

# The eightfold path to non-enzymatic RNA replication

Jack W Szostak

## Abstract

The first RNA World models were based on the concept of an RNA replicase - a ribozyme that was a good enough RNA polymerase that it could catalyze its own replication. Although several RNA polymerase ribozymes have been evolved *in vitro*, the creation of a true replicase remains a great experimental challenge. At first glance the alternative, in which RNA replication is driven purely by chemical and physical processes, seems even more challenging, given that so many unsolved problems appear to stand in the way of repeated cycles of non-enzymatic RNA replication. Nevertheless the idea of non-enzymatic RNA replication is attractive, because it implies that the first heritable functional RNA need not have improved replication, but could have been a metabolic ribozyme or structural RNA that conferred any function that enhanced protocell reproduction or survival. In this review, I discuss recent findings that suggest that chemically driven RNA replication may not be completely impossible.

**Keywords:** Non-enzymatic replication, RNA World, protocell, origin of life, prebiotic chemistry, fidelity

## 1. Introduction

Considerable evidence supports the RNA World hypothesis of an early phase of life, prior to the evolution of coded protein synthesis, in which cellular biochemistry and genetics were largely the province of RNA (reviewed in [1,2]). The central question concerning the origin of life has thus become: how did the RNA World emerge from the prebiotic chemistry of the early earth? The emergence of the RNA World can be divided, conceptually, into two phases: first, the prebiotic synthesis of ribonucleotides and of RNA molecules, and second, the replication of genomic RNA molecules within replicating protocells. As recently as 2004, Leslie Orgel, one of the great founding fathers of the field of prebiotic chemistry, suggested that “the abiotic synthesis of RNA is so difficult that it is unclear that the RNA World could have evolved *de novo* on the primitive earth” [3]. While a complete solution to the problem is not yet in hand, major advances have recently been made in this area, primarily from the Sutherland lab [4-6], such that there is now considerable optimism that a robust pathway to the

assembly of ribonucleotides and RNA oligomers will eventually be found.

Unfortunately, even given prebiotically generated RNA templates and abundant ribonucleotides, we do not understand how cycles of template-directed RNA replication could occur. RNA replication is fundamental to the RNA World model of early life, as it is essential for the transmission of useful genetic information (and variation) from mother cells to daughter cells. Although early discussions of the RNA World neglected the cellular nature of life, it is clear that the emergence of Darwinian evolution required spatial localization, so that potentially useful RNAs could accrue an advantage for themselves [7,8]. A cellular (i.e. membrane delimited) structure is a simple way to accomplish this, and provides a direct connection to the cellular structure of all modern life. At first glance, localization on the outer surface of vesicles (or on the surface of mineral particles, within porous rocks etc. etc.) would seem to pose less of a problem with respect to accessibility to oligonucleotide or activated monomer building blocks, as well as primers and divalent cations. However, if an evolving genetic system became adapted to and dependent on such an environment, the subsequent transition to a membrane based cellular structure would have been very difficult if not impossible. Although a

Correspondence: szostak@molbio.mgh.harvard.edu  
Howard Hughes Medical Institute and Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Simches Research Center, 185 Cambridge Street, Boston, Massachusetts 02114, USA

protocellular structure poses more problems initially, it is actually simpler to solve these problems up front rather than leave them till later when they could become completely intractable. In this Perspective, I will therefore focus on RNA replication in the context of a cellular environment, as described in recent model studies [9,10].

The original RNA World model envisaged a central role for RNA mediated catalysis in driving RNA replication, a seemingly straightforward extrapolation from the then recently discovered activities of biological ribozymes [11]. Despite recent progress [12,13], the experimental demonstration of RNA-catalyzed RNA copying still requires significant increases in efficiency and accuracy before self-replication can be achieved. In addition, replication cycles would require some means of strand separation, which is in itself a formidable challenge as a result of the high melting temperature of long RNA duplexes.

In light of the above problems, I suggest that it is worthwhile to revisit the much earlier model of chemical (i.e., non-enzymatic) RNA replication. Orgel, along with many of his students and colleagues, worked on this problem for many years and made considerable progress by demonstrating the partial copying of short RNA templates in the presence of suitably activated nucleotide substrates. In one of the most impressive examples of such chemistry, a full length complement of a 14-mer all GC template was generated in the presence of activated G and C monomers, albeit in low yield (< 2%) [14]. Despite such achievements, the efficient and accurate copying of arbitrary template sequences has remained frustratingly out of reach. The problems standing in the way of robust RNA template copying have been reviewed in detail [15]. Furthermore, as a result of the historical focus on template copying chemistry, the additional issues related to continuing cycles of replication such as strand separation and strand reannealing have seldom been considered. In addition, template copying experiments have been, with few exceptions [9], performed in solution, so that problems related to the compatibility of RNA replication and protocell membranes have also been neglected. Here I will review recent conceptual and experimental advances that point the way to a possible experimental realization of multiple cycles of chemically driven (i.e., non-enzymatic) RNA replication in a manner that could be integrated into a protocell model. I will also discuss the implications of chemical RNA replication for the emergence of the RNA World from the prebiotic chemistry of the early earth.

## 2. Discussion

### 2.1. Eight major problems in the way of cycles of chemical RNA replication

The most significant problems standing in the way of an experimental demonstration of multiple cycles of RNA replication can be summed up in the following list of 8

problems, each of which must be solved to enable continuing genomic replication within reproducing protocells.

1. Regiospecificity: chemical template copying generates complementary strands with a heterogeneous backbone containing randomly interspersed 2'-5' and 3'-5' linkages.

2. High  $T_m$  of long RNA duplexes: because of the high melting temperature of RNA duplexes, the accurate copying of an RNA template will generate a dead-end duplex product.

3. Fidelity of template copying chemistry: the accuracy of chemical replication is insufficient to allow for the propagation of functional genetic information.

4. Rate of template copying chemistry: chemical template copying occurs on the same timescale as template and substrate degradation.

5. Reactivation chemistry: the efficiency of template copying is limited by substrate hydrolysis, but current means of re-activating hydrolyzed substrates lead to damaging side reactions that would destroy both templates and substrates.

6. Divalent metal ions: the high concentrations of divalent cations required for RNA template-copying reactions catalyze RNA degradation and are incompatible with known vesicle replication systems.

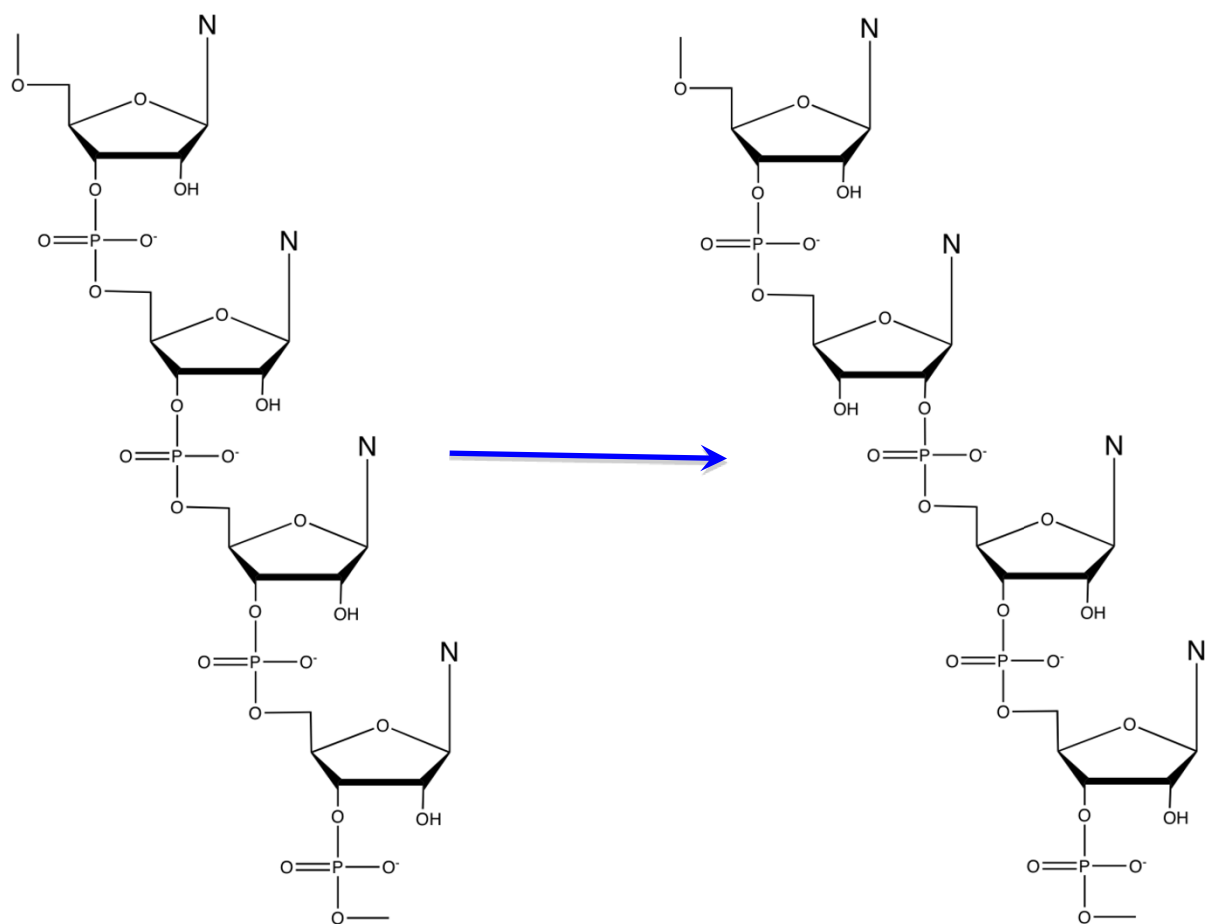
7. Primer-independent RNA replication: replication based on primer-extension is incompatible with a protocell model system, because protocells cannot take up exogenous primers.

