RNA polymerisation without catalyst from 2′,3′-cyclic nucleotides by drying at air-water interfaces


[A] A V, Dass; S, Wunnava; J, Langlais, M, Krusche; L, Ufer; N, Chrisam; C F, Dirscherl, D, Braun
Faculty of Physics, Systems Biophysics
Ludwig-Maximilians-Universität München
AmalienStr. 54, 80799, Munich, Germany
E-mail: dieter.braun@lmu.de
[B] B, von der Esch; C, Ochsenfeld
Chair of Theoretical Chemistry, Department of Chemistry,
Ludwig-Maximilians-Universität München
Butenandtstraße 5-13, 81377, Munich, Germany
[C] C, Ochsenfeld
Max Planck Institute for Solid State Research, Heisenbergstr. 1, D-70569 Stuttgart, Germany
[D] R C A, Dubini; P, Rovó
Faculty of Chemistry and Pharmacy,
Ludwig-Maximilians-Universität München
Butenandtstraße 5-13, 81377, Munich, Germany
[E] P, Rovó (Current address)
Institute of Science and Technology Austria
c/o NMR Facility, Am Campus 1, A-3400 Klosterneuburg, Austria
[F] F, Gartner; S, Angerpointner; E, Frey
Faculty of Physics, Statistical and biological physics,
Ludwig-Maximilians-Universität München,
Theresienstraße. 37, D-80333, Munich, Germany
[G] J, Špöner
Institute of Biophysics Academy of Sciences of the Czech Republic
Kolowrotská 135, 61265 Brno, Czech Republic

Supporting information for this article is given via a link at the end of the document.

Abstract: For the emergence of life, the abiotic synthesis of RNA from its monomers is a central step. We found alkaline, uncatalysed drying conditions in bulk and at heated air-water interfaces where 2′,3′-cyclic nucleotides polymerised, forming up to 10-mers within a day. The polymerisation proceeded at a pH range of 7-12 at temperatures between 40-80 °C and was marginally enhanced by K+ ions. Among the canonical ribonucleotides, cGMP polymerised most efficiently. Quantification was performed using HPLC coupled to ESI-TOF by fitting the isotope distribution to the mass spectra. Our study suggests a polymerisation mechanism where cGMP aids the incorporation of the relatively unreactive nucleotides C, A and U. The 2′,3′-cyclic nucleotides are byproducts of prebiotic phosphorylation, nucleotide syntheses and RNA hydrolysis, indicating direct recycling pathways. The simple reaction condition offers a plausible entry point for RNA to the evolution of life on early Earth.

Introduction

The central and multifunctional role of RNA within biology points towards RNA as a chief informational biopolymer for the onset of molecular evolution.[11] Polymerisation involving more than a single type of canonical nucleotide, generating a varied pool of RNA strands, has not been achieved under aqueous conditions.[2-4] Chemical activation strategies to trigger RNA polymerisation[3,7,8] and template-directed primer extension of sequences.[9,10] In the earliest self-replicating systems, the formation of complimentary strands for replication and transfer of genetic information by non-enzymatic processes is believed to be important and homopolymers are not considered very useful as genes.[11] Short RNA strands, especially from dimers[11] to tetramers[12,13] have been shown to enhance the copying of mixed-sequence templates in comparison to monomers. Thus, it is necessary to have a polymerisation mechanism that is able to generate short mixed-sequences that later function as primers and templates for copying of sequences.

We base this study on 2′,3′-cyclic mononucleotides (cNMP) which (a) possess an intrinsically activated phosphate; (b) are products of several prebiotic phosphorylation and nucleotide syntheses,[14-18] and (c) are products of neutral to alkaline chemical and enzymatic hydrolyses of RNA.[19-23] In comparison, the dry polymerisation of 3′,5′-cGMP[24-26] did not foster the polymerisation of the other ribonucleotides.[26] Orgel and coworkers, reported conditions for 2′,3′-cAMP polymerisation by drying for 40 days with a 5-fold excess of ethane-1,2-diamine and yields up to 0.67% of 14-mers.[4,6] Other catalysts such as imidazole or urea required temperatures up to 85°C and offered lower yields.[4,6]

