J Mol Evol
DOI 10.1007/s00239-013-9584-x

LETTER TO THE EDITOR

The Paradox of Dual Roles in the RNA World: Resolving the Conflict Between Stable Folding and Templating Ability

Nikola A. Ivica · Benedikt Obermayer · Gregory W. Campbell · Sudha Rajamani · Ulrich Gerland · Irene A. Chen

Received: 10 June 2013 / Accepted: 12 September 2013

Abstract The hypothesized dual roles of RNA as both information carrier and biocatalyst during the earliest stages of life require a combination of features: good templating ability (for replication) and stable folding (for ribozymes). However, this poses the following paradox: well-folded sequences are poor templates for copying, but poorly folded sequences are unlikely to be good ribozymes. Here, we describe a strategy to overcome this dilemma through G:U wobble pairing in RNA. Unlike Watson–Crick base pairs, wobble pairs contribute highly to the energetic stability of the folded structure of their sequence, but only slightly, if at all, to the stability of the folded reverse complement. Sequences in the RNA World might thereby combine stable folding of the ribozyme with an unstructured, reverse-complementary genome, resulting in a "division of labor" between the strands. We demonstrate this strategy using computational simulations of RNA folding and an experimental model of early replication, nonenzymatic template-directed RNA primer extension. Additional study is needed to solve other problems associated with a complete replication cycle, including separation of strands after copying. Interestingly, viroid RNA sequences, which have been suggested to be relics of an RNA World (Diener, Proc Natl Acad Sci USA 86:9370–9374, 1989), also show significant asymmetry in folding energy between the infectious (+) and template (−) strands due to G:U pairing, suggesting that this strategy may even be used by replicators in the present day.

Keywords Origin of life · RNA World · Nonenzymatic polymerization · Replication

Electronic supplementary material The online version of this article (doi:10.1007/s00239-013-9584-x) contains supplementary material, which is available to authorized users.

N. A. Ivica · S. Rajamani · I. A. Chen (✉)
FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA
e-mail: chen@chem.ucsb.edu

B. Obermayer
Department of Physics, Harvard University, Cambridge, MA, USA

Present Address:
B. Obermayer
Max-Delbruck Center for Molecular Medicine, Berlin, Germany

G. W. Campbell · I. A. Chen
Department of Chemistry and Biochemistry and Program in Biomolecular Sciences and Engineering, University of California, Santa Barbara, CA, USA

U. Gerland
Department of Physics, Ludwig-Maximillian University, Munich, Germany

Introduction

The ability of RNA to act both as genetic material and as a biocatalyst suggests that a simple version of life might be composed primarily of RNA. The origin of life is believed to have progressed through such a stage (the "RNA World") (Crick 1968; Orgel 1968; Gesteland et al. 2006). The plausibility of this theory is bolstered by several experimental findings, including the centrality of RNA nucleotides in metabolism, the nature of the ribosome (a ribozyme) (Ban et al. 2000; Wimberly et al. 2000; Yusupov et al. 2001), and the abiotic synthesis of RNA nucleosides (Powner et al. 2009). RNA-based biochemistry also represents an attractive platform for minimal synthetic self-replicating life since there is no need to support DNA and