8. Strand reannealing: the reannealing of separated strands prevents template copying, but the rate of strand reannealing is orders of magnitude faster than current copying chemistry.

Taken together, the above issues might suggest that the chemical replication of RNA templates within protocells is impossible. However, in the following sections, I will discuss each point in turn, and show that there are potential solutions to all of these problems. If such solutions can be experimentally demonstrated, it may yet be possible to implement chemically driven RNA replication inside replicating protocells.

### 2.2. Regiospecificity

Ribonucleotides have two nucleophiles, the 2' and 3' hydroxyls, that can potentially attack the activated 5'-phosphate of an adjacent nucleotide during RNA polymerization. Thus RNA template copying reactions generate products containing a substantial proportion of 2'-5' phosphodiester linkages, in place of the natural 3'-5' linkages (Figure 1). The ratio of 2'-5' to 3'-5' linkages can be shifted in favor of the desired 3'-5' product through the use of specific metal ions as catalysts, e.g.  $Zn^{2+}$  in place of  $Mg^{2+}$  [16], and the use of specific leaving groups on the



**Figure 1** Chemical copying of an RNA template generates a complementary strand containing a mixture of 2'-5' and 3'-5' linkages. (left) Normal all 3'-5' linked RNA template; (right) product strand containing a 2'-5' linkage.

activated nucleotides, e.g. 2-methyl-imidazole in place of imidazole [17]. The reasons for the changes in regioselectivity are unknown, but in any case 2'-5' linkages still make up 10-30% of the backbone of the newly synthesized complementary strand. It has long been thought that this is a serious, even fatal, problem with respect to purely chemical RNA replication, because such backbone heterogeneity has been assumed to preclude the evolution of heritable, functional structures such as ribozymes.

Recent findings suggest that this problem may be less severe than originally anticipated, and there are even reasons to think that imperfect regioselectivity may be essential for continued cycles of chemical RNA replication. Consider first the question of functional sequences. We recently showed [18] that non-heritable backbone heterogeneity involving ribo- vs. deoxyribo- sugars does not prevent the *in vitro* evolution of functional aptamers. We used a mutant form of T7 RNA polymerase to generate libraries of random sequences in which every position in every transcript had a roughly 50/50 chance of being

either a ribonucleotide or a deoxyribonucleotide. We then selected for binding to either of two small molecule targets, ATP or GTP, and carried out cycles of amplification and selection until the pool was dominated by target binding sequences. Despite the shuffling of the backbone ribo/deoxyribo heterogeneity during each round of amplification by RT-PCR-transcription, we were able to evolve aptamers with high specificity for their target ligands, although binding affinity was weaker than seen in previously isolated RNA or DNA aptamers for the same ligands. Thus, sequences exist that encode structures that are relatively insensitive to this particular type of backbone heterogeneity. These results are consistent with previous experiments, including occasional examples of aptamers that function whether they are made of RNA or DNA [19], and the observation that in other aptamers relatively few positions must be either ribo- or deoxyribo-nucleotides [20,21].

The finding that non-heritable ribo/deoxyribo backbone heterogeneity does not preclude heritable function

raises the question of whether other types of backbone heterogeneity, including 2'-5' linkages in RNA, are consistent with heritable function. Unfortunately it is not currently possible to enzymatically generate RNA transcripts that contain significant levels of 2'-5' linkages, because all known RNA polymerases are highly specific for the synthesis of 3'-5' linkages. However, the chemical synthesis of such mixed transcripts is quite simple, and ongoing experiments should reveal the extent to which 2'-5' linkages can be tolerated within known ribozymes and aptamers. If significant levels of 2'-5' linkages are tolerable in at least some structures, then it may well be that perfect or even high regiospecificity was simply not a requirement of the primordial RNA replication process.

### 2.3. High $T_m$ of long RNA duplexes

The second major problem with RNA replication is the difficulty of strand separation. RNA duplexes of significant length have very high melting temperatures, and duplexes of even 30-50 base-pairs may be impossible to thermally denature under the conditions required for replication (extrapolated from data in [22]). Thus, given an efficient and accurate copying chemistry, the conversion of a single-stranded RNA template into an RNA duplex would simply lead to a dead-end product. Without a simple means of separating the strands, such as thermal cycling, there is no way to continue to the next generation of replication. One way to lower the  $T_m$  of an RNA duplex is through errors in the copying chemistry, i.e. the introduction of mismatches. However, a reasonably low error rate (< 2% per position per strand) is thought to be necessary to allow for the inheritance of functional RNAs such as ribozymes (see Section 2.4 below). Such a low error rate is unlikely to result in sufficient RNA duplex destabilization to solve the strand separation problem, suggesting that we should look for additional destabilizing mechanisms.

It has been known for many years that 2'-5' linked RNA duplexes have a much lower  $T_m$  than natural RNA duplexes [23]. A few studies indicate that even a small proportion of 2'-5' linkages in an otherwise all 3'-5' RNA duplex can greatly lower the  $T_m$  [24,25]. Thus a non-enzymatic chemical copying process that generates a mixed-backbone product would be amenable to continued replication cycles through thermal strand separation, whereas a perfectly regiospecific system would not.

The imperfect regiospecificity of chemically copied RNA may be important in a more subtle way than simply enabling strand separation. The heterogeneous mixture of products generated by repeated copying of a given sequence that encodes, for example, a ribozyme, may include some strands that retain excellent functionality because no particular 2'-5' linkage disrupts the

folded structure of the ribozyme. Such strands are likely to be poor templates because of the energetic barrier to unfolding, but other strands in which 2'-5' linkages do disrupt the folded structure will be better templates, and they could be used preferentially for subsequent rounds of replication. Thus a heterogeneous set of products may support both superior function and superior replication compared to a homogeneous product. In summary, the imperfect regiospecificity resulting from chemical RNA replication may be precisely what allowed RNA to emerge as the first replicating polymer on the early earth.

### 2.4. Fidelity of template copying chemistry

In order for RNA to have emerged as the genetic polymer that enabled protocells to evolve in a Darwinian manner, the process of RNA replication must have been accurate enough to allow for the transmission of useful information from generation to generation, indefinitely [26]. To a first approximation, the minimum fidelity that must be exceeded corresponds to an error rate less than the reciprocal of the number of functionally important bases, i.e.  $1/n$  where  $n$  is the effective genome size. Greater error rates lead to an "error catastrophe", the inability to transmit useful information, no matter how strong the selective advantage conferred by that information. Useful information in this context is generally thought to imply sequences with ribozyme activity, where the ribozyme confers an advantage to the cell as a whole, e.g., by the catalytic synthesis of a useful metabolite. For example, we have recently shown that catalyzed phospholipid synthesis could confer such a selective advantage by driving protocell membrane growth through the adsorption of fatty acid molecules from surrounding vesicles [27]. It is possible that other metabolic ribozymes or even structural RNAs could contribute to cell survival or reproduction. Since the smallest highly active ribozymes are in the size range of ~50 nucleotides [28], the error rate during template copying should be less than ~2% at each position, assuming that only about half of the nucleotide positions need to be specified, but each position must be copied twice for full replication [29]. A recent study [30] suggests that the average error rate during RNA copying is ~17%, although this could potentially be reduced to under ~10% by using optimized nucleotide ratios, and perhaps further reduced to ~5% on GC-rich templates since U residues in the template contribute most to the overall error rate. Clearly, a robust means of further reducing the error rate is critical if non-enzymatic RNA replication is to serve as the means for initiating Darwinian evolutionary processes.

A recently described phenomenon that increases the effective fidelity of template copying follows from the observation that fidelity influences the rate of template

copying. Primer-extension following a mismatch can be much slower than following a Watson-Crick base-pair. Such post-mismatch stalling was originally noticed in enzymatic reactions [31,32] but is also significant in a non-enzymatic DNA template-copying model system [33]. The beneficial effect of post-mismatch stalling on fidelity is a consequence of the fact that those templates that are completed first tend to be the most accurately copied. If those rapidly completed accurate copies are immediately available for use as new templates (i.e., if strand separation is fast), then the average fidelity of replication is increased. Under reasonable circumstances, the effective fidelity can be 2-5 times as great as that calculated in the absence of the stalling effect (i.e. when all templates are fully copied once per cell cycle). The observation that fidelity is a time-dependent variable can therefore lead to situations in which the Eigen error threshold can be circumvented, and information can be transmitted in the face of error rates that, classically considered, would lead inexorably to the decay of useful information.

