

# Nonenzymatic, Template-Directed Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3′–5′ Phosphodiester Bonds

Rajat Rohatgi, David P. Bartel,<sup>†</sup> and Jack W. Szostak\*

Contribution from the Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114

Received November 3, 1995<sup>⊗</sup>

**Abstract:** We have found that nonenzymatic, template-directed ligation reactions of oligoribonucleotides display high selectivity for the formation of 3′–5′ rather than 2′–5′ phosphodiester bonds. Formation of the 3′–5′-linked product is favored regardless of the metal ion catalyst or the leaving group, and for several different ligation junction sequences. The degree of selectivity depends on the leaving group: the ratio of 3′–5′- to 2′–5′-linked products was 10–15:1 when the 5′-phosphate was activated as the imidazolide, and 60–80:1 when the 5′-phosphate was activated by formation of a 5′-triphosphate. Comparison of oligonucleotide ligation reactions with previously characterized single nucleotide primer extension reactions suggests that the strong preference for 3′–5′-linkages in oligonucleotide ligation is primarily due to the occurrence of ligation within the context of an extended Watson–Crick duplex. The ability of RNA to correctly self-assemble by template-directed ligation is an intrinsic consequence of its chemical structure and need not be imposed by an external catalyst (i.e., an enzyme polymerase); RNA therefore provides a reasonable structural basis for a self-replicating system in a prebiological world.

Nonenzymatic, template-directed copying reactions are thought to have played an important role in the early chemical evolution that culminated in the origin of life.<sup>1</sup> Template-directed condensation reactions of ribonucleotides and oligoribonucleotides have provided a useful experimental system for examining nonenzymatic, template-directed synthesis of double-stranded RNA. One problem with such reactions is that a template containing only 3′–5′ phosphodiester bonds can direct the synthesis of complementary products that contain a large proportion of 2′–5′ linkages.<sup>1–3</sup> These 2′–5′-linked products are not accurate copies of the template since they have failed to maintain its 3′–5′-linked backbone structure.

From a chemical perspective, the problem of regiocontrol is particularly severe for nonenzymatic RNA reactions because the 2′-hydroxyl of a ribonucleotide is 6–9 times more nucleophilic toward activated phosphate esters than is the 3′-hydroxyl.<sup>4</sup> In the case of template-directed reactions, the situation is further complicated because the regioselectivity is highly sensitive not just to the relative nucleophilicities of the two hydroxyls but also to the precise stereochemical orientation of the template-bound reactants. In the template-directed condensation reactions of activated mononucleotides studied extensively by Orgel and co-workers, the phosphodiester bonds in the products varied from predominantly 2′–5′-linked to predominantly 3′–5′-linked, depending upon the choice of ligation junction sequence, metal ion catalyst, and activating group. The poly(U) directed polymerization of adenosine 5′-phosphorimidazolide (ImpA) yields products that have a 3′–5′:2′–5′ ratio of 1:18, whereas the poly(C)/ImpG reaction gives a ratio of 1.5:1.<sup>5</sup> The presence

of different metal ions can have a large effect on these ratios. Although the addition of Pb<sup>2+</sup> decreases the 3′–5′:2′–5′ ratio for the poly(C)/ImpG reaction to 1:9, it has an opposite effect on the poly(U)/ImpA reaction, increasing the 3′–5′:2′–5′ ratio to 1:1.<sup>6–8</sup> In contrast, the addition of Zn<sup>2+</sup> to the poly(C)/ImpG reaction leads to products that have mostly 3′–5′ linkages.<sup>9</sup> Finally, if the activating group on the nucleoside 5′-monophosphate is changed from imidazole to 2-methylimidazole, the poly(C) directed polymerization of guanosine 5′-phosphoro(2-methyl)imidazolide (2-MeImpG) yields products that are >90% 3′–5′-linked.<sup>10</sup>

We have previously described the detection of a nonenzymatic ligation reaction between two oligoribonucleotides aligned on a template in which a terminal 3′- or 2′-hydroxyl of one oligonucleotide attacks the 5′-triphosphate of the other oligonucleotide, forming a new phosphodiester bond with concomitant release of pyrophosphate (Figure 1).<sup>11</sup> Analysis of the kinetics and mechanism of this reaction is reported in the preceding paper in this issue.<sup>12</sup> One of the most interesting properties of this ligation reaction is that it has a high preference for the formation of 3′–5′ rather than 2′–5′ phosphodiester bonds.<sup>11</sup> Here we report experiments designed to determine the contribution of the sequence of the ligation junction, the identity of the metal ion catalyst, and the identity of the leaving group to the large bias toward 3′–5′-linked products in template-directed oligonucleotide ligation.

