Cross-backbone templating; ribodinucleotides made on poly(C)

IRENE MAJERFELD, DEEPA PUTHENVEDU, and MICHAEL YARUS

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347, USA

ABSTRACT

G^{5'}pp^{5'}G synthesis from pG and chemically activated 2MeImpG is accelerated by the addition of complementary poly(C), but affected only slightly by poly(G) and not at all by poly(U) and poly(A). This suggests that 3'–5' poly(C) is a template for uncatalyzed synthesis of 5'–5' GppG, as was poly(U) for AppA synthesis, previously. The reaction occurs at 50 mM mono- and divalent ion concentrations, at moderate temperatures, and near pH 7. The reactive complex at the site of enhanced synthesis of 5'–5' GppG seems to contain a single pG, a single phosphate-activated nucleotide 2MeImpG, and a single strand of poly(C). Most likely this structure is base-paired, as the poly(C)-enhanced reaction is completely disrupted between 30 and 37°C, whereas slower, untemplated synthesis of GppG accelerates. More specifically, the reactive center acts as would be expected for short, isolated G nucleotide stacks expanded and ordered by added poly(C). For example, poly(C)-mediated GppG production is very nonlinear in overall nucleotide concentration. Uncatalyzed NppN synthesis is now known for two polymers and their complementary free nucleotides. These data suggest that varied, simple, primordial 3'–5' RNA sequences could express a specific chemical phenotype by encoding synthesis of complementary, reactive, coenzyme-like 5'–5' ribodinucleotides.

Keywords: cofactor; coenzyme; RNA gene; template; phenotype

INTRODUCTION

Chemically capable RNAs that reproduce are candidates for a role in life's origin. Computations demonstrate that small oligoribonucleotides in an aqueous, sporadically fed pool can carry out reactions required to replicate and possibly show a chemical phenotype (Yarus 2012). This remains true even if chemical supplies are undependable, lifetime of the pool is limited, and all nucleic acids themselves are unstable. This contrasts significantly with the parallel difficulties that appear for larger RNAs (Szostak 2012). Moreover, sporadic, unguided oligonucleotide synthesis displays unexpected aptitudes; for example, it can repetitively execute a complex reaction sequence (Yarus 2013). In addition, chemically capable RNAs can be very small. A pentaribonucleotide ribozyme suffices to perform highly regiospecific aminoacyl transfer (Chumachenko et al. 2009; Yarus 2011b). RNAs as small as two nucleotides have chemical phenotypes, even today (Yarus 2011a). Accordingly, we have searched for oligonucleotides that might have simple means of propagation, as well as intrinsic chemical competence.

Previously, we showed (Puthenvedu et al. 2015) that cross-backbone templated (or cross-templated) synthesis of the coenzyme congener A^{5'}pp^{5'}A by poly(U) is an easily observed reaction:

where the 2-methylimidazole phosphate activating group (Lohrmann et al. 1980) is indicated, 2MeIm. Chemical formation of AppA without template instruction also occurs as a slower, constantly observed background:

At millimolar concentrations of poly(U) phosphate, only one strand of poly(U) appeared to be required to form the template complex, and AppA was by far the prevalent product detected containing [³²P]pA. AppA synthesis likely proceeds via intermediate stacks of pA and 2MeImpA. Such polymer-bound stacks can be both more abundant and more favorable for reaction than free stacks, thus potentially accounting for stimulation by poly(U).

To generalize AppA synthesis, we have examined a parallel reaction using pG nucleotides in the presence of poly(C).

Corresponding author: yarus@colorado.edu

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.054866.115>. Freely available online through the RNA Open Access option.

© 2016 Majerfeld et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

This is independently interesting because of the excellent stacking of G nucleotides (Ts'o 1969; Freier *et al.* 1986). To make such experiments possible, we have inhibited the formation of extended G-nucleotide structures based on pG quadruplexes (Gellert *et al.* 1962; Wong *et al.* 2005) by using lithium as the major monovalent ion. Li^+ cannot effectively act as the central ion supporting G tetrad superstructures (Pinnavaia *et al.* 1978). In these Li^+ solutions, there are apparent differences of a few-fold between the systems, but poly(C) stimulates GppG synthesis, probably again via a base-paired, stacked complex. This second reaction strengthens the possibility (see Discussion section) that primordial genes might be as simple as pyrimidine homopolymer tracts. Such elementary sequences might have phenotypes based on cross-templated complementary 5'-5' purine ribodinucleotides decorated with reactive groups. Such synthesis could initiate a chemical pathway that has descended to modern co-enzymes (White 1976; Yarus 2011a; see Discussion below).

RESULTS

Resolution of reaction products

Figure 1 shows our usual TLC resolution of isotopic [^{32}P]pG products after incubations containing 2MeImpG (guanosine 5'-phospho-2-methyl imidazolid) and [^{32}P]pG, with and without poly(C) for 24 h at 12°C. As is clear from the distribution of radioactivity, the major pG-containing product observed is GppG, the 5'-5' linked dinucleotide of GMP. Synthesized GppG is markedly increased when poly(C) is present. This phosphorimaged TLC is accurate in suggesting that the major product is GppG; though not displayed, more highly resolving HPLC finds only a small, very poly(C) dependent, 3'-5' pGpG dinucleotide product. Thus, in both polymer-instructed and -uninstructed reactions, specific production of GppG in preference to other chemically possible dinucleotides is notable.

The origin spot at the right appears when poly(C) is in the reactions, and not otherwise. It is likely caused by nucleotide adherence to dried poly(C). This spot (~2% of cpm) does not increase regularly with the time of incubation. It is therefore included in total cpm, but not analyzed as a product.

