Common and Potentially Prebiotic Origin for Precursors of Nucleotide Synthesis and Activation

Albert C. Fahrenbach,[†,§,∇] Constantin Giurgiu,[†,‡,∇] Chun Pong Tam,[†,‡] Li Li,[†] Yayoi Hongo,[§] Masashi Aono,[§,⊥] and Jack W. Szostak*[,†,‡,§]

[†]Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, United States

[‡]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

[§]Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1-IE-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan

[⊥]Faculty of Environment and Information Studies, Keio University, 5322 Endo, Fujisawa, Kanagawa 252-0882, Japan

S *Supporting Information*

ABSTRACT: We have recently shown that 2-amino-imidazole is a superior nucleotide activating group for nonenzymatic RNA copying. Here we describe a prebiotic synthesis of 2-aminoimidazole that shares a common mechanistic pathway with that of 2-aminooxazole, a previously described key intermediate in prebiotic nucleotide synthesis. In the presence of glycolaldehyde, cyanamide, phosphate and ammonium ion, both 2-aminoimidazole and 2-aminooxazole are produced, with higher concentrations of ammonium ion and acidic pH favoring the former. Given a 1:1 mixture of 2-amino-imidazole and 2-aminooxazole, glyceraldehyde preferentially reacts and cyclizes with the latter, forming a mixture of pentose aminooxazolines, and leaving free 2-amino-imidazole available for nucleotide activation. The common synthetic origin of 2-aminoimidazole and 2-aminooxazole and their distinct reactivities are suggestive of a reaction network that could lead to both the synthesis of RNA monomers and to their subsequent chemical activation.

Imidazoles are thought to have played[1] important roles in prebiotic chemistry prior to their current biological significance as components of the histidine residues in proteins, where they function as, for example, charge-relay agents in catalytic triads.[2] One of the most important potential roles of imidazoles in prebiotic chemistry came to light through the efforts of Orgel and co-workers,[3,4] who demonstrated that nucleoside 5′-phosphoro-imidazolides, and especially nucleotides activated with 2-methylimidazole, allow for nonenzymatic template-directed[5,6] RNA synthesis yielding predominantly the canonical 3′-5′ phosphodiester linkage. These phosphoro-imidazolides are more reactive than nucleoside triphosphates as a result of their labile P–N bonds[7] and their propensity to form a reactive imidazolium-bridged[8] dinucleotide. In addition to prebiotic activation chemistry, imidazoles have also been suggested to play catalytic roles,[9] assisting in the oligomerization of amino acids,[10] phosphates[11,12] and nucleotides,[13,14] and a number of plausible mechanisms for the prebiotic synthesis of imidazoles have been reported[1,9] (further discussion in SI).

Scheme 1. Common Prebiotic Synthetic Pathway for 2-Aminooxazole (2NH₂Ox) and 2-Aminoimidazole (2NH₂Im)[a]
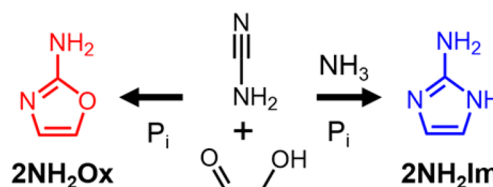
[a]P_i: inorganic phosphate.

Other than prebiotic chemistry, compounds containing imidazole motifs, including 2-aminoimidazole and 2-thioimidazole, are known for their medicinal properties.[15]

