

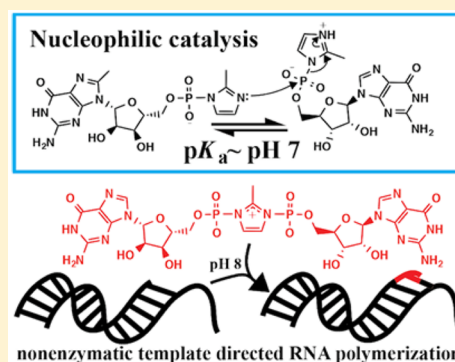
A Highly Reactive Imidazolium-Bridged Dinucleotide Intermediate in Nonenzymatic RNA Primer Extension

Travis Walton and Jack W. Szostak*

Howard Hughes Medical Institute, Department of Molecular Biology, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, United States

S Supporting Information

ABSTRACT: Because of its importance for the origin of life, the nonenzymatic copying of RNA templates has been the subject of intense study for several decades. Previous characterizations of template-directed primer extension using 5'-phosphoryl-2-methylimidazole-activated nucleotides (2-MeImpNs) as substrates have assumed a classical in-line nucleophilic substitution mechanism, in which the 3'-hydroxyl of the primer attacks the phosphate of the incoming monomer, displacing the 2-methylimidazole leaving group. However, we have found that the initial rate of primer extension depends on the pH and concentration at which the activated monomer is maintained prior to the primer extension reaction. These and other results suggest an alternative mechanism, in which two monomers react with each other to form an imidazolium-bridged dinucleotide intermediate, which then binds to the template. Subsequent attack of the 3'-hydroxyl of the primer displaces an activated nucleotide as the leaving group and results in extension of the primer by one nucleotide. Analysis of monomer solutions by NMR indicates formation of the proposed imidazolium-bridged dinucleotide in the expected pH-dependent manner. We have used synthetic methods to prepare material that is enriched in this proposed intermediate and show that it is a highly reactive substrate for primer extension. The formation of an imidazolium-bridged dinucleotide intermediate provides a mechanistic interpretation of previously observed catalysis by an activated nucleotide located downstream from the site of primer extension.



INTRODUCTION

A long-standing goal of synthetic biology and origin of life research is the creation of an artificial life form capable of evolution. A simplified model of an artificial cell is a protocell, consisting of short functional RNA oligonucleotides within a fatty acid vesicle.¹ To evolve, protocells must be able to grow, divide, and transmit functional RNAs to daughter protocells. While fatty acid vesicles readily undergo multiple rounds of growth and division,² the replication of RNA without evolved catalysts remains a challenge. Nonenzymatic template-directed RNA polymerization using chemically activated nucleotides, such as nucleotide-5'-phosphoryl-2-methylimidazole (2-MeImpN) monomers, has been intensively studied as a means of copying RNA templates.³ Nonenzymatic template-directed RNA polymerization can be compatible with fatty acid vesicles, suggesting that this approach may enable the replication of short RNA sequences within a protocell.⁴

Current systems of nonenzymatic template-directed RNA polymerization exhibit several limitations that prevent efficient copying of arbitrary sequences.⁵ For example, the rate of nonenzymatic template-directed RNA polymerization is sequence-specific and in some cases occurs on the same time-scale as RNA degradation. Continued optimization of nonenzymatic template-directed RNA polymerization depends critically on a complete understanding of the chemical mechanism of the reaction, so that the obstacles to fast polymerization on arbitrary

templates can be identified. For instance, the mechanism of nonenzymatic RNA polymerization is believed to begin with monomer binding to the template, followed by an S_N2 reaction with the 3'- (or 2'-) hydroxyl group on the elongating primer.^{6–8} Thus, efforts to increase the rate and sequence generality of nonenzymatic polymerization have focused on improving the binding of monomers to the template or have replaced the hydroxyls of RNA with the more nucleophilic amino group.^{9,10}

Previous reports suggest that the mechanism of nonenzymatic template-directed RNA polymerization is more complicated than just the expected S_N2 reaction between the monomer and elongating strand.^{11,12} These observations define a strong catalytic role for the 2-methylimidazole group of a mono- or oligonucleotide located immediately downstream of the monomer being added to the primer. The mechanism by which the downstream 2-methylimidazole-activated nucleotide exerts its catalytic effect on primer extension has been suggested to result from either a noncovalent interaction that orients the leaving group of the monomer adjacent to the primer so as to facilitate its in-line S_N2 reaction with the 3'-OH of the primer or acid–base catalysis.¹¹

Through kinetic analysis of primer extension reactions, we demonstrate that the rate of nonenzymatic RNA polymerization

Received: August 1, 2016

Published: August 23, 2016

is consistent with a reaction between the 2-MeImpN anion and zwitterion. Here we report evidence that the product of this reaction is a highly reactive imidazolium-bridged dinucleotide. We use site-specific ^{13}C -labeling of 2-MeImpN monomers to demonstrate by ^{13}C and ^{31}P NMR that the proposed intermediate is formed in aqueous solutions, and we show that the partially purified compound results in fast rates of primer extension. We propose that a downstream mono- or oligonucleotide accelerates the rate of primer extension through nucleophilic catalysis and formation of an imidazolium-bridged dinucleotide intermediate.

RESULTS

Kinetic Analysis of the Primer Extension Reaction.