## Experimental Section

**Reagents.** Unlabeled nucleoside triphosphates and Sephadex G-25 gel filtration resin were purchased from Pharmacia, ribonucleoside

(5) Ninio, J.; Orgel, L. E. *J. Mol. Evol.* **1978**, *12*, 91–99.

(6) Sleeper, H. L.; Lohrmann, R.; Orgel, L. E. *J. Mol. Evol.* **1979**, *13*, 203–214.

(7) Lohrmann, R.; Orgel, L. E. *J. Mol. Biol.* **1980**, *142*, 555–567.

(8) van Roode, J. H. G.; Orgel, L. E. *J. Mol. Biol.* **1980**, *144*, 579–585.

(9) Bridson, P. K.; Orgel, L. E. *J. Mol. Biol.* **1980**, *144*, 567–577.

(10) Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201–217.

(11) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411–1418.

(12) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 0000.

<sup>†</sup> Present address: Whitehead Institute for Biomedical Research, Cambridge, MA 02142.

<sup>⊗</sup> Abstract published in *Advance ACS Abstracts*, March 1, 1996.

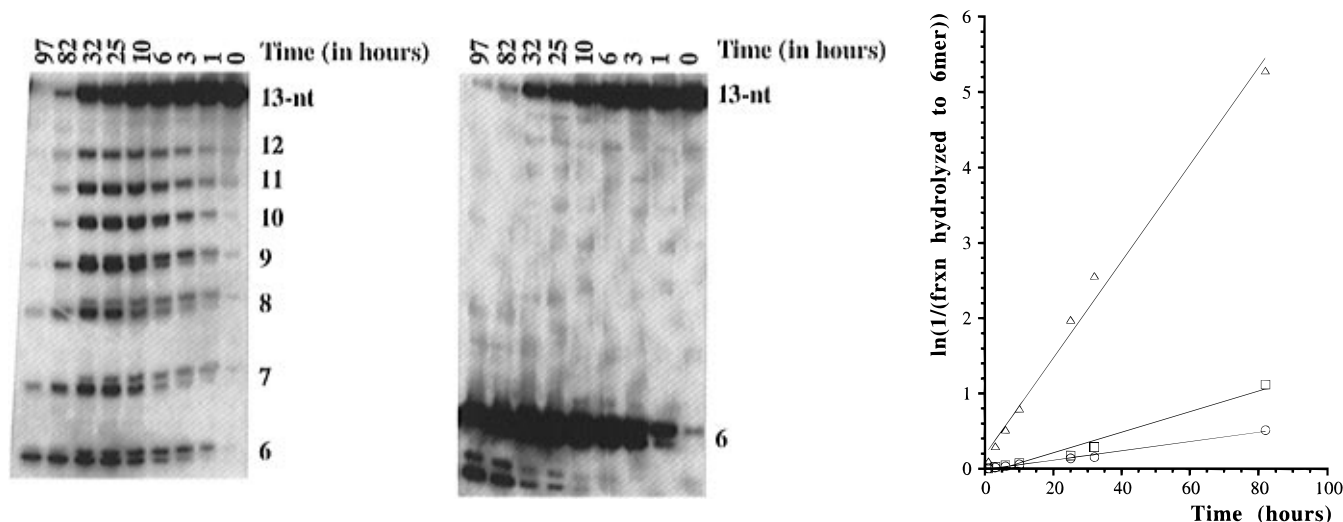
(1) Joyce, G. F.; Orgel, L. E. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1993; pp 1–25.

(2) Joyce, G. F. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 41–51.