We found that 2′,3′-cGMP polymerised spontaneously under alkaline (pH 7-12), uncatalysed drying conditions, (40-80 °C) within a day. The other canonical cNMP were relatively inert under similar conditions. Our observations of cAMP and cUMP forming up to trimers are consistent with literature.[27,28] The polymerisation is demonstrated in the presence of bulk water at the air-water interface, within a microfluidic thermal chamber. The chamber mimics conditions of a heated, water filled volcanic rock pore that includes a gas bubble.
In a polymerisation mixture of cGMP, we observed polymers rich in G nucleotides, but with C, A and U incorporated at lower concentrations. Computational and modelling results suggest that the polymerised oligomers of cGMP form a self-assembled scaffold in the dry state, which could incorporate the nucleotides C, A and U to form short mixed-sequence oligomers.

Results

Polymerisation of cGMP

An aqueous solution of the sodium salt of 2’,3’-cGMP (20 mM) was dried for 18 hours at 40 °C in the presence of 40 mM KCl. Since the monomers are monosodium forms, they have an equal concentration of Na+ ions when in solution (20 mM). All the reported concentrations throughout the article are calculated for a volume of 100 µL. The total concentration of each n-mer (oligomer) is a sum of the linear-phosphate (-P) and the cyclic-phosphate (-cP) on the n-mer terminus. Both endpoints are well discriminated by HPLC as the n-mer-cP is eluted before a n-mer-P of the same length (S2d). Typically, about 90% of the n-mers consisted of -P endings (S5d). The non-covalent stacked n-mers (eg. two 4-mers) are discriminated from covalent n-mers (eg. an 8-mer) due to the higher mass of the stacked n-mers by one H2O and the corresponding HPLC retention times of n-mers under denaturing HPLC conditions.[29-31]

The denaturing conditions of the HPLC column at 60°C efficiently separated a mixture of synthetic polyG n-mers without signs of aggregation, as shown in Figure 1c. It must be noted that n-mer-cP and a cyclised n-mer of the same length would have the same mass, but are unlikely not to be discriminated by the HPLC retention times. The presence of n-mer-cP is established from the 31P NMR peak at ~20 ppm in Figure 1d. Oligomers from 2 to 15mers (S8a) were detected by HPLC-MS (S8a) for cGMP polymerisation. For quantification, only 2 to 10-mers were considered throughout the study.

The error bars can be estimated based on plots of cGMP polymerisation (5 replicates) in S8a, with a mean standard deviation of 2.95 µM between independent runs of the experiment. The error bars are not indicated in the figures as they would appear insignificant on the log scale. For quantification, the HPLC retention times of the oligomer standards of G were first optimised on an RP C-18 HPLC column coupled to ESI-MS. Figure 1c shows the HPLC chromatogram of 2 to 10-mers for polyG standards (with 1 equiv. of KCI) with their respective retention times. We found efficient separation and no evidence for the formation of aggregates. The ion counts of the n-mer with their HPLC retention times are shown in S2b. By comparing the ion counts, we confirmed the high efficiency of the post-polymerisation ethanol precipitation protocol in S3, confirming negligible influence. However, the precipitation was used to remove excess monomers which otherwise saturated the HPLC column, yielding a robust method for the quantification of the complex oligonucleotide mixtures (S2c).

The calculated isotope probabilities of the n-mers in the various charge states were fitted to the raw mass spectra using a self-written LabView program. This allowed us to identify salt adducts formed in the mass spectrometer and to fit overlapping isotope patterns. The retention times of the PolyG standards were used to obtain time-brackets to average the mass spectra. Further details on the calibration used for the quantification within the program and the functional modes of the program are elaborated in S1-S6. Based on preliminary enzymatic digestion experiments, we estimated that the formed G polymers were linked by 2’-5’ and 3’-5’ phosphodiester linkages in about 1:1 ratio (S21-S23). The linkage type in the polymerisation was also confirmed by 31P NMR (Figure 1d) and the peaks were assigned based on the literature values.[32,33]

Figure 1a and 1b compare the effect of pH on the lengths and concentrations of the n-mers formed by drying with KCl (Cl) and imidazole respectively. We determined the optimal reaction temperature to be 40 °C (S5, S8b). Imidazole and its derivatives are used in the literature as nucleotide activation agents for templated primer extension reactions,[9] as a buffering agent and a catalyst for polymerisation.[4] The addition of imidazole did not enhance the length and concentration of n-mers in comparison to polymerisation with KCl.

