-purine nucleotide or 5′-pyrimidine nucleotide adjacent to the 3′-nucleotide in the dimer. The greater reactivity of pA$^{3'}$pA versus pA$^{2'}$pA and pA$^{3'}$pC versus pA$^{2'}$pC demonstrated that a 2′,5′ -link next to the 3′-nucleotide decreases the reactivity of the 3′-nucleotide 2−3-fold over the corresponding 3′,5′-linked 3′-nucleotide.

The following general rules for RNA elongation on montmorillonite were derived from these reactivity studies. (1) Purine nucleotides will be at the 5′-end of the RNA. (2) The reactivity for elongation depends mainly on the nucleotide and the regiospecificity of its phosphodiester bond at the 3′-end of the RNA. The extent of elongation decreases in the order: $-^{3'}$p(Pu) > $-^{2'}$p(Pu) > $-^{3'}$p(Py) > $-^{2'}$p(Py). (3) The low reactivity of $-^{2'}$p(Py) indicates it is essentially a chain-terminating group. (4) The $-^{3'}$Py group tends to be a chain-terminating group because it elongates mainly to $-^{3'}$Py$^{2'}$Py. As noted in (3) above, the $-^{2'}$Py is a chain-terminating group. (5) In the elongation of RNA oligomers, Imp(Py) reacts faster than Imp(Pu). (6) When activated nucleotides react with a 3′-terminal-$^{3'}$p(Pu), -$^{2'}$p(Pu), or -$^{2'}$p(Py), the adduct with a 3′,5′-phosphodiester bond is the major reaction product. (7) For the addition of an activated nucleotide to an oligomer with a 3′-terminal-$^{3'}$p(Py), the adduct with a 2′,5′-linkage predominates.

Two opposing factors are used by the montmorillonite clay to control sequence selectivity. The first factor is the greater reactivity of the 3′-purine nucleotide over the corresponding 3′-pyrimidine nucleotide for the addition of the next nucleotide to the chain. The second factor is the greater reactivity of the activated pyrimidine nucleotide over the corresponding purine derivative in the reaction with the 3′-terminal nucleotide on an RNA. In the present study, the formation of RNAs that have a high proportion purine nucleotides is favored because the difference in reactivity of 3′-purine nucleotides over pyrimidine nucleotides is greater than the higher reactivity of activated pyrimidine over activated purine. It is possible that RNAs that contain a different proportion of purine and pyrimidine nucleotides and 3′, 5′-phosphodiester bonds than the RNAs prepared in our present studies will be obtained. This could occur when: (1) The reaction conditions differ from those used in this study. (2) Groups other than imidazole activate the 5′ -nucleotides.

(9) Sawai, H.; Kuroda, K.; Hojo, T. *Bull. Chem. Soc. Jpn.* **1989**, *62*, 2018−2023.

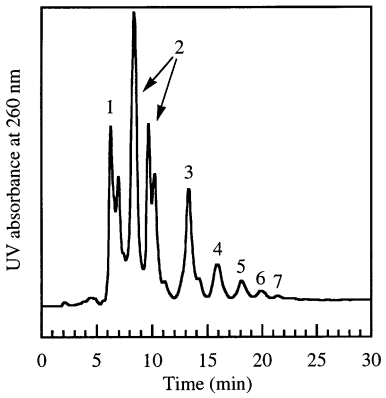

Figure 2. Anion-exchange HPLC of the reaction of ImpA with ImpC. The numbers indicate the number of monomers in each oligomer.

(3) The 3′-ends of longer RNAs have an altered orientation in the catalytic site than do the dimers used here. (4) Nucleotides with purine or pyrimidine bases with reactive functional groups are used that change the course of the reaction.

These rules will predict major but not minor reaction products. For example, they appear to preclude the incorporation of pyrimidine nucleotides in the growing RNA chain. This would be the case if no chains containing a $Py^{3'}Pu$ at the 3′-end of the RNA were formed. The $-Py^{3'}Pu$ will form slowly, but once formed the 3′-Pu will elongate efficiently. This will result in the incorporation of pyrimidine nucleotides in some of the RNAs.