(3) Orgel, L. E. *J. Theor. Biol.* **1986**, *123*, 127–149.

(4) Lohrmann, R. *Tetrahedron* **1978**, *34*, 853–855.





**Figure 2.** Differential hydrolysis rates of isolated 2'-5' and 3'-5' linkages within an RNA double helix. (A, left) The uniform hydrolysis of end-labeled 5'\*pCCAGUC<sup>2'</sup>p<sup>5'</sup>GGUUCUC in the absence of a complementary template in standard conditions at pH 9.0. (B, middle) Preferential hydrolysis at the 2'-5' phosphodiester bond in the presence of a template to give the 6-nt product 5'\*pCCAGUCp. The bands corresponding to the hydrolysis products on the gels are doublets because hydrolysis proceeds via a 2'-3' cyclic phosphate intermediate, which slowly converts to the faster-migrating 3'(2')-phosphate. (C, right) Hydrolysis of the bond between C6 and G7 within 5'\*pCCAGUCpGGUUCUC in the presence and absence of template. The 2'-5'-linked RNA was incubated in ligation conditions at pH 9.0 in the presence (triangles) and absence (squares) of template. Hydrolysis of 3'-5'-linked RNA was also assayed in the presence and absence (circles) of template (hydrolysis of the 3'-5'-linked RNA in the presence of template is not shown because it was too slow to measure accurately). The fraction hydrolyzed to a 6-mer was calculated from the total radioactivity in each lane (equivalent to the amount of 5'\*pCCAGUCpGGUUCUC at time = 0) and the radioactivity in the 6-mer (amount hydrolyzed to 5'\*pCCAGUCp at time = *t*). The slope of each plot is equal to  $k_{\text{hydr}}$ , the pseudo-first-order rate constant for hydrolysis of the particular bond (Table 1).

that would be obtained if P<sup>3'-5'</sup> and P<sup>2'-5'</sup> did not hydrolyze over the course of the ligation reaction. Since the rate of hydrolysis of a 3'-5' linkage in an RNA duplex is negligible,  $[P^{3'-5'}]_{\text{obs}} = [P^{3'-5'}]_{\text{corr}}$ . The following differential equation can be used to convert  $[P^{2'-5'}]_{\text{obs}}$  into  $[P^{2'-5'}]_{\text{corr}}$  (*t* = time):

$$\partial[P^{2'-5'}]_{\text{obs}}/\partial t = k_{\text{ign}2}[\text{ligation complex}] - k_{\text{hydr}2}[P^{2'-5'}]_{\text{obs}} \quad (1)$$

Because depletion of the ligation complex is negligible, solving for  $[P^{2'-5'}]_{\text{obs}}$  as a function of time yields

$$[P^{2'-5'}]_{\text{obs}}(t) = k_{\text{ign}2}[\text{ligation complex}]\{1 - \exp(-k_{\text{hydr}2}t)\}/(k_{\text{hydr}2}) \quad (2)$$

Because  $[P^{2'-5'}]_{\text{corr}}(t)$  (the concentration of P<sup>2'-5'</sup> obtained if its hydrolysis rate were zero) equals  $k_{\text{ign}2}[\text{ligation complex}]t$ ,

$$[P^{2'-5'}]_{\text{corr}}(t) = [P^{2'-5'}]_{\text{obs}}(t)\{k_{\text{hydr}2}t\}/\{1 - \exp(-k_{\text{hydr}2}t)\} \quad (3)$$

Since  $[P^{3'-5'}]_{\text{obs}} = [P^{3'-5'}]_{\text{corr}}$ ,

$$([P^{2'-5'}]_{\text{corr}}/[P^{3'-5'}]_{\text{corr}}) = ([P^{2'-5'}]_{\text{obs}}/[P^{3'-5'}]_{\text{obs}})\{k_{\text{hydr}2}t\}/\{1 - \exp(-k_{\text{hydr}2}t)\} \quad (4)$$

Equation 4 allows the conversion of  $[P^{3'-5'}]_{\text{obs}}:[P^{2'-5'}]_{\text{obs}}$  to  $[P^{3'-5'}]_{\text{corr}}:[P^{2'-5'}]_{\text{corr}}$  using the time of incubation (*t*) and the value of  $k_{\text{hydr}2}$  measured under ligation conditions (Table 1).