Figure 1. Polymerisation of Guanosine-2’,3’-cyclic monophosphate (cGMP.Na). A 20 mM cGMP.Na solution was heat-dried with 40 mM KCl at 40 °C for 18 hours, under ambient pressure in 100 µL volume. (a) Polymerisation was screened over a range of pH 3-12. The reported concentrations were the sum of terminal cyclic (cP) and linear phosphate (P). Oligomers without phosphates were not detected. Polymerisation was optimal at pH 10 with total oligomer yields of ~3.5% (inset). The solid line shows results of the polymerisation model based on stacked assembly (S1b). (b) pH screen with 100 mM imidazole under similar conditions. No significant increase in polymerisation was found by adding the organic catalyst. (c) Diode array detector (DAD) absorbance at 260nm for 50µM polyG standards (-P endings) and 100µM KCl used for confirming HPLC separation and the determination of retention time for quantification with ion counts. (d) 31P proton-decoupled NMR spectrum (10% D2O, pH 10), of polymerised G sample; the signals corresponding to phosphodiester linkages for both 3’-5’ and 2’-5’ are between -0.8 to -1.1 ppm.
Polymerisation from cNMP

We also tested the polymerisation tendencies of cAMP, cUMP, and cCMP under the same heat-drying conditions and found that these monomers did not polymerise to the same lengths and concentrations as cGMP. Figure 2a shows that the polymerisation trend decreases in the order cGMP > cUMP > cAMP > cCMP. The dominance of G-polymerisation prompted us to investigate the copolymerisation of these moderately reactive mononucleotides under the influence of the well polymerising cGMP. We found that a mixture of two or four different monomers was capable of generating mixed sequence polymers, where the majority of the mixed polymers were rich in G. We probed if the polymerisation of a G and C mixture could reach levels where hybridisation between strands could be possible. Thus, we polymerised a binary mixture of cGMP and cCMP (20 mM each), under heat-drying conditions (40 °C) in the presence of 40 mM KCl. Comparing quantities of G2 in Figure 2a and 2b, the concentration of C2 is enhanced 2 fold and C2 became detectable; besides the fact that mixed GC oligomers are formed (Figure 2b). The detailed sequence composition for GC mixed polymerisation is seen in Figure 2e, showing that the G2 to G3 contribute to the bulk of the oligomers formed in the polymerisation mixture. Up to two C’s were incorporated into polymers ≤4-mers, one C is incorporated into 5-mers and none were detectable beyond them. A similar analysis of GA and GU binary mixtures is available in S9a, b.

GC mixed polymerisation was favoured at temperatures ranging from 40 °C to 80 °C (Figure 2c), similar to cGMP (S8b, S8b), It must be noted that in reactions at 30 °C for 18 hours, the drying was incomplete within the polypropylene tubes used for the experiment and the reaction kinetics in the dry state was reduced. Higher temperatures on the other hand possibly contributed to the degradation of the monomers (S4c) and the formed oligomers as seen in the trace comparisons under 80, 60 and 40 °C in Figure 2c.

Specific cations also influenced cNMP polymerisation. We found that K+ ions yielded higher concentrations and lengths of the oligomers in comparison to Na+ ions at the same concentrations. The presence of Mg2+ ions in the reaction mixture inhibited polymerisation (Figure 2d). The dependence of polymerisation on K+, Na+ and Mg2+ salt concentrations is shown in S10, indicating that 1-3 equiv. of the same cation display similar results, but the type of cation affected the efficiency of polymerisation.