Previous characterizations of the RNA primer extension reaction assumed pseudo-first-order kinetics because the concentration of 2-MeImpN (**1**) is >1000 times the concentrations of primer and template (Figure 1a).^{4,13} Under this assumption,

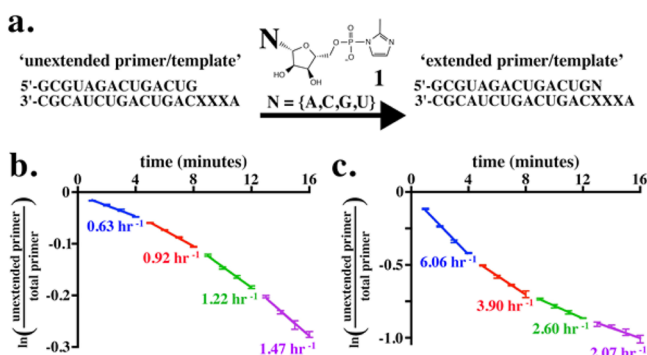


Figure 1. Rate of primer extension increases or decreases with time, depending on the pH of the monomer stock solution. (a) The primer extension assay measures the rate at which the primer in an RNA primer template duplex is converted to longer extension products by nonenzymatic reaction with 2-MeImpN monomers (**1**). (b) Using a 2-MeImpG stock at pH 9.88, the k_{obs} of primer extension increases over time. (c) Using a 2-MeImpG stock at pH 7.79, the k_{obs} of primer extension decreases over time. Error bars indicate \pm one SD.

the fraction of unreacted primer should decline exponentially with time, and the rate constant k_{obs} is obtained from the slope of a plot of the log of the fraction of unreacted primer versus time. While investigating the kinetics of primer extension, we occasionally observed that k_{obs} increased or decreased by 2–3 times during experimental time courses using 2-MeImpG on a template with a 3'-CCCA-5' single-stranded overhang (Figure 1b,c). In the course of studying these surprising kinetic profiles, we discovered that the deviations from pseudo-first-order kinetics were related to the pH of the 2-MeImpG stock solution, prior to its addition to the primer extension reaction mixture. If the 2-MeImpG stock solution was more basic than the pH 8 of the primer extension reaction, k_{obs} increased over time, but if the 2-MeImpG stock solution was more acidic than the pH of the primer extension reaction, k_{obs} decreased over time. We hypothesized that these effects might reflect the formation and/or decay of an unstable intermediate in the primer-extension reaction.

To further investigate this phenomenon, we systematically altered the pH of 2-MeImpN stock solutions and determined the effect of monomer pH on the initial rate of primer extension measured during the first 3 min of reaction at a constant pH of 8.0. We observed that incubation of the monomer

solution at pH \sim 6.9 produced the fastest initial rates of primer extension for 2-MeImpG, 2-MeImpC, and 2-MeImpA, indicating that optimal monomer pH does not depend on the identity of the nitrogenous base (Figure 2a); rates for 2-MeImpU

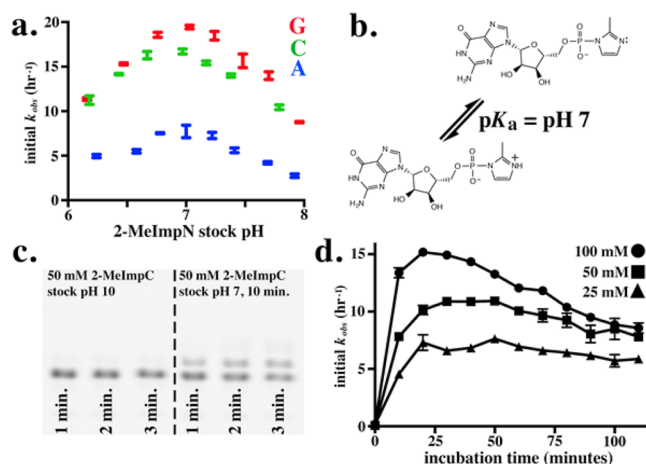


Figure 2. Initial rate of primer extension depends on both the pH and concentration of 2-MeImpN stock solutions. (a) The fastest initial rate of primer extension occurred with 2-MeImpN solutions preincubated at pH 6.8–7.0. (b) The 2-methylimidazole group of 2-MeImpG has a previously determined $\text{p}K_{\text{a}}$ of $\text{pH } 7.09 \pm 0.20$.¹⁴ (c) Representative gel image used to determine the initial rate of primer extension reactions initiated with a 50 mM 2-MeImpC stock at pH 10 or 7. (d) The initial rate of primer extension depends on the concentration of the 2-MeImpC stock solution and incubation time at pH 7. k_{obs} at time zero was determined using a 2-MeImpC stock at pH 10.4. Error bars indicate \pm one SD.

were too slow to accurately determine an initial rate. We noticed that the optimal stock pH corresponds to the $\text{p}K_{\text{a}}$ of the 2-methylimidazole moiety on 2-MeImpG (Figure 2b).¹⁴ Since solutions at a pH equal to the $\text{p}K_{\text{a}}$ contain equal amounts of conjugate acid and base, these results suggested that the rate of primer extension might be affected by an unknown interaction between the 2-MeImpN anion and zwitterion (i.e., unprotonated and protonated 2-methylimidazole leaving group).

The hypothesis of an intermolecular interaction between two monomer molecules prior to primer extension led us to predict that the rate of primer extension should also depend on the concentration of the 2-MeImpN stock solution. In particular, decreasing the stock concentration would be expected to lessen the interaction between 2-MeImpN anions and zwitterions at pH 7 and subsequently lower the initial rate of primer extension. To test this hypothesis, we monitored the effect of incubation time at pH 7 for three different stock concentrations of 2-MeImpC, because altering the stock concentration could also affect the optimal incubation time. At time zero, a 200 mM 2-MeImpC solution at pH 10.4 was diluted with dilute hydrochloric acid to neutral pH and 100, 50, or 25 mM concentration. Every 10 min, aliquots of the diluted stock solutions were used to initiate primer extension reactions at pH 8.0 (Figure 2c). We adjusted the volume of stock solution added such that all primer extension reactions had the same final concentration of 2-MeImpC (10 mM).