## Results

In our earlier analysis of the regioselectivity of diphosphate-activated oligonucleotide condensation, we did not detect any 2'-5'-linked product.<sup>11</sup> However, in that experiment the ligated product was accumulated over a lengthy (23-day) incubation.<sup>11</sup> Usher and McHale have reported that an isolated 2'-5' linkage in the context of a 3'-5'-linked duplex hydrolyzes 900 times faster than a 3'-5' linkage.<sup>16,17</sup> Thus, we were concerned that the reported  $\geq 200:1$  preference for 3'-5' linkages may have overestimated the intrinsic regioselectivity of the ligation

reaction because the preferential hydrolysis of 2'-5'-linked products was not taken into account.

In order to correct for the differential hydrolysis of the two products, we determined the rate of hydrolysis, under ligation conditions, of an oligoribonucleotide representing the ligation product with a 2'-5' linkage at the position corresponding to the ligation junction (5'-pCCAGUC<sup>2'</sup>p<sup>5'</sup>GGUUCUC). Similar analysis was performed with an entirely 3'-5'-linked control oligonucleotide. As expected, the hydrolysis reactions followed pseudo-first-order kinetics under all conditions tested (Figure 2C). In the absence of a complementary template, the 3'-5' and 2'-5' linkages hydrolyzed at similar rates (Table 1). However, when a 3'-5'-linked complementary template was added, the oligonucleotide with the isolated 2'-5' linkage showed selective hydrolysis at this particular phosphodiester bond (compare parts A and B of Figure 2). Placing the 2'-5' bond in the context of a double helix accelerated its hydrolysis by a factor of 5–10 (Table 1), while placing the corresponding 3'-5' linkage in a duplex decreased its hydrolysis rate by a factor of at least 10 (Table 1). Thus, with the sequences and conditions used in this study, the hydrolysis of a 2'-5' linkage within a double helix is at least 50–100-fold faster than that of a corresponding 3'-5' linkage. These results are consistent with, although somewhat smaller in magnitude than, Usher and McHale's report<sup>16</sup> that a double helical conformation stabilizes 3'-5' phosphodiester linkages but destabilizes isolated 2'-5'-linked bonds.

Given the dramatic difference in hydrolysis rates of the two ligation products, we chose to examine ligated product from a 60-h incubation, the shortest incubation of a pyrophosphate-activated ligation reaction that yielded sufficient material for linkage analysis. A small amount of 2'-5'-linked product was detected in this analysis (Figure 4). However, even when correcting for the differential hydrolysis rates of the 3'-5' and

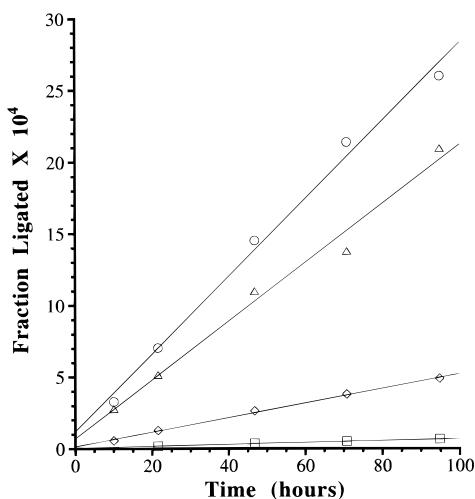
(16) Usher, D. A.; McHale, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1149–1153.

(17) Usher, D. A. *Science* **1977**, *196*, 311–313.