We became interested in finding a prebiotic synthesis of 2-aminoimidazole (**2NH₂Im**), because we recently demonstrated[16] that ribonucleoside 5′-monophosphates activated with **2NH₂Im** offer a 10–100-fold enhancement in the rate of nonenzymatic template-directed RNA primer extension compared to those activated with 2-methylimidazole, i.e., in this context **2NH₂Im** is the most effective leaving group known to date. Furthermore, the conjugate base of **2NH₂Im** is isosteric and isoelectronic with 2-aminooxazole[17] (**2NH₂Ox**) − a key intermediate in the prebiotic synthesis of cytidine and uridine 2′,3′-cyclic phosphates, as reported by Sutherland and co-workers.[18] The prebiotic synthesis[18] of **2NH₂Ox** is known, and begins with the addition of glycolaldehyde and cyanamide, followed by a series of steps that are facilitated by inorganic phosphate acting as both a pH buffer and general-base catalyst. In this pathway, cyanamide first undergoes addition to the carbonyl of glycolaldehyde. Next, an intramolecular general-base-catalyzed attack of the glycolaldehyde-derived hydroxyl on the cyanamide-derived nitrile carbon leads to a five-membered ring. General-base catalysis by phosphate also likely accelerates C–H deprotonation in the following dehydration step leading to the aromatic **2NH₂Ox**.
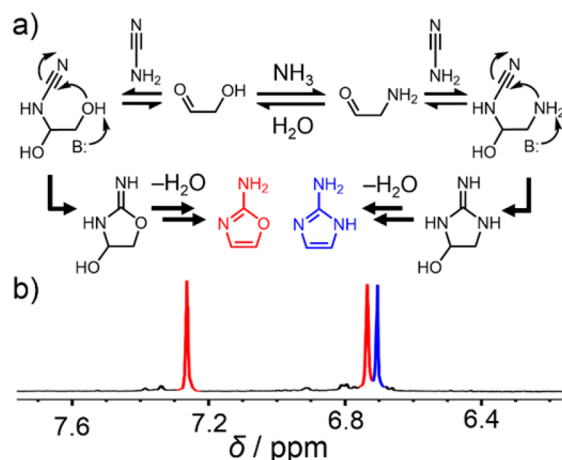
A

Figure 1. (a) Possible mechanism for the synthesis of **2NH$_2$Ox** and **2NH$_2$Im**. The initial formation of α-aminoacetaldehyde is thought to be a potential intermediate on the path to **2NH$_2$Im**, although this exchange reaction may also occur after addition of cyanamide. (b) Partial ^{1}H NMR spectrum (400 MHz, D$_2$O) showing the H4/H5 aromatic resonances for **2NH$_2$Ox** (red) and **2NH$_2$Im** (blue) after reaction of glycolaldehyde, cyanamide, sodium phosphate and ammonium chloride, all at 1 M, for 3 h at pH 7 and 60 °C.

Given their structural similarities, we wondered whether **2NH$_2$Im** and **2NH$_2$Ox** could share a common prebiotic synthetic pathway (Scheme 1). Previously reported (nonprebiotic) syntheses of **2NH$_2$Im** make use of the dimethyl- or diethyl-acetal of α-aminoacetaldehyde,[19−21] which presumably stabilizes the amino group by preventing enolization and subsequent exchange with water. We reasoned that in order to synthesize **2NH$_2$Im**, we would need to find a plausible route for the synthesis of α-aminoacetaldehyde. We hypothesized simply adding a sufficiently high concentration of an ammonium salt to the mixture of glycolaldehyde, cyanamide and inorganic phosphate, would lead to the formation of at least some of this intermediate at equilibrium through an Amadori-type enol-mediated exchange mechanism (Figure 1a, Scheme S1). Alternatively, exchange of the glycolaldehyde-derived hydroxyl to an amine could take place after the addition of cyanamide to glycolaldehyde.

To test this hypothesis, we carried out reactions with aqueous solutions of cyanamide and glycolaldehyde in the presence of 1 M phosphate at varying concentrations of NH$_4$Cl at pH 7, 60 °C for 3 h. In the absence of NH$_4$Cl, analysis of the reaction mixture by ^{1}H NMR spectroscopy revealed that **2NH$_2$Ox** is produced nearly exclusively (Figures S1 and S2), as expected based on previous reports by Sutherland et al.[17,18] The addition of 1 M NH$_4$Cl (Figure 1b) resulted in the appearance of another resonance in the aromatic region of the proton NMR spectrum. Addition of an authentic standard of **2NH$_2$Im** confirmed that this resonance arises from the H4 and H5 protons of **2NH$_2$Im**. Increasing the concentration of NH$_4$Cl to 5 M resulted in the almost exclusive formation of **2NH$_2$Im**. Quantification of the yield was carried out using a calibrated solution of 5′-cytidine monophosphate (CMP), which was added directly to the NMR tube; we used its H5 and H6 resonances as standards for integration. After 1 h, the reaction was complete, and the yield was determined to be 15%. Although this yield is not as high as that previously reported[18] for **2NH$_2$Ox** (>80%), it is important to note that no other major products could be observed in the ^{1}H NMR spectrum.