Published online: 29 September 2013

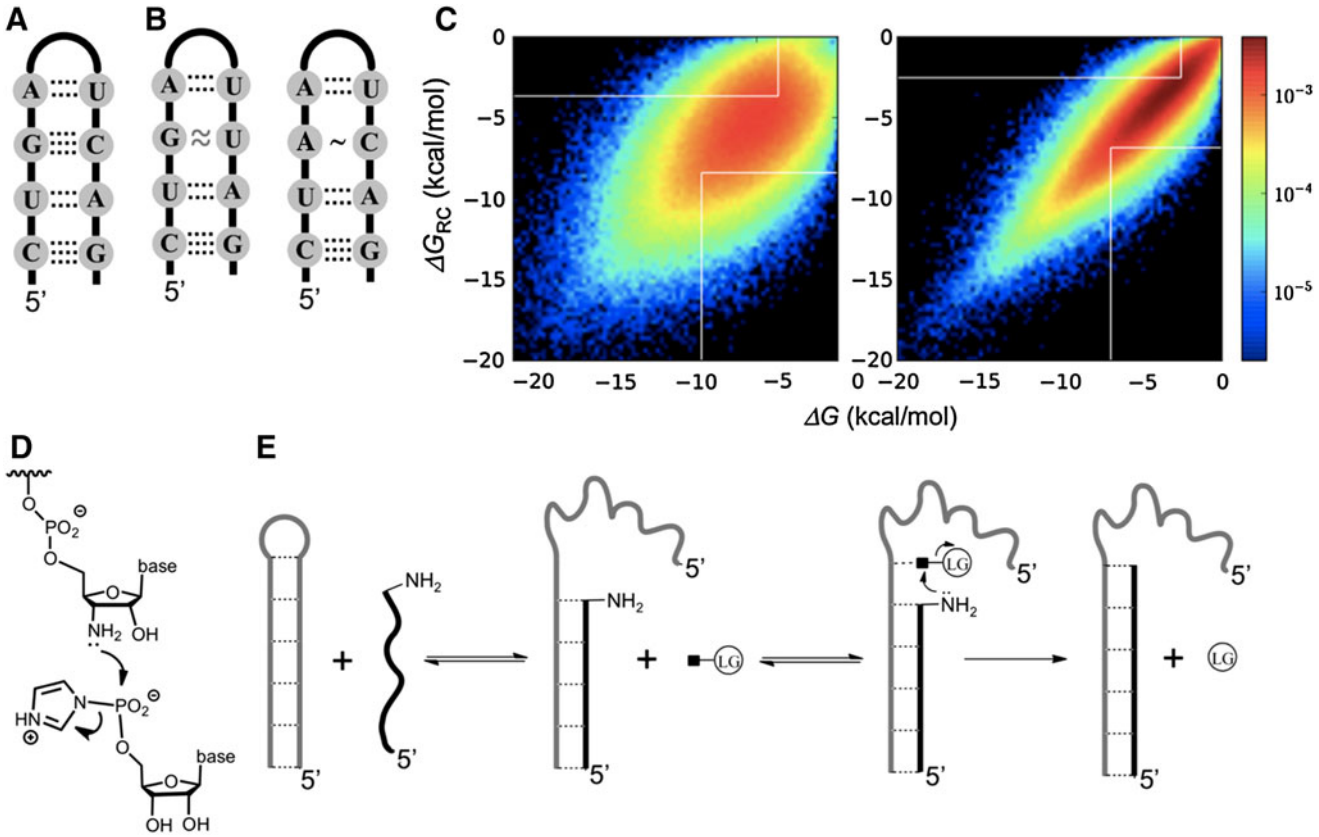

Fig. 1 Asymmetry in self-folding energies of reverse-complementary RNA strands and experimental system. **a** Palindromic stem obeying only Watson–Crick pairing rules has a reverse complement with equivalent sequence and folding energy. **b** Reverse complements; strand on left contains a G:U wobble pair. **c** Heat map histogram calculated for random 35-mers folding in silico. The *x*-axis is the ΔG of one strand; the *y*-axis is the ΔG of the reverse complement (ΔG_{RC}). Histogram color represents the fraction of sequences with the given ΔG and ΔG_{RC}; *red* is high, *black* is low (see legend). G:U wobble pairs were either allowed (*left*) or disallowed (*right*). The *boxed areas* of each plot correspond to asymmetric sequences. **d** Experimental system for nonenzymatic template-directed primer extension. Reactants: primer and activated nucleotide. Reaction proceeds much faster in the presence of a template. **e** Diagram of reaction using a structured template. The free energy of the first step is ΔG_1. *LG* leaving group

protein synthesis (Szostak et al. 2001; Muller 2006; Joyce 2009; Cheng and Unrau 2010). Several challenges must be overcome to create an experimental model of RNA World replication, including the so-called "error catastrophe" and the difficulty of separating strands of moderate length (Szostak 2013). One unresolved paradox for the dual roles of RNA in RNA-based life arises from a conflict between templating ability and stable folding (Szabo et al. 2002; Muller 2006). Well-folded structures are assumed to be a prerequisite for catalytic activity. On the other hand, an RNA sequence must presumably exist in an extended conformation that is accessible to a primer and substrate nucleotides in order to act as a template for copying. Thus, good template sequences lack structure, while ribozymes tend to adopt well-folded structures, so these two desirable properties seem to be mutually exclusive.

One possible strategy to overcome this paradox is "division of labor" between the two strands of the RNA, with one strand folding into a ribozyme and the other functioning as genome (Chen et al. 2006). However, at first glance, the folding capacity of a given RNA sequence and its complement are likely to be similar at the secondary structure level, since a given structural element is palindromic in the stem portion (Fig. 1a). Therefore, the complement of a well-folded ribozyme would also be expected to fold well. We hypothesized that this symmetry could be broken by wobble base pairing, particularly G:U, whose reverse complement would produce the less stable C:A mispair (Turner and Mathews 2010) (Fig. 1b). We verified this hypothesis with folding simulations. In addition, we used an experimental model for early RNA replication, nonenzymatic template-directed synthesis, to demonstrate that although folding is anti-correlated with templating ability, primer extension using the reverse complement of a well-folded sequence as template was $\sim$100-fold faster given high wobble pairing content. This could increase fitness compared to genome–ribozyme systems with symmetric low folding energies. Finally, we test our hypothesis

on a modern-day version of RNA replicators (viroids) and confirm that these parasitic sequences exhibit folding asymmetry between (+) and (−) strands due to G:U wobble pairing. Although other challenges remain for an experimental model of RNA World replication, including the related problem of strand separation, our results suggest that the paradox of ribozyme templates can be alleviated by wobble pairing, a prominent feature of RNA.

Results and Discussion

Folding Simulations of RNA Sequences and Their Complements

Using the Vienna RNA Package (Hofacker et al. 1994), we folded 500,000 random 35 mers with equal representation of all four nucleotides and recorded the free energy (ΔG) of the ensemble of structures for each sequence relative to the unfolded state (McCaskill 1990). To test our hypothesis that wobble pairing enhances folding asymmetry, we compared the histogram of ΔG for sequences and their complements, in which G:U pairing is either allowed or disallowed (Fig. 1c). Greater density can be found in the top left and bottom right regions, corresponding to asymmetric sequences, and greater dispersion of values is found in general, when G:U pairing is allowed. "Division of labor" might occur in an asymmetric sequence, which has a stable structure but whose reverse complement has an unstable structure. If we define the 20 % with lowest free energy as stably folded sequences ($\Delta G_{low} < -8.4$ kcal/mol) and the 20 % with highest free energy as unstable sequences ($\Delta G_{high} > -3.7$ kcal/mol), then about 1.5 % of random sequences are asymmetric by this definition. However, if we disallow G:U wobble pairs in the folding algorithm, the proportion of asymmetric sequences is 20-fold smaller (0.08 %, by the percentile definition). Alternatively, if we define asymmetry by ΔG_{low} and ΔG_{high} (rather than percentiles), then 0.05 % of sequences are asymmetric if G:U pairs are disallowed.