Unfortunately it has not yet been possible to measure stalling parameters in an all RNA template-copying system, due to the low rates of post-mismatch primer extension reactions. In order to assess the possible role of stalling in increasing the fidelity of RNA replication, RNA stalling parameters should be measured as means are developed to increase rates of RNA copying chemistry. In the case of RNA, the major source of copying errors is wobble base-pairing, primarily G:U mismatch formation [30], as expected from the geometric similarity of Watson-Crick and wobble base-pairs together with thermodynamic base-pair stability data [22]. Since the stalling effect on primer-extension is likely to be smaller following wobble base-pairs (G:U or A:C) than more disruptive mismatches, it is possible that stalling may only marginally increase the fidelity of RNA replication, in which case other means of increasing fidelity must be identified. Macromolecular catalysts solve this problem by helping to properly position the incoming monomers, and small molecule catalysts, including short peptides and intercalators, may also be able to minimize wobble mismatch formation through such steric effects. Screens for such catalysts are an important future research direction.

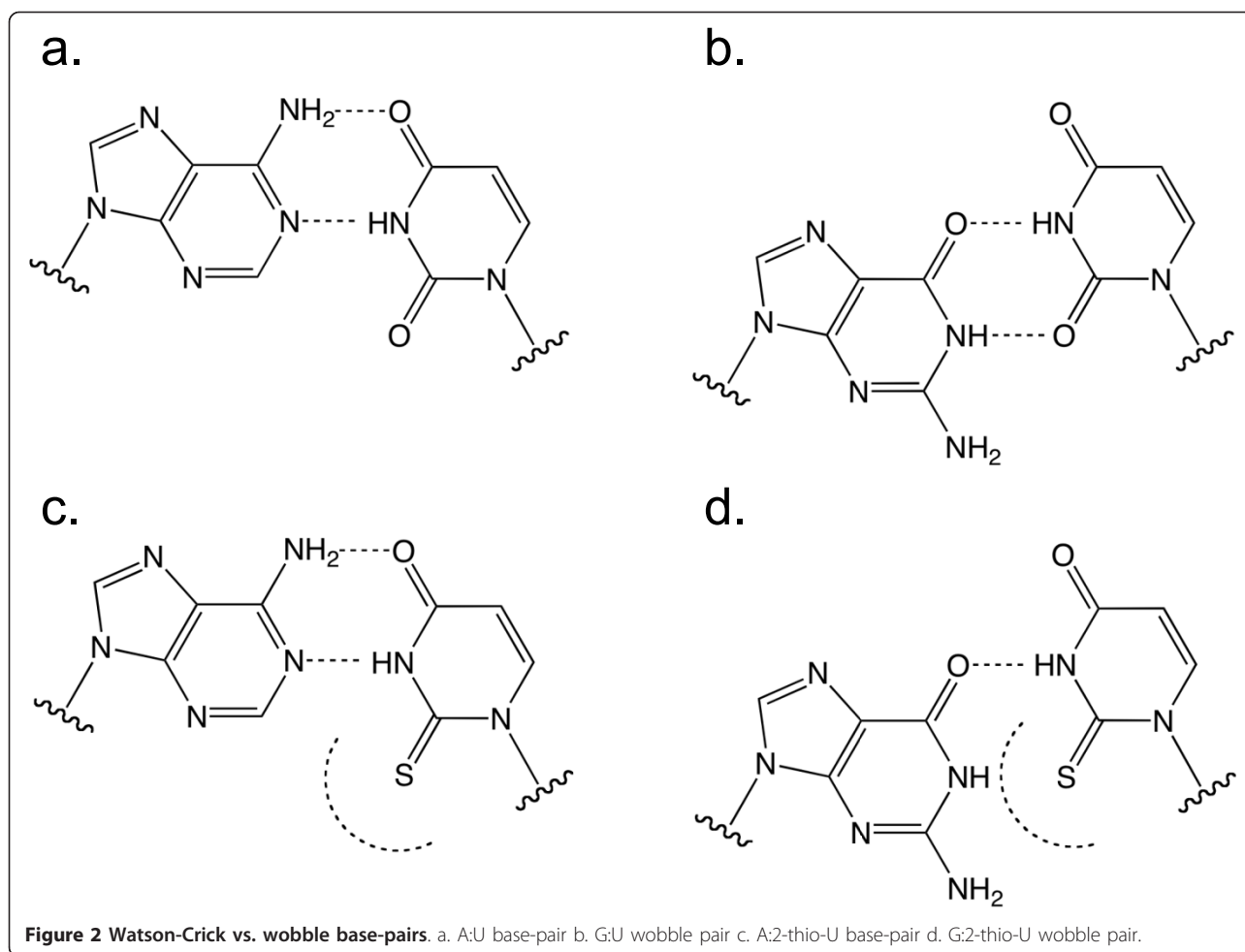
An alternative approach to improving fidelity is to consider nucleobase analogs that may increase the energetic discrimination between Watson-Crick and wobble base-pairs (Figure 2). A promising example is 2-thio-U (or 2-thio-T or other 5-substituted forms of 2-thio-U), based on thermodynamic data [34] that indicate that 2-thio-U forms a stronger base-pair with A than does standard U (by about 1 kcal/mol), while slightly destabilizing G:U pairing. This substitution could therefore

significantly increase the fidelity of RNA copying. It is also possible that the 2-thio-U substitution may further distort the geometry of a G:2-thio-U mismatch, thus increasing the stalling effect and further enhancing fidelity. The even bulkier 2-seleno-U substitution has also been proposed as a means of enhancing the fidelity of replication [35,36]. In nature, both 2-thio-U and 2-seleno-U are found in the wobble position of certain tRNAs where translational fidelity demands discrimination between Watson-Crick and wobble pairing in the third position of the codon [35,37]. It will be interesting to see if activated 2-thio-U (or 2-seleno-U) leads to enhanced fidelity in RNA template copying experiments. This connection to modern biology makes these substitutions particularly interesting as a possible relic of the emergence of the RNA World. Since any prebiotic chemistry leading to 2-thio-U is also likely to generate 2-thio-C, it will be important to assess the effect of this modification on the rate and fidelity of copying of RNA templates. Interestingly, the G:2-thio-C base-pair is only 0.6 kcal/mol less stable than a standard G:C base-pair at an internal position in a RNA duplex, and replacement of G with inosine (I) to generate an I:2-thio-C base-pair has almost no effect [38], suggesting that RNA templates that include either G or I residues should also be examined.

### 2.5. Rate of template copying chemistry

The copying of short RNA templates by the polymerization of activated monomers is typically examined over a time-course of days. In order to achieve optimal copying, high salt and divalent cation (usually  $Mg^{2+}$ , from 0.05 to 0.75 M) concentrations are used. However, in the presence of such high  $Mg^{2+}$  concentrations, single-stranded RNA templates experience significant degradation over a timescale similar to that of polymerization, while the activated monomers degrade by hydrolysis and cyclization (Figure 3). It would clearly be helpful to find conditions or catalysts that specifically increase the rate of polymerization relative to the rates of degradative processes. Furthermore, the rate of template copying varies greatly with template composition. The polymerization of activated GMP on poly-C templates generates oligomers up to 50 nucleotides in length over the course of a week [15], so an increase in the rate of the individual primer-extension step of only 10-fold would be sufficient to lead to efficient copying of moderately long templates. However, the copying of other bases interspersed within a C-rich template is much slower, so chemistry that would lead to equivalently rapid polymerization of all four nucleotides must be identified.

A-rich templates are among the most difficult RNA sequences to copy, most likely because of the relatively weak A:U base-pair combined with the low stacking