The resulting plot of incubation time of the 2-MeImpC stock solution at neutral pH versus the k_{obs} of primer extension showed that the initial rate of polymerization varied nonlinearly with the incubation time of 2-MeImpC at neutral pH (Figure 2d). For all three concentrations of the 2-MeImpC stock solution,

k_{obs} for primer extension increased sharply during the first 20 min of preincubation at pH 7.0, followed by a plateau or a steady decrease after extended incubation at that pH. The steady decrease of k_{obs} cannot be fully explained by the mild acidification (<0.1 pH units) we observed by the end of the second hour. Instead, we suggest that the decrease of k_{obs} might result from formation of inhibitors of primer extension, such as nucleoside-5'-monophosphates.¹⁵

Comparing the time courses of preincubation of the activated C monomer at 100, 50, and 25 mM stock concentrations showed that the initial rate of primer extension increases with increasing concentration of the 2-MeImpC stock solution during its pH 7 incubation. This result was especially prominent during the first hour of the time course. This concentration dependence during the 2-MeImpC incubation at pH \sim pK_a provides additional strong evidence that at least two monomers interact to generate a reactive intermediate that increases the initial rate of primer extension.

Identification of Proposed Intermediate. We wished to identify the structure of the intermediate formed by incubation at pH 7 to help explain our kinetic results and to further our understanding of the chemical mechanism of nonenzymatic template-directed RNA polymerization. In the course of our search for potential intermediates, we became interested in a previous report showing that two molecules of 1-phosphoimidazole can react to form one molecule of free imidazole and one molecule of 1,3-diphosphoimidazolium.¹⁶ 1,3-diphosphoimidazolium is a phosphorylating agent, and its structure has been confirmed by X-ray crystallography.¹⁷ We hypothesized that 2-MeImpG (**1a**) undergoes a similar reaction when incubated at pH 7, to form the 1,3-dinucleotide-2-methylimidazolium **2a** and free 2-methylimidazole (**3**) (Figure 3a). We propose that **2a** is formed by attack of the nucleophilic N3 of the deprotonated 2-methylimidazole moiety of one monomer on the phosphate of a second monomer with a protonated 2-methylimidazole leaving group. This mechanism explains why the optimal pH for monomer preincubation is equal to the pK_a of the 2-methylimidazole group of 2-MeImpG.

We began to investigate the possible formation of the proposed imidazolium-bridged dinucleotide intermediate by using NMR to analyze a 50 mM solution of 2-MeImpG in D₂O following a change from an initial pD of \sim 8.6 to either pD 7.06 or 4.78. When the sample pD was dropped to 7.06 (but not to 4.78) (Figures S1–S3, SI), we observed a 5–6 times increase of intensity for a peak in the ³¹P NMR spectrum at -13.83 ppm, distinct from the 2-MeImpG phosphorus resonance at -12.42 ppm (Figure 3b). The ¹H NMR spectrum of the same solution also revealed new peaks near the H1', H8, and the 2-methylimidazole aromatic protons (H_{imid}) of 2-MeImpG (Figure 3c). Since the increased intensity of these peaks correlates with the increased initial rate of primer extension, we have tentatively assigned them to an intermediate of this reaction.

We obtained similar NMR spectra for 2-MeImpA, 2-MeImpU, and 2-MeImpC near neutral pD (Figures S4–S6, SI), indicating that the increase in the intensity of these newly assigned peaks at pH \sim pK_a does not depend on the identity of the nitrogenous base. In addition, we also observed similar peaks in the NMR spectra of 2',3'-dideoxy-2-MeImpC and 2',3'-dideoxy-2-MeImpU during incubation at neutral pD, indicating that this result does not require formation of phosphodiester bonds (Figures S7–S9, SI). These observations are consistent with the structure of the proposed imidazolium-bridged dinucleotide. Due to its symmetry,

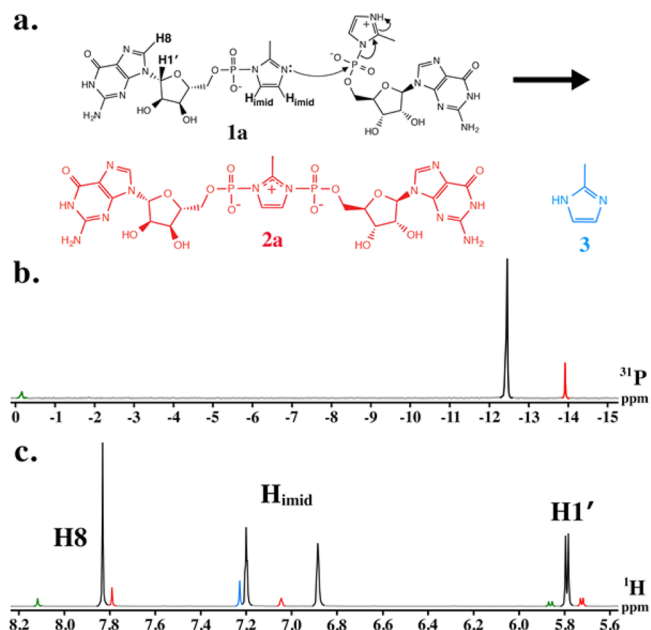


Figure 3. Observation of NMR signals from the putative diguanosine intermediate **2a** at neutral pD. (a) The proposed reaction of a 2-MeImpG (**1a**) anion and zwitterion produces the diguanosine intermediate **2a** and free 2-methylimidazole (**3**). H_{imid} represents the aromatic 2-methylimidazole hydrogens. H1', H8, and H_{imid} are observed via ¹H NMR. (b) ³¹P (161 MHz) and (c) ¹H (400 MHz) NMR spectra of 50 mM 2-MeImpG incubated at pD 7.06 between 55 and 60 min. Signals from GMP are highlighted in green, from 2-MeImpG in black, and from 2-methylimidazole in blue. Putative 1,3-di-GMP-2-methylimidazolium signals are highlighted in red. Referenced ³¹P NMR spectra are included in the Supporting Information (SI).

we expected that an imidazolium-bridged dinucleotide would produce a single ³¹P NMR singlet peak. This interpretation is supported by the observation of this peak for multiple nucleobases and sugars, which is predicted by the imidazolium-bridged structure.