Nucleotide concentrations

Figure 2 shows the initial rate of synthesis at 12°C (in Molar GppG/h) as total substrate nucleotide increases, while maintaining pG and 2MeImpG at equal concentrations. Thus, the abscissa at 0.002 M means 2 mM pG plus 2 mM 2MeImpG. The two curves represent synthesis with 5 mM template poly(C) phosphate (triangles, above) and without added polymer (circles, below). Both rates are concave upward; that is, the rate of GppG synthesis increases more-than-linearly with increasing precursor supplies. We show later that this is as expected when stacks of G provide the reactive centers. Even

FIGURE 1. Cellulose TLC resolution of standard reactions. 0, zero time, no incubation. 24, after 24 h at 12°C. 24 h poly(C), 24 h incubation with 5 mM poly(C). 3'-5' pGpG marker was from Thermo Fisher Scientific.

without further analysis, these data show that 5 mM poly(C) in these standard reactions is not saturated (fully paired with) G nucleotide stacks. If saturation of the poly(C) template were near, the upper poly(C) curve would be expected to bend to the near-horizontal, and assume the slope characteristic of polymer-free synthesis (as documented by the lower curve), because added nucleotides would necessarily form polymer-free stacks. In particular, the Discussion section shows that this curvature of the free stacked G and polymer-bound stacks is expected at low, nonsaturating nucleotide concentrations.

FIGURE 2. Initial rate of GppG synthesis with variation in nucleotide substrate concentration. Triangles, with poly(C); circles, without polymer. Rates are those from least squares slopes from lines fitted to GppG measured six times in the initial 8 h at 12°C.

Effects of poly(C) concentration

Figure 3 shows the response of the initial velocity of GppG production to poly(C) concentration, where concentration is shown as polymer nucleotide phosphate. Such data reproducibly show a linear increase in rate, followed at higher poly(C) concentrations by plateau and then decrease in rate. Fitting to the kinetic data suggests that the rate of templated synthesis is the source of the decline at high poly(C). It seems likely that this departure from linearity is due to polymer self-structure or interaction with G nucleotide, and resulting lowered availability of the active poly(C)–G nucleotide binding sites at higher polymer concentrations. Notably, below ~5 mM the polymer phosphate rate is approximately linear and therefore resembles that of AppA synthesis on poly(U) (Puthenvedu et al. 2015). Therefore, it is likely that the transition state for GppG synthesis contains one molecule of poly(C) template in the linear region (compare Fig. 2), and this is implemented in the full model when used for fitting data at lower poly(C) concentrations.

$$\frac{d[\text{GppG}]}{dt} = k_s(\text{poly(C)})(pG)(2\text{MeImpG}).$$

Note that the rate “constant” k_s is a convenience for discussion, and would have a different value at different nucleotide concentrations (Fig. 2, compare Discussion section). In particular, synthesis proportionate to poly(C) template concentration in standard reactions (Fig. 3, next below) implies synthesis in widely separated poly(C)-bound stacks of G, with intervening unpaired C sites in the majority. These active stacks apparently form proportionately to the total template space provided.

Response to varying ratios of nucleotides

The low concentration data above suggest that there is a single poly(C) molecule in the transition state for the GppG synthesis reaction, though this is somewhat complicated by non-ideal behavior of the poly(C) at higher concentrations. The same question of response to concentration can usefully be asked of pG and 2MeImpG, to determine how they participate in the reaction rate-determining complex. We determined initial rate of reaction at different nucleotide ratios, while keeping total nucleotide constant at $[pG] + [2\text{MeImpG}] = 20 \text{ mM}$. Varying the ratio $[2\text{MeImpG}]/[pG]$ over almost two orders of magnitude determines the reaction’s response to substrate nucleotide variation, while maintaining total nucleotide constant in all reactions tends to minimize the effect of experi-

mental variations on other common nucleotide reactions, such as metal binding or nucleotide stacking.

Expected behavior is shown by dotted and dashed lines in Figure 4. In this experimental design, peak velocity occurs when the ratio of substrates is equal to the ratio of substrate exponents in the governing kinetic expression. For example, if the rate of GppG synthesis $\propto [pG][2\text{MeImpG}]^2$, then when $[2\text{MeImpG}]/[pG] = 2$, maximal rate will occur. This is illustrated in the figure using the dashed line (curve labeled 2:1), which has a maximum at $[2\text{MeImpG}]/[pG] = 2$. Perhaps even more easily seen, this kind of response is distinctly asymmetric, with greater rates on the right of the figure, nearer the maximum. By the same argument, if maximal rates are seen when $[2\text{MeImpG}] = [pG]$ (curves labeled 1:1) then the exponents of the substrates in the rate expression are equal:

FIGURE 3. Response of GppG synthesis to poly(C) concentration; triangles are initial velocities from least squares linear fitting to early GppG concentrations versus time.

rate $\propto [\text{pG}][2\text{MeImpG}]$ or rate $\propto [\text{pG}]^{2*}[2\text{MeImpG}]^2$, or the like. This is illustrated in the figure by calculated (dashed) least squares lines placed near the data for the total reaction, and the chemical one (below). This exponent = 1 class of symmetric reactions (second order) can be discriminated because rate $\propto [\text{pG}][2\text{MeImpG}]$ yields a less sharply peaked curve than rate $\propto [\text{pG}]^{2*}[2\text{MeImpG}]^2$, or any larger common exponent (Puthenvedu et al. 2015).

In Figure 4 the chemical background and total poly(C)-stimulated reaction data (points) are similar, centered, and symmetric. Thus, templated and untemplated GppG production have equal nucleotide exponents. Moreover, the flatness of the observed rates over almost two orders in $[2\text{MeImpG}]/[\text{pG}]$ is as expected only for rate $\propto [\text{pG}][2\text{MeImpG}]$. That is, both the chemical and poly(C)-mediated reactions are first order in each nucleotide, or second order in mononucleotide concentrations overall. As explained later in the Discussion section, this conclusion is valid even though the reaction is not taking place free in solution, but mostly in stacks of heterogeneous length, paired to poly(C). Thus, the rate of total GppG synthesis is

$$\frac{d[\text{GppG}]}{dt} = k_c [\text{pG}][2\text{MeImpG}] + k_s [\text{poly(C)}][\text{pG}] \times [2\text{MeImpG}],$$

where the left-hand term is the chemical reaction (k_c is its apparent second order rate constant), and the right-hand term is the poly(C)-templated reaction. This equation, with poly(C) = 0 or 0.005 M (the polymer phosphate concentration), can be integrated to produce rate constants. But simultaneously,

$$\frac{d[2\text{MeImpG}]}{dt} = -k_d[2\text{MeImpG}],$$

which accounts for the disappearance of the activated nucleotide via hydrolysis:

and also for the simultaneous supplementation of the other nucleotide reactant pG (5'GMP). When reactions are modeled and all kinetic constants derived, it is via numerical integration and fitting of this system of differential equations to kinetic data, varying k_c , k_s , and k_d .