Oligomers Formed from the ImpA−ImpC Reaction. The rules derived from the above studies were tested in the binary reaction of ImpA with ImpC. The precise analysis of the products from the formation of 4-mers from pNpNpN and four activated monomers was not undertaken because of the absence of authentic samples, so the rules were tested by the analysis of the 2−5-mers formed from ImpA and ImpC. The 2−5-mers, labeled 2−5 in Figure 2, were separated on an anion-exchange HPLC. These compounds were collected and analyzed by reversed-phase HPLC (Figure 3a). They were then treated with alkaline phosphatase (APH) to cleave the 5′-terminal phosphate groupings and again analyzed by reversed-phase HPLC (Figure 3b). The peaks that shifted to longer retention times in Figure 3b are linear RNAs that had terminal phosphate groups. The compounds with the longer retention times were collected, and their structures were determined (Table 3).

The 39% yield of $pA^{3'}pC$ in the dimer fraction is consistent with its rapid formation and slow elongation, while the low yield of $pA^{3'}pA$ is consistent with its rapid formation and subsequent rapid elongation. The relative yields of the dimers formed were in good agreement with those reported previously.[7]

Analysis of the trimer fraction revealed that $pA^{3'}pA^{2'}pC$ was formed in 48% yield as would be expected if the $pA^{3'}pA$ elongated rapidly and the $-^{2'}pC$ group slowed its subsequent rate of elongation. $pA^{3'}pC^{2'}pC$ was observed because $pA^{3'}pC$ is formed in large amounts and its elongation product, $pA^{3'}pC^{2'}pC$, elongates more slowly due to its a 3′-terminal-$^{2'}pC$ grouping. The yield of $pA^{3'}pA^{2'}pA$ is less than $pA^{3'}pC^{2'}pC$ because the amount of $pA^{3'}pA$ formed is less than $pA^{3'}pC$ and $pA^{3'}pA^{2'}pA$ elongates more rapidly than $pA^{3'}pC^{2'}$-pC. Similar arguments explain the relative proportions of trimers formed.

The tetramers formed had $pA^{3'}pA^{3'}pA^{2'}pC$ and $pA^{3'}pA^{2'}pA^{3'}pC$ as major products, a result consistent with the rapid elongation of $pA^{3'}pApA$ and the slower elongation of products with 3′-terminal pyrimidine nucleotides. The oligomers $pA^{3'}pA^{2'}pC^{3'}pC$ and $pA^{3'}pC^{2'}pC^{3'}pC$ were detected because $pA^{3'}pA^{2'}pC$ and $pA^{3'}pC^{2'}pC$ are abundant in the trimer fraction and these compounds have 3′-terminal pyrimidines which elongate slowly. The pentamers observed all have $-^{2'}pC$ groups indicative of their slow elongation rates. The trimer $pA^{3'}pC^{2'}pA$ elongated to the pentamer $pA^{3'}pC^{2'}pA^{3'}pA^{2'}pC$, showing that if the oligomerization is able to proceed past the $-^{3'}pC$ reaction barrier further elongation will proceed efficiently. This observation demonstrates that the $-^{3'}pC$ group is not an absolute chain-terminating group but rather a group that elongates very slowly.

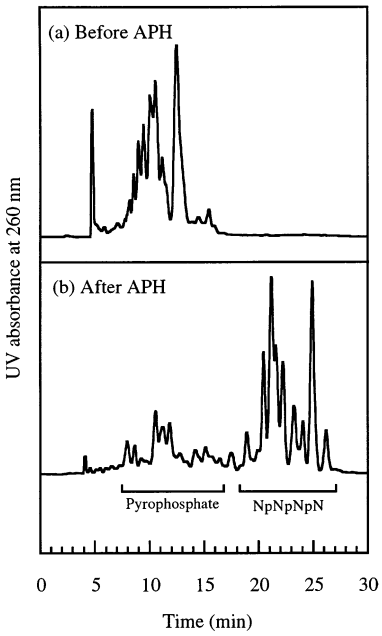

Figure 3. Reversed-phase HPLC of the 4-mers synthesized in the reaction of ImpA with ImpC. (a) 4-mers collected from the anion-exchange HPLC. (b) 4-mers hydrolyzed with alkaline phosphatase (APH).

$A^{3'}pC^{3'}p$ cyclic was the most abundant product with the yield of 14% based on all monomers and oligomers. It is formed by via $ImpA^{3'}pC$ which then cyclizes by the reaction of the 3′-hydroxyl group of cytidine with the 5′-phosphorimidazolide of adenosine. The yields of other cyclic dimer products are less than 2.2%. The high yield of $pA^{2'}pC$ and the low yields of other cyclic dimers are consistent with the rate of cyclization of $ImpA^{2'}pC$ being much slower than that of $ImpA^{3'}pC$ on montmorillonite.