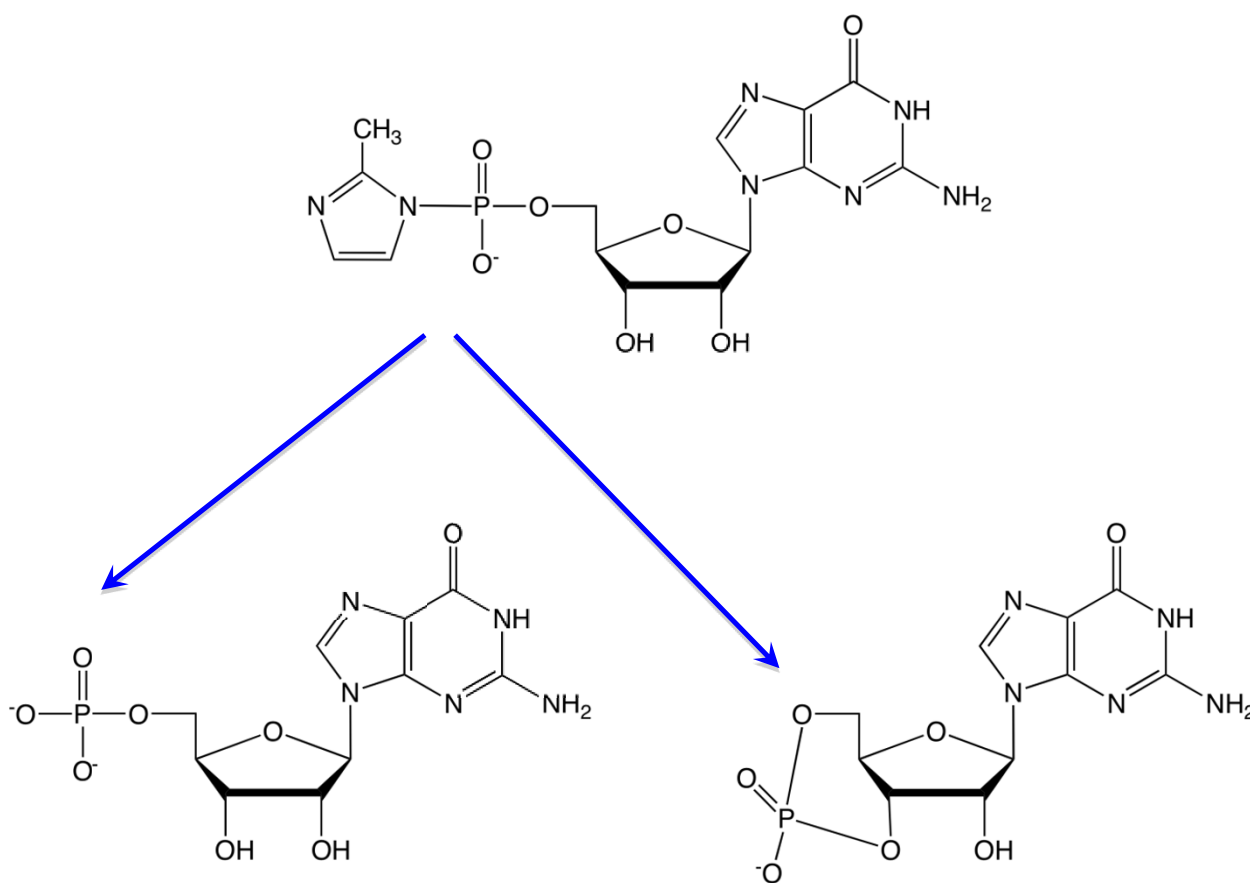


energy of the incoming U monomer. Recent work from the Richert lab addresses this problem [39,40]. Downstream helper oligos, designed to leave a single-nucleotide gap between primer and helper oligo, into which the incoming U monomer could fit, were found to contribute significantly to the rate and extent of primer extension. In a second advance, the hydroxyazabenzotriazole (OAt) leaving group was found to lead to significantly improved rates of primer-extension compared to 2-methylimidazole activation when the reactions were carried out at elevated pH (8.9 vs. 7.7). This is probably due at least in part to the higher  $pK_a$  of N7 of the OAt group compared to that of 2-methylimidazole, such that the optimal reaction pH increases from  $\sim 7.5$  to  $\sim 8.9$ . While increasing pH speeds up polymerization, it also speeds up degradative processes, so it is unclear whether there is a net benefit from this approach other than decreasing the overall time frame of the experiments. Moderately increased pH is compatible with fatty acid vesicle systems, many of which exhibit optimal stability near pH 8.5. A further increase in the overall extent of

primer extension was observed when the reaction mixture was frozen at  $-20^\circ\text{C}$ , probably due to the increased monomer concentration in the thin layers of eutectic fluid between the ice crystals. Unfortunately freezing is incompatible with vesicle encapsulated reactions due to vesicle rupture by ice crystals.

An alternative means of enhancing the copying of AU-rich sequences would be to take advantage of the fact that 2-thio-U forms a significantly stronger A:U base-pair than U. It will be of great interest to see if activated 2-thio-UMP allows for more rapid and efficient copying of A residues in RNA templates, either as a result of better template binding, or as a consequence of the conformational constraint of the 2-thio-U monomer in the 3'-endo configuration [41], which may increase the rate of the chemical step. If so, this effect, combined with the possibility of enhanced fidelity, would be a strong argument in favor of a role for this simple nucleobase modification in a primordial system for chemical RNA replication.

Template copying by monomer polymerization requires the maximum number of steps for the generation of a

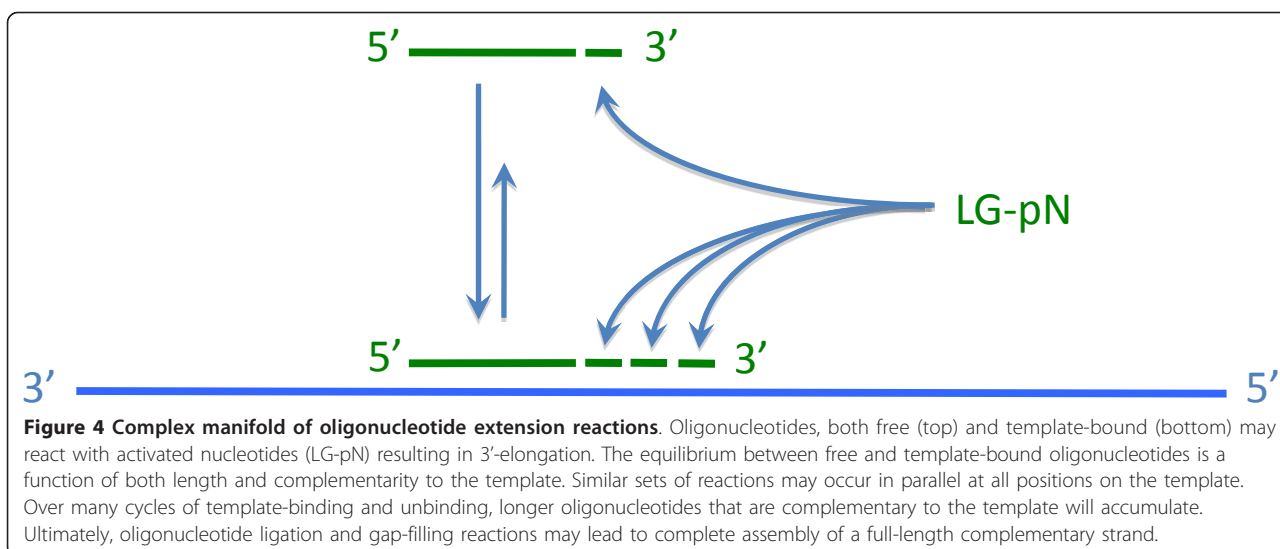


**Figure 3 Activated nucleotides are subject to spontaneous degradation.** Hydrolysis (left) and cyclization (right) reactions decrease substrate availability and limit the extent of template copying.

complementary strand; the use of pre-formed oligonucleotide substrates should therefore lead to faster template copying. Due to slow membrane permeation, oligonucleotide substrates must be formed inside the protocell from monomers that have diffused into the cell from the external environment. Given a pool of activated nucleotides, oligomers will inevitably begin to assemble, both on and off the template (Figure 4). If the template concentration inside a vesicle is high, most of these oligonucleotides will be generated on the template. On the other hand, a low initial template concentration coupled with a high concentration of activated monomers would favor the non-template-directed formation of short oligomers, which could lower the fidelity of template copying. Short oligomers formed on the template will usually dissociate from the template before being further extended, but may re-bind and react with adjacent monomers or oligomers to generate longer oligomers. In a vesicle-encapsulated replication process, such longer oligomers would accumulate inside the vesicle, increasing in concentration with time. Oligomers of 4 nucleotides and longer are probably permanently trapped inside the vesicle where they were formed,

and even dimers and trimers only slowly leak out at high temperature [42]. Thus, a pool of short and long oligonucleotides, complementary to the template sequence would build up with time, and they could be the major substrates for complete assembly of full-length complementary strands. In this scenario, oligonucleotide ligation together with gap filling reactions involving monomers and shorter oligomers, would lead to product strand synthesis. The final stages of product strand assembly could therefore involve only a few bond-forming steps, and could occur quite rapidly (for an excellent example in a non-RNA model system, see [43,44]). The fidelity of complementary strand synthesis in this manner is unclear. In model experiments with random sequence hexamer DNA oligonucleotides, chemical ligation on a template was found to be surprisingly accurate [45], especially at higher reaction temperatures. Clearly, detailed model studies with RNA should be performed to address the fidelity of RNA template copying with RNA oligonucleotide substrates.

The hierarchical assembly of full-length complementary strands will influence many aspects of the replication process. For example, if oligonucleotide substrates



gradually assemble over cycles of binding, reaction and dissociation, the regiospecificity of longer oligonucleotides may be greater than that of the intrinsic chemical step, as a result of the greater duplex stability of 3'-5' linked RNA and hence preferential incorporation of all 3'-5' oligomers into the final product. Conversely, longer oligonucleotides may be necessary in order to copy template regions containing 2'-5' linkages. Due to the complex mixtures of substrates and products involved, traditional analytical methods (e.g., gel electrophoresis, HPLC) will not be sufficient to characterize such replication reactions; instead, modern high-resolution mass spectrometry methods will be essential.