Other variations in reaction conditions

We have explored other variations in reaction composition. Briefly, pH does not greatly change the rate of GppG synthesis for the untemplated or the templated reaction over the range 6.4 to 7.6. However, though not otherwise significant (compare Fig. 1), at the highest pH, 2'-5' pGpG becomes a major product. Li^+ is slightly inhibitory to the templated production of GppG, but not to the chemical background; the standard concentration is a reasonable compromise for both (see Materials and Methods). Mg^{2+} is required; and we use its chelation to stop reactions (see Materials and

Methods). However, there are only relatively small increases in velocity above standard Mg^{2+} concentrations of 50 mM, which represents 2 mol of divalent ion for each mol of reaction nucleotide phosphate (pG plus 2MeImpG plus poly C).

Effect of temperature on relative rates of chemical and poly(C)-stimulated reactions

Reactions with and without poly(C) were carried out at temperatures from 8°C to 37°C, under otherwise standard conditions (see Materials and Methods), and the apparent rate constants for the chemical reaction (k_c ; circles in Fig. 5) and for the poly(C)-mediated reaction (k_s ; triangles in Fig. 5) were determined separately by fitting to the model above.

The majority of chemical reactions increase rapidly in rate with temperature. This is true for the chemical reaction of pG with 2MeImpG to yield GppG, as shown by the dotted line and data points (circles). The expected behavior of this reaction is, in fact, emphasized by the origins of the ascending dotted line. This line is not arbitrary, but is the best least squares fit of k_c to the Eyring–Polanyi equation, which describes rate–temperature variation in terms of an enthalpy of activation for reaction, $\Delta H^\ddagger$ (here = 13 kcal/mol), and an entropy of the activation reaction, $\Delta S^\ddagger$ (here = −17 cal/mol-degree).

$$\ln \frac{k}{T} = -\frac{\Delta H^\ddagger}{RT} + \ln \left\{ \frac{k_B}{h} \right\} + \frac{\Delta S^\ddagger}{R},$$

where k is the reaction rate, T is the Kelvin temperature, R the gas constant, k_B the Boltzmann constant, and h the Planck constant.

FIGURE 4. Initial rates of GppG synthesis, at different ratios of substrate nucleotides, with and without poly(C). Triangles indicate data, poly(C) added; circles denote data without polymer. Dashed lines are calculated, matching the overall velocity and assuming rate $\propto [2\text{MeImpG}][\text{pG}]$ (curves marked 1:1). Dotted lines are calculated, matching overall velocity and assuming rate $\propto [2\text{MeImpG}]^{2*}[\text{pG}]$ (curve marked 2:1), or rate $\propto [2\text{MeImpG}][\text{pG}]^2$ (curve marked 1:2).

FIGURE 5. Fitted rate constants for chemical and poly(C)-stimulated synthesis of GppG at different temperatures. Dashed line (triangles) poly(C)-stimulated synthesis; dotted line (circles) least squares fit of Eyring–Polanyi equation to chemical reactions.

The resulting numbers plausibly say ($\Delta H^\ddagger$) that bonds must be broken to react; possibly nucleotide stacking interactions must depart from their most-preferred conformation. In addition, the reactive complex is somewhat more ordered than the initial reactants ($\Delta S^\ddagger$); possibly stacked G nucleotides must reside within a smaller sector of possible orientations in order to yield an inter-nucleotide 5'–5' bond (compare Puthenvedu et al. 2015). Thus, the uninstructed chemical reaction increases its rate with temperature as expected from basic chemical considerations, except possibly at the highest temperature used.

However, poly(C)-mediated GppG synthesis behaves in a notably different way. While the rate constant (k_s) may increase at low temperatures, it enters a broad plateau around 10°C. However, this ends abruptly above 30°C, where the rate of the poly(C)-stimulated reaction decreases sharply, becoming indistinguishable from the chemical background at 37°C. Thus, in particular: While the chemical reaction is accelerating between 30°C and 37°C, the additional reaction mediated by poly(C) vanishes. This likely means that the crucial poly(C)–G nucleotide complex, which accounts for the acceleration of GppG synthesis, melts between 30°C and 37°C. This can be compared to the similar complex of poly(U) and -(A) nucleotides (yielding templated AppA), which is disrupted between 10°C and 20°C (Puthenvedu et al. 2015). The observed greater stability of the reactive poly(C–pG–2MeImpG) complex presumably embodies the greater stability of

C–G nucleotide stacks and pairs (Freier et al. 1986).

In short, the activation thermodynamics of stacked free G nucleotides suggests that (re)orientation for reaction of nucleotides is crucial. This leads naturally to the hypothesis that stimulation of GppG synthesis by poly(C) results from facilitation of the same changes as in thermal activation: the approach of 5' phosphates in paired-and-stacked nucleotides (molecular model presented in Puthenvedu et al. 2015).