The rules derived from the reactions of pN + ImpN and pNpN + 4 ImpNs appear to give good explanations to the sequence selectivity observed in the reaction of ImpA with ImpC. Most of the oligomers detected in the ImpA reaction with ImpC have an adenosine nucleotide at the 5′-end because the adenosine nucleotide is more reactive than the cytidine nucleotide. Most of the oligomers formed have a cytidine at the 3′- end because ImpC is more reactive than ImpA and oligomers with a 3′-terminal pC do not elongate as efficiently as those with a 3′-terminal pA. Of the two opposing factors that control the sequence selectivity, factor 1 is more important than factor 2 under the present experimental conditions consistent with the favored formation of adenosine-rich oligomers.

Table 3. Yields of Oligomers Synthesized by the Reaction of ImpA with ImpC (%)

	pNpN type[a]	all[b]
$A^{3'}pC^{3'}p$ cyclic		14
other cyclics		<2.2
	Dimers	
$pA^{3'}pC$	39	2.9
$pA^{2'}pC$	31	2.3
$pA^{2'}pA$	10	0.75
$pC^{2'}pC$	7.2	0.55
$pA^{3'}pA$	5.3	0.39
$pC^{3'}pC$	3.5	0.26
$pC^{2'}pA$	1.6	0.12
$pC^{3'}pA$	0.63	0.046
	Trimers	
$pA^{3'}pA^{2'}pC$	48	1.5
$pA^{3'}pC^{2'}pC$	11	0.37
$pA^{3'}pA^{2'}pA$	9.9	0.30
$pA^{3'}pA^{3'}pC$	8.0	0.25
$pA^{2'}pA^{3'}pC$	3.4	0.11
$pA^{2'}pA^{2'}pC$ or $pA^{2'}pC^{2'}pA$	2.5	0.08
$pA^{3'}pC^{3'}pC$	2.4	0.077
$pA^{3'}pA^{3'}pA$	2.0	0.063
$pC^{3'}pA^{2'}pC$	1.4	0.043
$pA^{3'}pC^{2'}pA$	1.1	0.033
other major peaks	<1.0	<0.033
	Tetramers	
$pA^{3'}pA^{3'}pA^{2'}pC$	21	0.18
$pA^{3'}pA^{2'}pA^{3'}pC$	17	0.15
$pA^{3'}pA^{2'}pC^{3'}pC$	8.6	0.075
$pA^{3'}pC^{2'}pC^{3'}pC$	5.7	0.05
$pA^{3'}pA^{3'}pC^{2'}pC$	5.5	0.048
other major peaks	<4.0	<0.035
	Pentamers	
$pA^{3'}pA^{2'}pA^{3'}pA^{2'}pC$	13	0.052
$pA^{3'}pA^{2'}pA^{3'}pC^{2'}pC$	6.6	0.026
$pA^{3'}pA^{3'}pA^{3'}pA^{2'}pC$	4.8	0.019
$pA^{3'}pC^{2'}pA^{3'}pA^{2'}pC$	4.3	0.017
other major peaks	<3.8	<0.015

[a] Based on pNpN type oligomers. Values not corrected for the molar extinction coefficients. [b] Based on all of the compounds detected. The percentages were obtained by dividing the percentage obtained from the HPLC absorbances by the oligonucleotide length. The values are not corrected for the molar extinction coefficient.

Conclusions

The need for catalysis in forming RNA oligomers has been demonstrated dramatically by the extensive studies of the reactions of ImpN derivatives that were not catalyzed by minerals.[10] It was observed that in reactions where concentrations of individual and mixtures of ImpNs of 1 M and greater were allowed to react, comparable amounts of all possible dimers and "...only traces of longer oligomers" are formed. The overall regioselectivity ratio for 3′,5′- to 2′,5′-phosphodiester bond formation in the uncatalyzed formation of the dimers was 0.4:1, while a 2:1 ratio was observed in the montmorillonite-catalyzed formation of dimers (Table 1). Clearly, the montmorillonite catalyst is changing the course of the reaction from that which is observed in the absence of montmorillonite.

The sequence selectivity in the reactions of ImpNs on montmorillonite is regulated by two opposing factors. The rate of elongation of an RNA oligomer is determined by the 3′-nucleotide and the regiochemistry of its phosphodiester bond. The reaction rates are faster if there is a 3′-purine nucleotide and if this nucleotide is linked by a 3′,5′-phosphodiester bond.