The rate and extent of duplex assembly through oligonucleotide based ligation chemistry may be substantially increased by the presence of intercalating small molecules [46]. Intercalators can increase the effective affinity of short oligos for templates, acting in effect as a kind of glue that keeps an oligonucleotide substrate bound to a complementary template long enough for it to react with an adjacent monomer or oligonucleotide. A secondary consequence of this effect is a decrease in the rate of oligonucleotide cyclization, since only free single-stranded oligonucleotides can cyclize. Although these studies have been carried out on DNA model systems, it is likely that planar polycyclic aromatic hydrocarbons, perhaps decorated with positively charged functional groups, could act similarly on RNA to enhance the rate of template copying. The model studies described above suggest that template copying might be achieved through a very different pathway than the standard assembly of a complementary strand on the template. If oligonucleotides of varying lengths are gradually built up through partial copying of both complementary template strands, dissociation of these oligonucleotides from

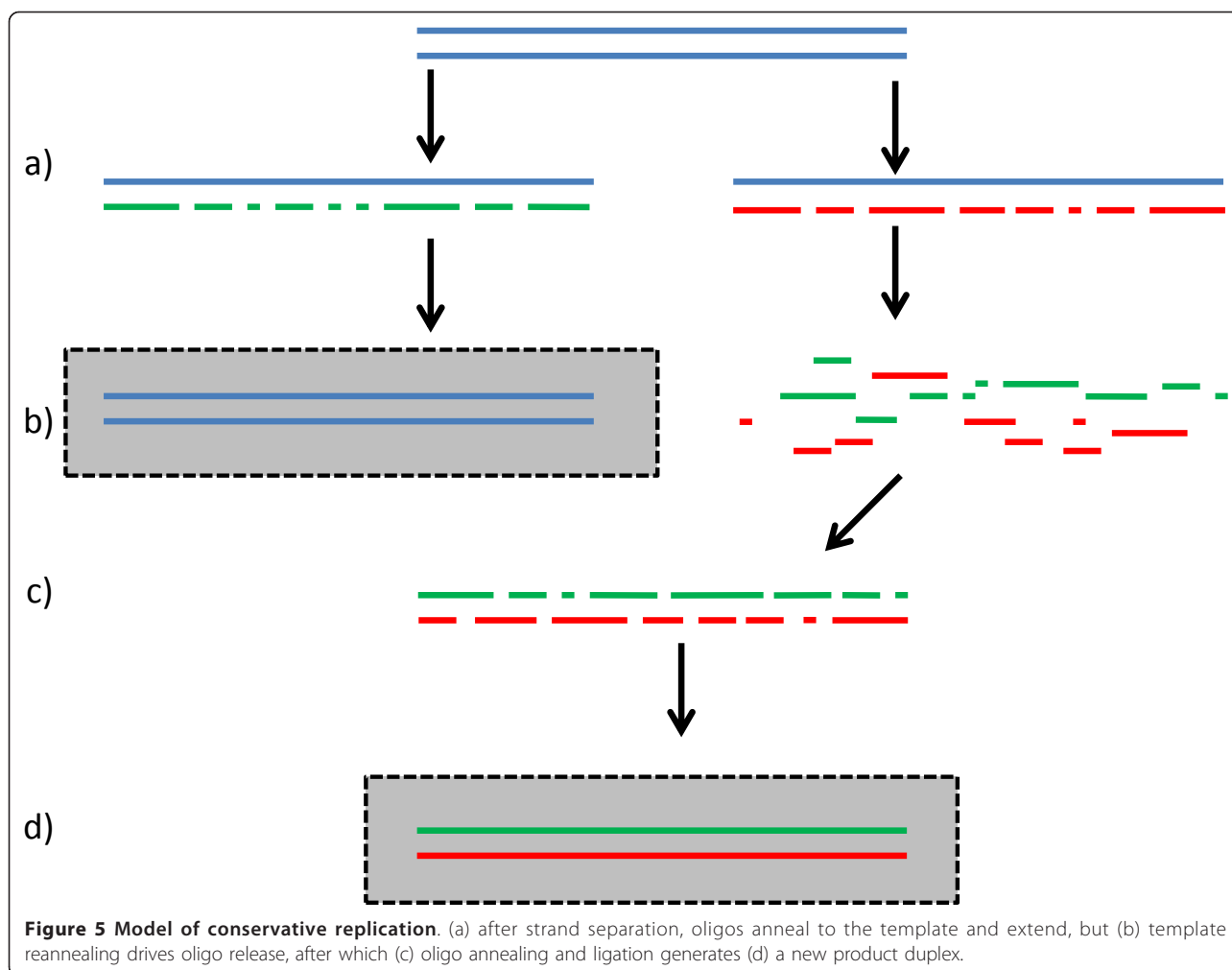
the template, followed by hybridization with each other, and finally ligation, could generate a duplex copy of the original template sequence (Figure 5). This model, for conservative as opposed to semi-conservative replication, has two advantages: first, most oligonucleotides are base-paired most of the time, and are thus protected from alkylation by reactive electrophilic compounds, and second, slow chemical assembly of the duplex need not compete with strand reannealing (see below).

Conformational constraint may provide another route to accelerated template copying chemistry. In a comparison of primer-extension reactions on a series of different templates, the best templates were those that were more conformationally constrained in the direction of a canonical A-type helix with 3'-*endo* monomer units [47]. The polymerization of 2'-amino 5'-phosphorimidazole monomers was, for example, faster on LNA templates than on RNA templates, which were in turn better than DNA or 2'-5' linked DNA templates. Thus, small molecules that constrain RNA conformation to the correct helical geometry may help to properly position incoming monomer or oligonucleotide substrates, thus increasing the rate of template-copying. Furthermore, the cooperativity of double helix formation suggests that template saturation with either monomer or oligomer substrates may favor the correct geometry, leading to enhanced reaction rates, fidelity and regiospecificity [48].

## 2.6. Reactivation chemistry

The hydrolysis and cyclization of chemically activated mononucleotide and oligonucleotide substrates limits the extent of template copying that can be achieved in closed system reactions. For RNA substrates, the rates of hydrolysis and cyclization are comparable to the rates





of template-directed polymerization. Since the copying of a template by primer-extension with monomer substrates involves many sequential reaction steps, it is common in such experiments to use a large excess of monomers over template, to ensure that, even in late stages of the reaction, there is at least some activated monomer left. Unfortunately, hydrolyzed monomers act as competitive inhibitors of the copying reaction, since they can bind to the template but cannot be incorporated into a growing chain [49].

It is clear that a chemical process that could continuously reactivate hydrolyzed substrates, e.g. regenerate phosphorimidazolides from 5'-nucleoside monophosphates, would greatly increase the efficiency of copying reactions and would probably allow for the successful copying of much longer templates. To maintain the pool of substrates in a highly activated state, e.g. at a high ratio of 5'-phosphorimidazolides to 5'-phosphates, a mild and specific reagent that can be introduced continuously to the system is required. Known phosphate activation

chemistries that could operate in the context of an aqueous RNA replication reaction tend to involve highly reactive and non-specific activating agents such as carbodiimides that also alkylate nucleobases [50,51]. While the alkylation of G and U can be minimized by keeping the pH mildly acidic (so that N1 of G and N3 of U remain protonated), polymerization chemistry generally operates best at mildly alkaline pH to facilitate deprotonation of the attacking hydroxyl. Other reagents such as cyanoimidazole, di-imidazole imine, carbonyl-diimidazole, and cyanoacetylene [4] are even more reactive and tend to alkylate mildly nucleophilic groups on both the nucleobases and sugars of both substrates and templates.

An attractive alternative activation strategy is to use a leaving group bearing a thioester functionality that would react rapidly with a phosphate monoester dianion, for the intramolecular delivery of the actual activating group (M. Powner, pers comm.). Alternatively, amino acid N-carboxyanhydrides, generated by reaction of amino acids with cyanate [52] or carbonyl sulfide

[53], have been shown to react with nucleoside 5'-monophosphates to form aminoacyl-carboxyphosphate anhydrides [54]. Such intermediates, although susceptible to rapid hydrolysis, might be trapped by reaction with various substituted imidazoles to yield the more stable phosphorimidazolides, or might lead directly to oligonucleotide ligation or primer-extension in template-directed reactions.

A entirely different strategy is to side-step the issue of re-activation by going from a closed system to an open, flow system. The periodic introduction of fresh, activated substrates, coupled with the removal of both hydrolyzed and cyclized substrates, was used by Ferris et al. [55] to drive the non-template directed synthesis of long oligonucleotides and oligopeptides on mineral surfaces. More recently the same approach has been used to drive the copying of immobilized templates to completion. This has been beautifully demonstrated in recent work from the Richert lab [49], using RNA templates immobilized on a solid support. This strategy could be modified to be compatible with the replication of templates encapsulated within fatty acid vesicles, by using vesicles attached to a surface. Flow over the surface could then result in the exchange of vesicle contents as fresh activated monomers diffused in, while hydrolyzed and cyclized monomers diffused out.