We performed mass spectrometry to identify the mass of the new compound formed by incubation of 2-MeImpC at pH 7. Compared to incubation at pH 11, mass spectrometry of 2-MeImpC incubated at pH 7 resulted in an additional peak at $m/z = 691$, the predicted mass of the intermediate (Figure S10, SI). However, tandem mass spectrometry of the $m/z = 691$ peak resulted in a fragmentation spectrum that included a peak resulting from loss of 2-methylimidazole, which is inconsistent with the imidazolium-bridged structure. Similar results were obtained using liquid chromatography–mass spectrometry (LC–MS), suggesting that one or more structural isomers are being formed instead of or in addition to the imidazolium-bridged dinucleotide. However, our analysis of additional activated monomers, notably 2-MeImpddU, is consistent with the formation of an imidazolium-bridged dinucleotide due to the absence of peaks resulting from loss of 2-methylimidazole in the fragmentation spectra (Figures S11–S14 and Table S1, SI). Our mass spectrometry data suggest that multiple structural isomers exist in our samples of 2-MeImpC, but they may also be explained by rearrangement of the highly reactive imidazolium bridge during fragmentation (supplementary text, SI).

Analysis of Intermediate-Enriched Material. Since our analysis by mass spectrometry indicated that monomer stocks contain multiple dinucleotides activated by 2-methylimidazole,

we sought to confirm the role and structure of the intermediate formed by incubation at neutral pH. We attempted to directly synthesize the intermediate corresponding to the peak we observed at -13.8 ppm by ^{31}P NMR. Using a modification of the procedure for 2-MeImpC synthesis, we reacted a concentrated solution of cytidine-5'-monophosphate (CMP) and 2-methylimidazole at a $\sim 2:1$ ratio. By ^{31}P NMR, we observed that the major product formed by this reaction corresponded to the intermediate observed in 50 mM solutions of 2-MeImpC at neutral pD (Figure 4a). Rigorous isolation of the product

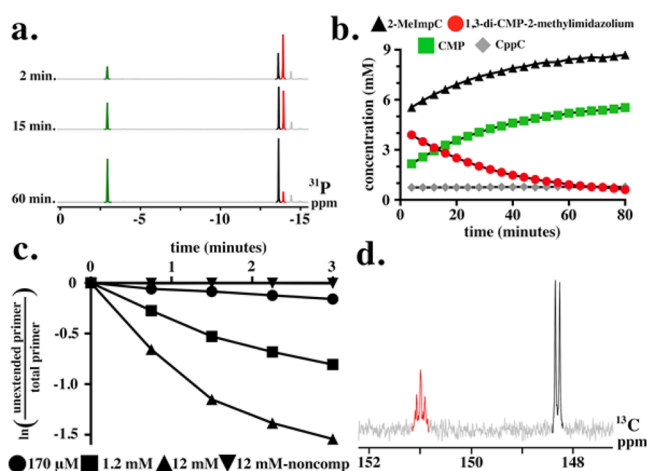


Figure 4. Characterization of the predicted imidazolium-bridged dinucleotide in enriched material. (a) Partially purified material quickly decays into CMP and 2-MeImpC. ^{31}P NMR (161 MHz) peaks from CMP are highlighted in green, from 2-MeImpC in black, from dicytidine-5',5'-pyrophosphate (CppC) in gray, and from putative imidazolium-bridged dinucleotide in red. CMP and 2-MeImpC were identified by addition of standards. Time begins with resuspension of enriched material in D_2O . (b) Quantification of proposed intermediate, CMP, CppC, and 2-MeImpC over time by ^1H NMR. (c) Primer extension reactions initiated using the partially purified material have fast rates, confirming its role in the reaction. Primer extension was not observed on a template noncomplementary to cytidine. (d) ^{13}C NMR (100 MHz) of the enriched material identifies a new triplet (red) at 151.0 ppm corresponding to an imidazolium-bridged structure. The doublet (black) at 148.3 ppm corresponds to C2 of the 2-methylimidazole group of 2-MeImpC monomer.

was impeded by its rapid hydrolysis into 2-MeImpC and CMP. However, we were able to obtain enriched material with the intermediate comprising $\sim 50\%$ of the C nucleotides. Since the intermediate is likely a dimer, this corresponds to a 0.35 mole fraction of the putative imidazolium-bridged dinucleotide.

Using this enriched material, we measured the half-life of the putative imidazolium-bridged dinucleotide at pD 4.5–4.8 by ^1H NMR. These values of pD provided a stable range during the hydrolysis studies without need of buffer, which could potentially react with our unstable intermediate. Under these conditions, we observed a half-life of 30 min (Figure 4b), much shorter than the ~ 16 h half-life of G monomer observed under similar conditions.¹⁴ In addition, we observed that the partially purified material hydrolyzed into equal amounts of CMP and 2-MeImpC, consistent with the expected composition of the imidazolium-bridged structure. The rapid decay of this material to form CMP and 2-MeImpC is not consistent with the structure of a phosphodiester-linked dimer containing a 5'-phosphoryl-2-methylimidazole group (2-MeImpCpC).

Instead, we expect that 2-MeImpCpC hydrolyzes to free 2-methylimidazole and a dinucleotide.