Understanding how the ratio of **2NH$_2$Im** to **2NH$_2$Ox** varies under different pH regimes and ammonium ion concentrations is important for evaluating the reaction in the context of potential geochemical scenarios. We systematically examined the reaction at pH 4, 5.5, 7 and 8.5 with NH$_4$Cl concentrations of 0, 0.5, 1, 2, 3, 4 and 5 M. All reactions were monitored by ^{1}H NMR after 1, 2 and 3 h, and the ratios of **2NH$_2$Im** to **2NH$_2$Ox** were measured by resonance integration (Table S1). The ratios after 3 h are plotted in Figure 2. For all pH values, the ratio of **2NH$_2$Im** to **2NH$_2$Ox** increases with increasing NH$_4$Cl, whereas mildly acidic pH also tends to increase this ratio. At least part of this effect of pH can be explained by the fact that increasing acidity does not tend to favor dehydration of the 4-hydroxy intermediate of **2NH$_2$Ox**, a minor species that can be detected under the acidic conditions tested. A maximum ratio of [**2NH$_2$Im**]/[**2NH$_2$Ox**] = 56 was obtained at a pH value of 5.5 with 5 M NH$_4$Cl after 3 h. Analysis by Q-TOF LC-MS confirmed these results (Table S2, Figures S3 and S4 for further details). Similar yields for both **2NH$_2$Ox** and **2NH$_2$Im** were obtained when either 100 or 50 mM concentrations of the starting materials were used. We carried out a preparative scale reaction for the synthesis of **2NH$_2$Im** at a pH of ∼5.3 employing 1 M NH$_4$H$_2$PO$_4$ and 5 M NH$_4$HCO$_2$ and obtained a 41% isolated yield (Scheme S2). Given the general mechanistic features of the above synthetic reactions, we anticipated the prebiotic formation of imidazoles may be rather general. We set out to make 2-thioimidazole (Figure S5). Heating a 1 M solution of NH$_4$SCN and glycolaldehyde at pH 4 in 4 M NH$_4$Cl at 60 °C for 24 h led to the formation of 2-thioimidazole with no other major products observable by ^{1}H NMR, although the yield was relatively low at 6.2%.

Knowing that **2NH$_2$Ox** and **2NH$_2$Im** can be made in the same reaction flask in approximately equimolar ratios at pH 7 and ∼1 M NH$_4$Cl, we asked whether these two compounds would display sufficiently different reactivities such that **2NH$_2$Ox** could be channeled toward nucleotide synthesis, whereas **2NH$_2$Im** would be preserved for later nucleoside 5′-phosphate activation. The next step on the pathway from