Polymer specificity is a crucial feature of the reaction

Is the effect of poly(C) on GppG synthesis specific, or can it be evoked by varied ribopolymers? In Figure 6 are the results

of incubation with homopolymers, along with curves fitted to the data using the integrated differential equations of the model above. Notably, poly(A) has no effect on the reaction, which is specifically stimulated by poly(C), when all polymers are used at 5 mM RNA phosphate. However, poly(G) gives a notably smaller, but reproducible, stimulation, whose source we do not know. It is possible that this poly(G) effect is attributable to other phenomena, such as exclusion of some reaction volume by residual gels of pG. However, for present purposes, we emphasize that poly(C) is reproducibly and by far the most effective stimulator of GppG synthesis (Fig. 6). This suggests that the active template RNA is base-paired to

FIGURE 6. Polymer specificity; molar GppG synthesized in the reaction, in the presence of poly(U) (hyphens), poly(C) (triangles), poly(A) (squares) and poly(G) (diamonds), as well as with no polymer (circles). Fine dots are points calculated from fitted least squares differential equations, described above.

G nucleotide(s), as seemed also true for poly(U)-stimulated AppA synthesis. In particular, note that poly(C), maximally active here, was previously completely inactive when incubated with A nucleotides (Puthenvedu et al. 2015).

Already existing work suggests that cross-templating may be further simplified. If both 2'-5' and 3'-5' internucleotide linkages are acceptable in the template polymer (as in Sheng et al. 2014), our coenzyme congeners will be even more accessible in a primitive environment. This is plausible, because only short spans of template will likely be relevant to an active center for dinucleotide cross-templating.

DISCUSSION

Behavior of a paired stacking system

To clarify interpretation, we now calculate the expected behavior of a system of nucleotides that can form indefinite stacks as free molecules, and, in addition, can form such stacks more easily when bound to a complementary polymer. Adjacent stacked molecules can react to yield a product (NppN), and we will ultimately assume that NppN synthesis rates are proportional to concentrations of stacked nucleotides. In actual experiments, there are two species of nucleotides, chemically activated and unactivated. But for notional simplicity, we first characterize the stacking reactions of a single type of nucleotide, N , then later introduce the idea that nucleotides can be of two types.

Nucleotide N stacks to any extent with, for simplicity, a uniform dissociation constant K :

$$K = \frac{(N)(N)}{N_2}; \quad N_2 = \frac{(N)(N)}{K} = \frac{(N)^2}{K},$$

$$A_3 = \frac{N_2(N)}{K} = \frac{(N)^3}{K^2},$$

$$N_i = \frac{(N)^i}{K^{i-1}}.$$

So, summing a geometric series in N/K gives the total concentration of N -containing species including all stacks, or of stack initiations (counting free N), or the concentration of 3' or 5' stack ends,

$$K \sum_{i=1}^{\infty} \frac{(N)^i}{K^i} = \frac{(N)^2}{K - N}.$$

Free nucleotide concentration (N) can be obtained by summing an arithmetico-geometric series which expresses mass conservation for total added nucleotide, N_0 :

$$K \sum_{i=1}^{\infty} i \frac{(N)^i}{K^i} = \frac{K^2(N)}{(K - N)^2} = N_0.$$

From the resulting quadratic equation for (N) ,

$$(N) = \frac{x - (x^2 - 4K^2)^{0.5}}{2}, \quad \text{where } x = \frac{K^2 + 2KN_0}{N_0},$$

the concentration of all nucleotide species can be calculated. In particular, the sum of all stacks (N_i with $i \geq 2$) is useful because such stacked nucleotides potentially react (Puthenvedu et al. 2015) to yield product NppN:

$$K \sum_{i=2}^{\infty} i \frac{(N)^i}{K^i} = N_0 - (N).$$

Now suppose that the nucleotide N can also pair and stack on an added polymer ΦU , whose nucleotide concentration is (ΦU) , with dissociation constant K_0 :

$$\Phi U \cdot N = (\Phi U)(N)/K_0.$$

Further nucleotides stack on the first N bound to polymer, and will do so with dissociation constant fK , where (as is most likely) binding and stacking is more favorable using specific bonds to polymer than for free nucleotides; thus probably, $f < 1$:

$$\Phi U \cdot N_2 = \frac{(\Phi U \cdot N)(N)}{K_0 fK},$$

$$\Phi U \cdot N_2 = \frac{(\Phi U)(N)^2}{K_0 fK},$$

$$\Phi U \cdot N_i = \frac{\Phi U(N)^i}{K_0 (fK)^{i-1}}.$$

While it is not urgent to consider the occupation of sites on ΦU for low N_0 , this can become essential as total added nucleotide concentrations increase. In experimental conditions, $N_0 > \Phi U$, so that complete filling of a polymer template is stoichiometrically possible (and can conceivably occur; see below). Consider a polymer (top, Scheme 1) relatively longer than a prospectively binding stack of nucleotides (bottom, Scheme 1).

The number of sites for binding the smaller stacks (below) is almost equal to the number of the nucleotides on the long polymer (neglecting end effects), so we can take (ΦU) to be, approximately, the maximum concentration of binding sites. But stacks lengthen and more than one stack can bind to a template. Thus, $(\Phi U \text{ minus bound } N)$ is a more accurate approximation for the number of available stack binding sites after substantial N_0 input, where some binding sites have been filled.

$$\Phi U \cdot N_i = \frac{(\Phi U - \text{bound } N)(N)^i}{K_0 (fK)^{i-1}}$$

$$= \frac{(\Phi U - N + [K^2(N)/(K - A)^2])(N)^i}{K_0 (fK)^{i-1}}.$$

SCHEME 1. The number of binding sites for a small stack of nucleotides (*below*) on a long polymer (*above*) is almost equal to the total number of nucleotides in the long polymer.