The instability of the material synthesized above led us to hypothesize that this putative intermediate would exhibit fast initial rates of primer extension. Primer extension reactions initiated by the addition of the partially purified intermediate slowed down even over the first few minutes, likely due to instability of the intermediate (Figure 4c). However, an approximate value of k_{obs} for primer extension using a sample of the enriched material containing 12 mM total C nucleotides (corresponding to ~ 3 mM of the intermediate) is >30 h^{-1} (Figure 4c, right side up triangles). This rate is higher than that observed simply by preincubating the 2-MeImpC stock solution at pH 7.0 (Figure 2a), presumably because the synthetic material contains a higher proportion of a reactive intermediate. We observed that even low concentrations of the enriched material, e.g. 170 μM (corresponding to ~ 43 μM intermediate) resulted in primer extension with a measured k_{obs} of 3 h^{-1} (Figure 4c, circles). Importantly, changing the single-stranded portion of the template to a sequence (3'-CCCA-5') that is not complementary to cytidine abolished primer extension, even at the highest concentration tested (12 mM total C nucleotide, corresponding to ~ 3 mM of the intermediate) (Figure 4c, upside down triangles). This result indicates that the fast rates of primer extension due to reaction with the imidazolium-bridged dinucleotide require template-binding.

Having verified that the partially purified material functions as a substrate for primer extension, we sought to identify unique spectroscopic signatures of this intermediate to test predictions of an imidazolium-bridged structure. We were drawn toward ^{13}C NMR because carbon–phosphorus coupling should produce unique signals for the imidazolium bridge of the intermediate. We took ^{13}C NMR spectra of a concentrated solution (~ 500 mM total C nucleotides, ~ 125 mM dicytidine intermediate) of the enriched material immediately after resuspension in D_2O and 1 h later. The high concentration of the enriched material was necessary to obtain a ^{13}C spectrum in the short time before the intermediate decayed. Peaks observed in the spectrum taken immediately after resuspension in D_2O , but absent 1 h later, were identified as candidate signals of the intermediate. Among such peaks, we observed a well-resolved triplet 2.7 ppm downfield from the doublet of the C2 imidazole carbon of 2-MeImpC (Figure 4d). We hypothesized that this triplet signal corresponded to the C2 imidazole carbon of the imidazolium bridge, coupled to both symmetrically disposed phosphorus atoms.

Confirmation of the Imidazolium Bridge by Site-Specific ^{13}C Labeling. To confirm our assignment of the triplet we observed in the ^{13}C NMR spectrum of the enriched material (Figure 4d), we labeled the C2 position of 2-methylimidazole moiety with ^{13}C . We were particularly interested in the C2 position of the 2-methylimidazole moiety because the ^{13}C NMR signal of this C2 carbon should display features diagnostic of the proposed imidazolium-bridged structure. Specifically, carbon–phosphorus coupling is expected to produce a doublet for C2 of the 2-methylimidazole ring of 2-MeImpC monomer (Figure 5a). However, the C2 imidazole carbon of the imidazolium-bridged intermediate would be coupled to two equivalent phosphorus atoms, producing a triplet (Figure 5b). In addition, we expected the sensitivity provided by an isotope label to allow us to observe the formation of the proposed intermediate by incubation at neutral pH in aqueous solution.

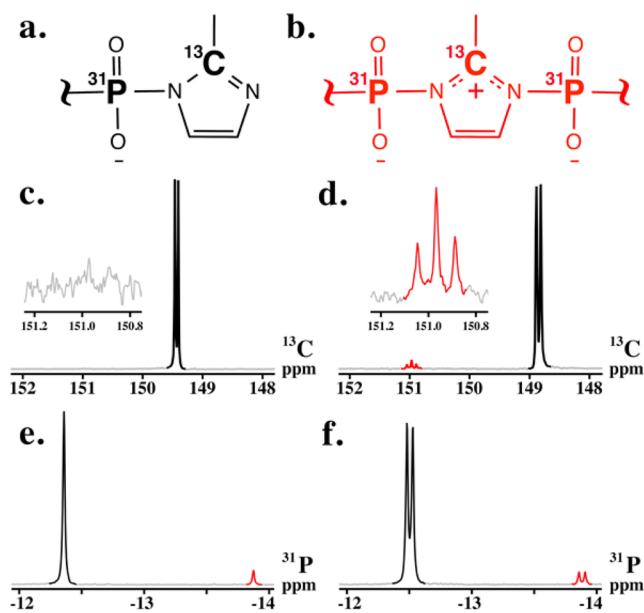


Figure 5. Imidazolium bridge structure predicts the effects of ^{13}C isotope label at C2 on NMR spectra. (a) Partial chemical structure of an activated monomer showing the location of the ^{13}C isotope label two bonds from the naturally abundant ^{31}P atom. (b) Partial chemical structure of the imidazolium bridge showing that the ^{13}C isotope label will be split by two equivalent ^{31}P atoms. (c) ^{13}C NMR (100 MHz) spectrum of 50 mM ^{13}C isotope labeled 2-MeImpC at pD 8.49. (d) ^{13}C NMR (100 MHz) spectrum of 50 mM ^{13}C isotope labeled 2-MeImpC at pD 6.93. The inset displays an observed triplet at 151.0 ppm. (e) ^{31}P NMR (161 MHz) spectrum of 50 mM 2-MeImpC without the isotope label at pD 7.02. (f) ^{31}P NMR (161 MHz) spectrum of 50 mM ^{13}C isotope labeled 2-MeImpC at pD 6.93. Peaks from 2-MeImpC are highlighted in black and from dicytidine intermediate in red.

Using monomer labeled at the C2 position of 2-methylimidazole with ^{13}C , we collected NMR spectra of 50 mM 2-MeImpC incubated at pD 8.49 or 6.93. In the ^{13}C spectrum of 2-MeImpC incubated at pD 8.49, we observed a pronounced doublet at 149.44 ppm that corresponds to the C2 carbon of 2-methylimidazole on the monomer 2-MeImpC (Figure 5c). When incubated at pD 6.93, we additionally observed a triplet in the ^{13}C spectrum at 150.97 ppm (Figure 5d), similar to that observed for the enriched material (Figure 4d). These results validated our assignment of the triplet peak observed in the ^{13}C NMR spectrum of the enriched material to an intermediate formed by incubation at pH 7. Importantly, the triplet C2 resonance provided strong evidence for our proposed symmetrical structure.