Table 4. Montmorillonite-Catalyzed Synthesis versus Theoretical Random Synthesis in the ImpA−ImpC Reaction

	catalyzed synthesis		random synthesis	
	number of isomers observed	yields of oligomers (%)	number of isomers predicted	yield of each isomer (%)
dimer	8	0.6−39	8	13
trimer	10	1.1−48	32	3.1
tetramer	5	5.5−34	128	0.78
pentamer	4	4.3−13	512	0.20

In contrast, an activated pyrimidine nucleotide adds more rapidly than the corresponding purine nucleotide to the 3′-end of the RNA oligomer.

The results of this study demonstrate that there is a strong bias away from the formation of all possible isomers in the RNAs formed by montmorillonite catalysis. This bias is clearly shown in the products formed in the reaction of ImpA with ImpC (Table 4). While the number of dimers formed equals the number of isomers predicted by the random synthesis of these compounds, the range in yields varies dramatically from the 13% expected in a random process. The number of isomers identified in the tetramer and pentamer fractions is 26−130 times smaller than the number expected by random synthesis, and the yields of individual oligomers range from 7 to 65 times greater than that predicted by a random synthetic processes. These findings support the view that longer oligomers form because the catalyst limits the number of reaction pathways and there are sufficient monomers to make the longer oligomers.

Experimental Section

Materials. A, G, C, U, pA, pG, pC, pU, $A^{3'}p$, $G^{3'}p$, $C^{3'}p$, $N^{2'}pN$, and $N^{3'}pN$ were obtained from Sigma. $C^{2'}pU$ was not available from a commercial source. Alkaline phosphatase and phosphodiesterase I were obtained from Worthington Biochemical, ribonuclease T_2 was from Sigma, and T4 polynucleotide kinase was from USB.

Reaction of pN with ImpN. An aqueous solution containing 7.5 mM pN, 7.5 mM ImpN, 200 mM NaCl, 75 mM $MgCl_2$, and 100 mM HEPES (pH 8) was kept in the presence of Na^+-montmorillonite for 3 days at room temperature. ImpA, ImpG, ImpC, and ImpU were prepared following the literature procedure.[11] The Na^+-montmorillonite was prepared from Volclay by titration method.[12] Because the ImpN hydrolyzes to pN with half-lives of about 9 h in the presence of clay,[13] most of the ImpN would have reacted or hydrolyzed after 3 days. Purities of ImpN synthesized were checked by anion-exchange HPLC prior to starting reactions to make sure that they were higher than 95% and the NppNs were lower than 1%. pN and NppN may be formed during storage. Because the NppN is a good initiator of elongation,[14] its contamination should be avoided. At the end of the reaction time, the mixture was centrifuged, and the supernatants were withdrawn. The clay was washed with 0.1 M ammonium acetate, except in the reaction of pG with ImpG, until most of the oligomers were extracted. The mixture solution of 0.1 M NaCl and 30% acetonitrile was used for extraction in the reaction of pG with ImpG. The supernatant and the washes were combined and heated at pH 4 and 37 °C for 2 h to hydrolyze the 5′-imidazolide groups.

Reaction of pNpN with 4 ImpNs. An aqueous solution containing 7.5 mM pNpN, 1.9 mM ImpA, 1.9 mM ImpG, 1.9 mM ImpC, 1.9 mM ImpU, 200 mM NaCl, 75 mM $MgCl_2$, and 100 mM HEPES (pH 8)

(10) Kanavarioti, A. *J. Mol. Evol.* **1998**, *46*, 622−632.
(11) Joyce, G. F.; Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1984**, *176*, 279−306.
(12) Banin, A.; Lawless, J. G.; Mazzurco, J.; Church, F. M.; Margulies, L.; Orenberg, J. B. *Origins Life Evol. Biosphere* **1985**, *15*, 89−101.
(13) Kawamura, K.; Ferris, J. P. *Origins Life Evol. Biosphere* **1999**, *29*, 563−591.
(14) Ferris, J. P.; Ertem, G. *Origins Life Evol. Biosphere* **1993**, *23*, 229−241.

was kept in the presence of Na^+-montmorillonite for 3 days at room temperature. At the end of the reaction time, the mixture was centrifuged, and the supernatants were withdrawn. The clay was washed with a mixture solution of 0.1 M NaCl and 30% acetonitrile. The wash was repeated several times until most of the oligomers were extracted as determined by anion-exchange HPLC. The supernatant and the washes were combined and heated at pH 4 and 37 °C for 2 h to hydrolyze the 5′-imidazolide groups.