### 2.7. Divalent metal ions

RNA template-directed copying reactions are typically carried out in the presence of  $\geq 0.1$  M  $Mg^{2+}$ . At lower concentrations, the reaction is slower, and little or no reaction is observed in the absence of  $Mg^{2+}$ . Even  $Zn^{2+}$  or  $Mn^{2+}$ , which strongly decrease the  $pK_a$  of bound water or a coordinated sugar hydroxyl, must be used in concentrations of 10-20 mM for optimal reaction. Such high concentrations are problematic because they lead to the degradation of RNA templates by transesterification mediated chain cleavage, i.e. attack of a 2'-OH on the adjacent phosphate, with formation of a 2'-3' cyclic phosphate terminated chain and a 5'-hydroxyl bearing downstream oligonucleotide [56]. In addition, high concentrations of divalent cations are incompatible with the existence of fatty acid based membranes, as a result of the rapid precipitation of the corresponding fatty acid salts at concentrations higher than 2-4 mM divalent  $Mg^{2+}$ .

How could RNA replication be induced to proceed in the presence of much lower concentrations of  $Mg^{2+}$  or other divalent metal ions? One interesting possibility comes from a consideration of the mechanism of ribozymes [57] and protein DNA polymerases [58], which act as phosphoryl transferases in part by binding and properly positioning metal ions so as to catalyze the desired reaction. The single chain viral RNA polymerases coordinate

two catalytic  $Mg^{2+}$  ions, in part through two conserved asp residues that are widely separated on the protein chain [59]. Interestingly, the multi-subunit cellular RNA polymerases also use two  $Mg^{2+}$  ions in the active site, of which the more tightly bound ion is coordinated to three asp residues in a highly conserved NADFDGD peptide sequence [60,61]. It is possible that short random asp/glu rich peptides could chelate  $Mg^{2+}$ , in such a manner that the complex could still function in the catalysis of polymerization. Some such complexes might even be active at a lower concentration because of favorable binding interactions between the chelator or peptide and the RNA reactants. Peptides or even small molecules such as di- or tri-carboxylic acids that coordinate  $Mg^{2+}$  and stabilize the optimal binding geometry may have a significantly enhanced catalytic effect through enhanced acid/base chemistry and/or stabilization of the transition state geometry.

It is notable that aspartic acid is one of the more abundant amino acids found in meteorites and synthesized in Miller-Urey type prebiotic simulations [62,63]. Short random sequence peptides are likely to have been abundant in local environments on the early earth, and may have contributed to the activity of primitive ribozymes. It is an important and open question as to whether such short peptides were in a sense the evolutionary precursors of modern enzymes. In the hope of identifying such catalytic peptides, we are currently screening libraries of short peptides for catalysts of RNA template-copying chemistry.

$Zn^{2+}$  is an interesting alternative to  $Mg^{2+}$  as a catalytic metal ion for RNA polymerization, in large part because many protein enzymes, such as the DNA polymerase 3'-5' exonuclease domain, use a bound  $Zn^{2+}$  ion in their active sites [64].  $Zn^{2+}$  was investigated early on in the Orgel lab, and was found to catalyze ImpG polymerization on oligoC templates with surprisingly good regiospecificity [16]. However,  $Zn^{2+}$  did not catalyze the polymerization of ImpA on polyU templates, possibly due to different coordination of the metal ion with the nucleobase. As a result, the use of  $Zn^{2+}$  has not been further investigated. However, the possibility remains that a  $Zn^{2+}$  chelate or peptide complex could act as a more general catalyst of RNA replication at a concentration low enough to be compatible with membrane integrity.

A completely different approach to the problem of the incompatibility of high divalent metal ion concentrations and fatty acid based membranes would be to examine different membrane compositions, based on non-ionic surfactants that would not be affected by divalent metal ions. For example, prebiotically plausible membranes might be formed from fatty alcohols, esterified to one or more glyceric acid units so as to build up a polyol headgroup. A slightly different amphiphile formed by esterification of diglycerol with myristic acid forms vesicles in water [65].

In addition, lysophospholipids mixed with fatty alcohols form membranes that are more resistant to divalent cations than fatty acid membranes. Clearly, the dynamic properties of such membranes must be examined to assess whether they could form the basis for a viable protocell model in the presence of high levels of divalent cations.

### 2.8. Primer-independent RNA replication

Most laboratory studies of template-copying chemistry make use of a primer, which is annealed to a template and then sequentially extended by monomer addition reactions. While such reactions are simple to analyze, they are prebiotically unrealistic. Even in laboratory studies of model protocells, primers cannot be added continuously because oligonucleotides of four residues or longer cannot cross fatty acid membranes. Furthermore, the initiation of template copying with a primer may be an inefficient approach to RNA replication, since synthesis must proceed sequentially from one end of the template to the other. A more realistic and possibly superior approach would involve nucleation of complementary strand synthesis in multiple locations (e.g., GC rich regions), which could greatly facilitate the copying of longer templates. Completion of replication by gap-filling processes could make use of a complex mixture of monomer and oligomer substrates, with the oligomers being synthesized by both template-directed and template-independent processes, as discussed above.

The initiation of template copying at multiple sites does raise several complicating issues however. Most importantly, the inactivation by hydrolysis of the 5'-end of a growing chain will be a fatal event, in the absence of ongoing re-activation chemistry (see above). Moreover, the reaction of the 5'-phosphate of a partial template copy with an activated monomer (or vice versa) will generate a 5'-5' pyrophosphate linked cap structure, which will terminate replication. It is therefore important to minimize this side reaction, if primer-independent copying is to work efficiently. The easiest way to accomplish this would seem to be through ongoing reactivation chemistry, so that the substrate pool is maintained in a highly activated state with few free 5'-phosphates. Alternatively, the presence of  $Zn^{2+}$ , albeit at fairly high concentrations, has been reported to prevent the hydrolysis of guanosine 5'-phosphorimidazole and the subsequent synthesis of 5'-5' pyrophosphate linked dinucleotides and cap structures [16]. This activity of  $Zn^{2+}$  may result from coordination with N3 of the phosphorimidazole, preventing protonation. If the inhibition of hydrolysis by  $Zn^{2+}$  applies to all nucleotides, and the required  $Zn^{2+}$  concentrations are not excessive, the efficiency of template copying could be significantly increased.

### 2.9. Strand re-annealing

Once an RNA template has been copied, the strands of the resulting duplex must be separated to allow for the next round of template copying. Strand separation could occur during transient high-temperature excursions, resulting from entrainment in the hot water emanating from a lacustrine hydrothermal vent, followed by rapid thermal quenching as the hot effluent mixes with surrounding cold lake or pond water [66,67]. However, the decreasing temperature will initiate strand reannealing, which competes with template copying. The rate of strand reannealing is second order with respect to strand concentration, so this process can in principle be slowed arbitrarily by dilution. In PCR reactions, strand reannealing generally limits product strand concentration to about 1  $\mu$ M; above that concentration, reannealing is faster than primer binding and enzymatic copying, which occur in seconds. The problem is much more severe for chemical replication, which may take hours to days. Unless chemical replication can be drastically speeded up, the maximum feasible strand concentration will be on the order of 1 nM. This might not be a problem for templates replicating in free solution, where the primary selection would be for optimal template structure. However, it is a major concern for vesicle-encapsulated templates, where strand concentrations of  $\sim$ 1 nM correspond to only a few molecules in a vesicle of 3-4 micrometers diameter. At such a low concentration, it is hard to imagine a primitive metabolic ribozyme being able to generate enough of any useful product to have a selectively advantageous effect. For example, the presence of phospholipids in a fatty acid based membrane confers a growth advantage over vesicles that do not contain phospholipids, so a ribozyme that could generate phospholipids could confer a very strong selective advantage [27]. However, for a ribozyme at 1 nM to generate 1 mM of phospholipid (a reasonable local concentration) would require  $10^6$  turnovers; for this to occur in the course of a day ( $\sim 10^5$  s) would require a  $k_{cat}$  of  $10$  s $^{-1}$ , which is not inconceivable but which is unlikely for a primordial, (i.e., unoptimized) ribozyme. On the other hand, if strand concentrations could approach 1  $\mu$ M, a much lower and more reasonable turnover number and  $k_{cat}$  would suffice to confer a selectable effect - but this would require some means of slowing down strand reannealing so as to allow replication to continue at such concentrations. Alternatively a very small selective advantage acting over a long period of time might be sufficient to lead to ribozyme optimization.

There are many possible means for slowing strand reannealing, but almost all of these would generate new problems for either replication or ribozyme activity. For example, template strands with a high degree of secondary structure could fold more rapidly than reannealing,

drastically slowing the rate of reannealing. However, chemical copying through such folded structures is likely to be much slower than on an open, unstructured template, so it is unclear that secondary structure would provide any net advantage.

A possible mechanism for slowing down strand reannealing rates without inhibiting replication would be the binding of complementary oligonucleotides generated during partial template-copying reactions, as discussed above. The rapid binding of such oligonucleotides to a template strand should inhibit reannealing, although the concentration dependence of this effect has not yet been experimentally investigated. This mechanism is appealing since the same oligonucleotides that would slow reannealing are also on-pathway replication intermediates, i.e. the same oligonucleotides could both slow reannealing of complementary strands and speed up the assembly of new full length copies of a template strand. On the other hand, it is also important to investigate experimentally whether high concentrations of such oligonucleotides would inhibit ribozyme function.