In addition to ^{13}C NMR, we also examined the ^{13}C isotope labeled samples by ^{31}P NMR. As described earlier in this report, the ^{31}P NMR spectrum of 50 mM 2-MeImpC lacking the isotope label displayed a singlet peak for both the monomer and intermediate when incubated at neutral pD (Figure 5e). However, introduction of the isotope label caused these peaks to become doublets (Figure 5f), indicating that the ^{31}P NMR peaks correspond to the 5'-phosphor-2-methylimidazolide. Coupling of the peak at -13.9 ppm by the ^{13}C isotope label excluded the assignment of this peak to other possible structural isomers involving bonds between phosphate and the nucleobase. This supports our assignment of the peak at -13.9 ppm to the imidazolium-bridged dinucleotide because our proposed structure predicts coupling of the ^{31}P peak by the ^{13}C isotope label at C2. Together, our analysis of the ^{13}C isotope

labeled 2-MeImpC and intermediate by NMR spectroscopy confirmed predictions based upon our proposed structure of an imidazolium-bridged dinucleotide.

DISCUSSION

Template-directed primer extension with 2-MeImpN substrates has long been assumed to result from a simple $\text{S}_{\text{N}}2$ reaction involving nucleophilic attack by the 2'- or 3'-hydroxyl of the primer on the phosphate of a 2-MeImpN monomer, with displacement of the 2-methylimidazole leaving group. We now propose an alternative mechanism for this reaction involving nucleophilic catalysis in which N3 of the imidazole moiety of one monomer attacks the phosphate of a second monomer, displacing its 2-methylimidazole leaving group and forming a reactive imidazolium-bridged dinucleotide intermediate. We propose that this intermediate binds to the template, forming two Watson–Crick base pairs (Figure 6a). At this

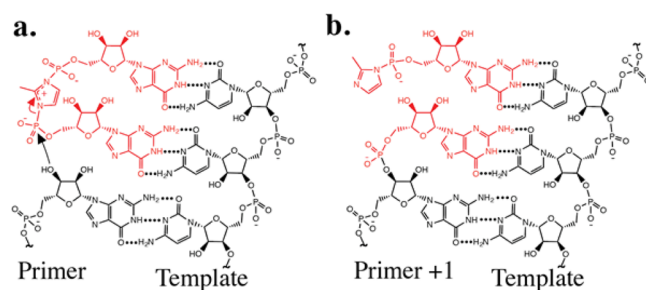


Figure 6. Proposed mechanism of primer extension by the imidazolium-bridged dinucleotide. (a) The imidazolium-bridged dinucleotide intermediate, where N = G (red) binds the template through Watson–Crick base pairs and reacts with the 3' hydroxyl of the primer to form the extension product, primer +1. (b) In addition to the primer +1 product, this reaction also regenerates the activated G monomer.

point, nucleophilic attack by the 2'- or 3'-hydroxyl of the primer on the phosphate of the adjacent nucleotide occurs, displacing the second downstream nucleotide, which is regenerated as a 2-methylimidazole-5'-phosphonucleotide (Figure 6b). The fast rates of primer extension observed with enriched preparations of this intermediate, taken together with the concentrations of intermediate present under primer extension conditions, suggest that our proposed intermediate can account for the observed rates of primer extension and is therefore the predominant mechanism of nonenzymatic primer extension.

One reason that our proposed mechanism has been previously overlooked may be the instability of the imidazolium-bridged intermediate, which makes structural characterization inherently difficult. To overcome this obstacle, we synthesized material that was sufficiently enriched that we were able to characterize its reactivity and identify putative signals by ^{13}C NMR. We then synthesized ^{13}C isotope labeled activated monomer, which greatly increased the sensitivity of detection of the proposed intermediate. Future studies of the primer extension reaction will be greatly aided by the synthesis of more stable analogs. We note that the 1,3-diphosphoimidazolium compound examined by X-ray crystallography is likely more stable than our intermediate due to enhanced charge shielding of water molecules by the increased negative charge over the six oxygen atoms.

The proposal of a covalent intermediate in the mechanism of primer extension explains several enigmatic features of

reaction. For instance, the formation of an imidazolium-bridged intermediate explains the previously described catalytic interaction between 2-methylimidazole groups of adjacent monomers. This catalytic effect has previously been attributed to a noncovalent interaction between the adjacent 2-methylimidazole groups. However, it would be extremely difficult to explain the changes in the rate of primer extension, which occur over minutes to 1 h following a change in monomer pH (Figure 2d), by the equilibration of noncovalent interactions; in contrast, these slow changes in rate are readily explained by the synthesis or decay of our proposed intermediate. In cases where the proposed intermediate cannot form, as in the ligation of two oligonucleotides bound to a template, the reaction must proceed through the classical direct S_N2 mechanism and is therefore extremely slow.^{11,12}

While prior models of primer extension have proposed a catalytic interaction between activated monomers bound to a template,^{11,12} our results indicate that the 2-MeImpNs can react to form the imidazolium-bridged intermediate in the absence of template. This suggests a model whereby primer extension is mediated by binding of an activated dimer that forms in solution. However, it remains possible that monomers can also react to form an intermediate when bound to the template. Understanding the contribution of template-dependent and template-independent formation of an intermediate to the rate of primer extension will help guide further optimization of nonenzymatic RNA polymerization.