Now, conserving N_0 mass that can be both free and bound:

$$N_0 = (N)$$

$$\times \left\{ \frac{K^2}{(K-N)^2} + \frac{(\Phi U - N_0 + [K^2(N)/(K-N)^2])}{K_0} \frac{(fK)^2}{(fK-N)^2} \right\},$$

where the first term in the rightward brackets is the free, and the second term in the rightward brackets is the polymer-bound, part of N_0 . This is an explicit equation for free nucleotide (N) in terms of known system constants, but is quartic in free nucleotide and not convenient for exploration. However, the equation was solved numerically using the generalized reduced gradient method (Frontline Systems, Incline Village, NV) within Microsoft Office Excel 2013. For example, below, for $K = 0.1$ M (for dissociation of free stacks), $f = 0.5$ (for the advantage of paired stacks), $K_0 = 0.5$ M (for dissociation of isolated, paired N from complementary polymer) and $\Phi U = 0.005$ M (concentration of added polymer), are some results calculated for equilibrium in reactions with 0 to 0.2 M total added nucleotide (N_0). The dissociation constants used are guesses, but bear some resemblance to available measurements (e.g., Solie and Schellman 1968).

We are particularly interested in free and polymer-bound nucleotide stacks (≥ 2 nucleotides), which are likely the productive precursors for NppN. The results (Fig. 7) show the accumulation of stacks of free nucleotide (circles), concave upward. It is more difficult to initiate stacks on an added polymer (triangles) because binding the first nucleotide without stacking support (via K_0) is unfavorable, so potentially reactive bound stacks remain at lower concentrations than the free ones at every concentration of added nucleotide (circles versus triangles, Fig. 7, top).

We have already referred to these calculations in Results. The strongly non-linear, increasing response of both free and polymer-mediated synthesis to total nucleotide concentration suggests that our standard conditions (as in Materials and Methods) are like those at the left in Figure 7. We are at “low” concentrations, meaning that poly(C) is far from binding saturation. This, among other

implications, justifies the model used for rate calculations, in that the initial polymer concentration is used in rate expressions without correction for binding of G nucleotide stacks. However, it is now clear from Figures 2 and 7 that derived rate constants and the stimulation by template depend on nucleotide substrate concentrations, so all rate calculations are tied to the particular reaction details used.

FIGURE 7. Free nucleotides stack or bind weakly to an added complementary polymer, where further stacking is enhanced by complementary base pairing. The figure shows species at equilibrium. (Top, right axis) Total free stacked N (circles; stacks ≥ 2 nt); polymer-bound stacked N (triangles). (Leftward axis) Circles indicate mean stack size for free N; diamonds refer to mean stack size for polymer-bound N. (Bottom, right axis) Paired (triangles) or free (circles) nucleotide concentration in polymer. (Left axis) Paired, stacked N/total stacked N (diamonds). Arrows beside curves point to the relevant y-axis.

Overall stacking behavior—polymer-bound stacks

New phenomena are suggested by these model calculations. The system reinforces a low level of polymer-bound stacks because there are only a limited number of polymer pairing sites (Fig. 7, bottom). The initial number of polymer sites (≈ 0.005 M on the left) decreases (circles), and such sites become completely filled at very high nucleotide concentrations. Filled polymer sites (triangles) end with more than 99.9% such sites occupied on the right of the figure. One can see the saturation of polymer-bound stacks (flattening of the line with triangles) at high N_0 concentrations on the right. Thus, there is a peak of fraction polymer-bound stacking at intermediate added N_0 , flanked by a lower proportion initially, and again a lower value because of polymer saturation at high N_0 . Stacked, polymer-bound nucleotides are only $\sim 3\%$ – 5% at maximum of total stacks with these chosen system constants (diamonds, Fig. 7, bottom).

Runaway stacks

There is another interesting aspect of saturation in the mean stack sizes (Fig. 7, top; left axis; diamonds—bound; circles—free stacks). At low concentrations, the minimal stack (two nucleotides) dominates both free and polymer–RNA bound stacks. However, as nucleotide input rises, stacks of free N lengthen sluggishly. Stacks are always longer among polymer-bound nucleotides (diamonds), because we have assumed stacking is favored there. In fact, at very high input N_0 (≥ 0.1 M nucleotide here), bound stacks quickly expand to very long sizes, and can fill the complementary polymer sites (diamonds) completely. The (off scale) bound stacks are almost 120 nucleotides long at the right hand limit (0.2 M total added nucleotide). Notably, such dramatic behavior requires only the seemingly moderate twofold (-0.4 kcal/mol) advantage assumed here for polymer-bound stacking, over free stacking.

Further, such extensive stacking is not solely hypothetical, but resembles real cases; completely stacked complexes have been observed for free nucleotides bound in three-stranded helices (Ts'o 1969). Further, base analogs modified for enhanced stacking form paired, free stacks thousands of residues in length (Cafferty et al. 2013). A completely paired template-nucleotide complex like the calculated one (Fig. 7) may be of interest as a replicative intermediate, especially under special conditions where nucleotide concentrations are forced to a high solubility limit (say, in a drying droplet). A completely filled template strand might roll on a catalytic surface, for example, to effect its own replication. Certainly, runaway stacking is of interest in this Discussion, because the ratio of polymer-paired-and-stacked to free stacked nucleotides must explain polymer enhancement of GppG synthesis. We now turn to this case.

Stacks versus solution

Nucleotides stack in all possible ways. Scheme 2 is a diagram representing all possible stacks of three nucleotides (stacks of three are used for definiteness; the results are the same for any stack length). We now assume that stacking of unactivated nucleotides (pG in our experiments) and activated nucleotides (2MeImpG in our experiments) is similar or equivalent. This is plausible because the activating group is distant from the stacking interface. Let the lightly shaded nucleotides in Scheme 2 below be the activated ones. There are 16 interfaces between the 24 nucleotides in the eight possible three-stacks.

Eight nucleotide interfaces (or 0.5 of the total) allow reaction because an unactivated and an activated nucleotide are apposed, and the activated nucleotide can be attacked by the unactivated one (compare Puthenvedu et al. 2015). This is true of any size stack: Given either type of nucleotide, the probability that the next will be the other type (when their concentrations are equal) is 0.5. So, we can just absorb this factor of 1/2 into the predicted rate constants of reaction for stacks of all sizes.