Polymerisation of cNMP in a heated rock pore mimic

Wet-drying cycling in surface-based geological settings are subjected to a drift in salt and pH conditions due to the imbalance caused by the evaporation of pure water and the rehydration of
the fluid that contains salt. But wet-dry cycling can also be in a closed chamber subjected to a temperature difference. The water that is evaporated on the warm side re-enters the fluid on its cold side. Wet-dry cycling is a result of interface shifts and the dew droplet dynamics on the cold side. Thus, offering wet-dry cycling under constant pH and salt conditions. The geological analogues of such a setting would be volcanic rock pores which are partially filled with fluids and are subjected to a thermal gradient. We have previously reported prebiotically important processes such as accumulation, phosphorylation, encapsulation, gelation, strand separation, enzymatic DNA replication and crystallisation within such settings.\textsuperscript{34–36}

For the polymerisation within this setting, we started with 20 mM total monomers (5 mM each of cG, cC, cA and cU). After the chamber was loaded with the monomer solution, a thermal gradient was applied which drove a continuous wet-dry cycle just above the air-water interface inside the chamber (Figure 3a). Over time, the meniscus of the bulk liquid receded in an oscillatory manner depending on how many dew droplets formed above the interface; and dried material precipitated on the warm side as a consequence (Figure 3b). The dew droplets grew at the cooler side of the chamber and upon contact with the warm side, the dew rehydrated the dried material and transports it back into the bulk mostly due to bubble fusion driven by surface tension. This phenomenon was allowed to continue for 18 hours, after which the setup was dismounted and the remaining bulk liquid and the dried flakes (after dissolution) were sampled for analysis.

The pH of the samples at the end the reaction was found to be lowered by a pH unit, indicating the formation of acidic species in the reaction mixture. A likely cause of the pH drop is the acidification by the ring opening of the cyclic phosphate in the mononucleotides and the oligomers (S4e, S5c, d). The pH drop was 1.5 to 2 pH units if polymerisation was performed at higher temperatures (S4e).

Despite the presence of bulk water; the polymerisation inside the simulated volcanic rock pore showed comparable yields as that of the heat-dried conditions. Therefore, the sustained polymerisation indicates that the heated interface can access conditions favourable for polymerisation similar to bulk dried polymerisation conditions. The constant feeding of monomers from the bulk fluid could also be an important factor. A length-selective enzymatic DNA replication was reported recently within this setting, indicating the possible continuity of prebiotic chemistry in such a setting.\textsuperscript{35}

We observed all the dimer sequence combinations and most of the trimers (Figure 3c). However, the tetramers and pentamers show sequences rich in G. The length selectivity of the HPLC allowed the detection of longer sequences, but the isotopic fit to the raw mass spectra provided by our LabView-based analysis showed that longer species with concentrations lower than 0.2 \mu M were lost in the background noise of the mass spectra. Moreover, different oligomers can have similar masses (eg. Table S3 and S4), so to avoid calling false positives, sequences with mass overlaps were not included here. This is in addition to the rigid selection criteria, based on fitting of the isotopic distribution (S12) and only considering mass spectra within the optimised n-mer retention times of the HPLC. A full sequence composition analysis for GC and GCAU mixtures with comparison between dry polymerisation and simulated rock-pore polymerisation is provided in S11. In comparison, CAU reaction

Figure 4. Possible supramolecular assemblies facilitating polymerisation of cNMP. Molecular dynamics simulations suggest polymerisation of A, U and C by intercalating into stacks of polyG. (a) Distances between bases in the complex of unpolymerised nucleotides show that polymerisation is disfavoured due to drifting away of the complex. (b) The assembly formed when the bases are templated by covalently linked G-G (2′-5′) dimers, forms the most stable complexes and make possible the incorporation of the G, C, A, and U within the n-mers observed in our experiments. (c) Snapshot after energy minimization of stacking interaction between a polymerised scaffold of cGMP (G-G) to template the cGMP and cUMP monomers (U,G). The dotted lines mark distances between the bases used to evaluate the stability of the complex. (d) However, the self-polymerisation of polyG could also be based on stacks of G-tetrads, stabilised by inner K⁺ ions, coinciding with the promotion of polyG polymerisation by K⁺ ions. R denotes the ribose of RNA.

mixture yielded only dimers (S9c), indicating again the central role of G in the copolymerisation process.