Folding is Anti-correlated with Extension Rate of Nonenzymatic Template-Directed Synthesis

To investigate the effect of folding on early RNA replication, we used an experimental model system for nonenzymatic RNA primer extension on an RNA template, which can occur on a laboratory time scale due to the increased nucleophilicity of the 3′-terminal residue of the primer (2′,3′-dideoxy, 3′-aminoguanosine) and increased lability of the leaving group of the activated nucleotide (ImpG) (Fig. 1d) (Weimann et al. 1968; Orgel 2004; Schrum et al. 2010; Leu et al. 2011; Leu et al. 2013). If the template itself can fold, the primer–template complex presumably exists in equilibrium with self-associated template and free primer according to the free energy difference (ΔG_1). Association of the activated nucleotide poises the complex for reaction, and bond formation is essentially irreversible (Fig. 1e). We designed 13 template sequences (I–XIII) to anneal to the same primer (P), covering a range of free energies from poorly folded sequences representing good templates, to well-folded sequences representing sequences with ribozyme potential (Table 1). We first compared the experimentally estimated proportion of primer in the hybridized state to the thermodynamic

Table 1 Predicted thermodynamic properties and experimental extension rate (k) for different template sequences

Template	ΔG (kcal/mol)	ΔG_1 (kcal/mol)	Hybridization (%)	Pre-reaction complex (%)	k (h^{-1})
I	−58.9	0.6	<1 %	<1 %	0.21 ± 0.02
II	−45.4	3.4	<1 %	<1 %	0.16 ± 0.05
III	−38.5	−1.3	<1 %	<1 %	0.11 ± 0.1
IV	−32.5	−8.8	57	<1 %	0.13 ± 0.02
V	−26.2	−9.7	78	<1 %	0.23 ± 0.03
VI	−15.3	−18.8	100	**71**	0.32 ± 0.04
VII	−13.4	−19.8	100	**99**	4.0 ± 0.1
VIII	−15.2	−20	100	**98**	4.9 ± 0.4
IX	−14.1	−12.5	98	**95**	2.0 ± 0.2
X	−26.1	−8.8	58	**52**	0.85 ± 0.07
XI	−13.8	−16	93	**55**	0.26 ± 0.08
XII	−12.2	−19	100	27	0.06 ± 0.05
XIII	−10.9	−19.7	100	**90**	1.1 ± 0.04
Asym	−25.6	−20.9	100	32	0.07 ± 0.04
Asym-RC	−11.2	−24.4	100	**90**	6.7 ± 1.5
V-RC	−25.6	−8	36	<1 %	0.17 ± 0.04

V-RC is the reverse complement of V. Asym is a sequence designed to have enhanced asymmetry in self-folding energy (ΔG) compared to its reverse complement, Asym-RC. Templates predicted to have "accessible" pre-reaction complexes are shown in bold. Standard deviation of multiple replicates is given for extension rates