Reaction of ImpA with ImpC. An aqueous solution containing 7.5 mM ImpA, 7.5 mM ImpC, 200 mM NaCl, 75 mM $MgCl_2$, and 100 mM HEPES (pH 8) was kept in the presence of Na^+-montmorillonite for 3 days at room temperature. At the end of the reaction time, the mixture was centrifuged, and the supernatants were withdrawn. The clay was washed with a mixture solution of 0.1 M NaCl and 30% acetonitrile. The wash was repeated several times until most of the oligomers were extracted as determined by anion-exchange HPLC. The supernatant and the washes were combined and heated at pH 4 and 37 °C for 2 h to hydrolyze the 5′-imidazolide groups.

Purification of Oligomers. Oligomers were first separated by anion-exchange HPLC (Biosphere GMB 1000 Q column, Melcor Tech, Inc.) with a pH 8, 2 mM Tris buffer (buffer A) and pH 8, 2 mM Tris and 400 mM $NaClO_4$ (buffer B). The oligomers were detected with a UV detector (L-4000, Hitachi) at 260 nm. A chromatogram of the reaction of ImpA with ImpC is shown in Figure 2. The fraction numbers correspond to the length of the oligomers. Fraction 1 is composed of pA and pC. Fractions 2, 3, 4, and 5 are composed of dimers, trimers, tetramers, and pentamers, respectively. Fractions 2−5 were collected, dialyzed in a membrane (Spectra/Pro 6, MWCO 1000, Spectrum Lab, Inc.) at 2 °C for 2 h, and freeze-dried. Each sample was then analyzed by reversed-phase HPLC (Alltima C18 column, Alltech) at 260 nm with a pH 2.7, 0.2% trifluoroacetic acid (buffer A) and a pH 2.7, 0.2% trifluoroacetic acid and 30% acetonitrile (buffer B).[10] The samples were hydrolyzed by alkaline phosphatase (APH) to remove a terminal phosphate. The peaks that shifted after APH hydrolysis were collected individually. The oligomers were repurified by reversed-phase HPLC at 260 nm with a pH 4, 10 mM ammonium acetate (buffer A) and a pH 4, 10 mM ammonium acetate and 30% acetonitrile (buffer B). Their purity was higher than 90%.

It was assumed that oligomers have the same length and similar retention times on the anion-exchange HPLC (Figure 2). It is true for a reaction in which only one type of activated nucleotide is used. However, it was not certain that it is the case for the oligomers formed in a binary reaction. Retention times of some trimers, for example, may be similar to those of some dimers, and they may overlap on its chromatogram. $pC^{2'}pC^{2'}pC$ and $pC^{2'}pC^{2'}pC^{2'}pC$, which are expected to have shortest retention times in trimers and 4-mers, respectively, were prepared and were co-injected with the sample prepared from a reaction of ImpA with ImpC. The $pC^{2'}pC^{2'}pC$ was correctly present in the trimer peak, and the $pC^{2'}pC^{2'}pC^{2'}pC$ was present in the tetramer peak. Co-injection of a binary reaction sample with a sample prepared from an ImpA reaction or with a sample prepared from an ImpC reaction also supports the premise that each peak on the anion-exchange HPLC chromatogram (Figure 2) represents a group of oligomers with the same chain length.