In our experiments above, the 2MeImpG/pG ratio is varied (Fig. 4), but total nucleotide concentrations are held constant. Thus total stacks may vary little, while 2MeImpG and pG vary within stacks according to their relative abundance. The finding (Fig. 4; Puthenvedu et al. 2015) that NppN synthesis is proportional to $(2\text{MeImpG})^*(\text{pG})$ reflects the molecularity of the chemical reaction, despite the limitation of the reaction to only free and polymer-bound nucleotide stacks. This is because the frequency of 2MeImpG/pG interfaces within stacks is proportional to (probability of occurrence of a nucleotide)*(probability that the next nucleotide differs). This in turn is:

$$P(\text{pG}) * P(2\text{MeImpG}) + P(2\text{MeImpG}) * P(\text{pG}),$$

where P is probability. Assuming that P is the concentration, the above becomes

$$\begin{aligned} & \frac{\text{pG}}{\text{pG} + 2\text{MeImpG}} * \frac{2\text{MeImpG}}{\text{pG} + 2\text{MeImpG}} + \frac{2\text{MeImpG}}{\text{pG} + 2\text{MeImpG}} \\ & * \frac{\text{pG}}{\text{pG} + 2\text{MeImpG}} \\ & = \frac{2 * \text{pG} * 2\text{MeImpG}}{\text{pG} + 2\text{MeImpG}}. \end{aligned}$$

The total reaction rate is proportional to $(\text{pG})^*(2\text{MeImpG})$, just as if reacting nucleotides were meeting in a second-order reaction in solution. Restriction of the reaction to stacks accordingly need have little or no effect on the determination of reaction order, as carried out in Figure 4.

SCHEME 2. All possible stacks of three nucleotides containing normal nucleotides (darker lines) and activated nucleotides (lighter lines).

Now we suppose that reaction at all interfaces between activated and unactivated nucleotides of the same structure have comparable rates. Thus, we interpret the increased rate of GppG formation with added poly(C) (e.g., Figs. 1, 2) as indicating that the uniform first-order rate at which polymer-bound and stacked nucleotides react is much greater than in free stacks. Because polymer-bound stacks are likely to be the minority stacked species (Fig. 7, bottom diamonds), polymer is expected to accelerate GppG synthesis at apposed nucleotides by approximately 140-fold.

GppG synthesis

We have characterized a second example of cross-backbone templating of a coenzyme-like dinucleotide, and clarified the role of nucleotide stacks in mediating these nucleotide reactions. Such stacks are likely facilitated by a complementary polymer, and these pairing-enhanced stacks are likely better oriented for synthesis (Figs. 1, 2, 5). One objective way to compare the two known reactions is to look at their rate constants, determined in both cases from the six substrate conditions in the nucleotide ratio experiment of Figure 4.

Values for the two tabulated cases (Table 1) would be altered by, e.g., use of differing substrate concentrations, as discussed just above. Nevertheless, nucleotides were similarly used in these two studies. The reaction of free 2MeImpG with free pG is not significantly different than the parallel reaction with A nucleotides (two-tailed *t*-test for difference of the means, $P = 0.15$), especially in view of a slightly increased temperature of incubation for GppG (cf. Fig. 5). What differs more significantly is the apparent rate constant in reactions with complementary polymer, where poly(C) is more effective (two-tailed *t*-test, $P = 0.0003$). Another way to say this is that stimulation by polymer is greater for poly(C) than for poly(U), despite similar chemical reactions using unstructured nucleotides. This is of interest, as it suggests that one way that cross-templating systems vary is in the effectiveness of alignment of polymer-paired nucleotides for reaction. Because nucleotide conformations required for cross-templating are probably accessible to all complementary reactants (Puthenvedu et al. 2015), we can hope that a search of complementary possibilities will find polymer-stimulated systems even more responsive than poly(C)-GppG.

TABLE 1. Table of rate constants

Quantity	Poly (U) → AppA	Poly (C) → GppG	GppG/ AppA
Chemical rate, $k_c \pm \text{SEM}$	0.06 ± 0.011	$0.10 \pm 0.02 \text{ M}^{-1} \text{ h}^{-1}$	1.7
Polymer rate, $k_s \pm \text{SEM}$	27 ± 6	$82 \pm 6 \text{ M}^{-2} \text{ h}^{-1}$	3

Reaction conditions are somewhat different—e.g., 10°C (AppA) versus 12°C (GppG). Values for AppA are taken from Puthenvedu et al. (2015).

GppG synthesis probably occurs in isolated stacks whose formation is enhanced by added complementary ribopolymer (as in Fig. 7). Even with polymer aid, however, it seems likely that reactive nucleotides are in a small minority (Fig. 7, bottom) among all nucleotides. This means that, to dominate overall synthesis as they do (e.g., Figs. 1, 2), polymer-bound stacks chemically amplify their effects via high synthetic rates, in comparison to free stacks. This implies that physical detection of active site structures could be challenging, because reactive sites are a minority among all paired nucleotides. Effective methods must be attuned to relevant minority properties. However, none of this implies that these are feeble reactions, viewed as chemical outcomes. In poly(C) incubations, order-of-magnitude stimulations by polymer can be attained (Fig. 1), and 40 μM pG is converted to GppG in an hour at 12°C; almost mM/day of a hypothetical catalyst (Figs. 1, 4, 6; see Materials and Methods).

Interpretation of cross-backbone templating in terms of evolutionary succession

Cross-templating suggests a fresh logic for transfer of genetic information via base-pairing, using a normal RNA template to yield a coenzyme-like product. Cross-templating accordingly suggests a simple coupling between the presence of oligonucleotides and phenotypic chemical changes in the environment of a gene (via reactive, coenzyme-like molecular products). Because cross-templating has markedly simpler requirements than either ribozyme action or modern biological chemistry, it might also be a primordial form of gene expression.