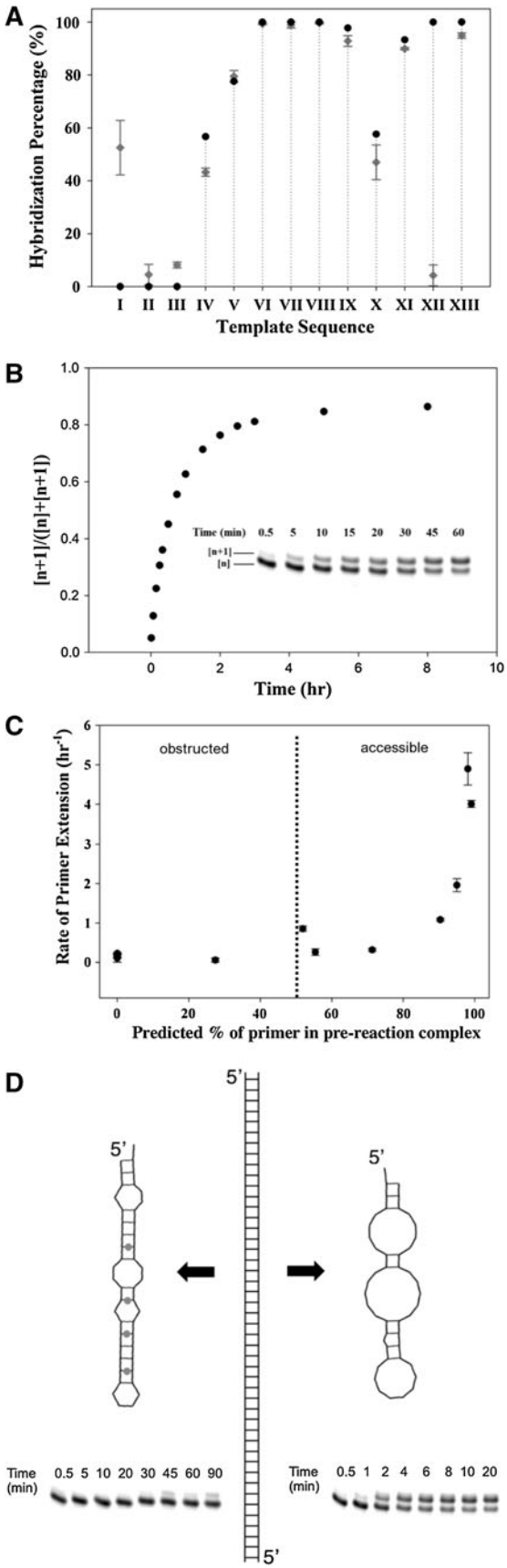

Fig. 2 Template–primer hybridization, nonenzymatic primer extension, and asymmetric sequence. **a** Experimentally determined percentage of primer in a complex (*gray diamonds*) and theoretically predicted hybridization percentage (*black circles*). *Error bars* represent standard deviation. **b** Example reaction of primer with template XIII; denaturing PAGE is shown in the inset. **c** Rate of primer extension on templates with different propensities for complex formation. **d** Experimentally tested asymmetric sequence. The predicted structure of Asym (*left*) and Asym-RC (*right*) are shown with *inset* gels of extension reactions. G:U wobble pairs are indicated by the *gray dots*

prediction (Fig. 2a). These were in general agreement; exceptions may be due to kinetic trapping or the presence of higher order complexes that are not considered in the RNA folding algorithm.

Using our model reaction system, we measured the extension rate for each template (I–XIII) with primer P (Fig. 2b; Table 1). Using the folding algorithm subject to constraints, we also calculated the thermodynamic expectation for the proportion of primer in the pre-reaction complex that should facilitate the extension reaction (i.e., the primer is fully hybridized to the template but the next template base is unpaired; Table 1). While computational RNA secondary structure folding is not sufficient to quantitatively predict extension rates, we found that the predicted proportion of primer in pre-reaction complexes for I–XIII correlated with extension rates (Fig. 2c; Spearman's rank correlation $\rho = 0.85$, $p = 2.2 \times 10^{-4}$), consistent with structural accessibility being a prerequisite for reaction. We categorized the templates into two classes based on their calculated thermodynamic properties: "obstructed" (<50 % of primer in pre-reaction complex) and "accessible" (≥50 % of primer in pre-reaction complex). A clear difference in experimentally observed extension rates was observed between these classes (Fig. 2c; Supplementary Fig. 1), although extension rates of "accessible" sequences are quite variable, presumably due to factors other than secondary structure (e.g., kinetic effects). In general, as expected, templates able to self-fold well had very low extension rates compared to relatively unfolded templates. Interestingly, even these slow rates were significantly higher than the reaction in the absence of template ($0.04 \pm 0.02\ h^{-1}$), suggesting that substantial template-directed synthesis can occur even on quite stably folded sequences.

Asymmetry in Copying Rates of Reverse Complements Based on Wobble Pairing

To test our hypothesis that increased G:U pairing in one strand would result in decreased relative stability, and therefore better templating, for its reverse complement, we designed a highly structured RNA sequence, Asym, with

four G:U wobble pairs ($\Delta G = -26$ kcal/mol; Fig. 2d). Its reverse complement, Asym-RC, had no G:U wobble pairs ($\Delta G = -11$ kcal/mol). While one could imagine a more extremely asymmetric sequence pair (e.g., entirely composed of G and U), lower complexity sequences are more prone to multiple misfolded states and might have been disfavored by prebiotic synthetic processes (Derr et al. 2012). The extension rates of Asym and Asym-RC, using primer P-Asym and P-Asym-RC, respectively, differed by a factor of 100 (Table 1; Fig. 2d), illustrating that the use of G:U pairs did result in a large asymmetry between the templating ability of the two strands. The extension rate of Asym was close to the non-templated rate of extension, suggesting that the observed difference is near the maximum that would be detectable in this assay. This difference implies that a good template sequence (genome) could encode a well-folded sequence (ribozyme) despite the palindromic nature of Watson–Crick base pairing. As a control, we compared template V ($\Delta G = -26$ kcal/mol; no G:U pairs) with its reverse complement V-RC ($\Delta G = -26$ kcal/mol; no G:U pairs); primer extension from P and P-V-RC, respectively, occurred at similar rates (Table 1).

Asymmetry in Folding Stability of RNA Viroid Strands Based on Wobble Pairing

Our hypothesis, that increased asymmetry in folding energy through G:U wobble pairing would improve fitness, should also apply to RNA-based replicators in modern biology. We tested this hypothesis on viroids, which are small, non-coding circular RNAs that infect plant cells (Navarro et al. 2012). In general, the (+) RNA strand of a viroid enters the host cell and is transcribed by host enzymes in rolling-circle amplification to produce an oligomeric (−) RNA strand. The (−) strand serves as the template (either directly or after processing to a monomeric species) for transcription to produce oligomeric (+) strands, which are processed to generate progeny infectious monomeric (+) strands. Therefore, we expect the (−) strand to be relatively optimized for templating while the (+) strand is optimized for folding-dependent functions related to infectivity.

We analyzed a set of 40 currently known viroids (Rocheleau and Pelchat 2006) to determine whether the (−) strand was poorly folded compared to the (+) strand. We calculated the difference in predicted folding energies ($\Delta\Delta G_{\text{strand}} = \Delta G_{(+)} - \Delta G_{(-)}$), with G–U pairing allowed, for each viroid and compared the distribution of $\Delta\Delta G_{\text{strand}}$ to that of control sets with identical base composition but randomized sequence (Fig. 3a). As expected, the $\Delta\Delta G_{\text{strand}}$ of viroids was shifted to the left and the distribution differed significantly from that of the controls (2-sample Kolmogorov–Smirnov test: $p = 0.0084$). Although the comparison between $\Delta\Delta G_{\text{strand}}$ for each viroid and its matched control sets can be noisy, the effect can be easily seen when summing $\Delta\Delta G_{\text{strand}}$ over the entire viroid set ($\Sigma(\Delta\Delta G_{\text{strand}})$ for viroids vs. controls; Fig. 3b).