Identification of Oligomers. Enzymatic hydrolyses were used to cleave the oligomers, and then the cleavage products formed were hydrolyzed by APH. The products were identified by comparison with standards by reversed-phase HPLC. Phosphodiesterase I was used to cleave the 2′,5′- and the 3′,5′-phosphodiesterase bonds to produce pNs and nucleosides. Quantitative data calibrated with the molar extinction coefficients indicated the ratio of number of different nucleotides in the oligomers. The nucleoside formed by hydrolysis is the 5′-nucleotide in the oligomer. Ribonuclease T_2 hydrolyzes only the 3′,5′-bond and was used to distinguish the 3′,5′-phosphodiesterase bond from the 2′,5′-bond. Although most of the oligomers obtained could be identified by the enzymatic hydrolyses described above, some tetramers and pentamers synthesized by the reaction of ImpA with ImpC could not. For example, $A^{3'}pA^{2'}pA^{3'}pC$ and $A^{2'}pA^{3'}pA^{3'}pC$ cannot be distinguished by the phosphodiesterase I hydrolysis and the ribonuclease T_2 hydrolysis. The $A^{3'}pA^{2'}pA^{3'}pC$ and the $A^{2'}pA^{3'}pA^{3'}pC$ are hydrolyzed with the phosphodiesterase I to form A, 2 pA, and pC. They are hydrolyzed with the ribonuclease T_2 to form $A^{3'}p$, $A^{2'}pA^{3'}p$, and C. To distinguish these isomers, T4 polynucleotide kinase and ATP were used to add a phosphate group to the 5′-end position of $A^{3'}pA^{2'}pA^{3'}pC$ to give $pA^{3'}pA^{2'}pA^{3'}pC$. $A^{2'}pA^{3'}pA^{3'}pC$ cannot be phosphorylated with the T4 polynucleotide kinase because its 5′-nucleotide is linked by a 2′,5′-bond.

Some samples were hydrolyzed with APH after phosphodiesterase I or ribonuclease T_2 to confirm the assigned structure. In the pH 2.7 reversed-phase HPLC, A, G, C, and U are separated. Therefore, the phosphodiesterase I hydrolysis and subsequent APH hydrolysis indicate the composition of oligomers and the ratios of A to G to C to U. The co-injection of standard samples of A, G, C, U, pA, pG, pC, pU, $A^{3'}p$, $G^{3'}p$, $C^{3'}p$, $N^{2'}pN$, and $N^{3'}pN$ was used to confirm the structures of the oligomer hydrolysis products. No standard sample of $C^{2'}pU$ was available.

Yields. Yields of dimers synthesized from the reaction of pN with ImpN were obtained from peak areas on reversed-phase HPLC chromatograms at 260 nm. The data were corrected with the molar extinction coefficients and hypochromicity. The molar extinction coefficients of A, pA, G, pG, C, pC, U, and pU are 14 300, 14 500, 11 750, 11 600, 6400, 6300, 9950, and 9900 at 260 nm under acidic conditions, respectively.[15] Because the hypochromicity of pNpN is not available, the hypochromicity of NpN was used.[8] For the unidentified products, the molar extinction coefficients were assumed to be the average of pNpN. The relative uncorrected yields were not significantly different from those that were corrected, indicating that it did not matter here whether the data were corrected or not.

The yields of trimers synthesized from the reaction of pNpN with ImpN were obtained from peak areas on HPLC chromatograms and were corrected with the molar extinction coefficients. It was assumed that the molar extinction coefficients of trimers and dimers identified were obtained by adding those of corresponding monomers. No correction was made for hypochromicity because these data are not available. Monomers and oligomers unidentified were assumed to be the average of pN and pNpN, respectively. The relative uncorrected yields for the dimers were not significantly different from those corrected.

The yields of oligomers synthesized from a reaction of ImpA with ImpC were obtained by dividing the percentage obtained from the HPLC absorbances by the oligonucleotide length. The values were not corrected for the molar extinction coefficients. All HPLC chromatograms of 4-mers and 5-mers were measured at pH 8 by anion-exchange HPLC at 271 nm and at pH 2.7 by reversed-phase HPLC at 280 nm. The increase in the areas of the HPLC peaks showed the oligomers contained cytidine because it has an absorption maximum at this wavelength.

Effects of Ammonia. The pNpNs used in the reaction of pNpN with 4 ImpNs were prepared as ammonium salts. The possible major effect of ammonia on the reaction is the formation of NH_2pN by the reaction of ammonium ion with the ImpNs. However, NH_2pN was not detected in the reaction mixture. In addition, the presence of 100 mM ammonium acetate in the reaction of $pA^{3'}pA$ with ImpA did not change the yields of $pA^{3'}pA^{3'}pA$ and $pA^{3'}pA^{2'}pA$.