**Table 2.** Effects of Ligation Junction Sequence, Leaving Group, and  $\text{Pb}^{2+}$  on the Regioselectivity of Oligonucleotide Ligation<sup>a</sup>

ligation junction sequence	3'-5':2'-5' ratio			
	pyrophosphate leaving group		imidazole leaving group	
	100 mM $\text{Mg}^{2+}$ , pH 8.0	100 mM $\text{Mg}^{2+}$ , 10 mM $\text{Pb}^{2+}$ , pH 7.0	100 mM $\text{Mg}^{2+}$ , pH 7.0	100 mM $\text{Mg}^{2+}$ , 10 mM $\text{Pb}^{2+}$ , pH 7.0
$\begin{array}{c} \text{5'---OH} \quad \text{P}^{\text{X}} \\ \text{3'---G---C---5'} \end{array}$	61:1, 65:1, 60:1	60:1, 60:1	15:1, 16:1	41:1, 40:1
$\begin{array}{c} \text{5'---OH} \quad \text{P}^{\text{X}} \\ \text{3'---U---C---5'} \end{array}$	85:1, 85:1, 80:1	70:1, 75:1	10:1, 10:1	nd
$\begin{array}{c} \text{5'---OH} \quad \text{P}^{\text{X}} \\ \text{3'---U---G---5'} \end{array}$	55:1, 65:1, 57:1	60:1, 68:1	11:1, 13:1	nd
$\begin{array}{c} \text{5'---OH} \quad \text{P}^{\text{X}} \\ \text{3'---G---C---5'} \end{array}$	nd	nd	13:1, 13:1	nd

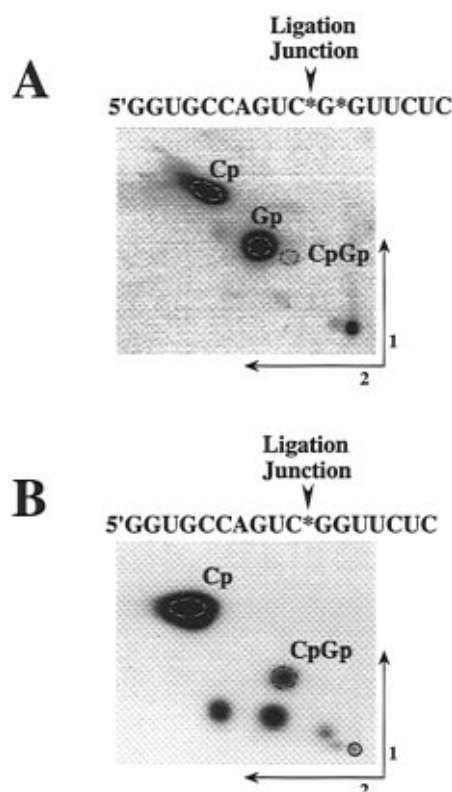
<sup>a</sup> All ratios have been adjusted for the differential hydrolysis of the 3'-5'- and 2'-5'-linked products. The pyrophosphate-activated reaction with the wobble base pair (U·G) could not be subjected to linkage analysis because the ligation rate was too slow to isolate adequate amounts of product.



**Figure 3.** Influence of the identity of the base pair flanking the ligation junction. The  $k_{\text{obs}}$  values obtained for the base pairs are as follows: C·G,  $2.1 \times 10^{-5} \text{ h}^{-1}$  (triangles); A·U,  $2.7 \times 10^{-5} \text{ h}^{-1}$  (circles); U·A,  $5.1 \times 10^{-6} \text{ h}^{-1}$  (diamonds); U·G,  $6.4 \times 10^{-7} \text{ h}^{-1}$  (squares). Reactions were performed under standard conditions at 25 °C, 100 mM  $\text{Mg}^{2+}$ , and pH 9.0.

2'-5' linkages, the reaction displayed marked regioselectivity, with the 3'-5' linkage forming 60 times faster than the 2'-5' linkage.