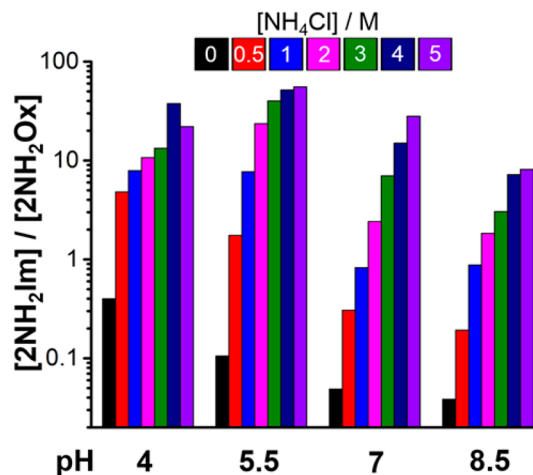**Figure 2.** Bar graph displaying the ratios of **2NH$_2$Im** to **2NH$_2$Ox** at varying pH and NH$_4$Cl concentrations. All ratios were determined by ^{1}H NMR spectroscopy from reactions that were carried out at 60 °C with 1 M sodium phosphate, and monitored over a period of 3 h.
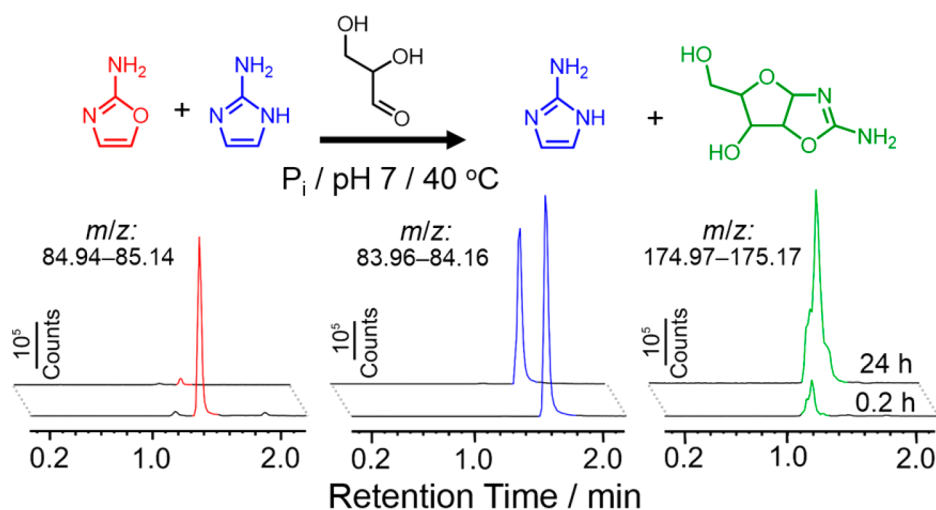
Figure 3. Selective cyclization of *rac*-glyceraldehyde with **2NH$_2$Ox** in the presence of **2NH$_2$Im**. The reaction was carried out at 1 M of each component and was monitored by high-resolution (Q-TOF) LC-MS using a C$_{18}$ column. All traces are extracted ion chromatograms for *m/z* values that correspond to the [M + H]$^+$ ions for **2NH$_2$Ox** (red), **2NH$_2$Im** (blue) and the mixture of aminooxazoline stereoisomers (green). All chromatograms were extracted with a tolerance ±0.1 Da. The 24 h chromatograms have been displayed offset by ∼10 s for clarity.

2NH$_2$Ox to the cytidine and uridine cyclic phosphates involves a cycloaddition reaction with glyceraldehyde, which generates a mixture of ribo- and arabino-furanosyl aminooxazolines[18] as the major products. For **2NH$_2$Im** to be preserved for subsequent nucleotide activation, it would have to be significantly less reactive with glyceraldehyde than **2NH$_2$Ox**. To address this question, we reacted a 1:1 mixture of **2NH$_2$Ox** and **2NH$_2$Im** with glyceraldehyde at 40 °C in the presence of inorganic phosphate for 24 h while monitoring the reaction mixture by high-resolution LC-MS (Figure 3). The mixture of pentose aminooxazolines was detected in the reaction even after only ∼10 min of reaction time, and their concentration reached a maximum after approximately 3 h (Figure S6). After 24 h, nearly all the **2NH$_2$Ox** had been depleted from the reaction mixture, whereas ∼80% of the **2NH$_2$Im** remained. Analysis by ^{1}H NMR spectroscopy confirmed this result (Figure S7). We also detected lesser amounts of a product with an *m/z* of 174.09, a value that is consistent with the [M + H]$^+$ mass of the analogous product but cyclized with **2NH$_2$Im**. Although it would appear that the **2NH$_2$Im** can also react in some fashion with glyceraldehyde, based on the initial rates measured from control experiments reacting **2NH$_2$Im** and **2NH$_2$Ox** with glyceraldehyde individually in separate solutions, we estimate that the reaction with **2NH$_2$Ox** is about 1 order of magnitude faster. Indeed, the reaction of **2NH$_2$Im** with glyceraldehyde is sufficiently slow such that the reaction stops after the consumption of only ∼20% of **2NH$_2$Im**. The likely reason for this effect is that the isomerization of glyceraldehyde to dihydroxyacetone is a competing process occurring at a similar rate, preventing the reaction from going to a completion. We suspect that the slower reaction kinetics of **2NH$_2$Im** compared to **2NH$_2$Ox** are at least in part a result of the greater aromatic stability of **2NH$_2$Im**, the greater nucleophilicity of **2NH$_2$Ox**, or both. We also examined the case of a "one-pot" reaction, in which a solution of 1 M cyanamide, glycolaldehyde, glyceraldehyde, ammonium chloride and sodium phosphate at pH 7 was heated to 40 °C and monitored by LC-MS and NMR over time (Figures S8 and S9). These results provide additional evidence that the greater stability of **2NH$_2$Im** allows it to persist and potentially accumulate for later activation chemistry.