To determine whether the observed asymmetry between viroid strands was due to increased G:U wobble pairing in the (+) strand, we calculated the free energy contribution of G:U wobble pairing ($\Delta G_{\text{with G:U}} - \Delta G_{\text{no G:U}} = \Delta\Delta G_{\text{G:U}}$) for each strand. If our hypothesis were correct, the (+) strand would exhibit greater stabilization from G:U pairing than the (−) strand. We calculated the stabilization ratio ($R_{\text{G:U}} = \Delta\Delta G_{\text{G:U,(+)}}/\Delta\Delta G_{\text{G:U,(−)}}$) and found that G:U wobble pairing indeed contributed greater stability to the (+) strand (Fig. 3c). Furthermore, the stabilization ratio correlated well with asymmetry between the strands (Fig. 3d), suggesting that G:U wobble pairing causes the asymmetry in folding between strands. If $\Delta\Delta G_{\text{strand}}$ is calculated without allowing G:U pairs, only a slight difference is seen between the distribution of viroids and that of controls, suggesting that while there may be a small degree of stabilization from other factors, most of the effect is attributable to G:U pairs (Supplementary Fig. 2).

Concluding Remarks

The perceived tradeoff between stable folding and templating ability has presented a paradox for the dual roles of RNA in the origin of life. The effect of base composition on folding energy can be deduced from nearest-neighbor models of RNA folding (Kennedy et al. 2010; Lorenz et al. 2011; Derr et al. 2012), and the fact that secondary structure disrupts primer binding and unfolding during extension is widely recognized (Peters et al. 2004; Nolan et al. 2006). However, to our knowledge, this information had not been previously applied to the paradox of RNA World templates or to understand modern-day RNA replicators. Here we suggest that the ability of RNA to form wobble pairs, particularly G:U, would enable a "division of labor" between the ribozyme strand and the genome strand. Our study indicates that the poor templating ability of a highly structured RNA sequence can be countered in its reverse complement through increased G:U wobble pairing.

Although we did not study copying of entire templates, one might expect that each base extension would follow the observed pattern, summing to a similar qualitative effect overall. While this effect could allow efficient transcription of the ribozymes from the genome, the reverse process of copying the ribozyme into the genome is still slow. However, increased efficiency of ribozyme production means a greater number of ribozymes will be present, presenting more opportunities for copying into the genome strand, potentially resulting in faster overall replication. In