The initial analysis of ligation regioselectivity was performed with a sequence that had a C·G base pair on the 5' side (the primer side) of the ligation junction and a G·C base pair on the 3' side (the ligator side) (Figure 1). In this system, the attacking hydroxyl is on the 3'-terminal C of the primer and the pyrophosphate leaving group is on the 5'-terminal G of the ligator. Because the base pairs at the ligation junction are expected to have the strongest effect on the orientation of donor and acceptor groups, we investigated the sensitivity of reaction regioselectivity to the sequence at the ligation junction. The C·G base pair on the 5' side of the ligation junction was changed to a U·A base pair (transition change), an A:U base pair (transversion change), and a U·G wobble base pair (Figure 1). The guanosine on the 3' side of the ligation junction could not be substituted because the ligator is transcribed using T7 RNA polymerase, which strongly prefers GTP as its initiating nucleotide. The base-pair changes on the 5' side of the ligation junction had significant effects on reaction rates (Figure 3), but negligible effects on ligation regioselectivity (Table 2). The rates measured with A·U and C·G base pairs were ~5 times faster than the rate with the U·A base pair and ~50 times faster than the rate with the U·G wobble pair (Figure 3). These effects probably reflect slight alterations of the relative orientation of



**Figure 4.** Representative RNase T2 digests of the labeled ligation products from the diphosphate-activated reaction (A) and the imidazole-activated reaction (B) as analyzed by two-dimensional TLC. Sequences of the ligated products are shown, and the positions of labeled phosphates are indicated (asterisks). The location of the digestion products from carrier RNA representing the 2'-5'-linked ligation product were visualized by UV shadowing (dashed circles). Although the radiolabeled CpGp dinucleotide in (A) is not visible in this reproduction of the autoradiogram, it was visible and quantifiable in the PhosphorImager scan. The two extra spots seen at the bottom of the TLC in (B) do not comigrate with any of the nucleoside 3'-monophosphates seen by UV shadowing, and may be base adducts formed during the synthesis of the imidazole.

acceptor and donor groups. The change to a wobble base pair, which led to the greatest alteration in the reaction rate, is expected to produce the most severe distortion in the geometry of the ligation junction. Despite the differences in rates, and the presumed differences in ligation junction geometry, the high degree of ligation regioselectivity did not change. With all three ligation junction sequences examined, 3'-5' linkage formation dominated by a factor of 60–80.

To assess the contribution of the leaving group to the observed regioselectivity, we replaced the diphosphate activating group

with an imidazole moiety (Figure 1). As discussed earlier, template-directed primer extension reactions with imidazole-activated mononucleotides yield products with a large proportion of 2'-5' internucleotide linkages. Although the rate of imidazole-activated oligonucleotide condensation could not be accurately measured because a significant portion of the phosphorimidazolide hydrolyzes during the course of the reaction, we estimate that it is at least 500 times faster than the rate of diphosphate-activated condensation. Imidazole-activated reactions yielded 3'-5':2'-5' ratios of 10-15:1 (Figure 4, Table 2). Comparison with the diphosphate-activated reactions indicates that some (4-8-fold) of the regioselectivity of the diphosphate-activated reaction can be attributed to the leaving group. As with the diphosphate-activated reaction, the regioselectivity of imidazole-activated ligation is relatively insensitive to changes in the ligation junction sequence (Table 2).

In template-directed ligation, the ligation junction is bracketed on both sides by a Watson-Crick double helix, whereas in single-nucleotide primer-extension reactions the reactive site is flanked by an extended Watson-Crick duplex only on the primer side. To assess the contribution of the extended Watson-Crick duplex to ligation regioselectivity, our imidazole-activated oligonucleotide ligation reaction can be compared to a similar template-directed primer extension reaction. Ninio and Orgel<sup>5</sup> reported that the condensation of A<sub>4</sub> with ImpG on a U<sub>9</sub>C<sub>1</sub> template yields products which are <10% 3'-5'-linked. In an oligonucleotide condensation reaction with the same ligation junction sequence and activating group (an A·U base pair on the 5' side of the junction, G·C base pair on the 3' side, imidazole activating group), we found that 90% of the linkages are 3'-5' (Table 2). This comparison demonstrates that the presence of duplex RNA on both sides of the ligation junction as opposed to only one side increases the 3'-5':2'-5' ratio by a factor of ~100, indicating that the extended Watson-Crick duplex is a more important determinant of oligonucleotide ligation regioselectivity than is the leaving group. Placing the ligation junction in the context of a double helix appears to orient the reacting groups such that the reaction is highly regioselective. To more completely dissect the effect of the helix from the effect of the leaving group, we would have to determine the regioselectivity of phosphodiester bond formation when a diphosphate-activated mononucleotide (NTP) condenses with a primer. However, the rate of such NTP addition is so slow that, despite extensive efforts, we have been unable to detect any product.