### 3. Conclusions

The arguments presented above suggest that some of the perceived difficulties with chemical RNA replication, such as regiospecificity and strand separation, will turn out not to be problems after all. Other difficulties, such as the fidelity of the replication process, have simple potential solutions, such as the substitution of U with 2-thio-U, while other problems, such as the slow rate of template copying, may yield to a combination of several distinct improvements. Although a great deal of experimental work must yet be done, all of the major apparent stumbling blocks on the path to RNA replication appear to be at least potentially amenable to fairly simple solutions. An interesting and somewhat surprising aspect of this conclusion is the frequent coupling between solutions to different problems. Perhaps most strikingly, accepting a heterogeneous backbone containing both 2' and 3' linkages solves the  $T_m$  problem. Similarly, replacing U with 2-thio-U has the potential to solve both the difficulty of copying A-rich templates as well as the poor fidelity caused by wobble pairing. However, the effect of 2'-5' linkages in a template strand on the rate and fidelity of template-directed copying must be evaluated. Finding a mild but effective re-activation chemistry should help not only by increasing the rate of template-copying, but also by minimizing termination events caused by 5'-pyrophosphate capping. An appropriate peptide chelator of  $Mg^{2+}$  or  $Zn^{2+}$  might solve not only the problem of compatibility with fatty acid protocell membranes, but could also decrease template and substrate degradation, and potentially increase rates of ligation or polymerization. Conversely, the failure to solve one problem is likely to

impact many aspects of the replication problem. Such a high degree of interaction and interdependence highlights the fact that the problem of chemical RNA replication is truly a problem in 'systems chemistry', in which all aspects of a complex system must be considered together.

The process of vesicle-encapsulated, chemically-driven RNA replication appears to be compatible with laboratory models of protocell replication, as well as with a specific geochemical scenario that we have previously proposed [67]. A geothermally active region of the early earth that was generally cold could contain numerous lakes and ponds, similar to Yellowstone lake in the USA, and many other environments on the modern earth, in which hydrothermal vents release plumes of hot water into cold lake water [66]. In such an environment, protocells would exist at low temperatures most of the time, during which template copying could occur, punctuated by short intervals at high temperature, leading to strand separation and an influx of nutrients such as nucleotides. Endorheic lakes or ponds could accumulate organic compounds to high levels, especially in geothermally active regions where fatty acids and related compounds might be synthesized by Fischer-Tropsch type chemistry, and high energy carbon-nitrogen compounds could be synthesized as a result of electrical discharges surrounding active volcanoes. Sulfurous exhalations such as COS and  $H_2S$  could be important for the synthesis of thioesters or N-carboxyanhydrides for re-activation chemistry, and for the synthesis of modified nucleosides such as 2-thio-U for improved rate and fidelity of RNA replication.

If the first replicating protocells relied on a chemically replicating RNA genome, it is likely that the first ribozyme to evolve carried out a metabolic function, such as phospholipid synthesis, that conferred an advantage on its host cell [27]. As soon as that occurred, there would be strong selective pressure for improved replication of the sequence encoding that ribozyme. It is interesting to speculate as to the types of ribozyme-catalyzed activities that could lead to enhanced replication efficiency in such a scenario [68]. For example, nuclease ribozymes that trimmed off terminal mismatched nucleotides from growing oligonucleotide chains would increase both fidelity and the rate of replication (because of the post mismatch stalling effect). Other nucleases that trimmed overlapping oligonucleotides would facilitate chemical ligation, again increasing both rate and fidelity. Phosphodiesterases that regenerated 5'-nucleoside monophosphates from 3'-5' cyclic nucleotides would increase the pool of monomer substrates, while nucleases that cleaved cyclic oligonucleotides would increase the available pools of oligonucleotide substrates. Other ribozymes might catalyze the synthesis of 5'-activated phosphates from available activating agents, while a hydrolase could recycle 5'-pp-5' dinucleotides into viable

substrates. Finally, of course, ligases and polymerases could contribute to the replication process, but there is no reason to think that such activities must have come first. The diversity of activities that could increase the efficiency of genomic replication suggest that selection for enhanced replication efficiency would have driven a rapid complexification of the biochemical machinery of early protocells.

By considering a series of factors that may have allowed for the chemical replication of RNA, we have come to a novel picture of the emergence of the RNA World from prebiotic chemistry. Most significantly, the RNA of the early RNA World was not modern RNA - it was a more primitive version, with a heterogeneous backbone containing mixed 3'-5' and 2'-5' linkages, and it may also have exhibited subtle chemical differences in its nucleobases, such as sulfur-substituted pyrimidines. Perhaps the long hypothesized progenitor nucleic acid [69] that would be easier to synthesize and to replicate than canonical RNA is really no more than a slightly modified version of standard RNA. This would solve the greatest problem with the progenitor nucleic acid hypothesis, which is the difficult nature of the genetic takeover of a distinct nucleic acid by RNA. With such a similar progenitor, it is much easier to understand how sequence-encoded functionality could be maintained during the takeover. At the same time the takeover itself is easier to understand, since all that would be required is a gradually more sophisticated replication apparatus, consisting of a replicase that would impose regiospecificity and fidelity, and enzymatic machinery for strand separation or strand-displacement synthesis. The advantages of producing a homogeneous genetic material from abundant building blocks could have driven the transition to modern RNA, in parallel with the increasing complexity of the metabolic and replication machinery.

#### Acknowledgements

I thank Noam Prywes, Aaron Englehart, Matt Powner, Anders Bjorkbom, Irene Chen and Itay Budin for helpful comments on the manuscript, and all of the current and former members of my laboratory for helpful discussions.

#### Competing interests

The authors declare that they have no competing interests.

Received: 23 November 2011 Accepted: 3 February 2012

Published: 3 February 2012

#### References

- Joyce GF: RNA evolution and the origins of life. *Nature* 1989, **338**:217-224.
- Joyce GF, Orgel LE: Prospects for understanding the origin of the RNA world. In *The RNA World*. Edited by: Gesteland RF, Atkins JF. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1993:1-25.
- Orgel LE: Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* 2004, **39**:99-123.
- Powner MW, Gerland B, Sutherland JD: Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 2009, **459**:239-242.
- Powner MW, Sutherland JD, Szostak JW: Chemoselective multicomponent one-pot assembly of purine precursors in water. *J Am Chem Soc* 2010, **132**:16677-88, Erratum in: *J Am Chem Soc* 2011, **133**:4149-4150.
- Powner MW, Sutherland JD: Prebiotic chemistry: a new modus operandi. *Philos Trans R Soc Lond B Biol Sci* 2011, **366**:2870-2877.
- Szostak JW, Bartel DP, Luisi PL: Synthesizing life. *Nature* 2001, **409**:387-390.
- Szabó P, Scheuring I, Czárán T, Szathmáry E: In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* 2002, **420**:340-343.
- Mansy SS, Schrum JP, Krishnamurthy M, Tobé S, Treco DA, Szostak JW: Template-directed synthesis of a genetic polymer in a model protocell. *Nature* 2008, **454**:122-125.
- Zhu TF, Szostak JW: Coupled growth and division of model protocell membranes. *J Am Chem Soc* 2009, **131**:5705-5713.
- Gilbert W: Origin of life: The RNA world. *Nature* 1986, **319**:618.
- Zaher HS, Unrau PJ: Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* 2007, **13**:1017-26.
- Wochner A, Attwater J, Coulson A, Holliger P: Ribozyme-catalyzed transcription of an active ribozyme. *Science* 2011, **332**:209-12.
- Acevedo OL, Orgel LE: Non-enzymatic transcription of an oligodeoxynucleotide 14 residues long. *J Mol Biol* 1987, **197**:187-193.
- Orgel LE: Molecular Replication. *Nature* 1992, **358**:203-209.
- Bridson PK, Orgel LE: Catalysis of Accurate Poly(C)-directed Synthesis of 3'-5'-linked Oligoguanylates by Zn<sup>2+</sup>. *J Mol Biol* 1980, **144**:567-577.
- Inoue T, Orgel LE: Oligomerization of (Guanosine 5'-phosphor)-2-methylimidazole on Poly(C): An RNA Polymerase Model. *J Mol Biol* 1982, **162**:201-217.
- Trevino SG, Zhang N, Elenko MP, Lupták A, Szostak JW: Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity. *Proc Natl Acad Sci USA* 2011, **108**:13492-13497.
- Lauhon CT, Szostak JW: RNA aptamers that bind flavin and nicotinamide redox cofactors. *J Am Chem Soc* 1995, **117**:1246-1257.
- Dieckmann T, Butcher SE, Sasanfar M, Szostak JW, Feigon J: Mutant ATP-binding RNA aptamers reveal the structural basis for ligand binding. *J Mol Biol* 1997, **273**:467-478.
- Travascio P, Bennet AJ, Wang DY, Sen D: A ribozyme and a catalytic DNA with peroxidase activity: Active sites versus cofactor-binding sites. *Chem Biol* 1999, **6**:779-787.
- Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH: Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci USA* 1986, **83**:9373-9377.
- Kierzek R, He L, Turner DH: Association of 2' - 5' oligoribonucleotides. *Nuc Acids Res* 1992, **20**:1685-1690.
- Giannaris PA, Damha MJ: Oligoribonucleotides containing 2', 5'-phosphodiester linkages exhibit binding selectivity for 3', 5'-RNA over 3', 5'-ssDNA. *Nuc Acids Res* 1993, **21**:4742-4749.
- Wasner M, Arion D, Borkow G, Noronha A, Uddin AH, Parniak MA, Damha MJ: Physicochemical and biochemical properties of 2', 5'-linked RNA and 2', 5'-RNA:3', 5'-RNA "hybrid" duplexes. *Biochem* 1998, **37**:7478-7486.
- Eigen M: Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 1971, **58**:465-523.
- Budin I, Szostak JW: Physical effects underlying the transition from primitive to modern cell membranes. *Proc Natl Acad Sci USA* 2011, **108**:5249-5254.
- Ferré-D'Amaré AR, Scott WG: Small self-cleaving ribozymes. *Cold Spring Harb Perspect Biol* 2010, **2**:a003574.
- Kun A, Santos M, Szathmáry E: Real ribozymes suggest a relaxed error threshold. *Nat Genet* 2005, **37**:1008-1011.
- Leu K, Obermayer B, Rajamani S, Gerland U, Chen IA: The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. *Nuc Acids Res* 2011, **39**:8135-8147.
- Ichida JK, Horhota A, Zou K, McLaughlin LW, Szostak JW: High fidelity TNA synthesis by Terminator polymerase. *Nuc Acids Res* 2005, **33**:5219-5225.
- Huang MM, Arnheim N, Goodman MF: Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nuc Acids Res* 1992, **20**:4567-4573.
- Rajamani S, Ichida JK, Antal T, Treco DA, Leu K, Nowak MA, Szostak JW, Chen IA: Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J Am Chem Soc* 2010, **132**:5880-5885.