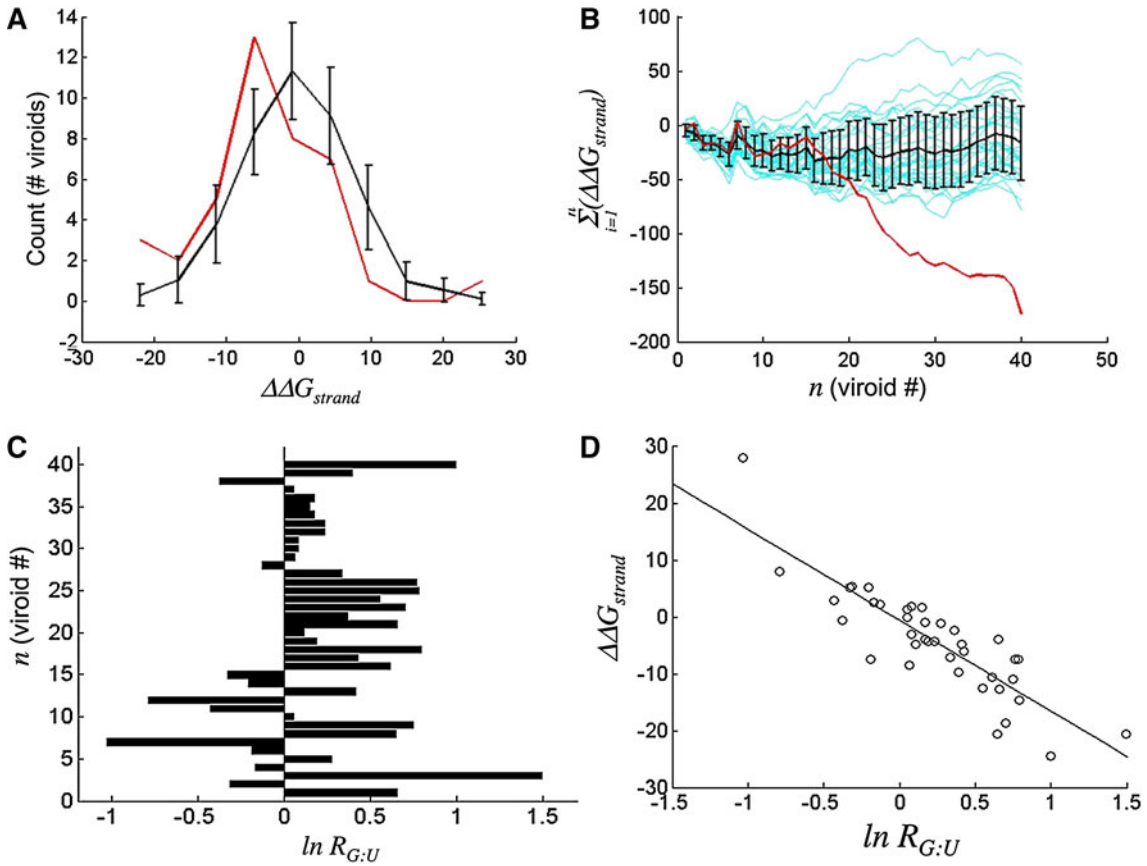

Fig. 3 Asymmetry between strands in RNA viroids. **a** Distribution of folding energy difference between (+) and (−) strands ($\Delta\Delta G_{\text{strand}}$) for viroids (*red*) and control sequences (*black*). Values were sorted into 10 bins. Leftward shift of viroids indicates asymmetry with the (+) strands being more well-folded than the (−) strands. **b** Cumulative sum of $\Delta\Delta G_{\text{strand}}$ over all 40 viroids (*red*) or over 30 analogous control sets (*blue*, with mean and standard deviation shown in *black*). Downward trend indicates a bias toward negative values of $\Delta\Delta G_{\text{strand}}$. The sum, $\Sigma(\Delta\Delta G_{\text{strand}})$, over viroids is significantly lower than $\Sigma(\Delta\Delta G_{\text{strand}})$ over controls ($p = 4.3 \times 10^{-6}$). Viroid sequences are numbered (n) as given in Supplementary Table 1. **c** Ratio of stabilization energy due to G:U wobbling for (+) versus (−) strand. The log ratio ($\ln R_{\text{G:U}}$) is given for each of 40 viroid sequences. The observed $\overline{\ln R_{\text{G:U}}} = 0.23$ is significantly different from the null expectation of identical stabilization from wobble pairing (i.e., $\ln R_{\text{G:U}} = 0; p = 0.0029$). **d** Correlation between $\Delta\Delta G_{\text{strand}}$ and $\ln R_{\text{G:U}}$, showing that greater G:U stabilization of the (+) strand corresponds to greater folding asymmetry (more negative $\Delta\Delta G_{\text{strand}}$). Linear regression line of best fit is shown ($r^2 = 0.755$)

addition, if the ribozyme provides a benefit to the ensemble (e.g., in a protocell), fitness might be further enhanced by the presence of more ribozymes.

To the extent that polymerization rates influence overall replication, one might speculate that wobble pairing would essentially decouple the folding energies of the ribozyme and its genome, perhaps allowing the combination of well-folded ribozymes and efficient genomic templates in RNA World life. Such non-Watson–Crick pairings are relatively common in RNA structures (Leontis et al. 2002), and wobble pairs other than G:U may display similar, albeit attenuated, advantages. Interestingly, another advantage of wobble pairing is suggested by the observation that G/T-rich DNA sequences improve the efficiency of an RNA polymerase ribozyme, presumably by increasing non-specific associations (Yao and Muller 2011). However, wobble pairing also results in mispairings that substantially lower fidelity, limiting the amount of information that can be propagated in such a system (Eigen 1971; Leu et al. 2011). Further "division of labor" segregating the genomic function into a specialized chemical form with reduced wobble pairing (DNA) would ultimately relieve the conflicting pressures of ribozyme folding and templating ability while improving the fidelity of information transfer.

An interesting postscript concerns whether strand asymmetry through G:U wobble pairing might have evolved in the modern day, since the general principle of wobble pairing leading to folding asymmetry should still apply. Our analysis of viroid RNA sequences suggests that this is the case, with the infectious (+) strand having lower free energy than the templating (−) strand due in large part to G:U pairs. This finding is remarkable since viroid sequences have probably evolved in response to numerous selection pressures ignored by our simplified "division of

labor" perspective (Flores et al. 2012). For example, viroids in family Avsunviroidae contain hammerhead self-cleavage motifs in both the (+) and (−) strands to process oligomeric intermediates into monomeric forms (Daros et al. 2006), but apparently this effect is weak relative to selection pressures favoring asymmetry. Since viroids undergo mutation at a very high rate (Gago et al. 2009), the observed asymmetry is likely to be the outcome of ongoing selection. Although it is unclear what causes the bias in viroids, one might speculate that the same principles that governed evolution during the origin of life also apply to modern RNA replicators.

Materials and Methods

Activated Nucleotide

Guanosine 5′-phosphorimidazolide (ImpG) was synthesized by GL Synthesis Inc. in Worcester, MA based on a previously published protocol (Rajamani et al. 2010). The purity of ImpG was found to be >93 %, determined by mass spectrometry and HPLC as previously described (Rajamani et al. 2010).

Oligonucleotides

The RNA primers were synthesized by reverse synthesis (W. M. Keck Biotechnology Resource Laboratory, Yale University, New Haven, CT), using 3′-*O*-tritylamino-*N*6-benzoyl-2′,3′-dideoxyguanosine-5′-cyanoethyl phosphoramidite at the 3′-terminus (indicated by $-NH_2$) and labeled with Cy3 at the 5′-terminus. The primers were PAGE-purified and masses were verified by MALDI-TOF. RNA template sequences were synthesized and PAGE-purified by UCDNA Services (Calgary, AB, Canada). RNA excess primers were synthesized by Dharmacon (Lafayette, CO). Templates were 46 mers with the exception of I (56-mer). Sequences were as follows:

- P: 5′-GGGAUUAAUACGACUCACUG-NH_2
- P-Asym: 5′-AGGCCCAGUCCAAUCG-NH_2
- P-Asym-RC: 5′-GGCGAGUUCUUUUUG-NH_2
- P-V-RC: 5′-UAAUAAUUACCACUG-NH_2
- I: 5′-GGGAUUAAUACGACUCACUGGAGAUCA AGUGAUCUCCAGUGAGUCGUAUUAAUCCC
- II: 5′-UAAUACGACUCACUGGAGAUCAAGUGA UCUCCAGUGAGUCGUAUUA
- III: 5′-UAAUACGACUAACUGGAGAUCAAGUGA UCUCCAGUGAGUCGUAUUA
- IV: 5′-UAAUACGAGAGACUGGAGAUCAAGUGA UCUCCAGUGAGUCGUAUUA
- V: 5′-UAAUAAUUACCACUGGAGAUGA AGUGAUCUCCAGUGAGUCGUAUUA
- VI: 5′-UAAUACCUGAGACUGAAGAUCAAGUCA UCUCCAGUGAGUCGUAUUA
- VII: 5′-UACCCUCGUUCUAGGACGAAUAAUAU UUGGCCAGUGAGUCGUAUUA
- VIII: 5′-ACCGGCGUGCCGAUUCCGGAUUUCCC AUCUCCAGUGAGUCGUAUUA
- IX: 5′-UAUGCGGCAAAUUCACUCUACACUCA UCUACCAGUGAGUCGUAUUA
- X: 5′-CUCAAUACAGACUCGUGGUUGAGUGUA CAGCCAGUGAGUCGUAUUA
- XI: 5′-UACAUUGCAUACAAAUCGAUCAGGGGC GCGCCAGUGAGUCGUAUUA
- XII: 5′-UAAUUCCUGAGACUGAUGAUCAAGUU AACUCCAGUGAGUCGUAUUA
- XIII: 5′-UAAGACCUAAGACAGAAGAUCACGUC AUCUCCAGUGAGUCGUAUUA
- Asym: 5′-GGCGAGUUCUUUUUGGGUUGUUGUC GACUCCGAUUGGACUGGGCCU
- Asym-RC: 5′-AGGCCCAGUCCAAUCGGAGUCG ACAACAACCCAAAAAGAACUCGCC
- V-RC: 5′-UAAUACGACUCACUGGAGAUCACU UCAUCUCCAGUGGUAAUUAUUA

Nonenzymatic Template-Directed Primer Extension

Fluorescently labeled primer (1.5 μM) and template strand (1.5 μM) were mixed in solution containing Tris–HCl (100 mM, pH 7.0) and NaCl (200 mM). The solution was heated to 95 °C for 5 min and cooled on the benchtop to room temperature for 15 min. After annealing, ImpG was added to 11.5 mM. At time points, 1 μL reaction mixture was removed and added to 4μL loading buffer containing 8 M urea, 100 mM Tris–HCl, and 50 mM EDTA. To prevent binding of the fluorescently labeled primer to the template strand in the polyacrylamide gel, we added unlabeled RNA primer of the same sequence to the loading buffer (40× excess unlabeled primer relative to labeled primer). The reaction was analyzed on a 20 % polyacrylamide urea gel. The gel was scanned using a GE Amersham Typhoon Trio Imager at 532 nm. The band intensities were analyzed using ImageQuant. The band intensity of the extended primer was divided by the total band intensity (unextended and extended primer) for each time point. The rate was estimated as the slope of the linear regression line passing through the initial phase of the reaction.

Template–Primer Hybridization

Fluorescently labeled primer (1.5 μM) was mixed with template (1.5 μM) in 100 mM Tris–HCl (pH 7.0) and 200 mM NaCl. The solution was annealed as described above and run on a 24 % nondenaturing polyacrylamide

gel at 4 °C. The loading buffer used was 80 % glycerol in ddH_2O v/v, 100 mM Tris–HCl, 50 mM EDTA. The gel was scanned as described above. The intensity of the band corresponding to primer alone was divided by the total intensity of the lane to estimate the fractions of free versus hybridized primer.

Calculation of Thermodynamic Properties

Each sequence used in the experiments together with the corresponding primer was folded using the cofolding routines of the Vienna package (Hofacker et al. 1994; Bernhart et al. 2006). The free energies of the possible structures (self-folding of template or primer, homo-duplexes of template or primer, and the primer–template hetero-duplex) as well as the free energy of binding (i.e., the free energy of the primer–template hetero-duplex minus the two self-folding energies of primer and template) were recorded from the output, and the expected proportion of primer associated with the template for experimental concentrations of template and primer was calculated (Bernhart et al. 2006). Additionally, we used constrained cofolding to estimate the proportion of template–primer complexes that would permit association of the activated nucleotide and thus enable the extension reaction, by requiring that the primer be fully hybridized to the template and the first template base downstream of the primer be unpaired. The proportion of pre-reaction complexes is then the product of the expected proportion of primer associated with the template and the proportion of template–primer complexes that attain a permissive structure. The latter is calculated from the exponential of the difference in free energy of primer–template hetero-duplex structures with and without the constraint (divided by RT, where R is the universal gas constant and T is the temperature).

Viroid Sequence Analysis

RNA sequences for 40 viroid species (+ strand) were obtained from the Subviral RNA Database (Rocheleau and Pelchat 2006). The first sequence entry for each species was taken as the representative (i.e., *0.001). 30 control sets were generated by random shuffling of bases for each viroid, to preserve base composition but not base order (Forsdyke 2007). Folded free energy prediction was carried out by RNAfold from the Vienna RNA package, specifying circular RNA (Hofacker and Stadler 2006), with or without allowance of G:U pairing. Data visualization and statistical tests were carried out using MATLAB 2011b.