## Discussion

It has been demonstrated that the regioselectivity of the template-directed condensation of activated mononucleotides is extremely sensitive to a variety of factors such as the ligation junction sequence, the identity of the metal ion catalyst, and the identity of the leaving group.<sup>1,2,5,9,10</sup> We have shown that template-directed oligonucleotide condensation is relatively insensitive to these factors. There is a high preference for the formation of 3'-5'-linked bonds with both pyrophosphate and imidazole leaving groups, and this regioselectivity is unaffected

by the ligation junction sequence changes that we have tested, although it remains possible that changes on the 3' (ligator) side will have larger effects. Moreover, although divalent cations like Pb<sup>2+</sup> can dramatically enhance the rates of condensation of both mononucleotides<sup>7,8,18</sup> and oligonucleotides,<sup>12</sup> Pb<sup>2+</sup> catalysis of oligonucleotide condensation does not come at the expense of regioselectivity (Table 2). These results are similar to, but more dramatic than, the trend to greater regioselectivity observed in studies of the template-directed condensation of dinucleotides.<sup>19-21</sup> The high regioselectivity of template-directed oligoribonucleotide ligation under a wide range of conditions shows that the capacity of RNA for self-assembly is intrinsic to the structure of RNA and need not be imposed by a structured external catalyst such as a ribozyme or a protein enzyme. The self-assembly of duplex RNA from oligonucleotide fragments thus follows the same principles, outlined by Eschenmoser, as the prebiotic self-assembly of biomolecules from simpler precursors.<sup>22</sup>

This study highlights factors that may have led to the accumulation of 3'-5'-linked RNA in a prebiotic chemical environment. Under our conditions for template-directed oligonucleotide ligation, 3'-5' phosphodiester bonds form at a faster rate than 2'-5' bonds and, once formed, are >100 times more stable against base hydrolysis. In addition, the preferred substrates for oligonucleotide condensation would be 3'-5'-linked because 3'-5'-linked oligonucleotides form more stable duplexes with 3'-5'-linked templates than do 2'-5'-linked oligonucleotides.<sup>23</sup> Such selection mechanisms may account for the exclusive survival of 3'-5' phosphodiester bonds in modern-day informational nucleic acids.

In terms of the elementary chemical step (the attack of a 3'-hydroxyl on a 5'-triphosphate), the chemical transformation catalyzed by modern-day template-dependent polymerases is identical to the oligonucleotide ligation reaction. Such polymerases exclusively catalyze the formation of 3'-5' phosphodiester bonds. Primitive ribozyme-based cells in the RNA world would have required ribozyme polymerases. Since template-directed oligonucleotide ligation displays a high preference for the formation of 3'-5' linkages, it is possible that the earliest ribozyme polymerases were enzymes that catalyzed oligonucleotide condensation reactions. More sophisticated ribozymes with the ability to bind and polymerize mononucleotides with high fidelity and regioselectivity may have evolved later.

**Acknowledgment.** We would like to thank Drs. Richard Roberts, Peter Lohse, and Joh Lorsch for helpful discussions and comments on the paper. This work was supported by a grant from NASA.

JA9537134

(18) Sleeper, H. L.; Orgel, L. E. *J. Mol. Evol.* **1979**, *12*, 357-364.

(19) Lohrmann, R.; Orgel, L. E. *J. Mol. Evol.* **1979**, *12*, 237-257.

(20) Lohrmann, R.; Orgel, L. E. *J. Mol. Evol.* **1979**, *14*, 243-250.

(21) Lohrmann, R.; Bridson, P. K.; Orgel, L. E. *J. Mol. Evol.* **1981**, *17*, 303-306.

(22) Eschenmoser, A.; Loewenthal, E. *Chem. Soc. Rev.* **1992**, 1-16.

(23) Giannaris, P. A.; Damha, M. J. *Nucleic Acids Res.* **1993**, *21*, 4742-4749.