Finally, we demonstrate that *N*-cyano-2-aminoimidazole (**2NH$_2$ImCN**) is capable of activating CMP to furnish (Scheme 2) cytidine-5′-phosphoro-(2-aminoimidazole) (**2NH$_2$ImpC**). Several reports in the literature suggest that *N*-cyanoimidazole can serve[22,23] as an activating agent for the formation of phosphodiester bonds. **2NH$_2$ImCN** appeared attractive as a potentially prebiotic activating agent, because it represents one of the simplest possible extensions of the chemistry reported here, i.e., oxidative coupling with hydrogen cyanide. We formed **2NH$_2$ImCN** by reacting[24] **2NH$_2$Im** with cyanogen bromide in acetone at room temperature for 15 min (Figure S10). The reaction was concentrated to near dryness, at which point an aqueous solution of CMP with 10% D$_2$O was added and the pH adjusted between 5.5 and 6. After about 10 min at room temperature, ^{31}P NMR spectroscopic analysis revealed ∼20% conversion to **2NH$_2$ImpC** (Figure S11 and S12). A maximum of about 40% conversion was obtained after 1.6 h. By addition of another freshly prepared batch of **2NH$_2$ImCN**, a final conversion of ∼75% was achieved. While the prebiotic synthesis of **2NH$_2$ImCN** has not yet been achieved, this synthesis of **2NH$_2$ImpC** serves as a proof of concept, highlighting the potential usefulness of *N*-cyanoimidazoles as prebiotic activating agents.

In summary, we have demonstrated a prebiotically plausible synthetic pathway for 2-aminoimidazole that shares a common origin with the synthesis of 2-aminooxazole. Recently, Powner et al. showed 2-aminothiazole can be efficiently synthesized in a

Scheme 2. Synthesis of Cytidine-5′-phosphoro-(2-aminoimidazole) Making Use of *N*-cyano-2-aminoimidazole
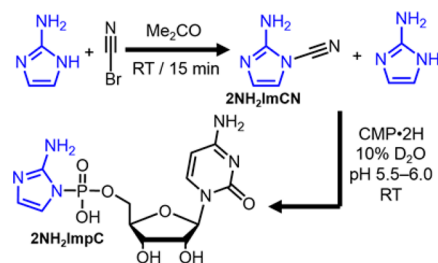
similar fashion as **2NH₂Ox** and **2NH₂Im**, starting from cyanamide and β-mercaptoacetaldehyde.[25] 2-Aminothiazole forms stable crystalline aminals with aldehydes, but not with ketones, allowing for the concomitant accumulation and purification of reactive aldehydes, and the chemical selection of proteinogenic amino acids. Remarkably, all three compounds in the series 2-amino- oxazole/imidazole/thiazole seem to have important potential prebiotic roles. In the shared pathway for **2NH₂Ox** and **2NH₂Im** production, the relative yield of each species depends on the pH and ammonium chloride concentration. At neutral pH, this proposed pathway for the prebiotic synthesis of **2NH₂Im** requires high concentrations of aqueous ammonia, on the order of 1 M to generate comparable amounts of **2NH₂Im** and **2NH₂Ox**. There are several scenarios in which ammonium ions could have been generated in a concentrated form within an ancient aqueous reservoir. In one scenario proposed by Sutherland,[26] cyanide produced in the atmosphere rains out and is captured by ferrous ions as ferrocyanide; salts of ferrocyanide then precipitate and accumulate over long periods of time. Subsequent thermal processing of deposits of magnesium ferrocyanide by magma or impacts would generate magnesium nitride, which upon moistening would hydrolyze, releasing ammonia. Other pathways to ammonia sources have also been suggested (see SI). Finally, the greater stability/slower reactivity of **2NH₂Im** in comparison to **2NH₂Ox** suggests that the former could accumulate over time even as the latter is continuously processed into intermediates on the path to nucleotide synthesis. Having a mechanism for the simultaneous production of **2NH₂Im** and **2NH₂Ox** suggests the tantalizing prospect of an ancient prebiotic reaction network that could have led to both nucleotide synthesis and to the subsequent chemical activation of those nucleotides in a manner suitable for efficient nonenzymatic template-directed replication. Although we have shown that *N*-cyano-2-aminoimidazole can serve as an activating agent, one of the great challenges ahead is understanding how such an activating agent, or another mechanism altogether for the prebiotic activation of nucleoside 5′-monophosphates, could have arisen from such a network.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b01562.