In addition, the identification of the imidazolium-bridged dinucleotide opens new questions regarding the conformation and affinity of the template-bound intermediate. Preliminary modeling has suggested that pyrophosphate-linked dinucleotides can interact with a template via Watson–Crick base pairing of both nucleotides.¹⁸ Assuming that both nucleotides of the imidazolium-bridged intermediate can also base-pair with the template simultaneously, the 5'-imidazolium 5'-linkage of the intermediate might constrain the conformation of the dimer so as to enhance the rate of reaction. We hypothesize that the conformation of the template-bound intermediate may position its phosphate preferentially for in-line nucleophilic attack by the 3'-hydroxyl of the primer. Such a conformational constraint could explain how the subtle change from imidazole to 2-methylimidazole on the monomer increases the proportion of canonical 3'-5' versus noncanonical 2'-5' phosphodiester linkages.¹⁹

Finally, measuring the affinity of the intermediate for the single-stranded template might provide insight into the sequence specificity of the rate of the primer extension reaction, which is not well correlated to the affinity of the monomer.^{6,7} The intermediate is a relatively minor component of monomer solutions, but it has a large effect on the rate of primer extension, suggesting that the intermediate might have strong affinity for the template or high intrinsic reactivity, or both. We also suggest that the rate enhancement by the dinucleotide intermediate would be decreased on a template with a single nucleotide-binding site due to being outcompeted by the more abundant monomer.

A recent report²⁰ by the Richert group proposed that an imidazolium-bridged dinucleotide is a chemical impurity in their crude synthetic preparations of 2-MeImpG. They also suggested that this compound resulted in very rapid primer extension, thus explaining previously puzzling observations of burst kinetics. Our results are consistent with those reported by the Richert group.

In summary, the identification and characterization of a highly reactive imidazolium-bridged dinucleotide intermediate clarifies our mechanistic understanding of the primer extension reaction. The dual roles for the 2-methylimidazole moiety of activated nucleotides as both nucleophilic catalyst and leaving group suggest that these roles might be separately optimized to increase the formation of the intermediate. Future work will explore the utilization as well as potential limitations of the intermediate for increasing the rate and sequence-generality of nonenzymatic template-directed RNA polymerization.

METHODS AND MATERIALS

Materials. All materials were purchased from Sigma-Aldrich and used without further purification unless noted below. Guanosine-5'-monophosphate, free acid, was purchased from Santa Cruz Biotechnology, Inc. Deuterium oxide was obtained from Cambridge Isotope Laboratories, acetone and diethyl ether were from Thermo-Fisher Scientific, 2,2'-dipyridyl disulfide was from Chem-Impex International, and 2',3'-dideoxycytidine was from Tokyo Chemical Industry Co. Solutions of 1 M $MgCl_2$, 1 M Tris buffer pH 8, and nuclease-free non-DEPC-treated water were purchased from Life Technologies Corp. RNA oligonucleotides were purchased from Integrated DNA Technologies, and one RNA/LNA oligonucleotide was purchased from Exiqon.

Synthesis of 2-MeImpN. Activated monomers were prepared as previously described²¹ with the following slight modifications. A 100 mg portion of free acid nucleotide-5'-monophosphate and 200 mg of 2-methylimidazole were dissolved in 10 mL of DMSO and 300 μL of triethylamine. Triphenylphosphine (1.5 g) and 2,2'-dipyridyl disulfide (1.3 g) were added, and the reaction was incubated at room temperature for 1–2 h before precipitation in a solution of 60 mL of diethyl ether, 120 mL of acetone, and 4.5 g of sodium perchlorate. Precipitated material was washed with acetone and desiccated overnight. 2-MeImpN was purified by column chromatography using a 30 g C18aq column on a Combiflash RF-200 from Teledyne Isco. Fractions of 2-MeImpN were eluted with a gradient of water from an EMD Millipore Milli-Q Reference purification system and 0% to 10% acetonitrile over 5 column volumes at a flow rate of 25 mL/min. Importantly, the pH of the fractions was adjusted to pH 10.4 using NaOH prior to flash freezing and lyophilization on a VirTis Freezemobile 2SEL from SP Scientific.

Synthesis of 2-MeImpG was performed as described above with the following modifications. First, a lyophilized mixture of 100 mg of GMP and 200 mg of 2-methylimidazole was prepared to improve the solubility of GMP in DMSO. Second, the reaction mixture was incubated overnight rather than for 1–2 h to increase the yield. Syntheses of 2',3'-dideoxy-2MeImpC and 2',3'-dideoxy-2MeImpU were performed by converting the nucleoside to the nucleotide-5'-monophosphate using an established protocol,²² followed by the above procedure. Synthesis of ¹³C isotope labeled 2-MeImpC began by synthesis of 2-methylimidazole labeled with ¹³C at the C2 position using acetaldehyde-1-¹³C, following a previously published synthetic procedure,²³ followed by nucleotide activation as described above.

Synthesis of the Imidazolium Bridged Dicytidine Intermediate. A 130 mg portion of cytidine-5'-monophosphate was dissolved in 1.5 mL of DMSO, 200 μL of triethylamine, and 550 μL of 480 mM 2-methylimidazole. Four minutes after adding 1 g of triphenylphosphine and 1 g of 2,2'-dipyridyl disulfide, the reaction was precipitated as described above. After desiccation, one-half of the precipitated material was resuspended in 3 mL of 0.1% hydrochloric acid to aid in separation by column chromatography. Enriched fractions (estimated purity 30%) were obtained using a 5.5 g C18aq column and an 8 mL/min flow rate using the Combiflash system with Milli-Q water and a 0–20% acetonitrile gradient over 7 column volumes, followed by 20–80% acetonitrile over 7 column volumes. Fractions were immediately flash frozen without pH adjustment and lyophilized at –20 °C on a VirTis AdVantage Plus EL-85 lyophilizer from SP Scientific. Upon resuspension in D₂O, the pD was 4.6.