In the primordial state, an archegene (left, Fig. 8) modifies its environment by accumulating reactive dinucleotides, synthesized on a chemically created template. Synthesis of linear 55-mer RNAs from activated nucleotides resembling those used here, stimulated by montmorillonite clay, is an established reaction (Ferris et al. 1996). Inheritance of such a template subsequently confers the ability to exert an enhanced or new chemical effect, based on a cross-templated product.

Later, the evolved system in the center of Figure 8 articulates a phenotype by ribozymic synthesis of coenzymes (Huang et al. 1998) or a cofactor-ribozyme (Huang et al. 2000; Jadhav and Yarus 2002). After the appearance of translation in an RNA world (Yarus 2001), at the right, genes of complex sequence are expressed via synthesis of RNAs producing coenzyme-like molecules or by translating capped messages (descended from pre-existing coenzyme-RNAs, as shown) to produce RppA-binding peptides. Such peptides form relatively modern coenzyme-assisted catalytic sites. We suggest that the contemporaneous shift to novel peptide catalysts may account for the shift from coenzyme diphosphate (left, center) to triphosphate backbones (right) in RNA caps. The evolutionary succession for RppA from the bottom center to the top right was first suggested by Harold White (White 1976).

FIGURE 8. A novel primordial gene, logically related to modern gene expression. (Left) A simple, perhaps even monotonous, sequence in a small, primordial RNA (the archegene) expresses a phenotype by cross-templating a cofactor-like nucleotide RppA molecule, using an activated nucleotide with the primordial leaving group L. The archegene is the predecessor of a later, more evolved genetic system (the RNA Gene, center) which can perform RNA-catalyzed 3'-5' coenzyme-RNA synthesis. Later yet, RNA world translation makes possible modern catalysts: coenzymes bound to peptides (from the Gene, on the right).

We are quite certain that early potentially coenzymatic nucleotides, perhaps appearing via cross-backbone templating, could be later adopted by ribozymes, as in Figure 8. Such reactions have been tested by construction of cofactor-ribozymes. Experiments confirm that dinucleotides with the backbone structure of cofactors can be synthesized by ribozymes in both covalently linked (Huang et al. 2000; Zaher et al. 2006) and free forms. In the latter case, two nucleotides are bound from solution, apposed and linked 5'-5' in an unusual ribozyme reaction (Huang et al. 1999). Further, coenzymatic dinucleotides can be attached to other RNAs with a particular 5' sequence by ribozymes, and these RNA bound cofactors subsequently used by a selected RNA for catalysis; for example, for the production of butyryl- and acetyl-CoA (Jadhav and Yarus 2002). A particularly interesting cofactor-ribozyme activity is oxidation and reduction (Tsukiji et al. 2003, 2004), because here redox chemistry is supplied by a ribozyme's coenzymatic NAD^+ , even though redox reactions are not native RNA abilities.

Present results recall the prediction that, if a template for 5'-5' coenzyme-like RNA structures were to exist, the template molecule might have been lost (Yarus 2011a). Remarkably, it now appears that both templates and cross-templated products may have survived gigayears of parallel evolutionary alteration, though their close relation has been obscure. It will be particularly useful to define the limits of cross-templating.

MATERIALS AND METHODS

Reactions

Nucleotide reactions, usually 12 μL , were carried out in 200 mM MOPS adjusted with LiOH to pH 6.7 (at room temperature),

50 mM LiCl and 50 mM MgCl_2 . Nucleotides were 10 mM 5' GMP-Li, spiked with $[^{32}\text{P}]$ 5' GMP (Hartmann Analytic GmbH) and 10 mM 2mImpG. Incubation was at 12°C. If present, poly(C) or other polyribonucleotides were held at a final concentration of 5 mM ribonucleotide phosphate. Radioisotopic nucleotide was always negligible in concentration, therefore nucleotide and product concentrations were calculated by measuring the fraction of total radioactivity in each spot, and multiplying by total pG.

Chromatography

Reaction samples were rapidly frozen on dry ice, with agitation, and held at -76°C prior to fractionation by thin layer chromatography on Cellulose 300 F254 (Selecto Scientific). Dilution of kinetic samples into an equal volume of 100 mM EDTA and freezing is essential, because otherwise, reactions apparently continue on the origin of the chromatogram. After spotting, chromatograms were immersed in methanol for 6 min and air dried (Bochner and Ames 1982). Methanol immersion sharpens the resolution, particularly of the GppG spot. No $[^{32}\text{P}]$ was detected in the methanol wash. Chromatographic eluant was *n*-butanol, acetic acid, water (3:2:5). Chromatograms were analyzed in a Bio Rad FX phosphorimager.

Nucleotides

NppN standards were synthesized as by Kanavarioti et al. (1991). 2MeImpN was synthesized as by Joyce et al. (1984), with the modifications previously described (Puthenvedu et al. 2015). 2MeImpG was $\geq 91\%$ pure by HPLC, with the major impurity being 5' pG.

The chromatographic spot of the major product was identified with GppG as follows:

1. It has the mass/charge of GppG (707.1 by electrospray mass spectrometry).
2. It has a pyrophosphate linkage because it is completely degraded to pG by Tobacco Acid Pyrophosphatase (Epicentre) under conditions in which pGpG (Thermo Fisher) is resistant.
3. It has no terminal phosphates; it is completely resistant to Shrimp Alkaline Phosphatase (US Biochemicals) under conditions that completely alter pGpG (Thermo Fisher).
4. It has no 3'-5' phosphodiester; it is completely resistant to T1 nuclease (Epicentre) under conditions that completely digest pGpG (Thermo Fisher).
5. It is coincident in this chromatographic system (Fig. 1) and on HPLC with GppG synthesized by a characterized method (Kanavarioti et al. 1991).

RNA homopolymers are from Sigma-Aldrich Corporation. The manufacturer characterizes them as high molecular weight (≥ 250 nt, mean length). Poly(C) concentration was measured in water at pH 7, using a molar extinction of 8150 at 270 nm. Similar results were obtained when synthetic homopolymers from Santa Cruz Biotechnology (Dallas, TX) were substituted for Sigma-Aldrich polymers.