34. Testa SM, Disney MD, Turner DH, Kierzek R: **Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: implications for antisense design and targeting a group I intron.** *Biochem* 1999, **38**:16655-16662.
35. Caton-Williams J, Huang Z: **Biochemistry of selenium-derivatized naturally occurring and unnatural nucleic acids.** *Chem Biodivers* 2008, **5**:396-407.
36. Hassan AE, Sheng J, Zhang W, Huang Z: **High fidelity of base pairing by 2-selenothymidine in DNA.** *J Am Chem Soc* 2010, **132**:2120-2121.
37. Ajitkumar P, Cherayil JD: **Thionucleosides in transfer ribonucleic acid: diversity, structure, biosynthesis, and function.** *Microbiol Rev* 1988, **52**:103-113.
38. Siegfried NA, Kierzek R, Bevilacqua PC: **Role of unsatisfied hydrogen bond acceptors in RNA energetics and specificity.** *J Am Chem Soc* 2010, **132**:5342-4.
39. Vogel SR, Deck C, Richert C: **Accelerating chemical replication steps of RNA involving activated ribonucleotides and downstream-binding elements.** *Chem Commun (Camb)* 2005, **39**:4922-4924.
40. Vogel SR, Richert C: **Adenosine residues in the template do not block spontaneous replication steps of RNA.** *Chem Commun (Camb)* 2007, **21**:1896-1898.
41. Diop-Frimpong B, Prakash TP, Rajeev KG, Manoharan M, Egli M: **Stabilizing contributions of sulfur-modified nucleotides: crystal structure of a DNA duplex with 2'-O-[2-(methoxy)ethyl]-2-thiothymidines.** *Nuc Acids Res* 2005, **33**:5297-307.
42. Mansy SS, Szostak JW: **Thermostability of model protocell membranes.** *Proc Natl Acad Sci USA* 2008, **105**:13351-13355.
43. Li X, Zhan ZY, Knipe R, Lynn DG: **DNA-catalyzed polymerization.** *J Am Chem Soc* 2002, **124**:746-747.
44. Li X, Hernandez AF, Grover MA, Hud NV, Lynn DG: **Step-growth control in template-directed polymerization.** *Heterocycles* 2011, **82**:1477-1488.
45. James KD, Ellington AD: **Surprising fidelity of template-directed chemical ligation of oligonucleotides.** *Chem Biol* 1997, **4**:595-605.
46. Horowitz ED, Engelhart AE, Chen MC, Quarles KA, Smith MW, Lynn DG, Hud NV: **Intercalation as a means to suppress cyclization and promote polymerization of base-pairing oligonucleotides in a prebiotic world.** *Proc Natl Acad Sci USA* 2010, **107**:5288-5293.
47. Schrum JP, Ricardo A, Krishnamurthy M, Blain JC, Szostak JW: **Efficient and rapid template-directed nucleic acid copying using 2'-amino-2', 3'-dideoxyribonucleoside-5'-phosphorimidazole monomers.** *J Am Chem Soc* 2009, **131**:14560-14570.
48. Rohatgi R, Bartel DP, Szostak JW: **Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3'-5' phosphodiester bonds.** *J Am Chem Soc* 1996, **118**:3340-3344.
49. Deck C, Jauker M, Richert C: **Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA.** *Nat Chem* 2011, **3**:603-8.
50. Gilham PT: **An addition reaction specific for uridine and guanosine nucleotides and its application to the modification of ribonuclease action.** *J Am Chem Soc* 1962, **84**:687-688.
51. Chu BCF, Wahl GM, Orgel LE: **Derivatization of unprotected polynucleotides.** *Nuc Acids Res* 1983, **11**:6513-6529.
52. Biron JP, Pascal R: **Amino acid N-carboxyanhydrides: activated peptide monomers behaving as phosphate-activating agents in aqueous solution.** *J Am Chem Soc* 2004, **126**:9198-9199.
53. Leman L, Orgel L, Ghadiri MR: **Carbonyl sulfide-mediated prebiotic formation of peptides.** *Science* 2004, **306**:283-286.
54. Leman LJ, Orgel LE, Ghadiri MR: **Amino acid dependent formation of phosphate anhydrides in water mediated by carbonyl sulfide.** *J Am Chem Soc* 2006, **128**:20-21.
55. Ferris JP, Hill AR Jr, Liu R, Orgel LE: **Synthesis of long prebiotic oligomers on mineral surfaces.** *Nature* 1996, **381**:59-61.
56. Butzow JJ, Eichhorn GL: **Interaction of metal ions with nucleic acids and related compounds. XVII. On the mechanism of degradation of polyribonucleotides and oligoribonucleotides by zinc(II) ions.** *Biochemistry* 1971, **10**:2019-27.
57. Steitz TA, Steitz JA: **A general two-metal-ion mechanism for catalytic RNA.** *Proc Natl Acad Sci USA* 1993, **90**:6498-6502.
58. Brautigam CA, Steitz TA: **Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes.** *Curr Opin Struct Biol* 1998, **8**:54-63.
59. Woody AY, Eaton SS, Osumi-Davis PA, Woody RW: **Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription.** *Biochem* 1996, **35**:144-152.
60. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA: **Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution.** *Cell* 1999, **98**:811-824.
61. Cramer P, Bushnell DA, Kornberg RD: **Structural basis of transcription: RNA polymerase II at 2.8 Å resolution.** *Science* 2001, **292**:1863-1876.
62. Glavin DP, Bada JL, Brinton KL, McDonald GD: **Amino acids in the Martian meteorite Nakhla.** *Proc Natl Acad Sci USA* 1999, **96**:8835-8838.
63. Miller SL: **Which organic compounds could have occurred on the prebiotic earth?** *Cold Spring Harb Symp Quant Biol* 1987, **52**:17-27.
64. Wang J, Yu P, Lin TC, Konigsberg WH, Steitz TA: **Crystal structures of an NH<sub>2</sub>-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions.** *Biochem* 1996, **35**:8110-8119.
65. Shrestha LK, Shrestha RG, Iwanaga T, Aramaki K: **Aqueous Phase Behavior of Diglycerol Fatty Acid Esters.** *J Disp Sci Tech* 2007, **28**:883-891.
66. Morgan LA, Shanks WC III, Lovalvo DA, Johnson SY, Stephenson WJ, Pierce KL, Harlan SS, Finn CA, Lee G, Webring M, Schulze B, Dühn J, Sweeney R, Balistrieri L: **Exploration and discovery in Yellowstone Lake: results from high-resolution sonar imaging, seismic reflection profiling, and submersible studies.** *J Volc Geo Res* 2003, **122**:221-242.
67. Ricardo A, Szostak JW: **Origin of life on earth.** *Sci Am* 2009, **301**:54-61.
68. Szostak JW: **An optimal degree of physical and chemical heterogeneity for the origin of life?** *Philos Trans R Soc Lond B Biol Sci* 2011, **366**:2894-2901.
69. Joyce GF, Schwartz AW, Miller SL, Orgel LE: **The case for an ancestral genetic system involving simple analogues of the nucleotides.** *Proc Natl Acad Sci USA* 1987, **84**:4398-4402.

doi:10.1186/1759-2208-3-2

**Cite this article as:** Szostak: The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry* 2012 **3**:2.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
<http://www.chemistrycentral.com/manuscript/>