General Characterization of 2-MeImpN Stock Solutions. The concentration of 2-MeImpN was determined by analysis of serial dilutions on a NanoDrop 2000c spectrophotometer. The pH of 2-MeImpN stock solutions was determined by pipetting 4 μL onto the tip of an Orion 9863BN micro pH probe equipped with a needle tip and a Mettler Toledo S220 SevenCompact meter.

Primer Extension Reactions. Primer extension experiments were carried out with 2 μM cyanine 3 (Cy3)-labeled RNA primer (5'-Cy3-GCGUAGACUGACUG-3'), 5 μM template (5'-ANNNCAGUC-AGUCUACGC-3', where NNN denotes either GGG or CCC), 100 mM MgCl_2 , and 250 mM Tris-Cl pH 8. Primer extension experiments with 2-MeImpA (Figure 2a) used an LNA/RNA template (5'-GTTTCAGUCAGUCUACGC-3', where the locked nucleic acid (LNA)²⁴ portion is italicized). Primer extension reactions in Figure 1 contained 40 mM 2-MeImpG. Primer extension reactions in Figure 2a contained 20 mM 2-MeImpN. Primer extension reactions in Figure 2d contained 10 mM 2-MeImpC. In all experiments, 1 μL reaction aliquots were quenched in 7 μL of 8 M urea, 100 mM Tris-Cl, 100 mM boric acid, and 75 mM EDTA before flash-freezing. After running the samples on 20% polyacrylamide gels at a constant power of 30 W for 20–30 min, gels were imaged on a Typhoon 9410 scanner and quantified using the accompanying ImageQuant TL software. All reactions were performed in triplicate or greater.

NMR Spectroscopy. All spectra were obtained using a Varian INOVA 400 MHz NMR spectrometer at 23 $^\circ\text{C}$ in D_2O . All spectra were analyzed using the MNova software. Quantifying the decay of the dicytidine intermediate in Figure 4d was performed in triplicate by taking ^1H spectra with 64 scans every 4 min. Referenced spectra using the following standards are included in the Supporting Information: internal residual HOD ($\delta = 4.79$ ppm) for ^1H (400 MHz), internal trimethyl phosphate ($\delta = 0.00$ ppm) for ^{31}P (161 MHz), and internal acetone ($\delta = 30.89$ ppm) for ^{13}C (100 MHz). ^{31}P and ^{13}C spectra were acquired using ^1H -decoupling. ^1H spectra were obtained using ^{13}C -decoupling.

Mass Spectrometry. All samples were diluted to 200 μM in Milli-Q water immediately prior to analysis. Spectra were obtained by direct injection on an Esquire 6000 mass spectrometer from Bruker Daltonics, operated in negative ion mode.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b07977.

Figures S1–S14, Table S1, and supplementary text, methods and references (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*szostak@molbio.mgh.harvard.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by a grant (290363) from the Simons Foundation to J.W.S. J.W.S. is an Investigator of the Howard Hughes Medical Institute. We thank Li Li, Tony Jia, Aaron Engelhart, Albert Fahrenbach, Nathalie Madern, and Noam Prywes for helpful discussions and comments on the manuscript. We acknowledge the assistance of Anders Björkbo and Victor Lelyveld with mass spectrometry. We acknowledge the assistance of Enver Izgu with NMR spectroscopy.

■ REFERENCES

- (1) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409*, 387.
- (2) Zhu, T. F.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 5705.
- (3) Blain, J. C.; Szostak, J. W. *Annu. Rev. Biochem.* **2014**, *83*, 615.

- (4) Adamala, K.; Szostak, J. W. *Science* **2013**, *342*, 1098.
- (5) Szostak, J. W. *J. Syst. Chem.* **2012**, *3*, 2.
- (6) Kervio, E. C.; Claasen, B.; Steiner, U. E.; Richert, C. *Nucleic Acids Res.* **2014**, *42*, 7409.
- (7) Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 6373.
- (8) Li, L.; Lelyveld, V. S.; Prywes, N.; Szostak, J. W. *J. Am. Chem. Soc.* **2016**, *138*, 3986.
- (9) Vogel, S. R.; Deck, C.; Richert, C. *Chem. Commun.* **2005**, *39*, 4922.
- (10) Schrum, J. P.; Ricardo, A.; Krishnamurthy, M.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 14560.
- (11) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 5496.
- (12) Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. *eLife* **2016**, *5*, e17756.
- (13) Heuberger, B. D.; Pal, A.; Del Frate, F.; Topkar, V. V.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 2769.
- (14) Kanavarioti, A.; Rosenbach, M. T. *J. Org. Chem.* **1991**, *56*, 1513.
- (15) Deck, C.; Jauker, M.; Richert, C. *Nat. Chem.* **2011**, *3*, 603.
- (16) Rathlev, T.; Rosenberg, T. *Arch. Biochem. Biophys.* **1956**, *65*, 319.
- (17) Beard, L. N.; Lenhart, P. G. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1968**, *24*, 1529.
- (18) Puthenvedu, D.; Janas, T.; Majerfeld, I.; Illangasekare, M.; Yarus, M. *RNA* **2015**, *21*, 1818.
- (19) Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1982**, *162*, 201.
- (20) Kervio, E.; Sosson, M.; Richert, C. *Nucleic Acids Res.* **2016**, *44*, 5504.
- (21) Joyce, G. F.; Inoue, T.; Orgel, L. E. *J. Mol. Biol.* **1984**, *176*, 279.
- (22) Nawrot, B.; Hoffmüller, P.; Sprinzl, M. *Chem. Papers* **1996**, *50*, 151.
- (23) Oró, J.; Basile, B.; Cortes, S.; Shen, C.; Yamrom, T. *Origins Life* **1984**, *14*, 237.
- (24) Singh, S. K.; Nielsen, P.; Koshkin, A. A.; Wengel, J. *Chem. Commun.* **1998**, 455.