Synthesis of $pA^{3'}pA$, $pA^{3'}pC$, $pA^{2'}pC$, and $pA^{3'}pU$. For the synthesis of $pA^{3'}pA$, an aqueous solution containing 100 mM adenosine 5′-monophosphoramidate (NH_2pA), 100 mM ImpA, 200 mM NaCl, 75 mM $MgCl_2$, and 100 mM HEPES (pH 8) was kept in the presence of Na^+-montmorillonite for 3 days at room temperature. The supernatant

(15) Dawson, R. M. C.; Elliot, D. C.; Elliot, W. H.; Jones, K. M. *Data for Biochemical Research*, 2nd ed.; Oxford University Press: New York, 1969.

was withdrawn after centrifugation. The clay was not washed because the wash contains a large amount of cyclic $A^{3'}pA^{3'}p$, which is difficult to separate from $pA^{3'}pA$. Heating at pH 2 and 37 °C for 1 day hydrolyzed the phosphoramidate group. The purity of the dimers was confirmed by pH 2.7 reversed-phase HPLC. The same procedure was used for the syntheses of $pA^{3'}pC$, $pA^{2'}pC$, and $pA^{3'}pU$, except that ImpC and ImpU were used instead of ImpA. The yields, uncorrected, of $pA^{3'}pA$, $pA^{3'}pC$, $pA^{2'}pC$, and $pA^{3'}pU$ were 15%, 31%, 14%, and 30%, respectively. The dimers were purified by reversed-phase HPLC (Alltima C18 preparative column, Alltech) at 260 nm with a pH 2.7, 0.2% trifluoroacetic acid (buffer A) and a pH 2.7, 0.2% trifluoroacetic acid and 30% acetonitrile (buffer B). The samples were desalted by reversed-phase HPLC with a pH 4, 10 mM ammonium acetate (buffer A) and a pH 4, 10 mM ammonium acetate and 30% acetonitrile (buffer B). The final product was an ammonium salt, and its purity was higher than 98%. Because the amount of dimers synthesized was too small to weigh accurately (several milligrams) and there were no available data of the molar extinction coefficients of the pNpN dimers, they were cleaved with phosphodiesterase I, and the amounts of products were determined by comparison with standard pA by HPLC analysis.

Synthesis of $pA^{3'}pG$ and $pG^{3'}pA$. Because the solubility of pG and ImpG is low, $pA^{3'}pG$ and $pG^{3'}pA$ were synthesized by a reaction of pA with ImpG and a reaction of pG with ImpA. For the synthesis of $pA^{3'}pG$, an aqueous solution containing 7.5 mM pA, 7.5 mM ImpG, 200 mM NaCl, 75 mM $MgCl_2$, and 100 mM HEPES (pH 8) was kept in the presence of Na^+-montmorillonite for 3 days at room temperature. The supernatant was withdrawn after centrifugation and was hydrolyzed at pH 4 and 37 °C for 3 h to remove the imidazole group. The uncorrected yields of $pA^{3'}pG$ and $pG^{3'}pA$ were about 10%. The dimers were purified by reversed-phase HPLC. The final products were obtained as their ammonium salts, and their purity was higher than 98%.

Synthesis of $pA^{2'}pA$. Because $pA^{2'}pA$ is not efficiently synthesized by the clay-catalyzed reaction, it was synthesized by catalysis of the reaction of ImpA with uranyl acetate.[9] The ammonium salt of $pA^{2'}pA$ was purified by reversed-phase HPLC, and its final purity exceeded 98%.

Synthesis of $pU^{3'}pA$. Because a $pU^{3'}pA$ is not efficiently synthesized by the clay-catalyzed reaction, the $pU^{3'}pA$ was synthesized by phosphorylation of a $U^{3'}pA$ from Sigma. A T4 polynucleotide kinase was used for phosphorylation. The $pU^{3'}pA$ was purified by reversed-phase HPLC. The final product was an ammonium salt, and its purity was higher than 98%.

Synthesis of $pC^{2'}pC^{2'}pC$ and $pC^{2'}pC^{2'}pC^{2'}pC$. $pC^{2'}pC^{2'}pC$ and $pC^{2'}pC^{2'}pC^{2'}pC$ were prepared by the clay-catalyzed reaction of ImpC.[16] These compounds were purified by anion-exchange HPLC and reversed-phase HPLC and were identified by APH hydrolysis and ribonuclease T_2 hydrolysis.

Acknowledgment. We thank Mr. Eric Hebrard for assistance with this research. Montmorillonite (Volclay) was a gift from the American Colloid Co. Research support from NSF grant CHE-0076281 and NASA grant NAG5-9607 to the NY Center for Studies on the Origins of Life is acknowledged.

JA034328E

(16) Ertem, G.; Ferris, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 7197−7201.