Computational study of the proposed intercalated stacked arrangement

Based on the hypothesis that a stack-assisted geometry is triggering the polymerisation of 3′,5′ cGMP,\textsuperscript{36} we studied the suitability of intercalated stack arrangements for the polymerisation of 2′,3′-cNMP. We explored the stability of the stack arrangements, and the incorporation of cNMP monomers into polymerised cGMP scaffold, based on minimum energy structures and molecular dynamics simulations (Figure 4a-c and the computation section in SI).

To investigate the suggested intercalated stack arrangement for several possible species, we have computed the stacking interaction energies and evaluated the minimum energy geometries obtained at wb97M-V/def2-TZVPD level of theory.\textsuperscript{37–40} All systems were studied in the gas phase as well as with implicit solvation (C-PCM).\textsuperscript{41} The quantum mechanical computations were performed using FermiONs++\textsuperscript{42–44} in combination with Chemshell.\textsuperscript{45}

These were complemented by GFN-FF molecular dynamics simulations,\textsuperscript{46} for the systems encapsulated in an explicit water
sphere using xtb.[47] The stacking of homogeneous monomers were tested (N/N / N,N) with N= A, U, G, C and the incorporation of monomers into a dimer and trimer scaffold of G was probed (N,N / G-G or N,N/ G-G-G). The 3'-5' linked G-G and G-G-G accommodate A, U and G monomers into the scaffold providing a stable arrangement for the initiation of polymerisation. For C an alternate arrangement involving hydrogen bonding with a G within the scaffold is observed (S28). We found that a 2'-5'polyG scaffold seemed to enhance the alignment (Figure 4a, b, c), confirmed both by static and dynamic computations (S30, S31).

**Theoretical model of cGMP polymerisation**

Additional evidence supporting a stacked polymerisation mechanism comes from the observed non-exponential length distribution of the oligomer concentrations. This supports the idea that the formation of dimers is the rate limiting step: the concentration drop from monomers to dimers was most significant. For the cGMP polymerisation in Figure 1a, the 20 mM monomer concentrations drop to 0.15 mM for G2, then forming a flat concentration plateau, in contrast to the typical exponential length distribution in homogeneous polymerisation.[48]

To test this idea, we fit the concentration distribution of G homopolymers with a stacked polymerisation model (solid line Figure 1a and 2a). The model assumed a three-step polymerisation reaction: i) a monomer of length i and a polyG scaffold k can stack together with rate ν, ii) the de-stacking rate δi,k decreases exponentially with the number of stacked bases ni,k, iii) another monomer of length j can stack to the complex. If the stacks persist long enough, the polymerisation reaction ligates the two monomers with rate ρ (see for details S18-S19). The model fits the experimental data, suggesting a rate limiting step for the formation of short polymers due to the required mutual alignment. It should be noted that it is difficult to distinguish between inter-base stacking or a plausible G-tetrad arrangement suggested based on the enhanced polymerisation observed with K+ (Figure 4d).

**Discussion**

Our data suggests that cGMP polymerises in dry state at moderate temperatures and pH. The polymerisation occurs over a range of temperatures (40-80 °C) and pH (7-12) and does not require additional catalysts. This makes RNA formation significantly more robust and probable under early Earth models.[49,50] We also showed polymerisation in the wet-dry cycling environment at a heated air-water interface, adding RNA polymerisation to the pool of prebiotic processes possible within such a setting.[34-36] Under the tested conditions, the reaction yielded polymers up to 15-mers. The formed polyG incorporated cCMP, cAMP and cUMP monomers, albeit in lower concentration, which did not homopolymerise significantly.