Numerical methods

When numerical integration of systems of differential rate equations was required to fit rate data in order to define rate constants, Berkeley Madonna v8.3.23.0 was used, as previously described

(Puthenvedu et al. 2015). Fitting of the chemical reaction yields k_c , then the same k_c is used to define k_s in parallel reactions containing poly(C), and carrying out both simultaneous syntheses. Numerical solution of the nucleotide binding model was performed by using the “solver” functions within Microsoft Excel 2013, confining solutions to physically realistic nucleotide concentrations that also conserved total nucleotide.

Received October 13, 2015; accepted December 2, 2015.

REFERENCES

- Bochner BR, Ames BN. 1982. Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. *J Biol Chem* **257**: 9759–9769.
- Cafferty BJ, Gállego I, Chen MC, Farley KI, Eritja R, Hud NV. 2013. Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues. *J Am Chem Soc* **135**: 2447–2450.
- Chumachenko NV, Novikov Y, Yarus M. 2009. Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *J Am Chem Soc* **131**: 5257–63.
- Ferris JP, Hill JAR, Liu R, Orgel LE. 1996. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**: 59–61.
- Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci* **83**: 9373–9377.
- Gellert M, Lipsett MN, Davies DR. 1962. Helix formation by guanylic acid. *Proc Natl Acad Sci* **48**: 2013–2018.
- Huang F, Yang Z, Yarus M. 1998. RNA enzymes with two small-molecule substrates. *Chem Biol* **5**: 669–678.
- Huang F, Yang Z, Yarus M. 1999. Self-capping RNA catalysts derived from selection-amplification. *Biol Bull* **196**: 320–321.
- Huang F, Bugg CW, Yarus M. 2000. RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry (Mosc)* **39**: 15548–15555.
- Jadhav VR, Yarus M. 2002. Acyl-CoAs from coenzyme ribozymes. *Biochemistry (Mosc)* **41**: 723–729.
- Joyce GF, Inoue T, Orgel LE. 1984. Non-enzymatic template-directed synthesis on RNA random copolymers. Poly(C, U) templates. *J Mol Biol* **176**: 279–306.
- Kanavarioti A, Lu J, Rosenbach MT, Hurley TB. 1991. Unexpectedly facile synthesis of symmetrical P1,P2-dinucleoside-5′ pyrophosphates. *Tetrahedron Lett* **32**: 6065–6068.
- Lohrmann R, Bridson PK, Orgel LE. 1980. Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **208**: 1464–1465.
- Pinnavaia TJ, Marshall CL, Mettler CM, Fisk CL, Miles HT, Becker ED. 1978. Alkali metal ion specificity in the solution ordering of a nucleotide, 5′-guanosine monophosphate. *J Am Chem Soc* **100**: 3625–3627.
- Puthenvedu D, Janas T, Majerfeld I, Illangasekare M, Yarus M. 2015. Poly(U) RNA-templated synthesis of AppA. *RNA* **21**: 1818–1825.
- Sheng J, Li L, Engelhart AE, Gan J, Wang J, Szostak JW. 2014. Structural insights into the effects of 2′-5′ linkages on the RNA duplex. *Proc Natl Acad Sci* **111**: 3050–3055.
- Solier TN, Schellman JA. 1968. The interaction of nucleosides in aqueous solution. *J Mol Biol* **33**: 61–77.
- Szostak J. 2012. The eightfold path to non-enzymatic RNA replication. *J Syst Chem* **3**: 2.
- Ts’o PO. 1969. The hydrophobic-stacking properties of the bases in nucleic acids. *Ann NY Acad Sci* **153**: 785–804.
- Tsukiji S, Pattnaik SB, Suga H. 2003. An alcohol dehydrogenase ribozyme. *Nat Struct Biol* **10**: 713–717.
- Tsukiji S, Pattnaik SB, Suga H. 2004. Reduction of an aldehyde by a NADH/Zn²⁺-dependent redox active ribozyme. *J Am Chem Soc* **126**: 5044–5045.
- White HB III. 1976. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**: 101–104.
- Wong A, Ida R, Spindler L, Wu G. 2005. Disodium guanosine 5′-monophosphate self-associates into nanoscale cylinders at pH 8: a combined diffusion NMR spectroscopy and dynamic light scattering study. *J Am Chem Soc* **127**: 6990–6998.
- Yarus M. 2001. On translation by RNAs alone. *Cold Spring Harb Symp Quant Biol* **66**: 207–215.
- Yarus M. 2011a. Getting past the RNA world: the initial Darwinian ancestor. In *RNA worlds: from life’s origins to diversity in gene regulation* (ed. Gesteland RF, Atkins JF, Cech TR), pp. 43–50. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Yarus M. 2011b. The meaning of a minuscule ribozyme. *Philos Trans R Soc Lond B Biol Sci* **366**: 2902–2909.
- Yarus M. 2012. Darwinian behavior in a cold, sporadically fed pool of ribonucleotides. *Astrobiology* **12**: 870–883.
- Yarus M. 2013. A ribonucleotide origin for life - fluctuation and near-ideal reactions. *Orig Life Evol Biosph* **43**: 19–30.
- Zaher HS, Watkins RA, Unrau PJ. 2006. Two independently selected capping ribozymes share similar substrate requirements. *RNA* **12**: 1949–1958.

Cross-backbone templating; ribodinucleotides made on poly(C)

Irene Majerfeld, Deepa Puthenvedu and Michael Yarus

RNA 2016 22: 397-407 originally published online January 12, 2016

Access the most recent version at doi:[10.1261/rna.054866.115](https://doi.org/10.1261/rna.054866.115)

References

This article cites 28 articles, 8 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/22/3/397.full.html#ref-list-1>

Open Access

Freely available online through the *RNA* Open Access option.

Creative Commons License

This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:

<http://rnajournal.cshlp.org/subscriptions>