  Experimental details (PDF)

AUTHOR INFORMATION

Corresponding Author
*szostak@molbio.mgh.harvard.edu

ORCID
Albert C. Fahrenbach: 0000-0002-8315-8836
Chun Pong Tam: 0000-0001-6381-9011
Jack W. Szostak: 0000-0003-4131-1203

Author Contributions
▽These authors contributed equally.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Vázquez-Salazar, A.; Tan, G.; Stockton, A.; Fani, R.; Becerra, A.; Lazcano, A. *Origins Life Evol. Biospheres* **2016**, DOI: 10.1007/s11084-016-9525-y.
(2) Dodson, G.; Wlodawer, A. *Trends Biochem. Sci.* **1998**, 23, 347.
(3) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, 114, 5496.
(4) Blain, J. C.; Szostak, J. W. *Annu. Rev. Biochem.* **2014**, 83, 615.
(5) Fahrenbach, A. C. *Pure Appl. Chem.* **2015**, 87, 205.
(6) 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.
(7) Li, L.; Lelyveld, V. S.; Prywes, N.; Szostak, J. W. *J. Am. Chem. Soc.* **2016**, 138, 3986.
(8) Walton, T.; Szostak, J. W. *J. Am. Chem. Soc.* **2016**, 138, 11996.
(9) Oró, J.; Basile, B.; Cortes, S.; Shen, C.; Yamrom, T. *Origins Life* **1984**, 14, 237.
(10) Lohrmann, R.; Orgel, L. E. *Nature* **1973**, 244, 418.
(11) Weber, A. L. *J. Mol. Evol.* **1981**, 18, 24.
(12) Keefe, A. D.; Miller, S. L. *Origins Life Evol. Biospheres* **1996**, 26, 15.
(13) Pongs, O.; Ts'o, P. O. P. *J. Am. Chem. Soc.* **1971**, 93, 5241.
(14) Oró, J.; Stephen-Sherwood, E. *Origins Life* **1974**, 5, 159.
(15) Zhang, L.; Peng, X.-M.; Damu, G. L. V.; Geng, R.-X.; Zhou, C.-H. *Med. Res. Rev.* **2014**, 34, 340.
(16) Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. *J. Am. Chem. Soc.* **2017**, 139, 1810.
(17) Anastasi, C.; Crowe, M. A.; Powner, M. W.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2006**, 45, 6176.
(18) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, 459, 239.
(19) Lawson, A. *J. Chem. Soc.* **1956**, 307.
(20) Lancini, G. C.; Lazzari, E. *J. Het. Chem.* **1966**, 3, 152.
(21) Weinmann, H.; Harre, M.; Koenig, K.; Merten, E.; Tilstam, U. *Tetrahedron Lett.* **2002**, 43, 593.
(22) Ferris, J. P.; Huang, C.-H.; Hagan, W. J., Jr. *Nucleosides Nucleotides* **1989**, 8, 407.
(23) Chen, H.; Du, F.; Chen, G.; Streckenbach, F.; Yasmeen, A.; Zhao, Y.; Tang, Z. *Sci. Rep.* **2014**, 4, 4595.
(24) McCallum, P. B. W.; Weavers, R. T.; Grimmett, M. R.; Blackman, A. G. *Aust. J. Chem.* **1999**, 52, 159.
(25) Islam, S.; Bučar, D.-K.; Powner, M. W. *Nat. Chem.* **2017**, 9, 584.
(26) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. *Nat. Chem.* **2015**, 7, 301.