An important feature of this polymerisation is that the 2',3'-cyclic phosphate group is sufficient to trigger polymerisation without external activation mechanism or catalysts under low salt conditions. Low salt conditions are interesting for RNA evolution since they notably help strand separation and reduce RNA degradation.[55]

cNMP polymerisation is found to be a relatively clean reaction under the conditions tested. In comparison, in situ EDC activation yields side products, especially at high temperatures.[51] We did not detect any major side products with ESI-TOF, other than the salt adducts of sodium and potassium.

The abiotic formation and recyclability of the cNMP monomers is feasible, as they are known to be produced under several phosphorylation conditions, nucleotide syntheses and are common degradation products of RNA.[19-23] Thus, a cycle of reactions involving polymerisation, oligomer extension, polymer hydrolysis, reactivation of monomers under early Earth conditions becomes conceivable. Furthermore, recombination and templated ligation involving 2',3'-cyclic ending oligomers[52] have been observed.

For our studies, we compared two monovalent ions (K+, Na+) and one divalent cation (Mg2+); they were chosen for their relevance in contemporary cytosolic media, their abundance on the early Earth[53] and for the role of Mg2+ in ribozyme activity.[64] Polymerisation is enhanced in the presence of K+ in comparison to Na+ ions. The inhibition by Mg2+ ions possibly occurs by a combination of base catalysis mechanism, the deactivation of -cP ends of the reactant, products and enhanced oligomer hydrolysis. Despite its role in ribozyme functionality, at high concentrations Mg2+ inhibits RNA replication by creating strong RNA duplexes, limiting thermal denaturation and enhancing temperature dependent hydrolysis.[55] It is also known that the presence of ~1.5 mM Mg2+ is sufficient to inhibit the membrane self-assembly of fatty acids and this has been considered an incompatible aspect for the co-emergence of RNA and fatty acid membranes.[56,57] However, under the discussed reaction conditions of cNMP polymerisation, RNA formation and encapsulation with fatty acids might be conceivable within freshwater locations on the primordial Earth. Moreover, we have shown that efficient strand separation can be achieved by low sodium concentrations, triggered by microscale water cycles within heated rock pores.[35,36]

Our very preliminary digestion studies suggest a considerable backbone heterogeneity (2'-5' and 3'-5') within the polymers. However, a full quantitative treatment is beyond the scope of this study. It has been demonstrated that the presence of 2'-5' linkages allow efficient strand separation by reducing the melting temperature (Tm) of oligomers, which is pertinent in the case of G-rich sequences that are observed in this polymerisation.[58] Lowering of Tm is critical to replication of sequences.[58,59] These studies also show that the presence of 2'-5' linkages allow the folding of RNA into three-dimensional structures, similar to native linkages and do not hinder the evolution of functional RNAs, such as ribozymes. The susceptibility towards enhanced hydrolysis of the 2'-5' over the 3'-5' linkages could select the latter in wet-dry cycling conditions, similar to the reported backbone selection of RNA and DNA.[58,60,61]

Mechanistically, molecular dynamics studies indicated that cGMP polymerisation could be due to the formation of intercalated stacks of cGMP as a consequence of hydrophobic interactions between the guanine bases. On attaining a stable intermolecular arrangement, the 5'-OH of a nucleotide can attack the cyclic phosphate of the neighbouring nucleotide. This could allow the formation of polymeric G-scaffolds (Figure 4c, d, S25). However, the formation of tetrad stacks over one another with a central K+ ion between the stacks could also promote polymerisation (Figure 4d).

The notion of multi-molecular assemblies is supported by the presence of slow-diffusing species observed in 1H, 31P
Acknowledgements

We would like to thank Ulrich Gerland, Tobias Göppl, Joachim Rosenberger and Bernhard Altaner for their helpful remarks and discussions; Thomas Mateux, Alexandra Kühnlein, Noël Yeh Martin and Maximilian Weingart for comments on the manuscript. The authors thank J. Kussmann (LMU Munich) for providing a development version of the FermiONs++ program package.

Keywords: RNA • Polymisation • 2’,3’-cyclic nucleotides • Non-equilibrium • Air-water interfaces

References