Acknowledgments The authors thank the anonymous reviewers and the editors for their comments. This work was supported by Grant 290356 from the Simons Foundation (I.A.C.), Grant RFP-12-05 from the Foundational Questions in Evolutionary Biology Fund of the John Templeton Foundation (I.A.C), a DAAD Grant (B.O.), DFG Grant GE 1098/3-1 (U.G.), and NIH Grant GM068763 to the Center for Modular Biology at Harvard. I.A.C. was a Bauer Fellow at Harvard University.

Conflict of interest The authors declare that they have no conflict of interest.

References

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289:905–920

Bernhart SH, Tafer H, Muckstein U, Flamm C, Stadler PF, Hofacker IL (2006) Partition function and base pairing probabilities of RNA heterodimers. Algorithms Mol Biol 1:3

Chen IA, Hanczyc MM, Sazani PL, Szostak JW (2006) Protocells: genetic polymers inside membrane vesicles. In: Gesteland R, Cech T, Atkins J (eds) The RNA World. Cold Spring Harbor Laboratory Press, New York

Cheng LK, Unrau PJ (2010) Closing the circle: replicating RNA with RNA. Cold Spring Harb Perspect Biol 2:a002204

Crick FH (1968) The origin of the genetic code. J Mol Biol 38:367–379

Daros JA, Elena SF, Flores R (2006) Viroids: an Ariadne's thread into the RNA labyrinth. EMBO Rep 7:593–598

Derr J, Manapat ML, Rajamani S, Leu K, Xulvi-Brunet R, Joseph I, Nowak MA, Chen IA (2012) Prebiotically plausible mechanisms increase compositional diversity of nucleic acid sequences. Nucleic Acids Res 40:4711–4722

Diener TO (1989) Circular RNAs: Relics of precellular evolution? Proc Natl Acad Sci USA 86:9370–9374

Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58:465–523

Flores R, Serra P, Minoia S, Di Serio F, Navarro B (2012) Viroids: from genotype to phenotype just relying on RNA sequence and structural motifs. Front Microbiol 3:217

Forsdyke DR (2007) Calculation of folding energies of single-stranded nucleic acid sequences: conceptual issues. J Theor Biol 248:745–753

Gago S, Elena SF, Flores R, Sanjuan R (2009) Extremely high mutation rate of a hammerhead viroid. Science 323:1308

Gesteland R, Cech T, Atkins J (eds) (2006) The RNA world. Cold Spring Harbor Laboratory Press, New York

Hofacker IL, Stadler PF (2006) Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics 22:1172–1176

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. Monatshefte Fur Chemie 125:167–188

Joyce GF (2009) Evolution in an RNA world. Cold Spring Harb Symp Quant Biol 74:17–23

Kennedy R, Lladser ME, Wu Z, Zhang C, Yarus M, De Sterck H, Knight R (2010) Natural and artificial RNAs occupy the same restricted region of sequence space. RNA 16:280–289

Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson–Crick base pairs and their associated isostericity matrices. Nucleic Acids Res 30:3497–3531

Leu K, Obermayer B, Rajamani S, Gerland U, Chen IA (2011) The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA. Nucleic Acids Res 39:8135–8147

Leu K, Kervio E, Obermayer B, Turk-MacLeod RM, Yuan C, Luevano JM Jr, Chen E, Gerland U, Richert C, Chen IA (2013) Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. J Am Chem Soc 135:354–366

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA package 2.0. Algorithms Mol Biol 6:26

McCaskill JS (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers 29:1105–1119

Muller UF (2006) Re-creating an RNA world. Cell Mol Life Sci 63:1278–1293

Navarro B, Gisel A, Rodio ME, Delgado S, Flores R, Di Serio F (2012) Viroids: how to infect a host and cause disease without encoding proteins. Biochimie 94:1474–1480

Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. Nat Protoc 1:1559–1582

Orgel LE (1968) Evolution of the genetic apparatus. J Mol Biol 38:381–393

Orgel LE (2004) Prebiotic chemistry and the origin of the RNA world. Crit Rev Biochem Mol Biol 39:99–123

Peters IR, Helps CR, Hall EJ, Day MJ (2004) Real-time RT-PCR: considerations for efficient and sensitive assay design. J Immunol Methods 286:203–217

Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459:239–242

Rajamani S, Ichida JK, Antal T, Treco DA, Leu K, Nowak MA, Szostak JW, Chen IA (2010) Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. J Am Chem Soc 132:5880–5885

Rocheleau L, Pelchat M (2006) The subviral RNA database: a toolbox for viroids, the hepatitis delta virus and satellite RNAs research. BMC Microbiol 6:24

Schrum JP, Zhu TF, Szostak JW (2010) The origins of cellular life. Cold Spring Harb Perspect Biol 2:a002212

Szabo P, Scheuring I, Czaran T, Szathmary E (2002) In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. Nature 420:340–343

Szostak JW (2013) The eightfold path to non-enzymatic RNA replication. J Syst Chem 3:2

Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. Nature 409:387–390

Turner DH, Mathews DH (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res 38:D280–D282

Weimann BJ, Lohrmann R, Orgel LE, Schneider-Bernloehr H, Sulston JE (1968) Template-directed synthesis with adenosine-5′-phosphorimidazolide. Science 161:387

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V (2000) Structure of the 30S ribosomal subunit. Nature 407:327–339

Yao C, Muller UF (2011) Polymerase ribozyme efficiency increased by G/T-rich DNA oligonucleotides. RNA 17:1274–1281

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF (2001) Crystal structure of the ribosome at 5.5 A resolution. Science 292:883–896