

Selection of self-priming molecular replicators

Daechan Park¹, Andrew D. Ellington^{2,*} and Cheulhee Jung^{3,*}¹Department of Biological Sciences, College of Natural Sciences, Ajou University, Suwon 16499, Republic of Korea,²Institute for Cellular and Molecular Biology, Department of Molecular Biosciences, University of Texas at Austin,Austin, TX 78712, USA and ³Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Republic of Korea

Received December 07, 2018; Revised January 14, 2019; Editorial Decision January 17, 2019; Accepted January 18, 2019

ABSTRACT

Self-priming amplification of oligonucleotides is possible based on foldback of 3' ends, self-priming, and concatemerization, especially in the presence of phosphorothioate linkages. Such a simple replicative mechanism may have led to the accumulation of specific replicators at or near the origin of life. To determine how early replicators may have competed with one another, we have carried out selections with phosphorothiolated hairpins appended to a short random sequence library (N10). Upon the addition of deoxynucleoside triphosphates and a polymerase, concatemers quickly formed, and those random sequences that templated the insertion of purines, especially during initiation, quickly predominated. Over several serial transfers, particular sequences accumulated, and in isolation these were shown to out-compete less efficient replicators.

INTRODUCTION

Replicators are fundamental to the origin of life and evolvability. While the literature is replete with analyses of the fidelity of replication, both catalyzed (1–3) and uncatalyzed (4), there have been few studies in which molecular replicators have been competed with one another, and where their relative fecundities have been compared. Nonetheless, it is likely that raw replicability would have been one of the first Darwinian constraints to come into play at or near the origin of life (5).

Ferris, Schwartz, Orgel, and other pioneers have shown that it is possible to produce simple nucleic acid templates starting from monomeric precursors (6). In particular, monomer extension via purely chemical mechanisms was originally elegantly demonstrated by Orgel *et al.* (7) and more recently elaborated by Chen (8) and Szostak (9,10).

While these are impressive demonstrations, virtually no replication mechanism exists that does not require an initial primer that can initiate polymerization or ligation. The

paradox of how to primer in a world where there were few, if any, primers can be resolved via foldback structures, similar to second strand priming in some retroviruses. In particular, it has already been demonstrated that self-priming via foldback structures can lead to the replication of very short (decamer) oligonucleotides, similar to those that might have been present at or near origins (11). However, foldback replication can be hampered by low self-folding efficiencies and a corresponding narrow set of reaction temperatures. To overcome the limitation, phosphorothioates (PS) was incorporated into the nucleic acid substrates. Boczkowska and co-workers have carried out thermodynamic studies on the stability of duplexes formed between PS-modified ssDNA (all-Rp, all-Sp and mixed Rp/Sp) and complementary phosphodiester (PO)-modified ssDNA (12), and found that the inclusion of any PS modifications substantially reduced the T_m of PS-PO dsDNA (13). Given that phosphorothioates can destabilize duplexes, we have now shown how strands with phosphorothioate linkages can foldback and self-prime, and thereby lead to remarkably stable amplification over a much wider range of temperature (14).

The invocation of foldback structures as a mechanism for early replicators also overcomes an additional barrier to the origin of life. The reassociation of replicated product strands during amplification reactions leads to parabolic replication and the 'survival of everyone' (15), ultimately thwarting exponential replication and Darwinian mechanisms for selection (16). This problem of strand dissociation grows not only in relation to concentration, but to length, such that double-stranded DNA of roughly 20 nucleotides in length will not dissociate into single strands at temperatures that are clement to living systems.

We have previously developed a simplistic model for self-priming nucleic acid replication that involves efficient foldback self-priming and concatemer formation, which we termed phosphorothioated-terminal hairpin formation and self-priming extension (PS-THSP) (14). The mechanism of PS-THSP is as shown in Figure 1. The substrate is a short oligonucleotide that forms hairpins at each end. The extension of the 3' end (D2*) leads to the formation of an extended hairpin (L) (i.e. the seeding state). In this extended

*To whom correspondence should be addressed. Tel: +82 2 3290 3023; Email: damo363@gmail.com
Correspondence may also be addressed to Andrew D. Ellington. Email: ellingtonlab@gmail.com

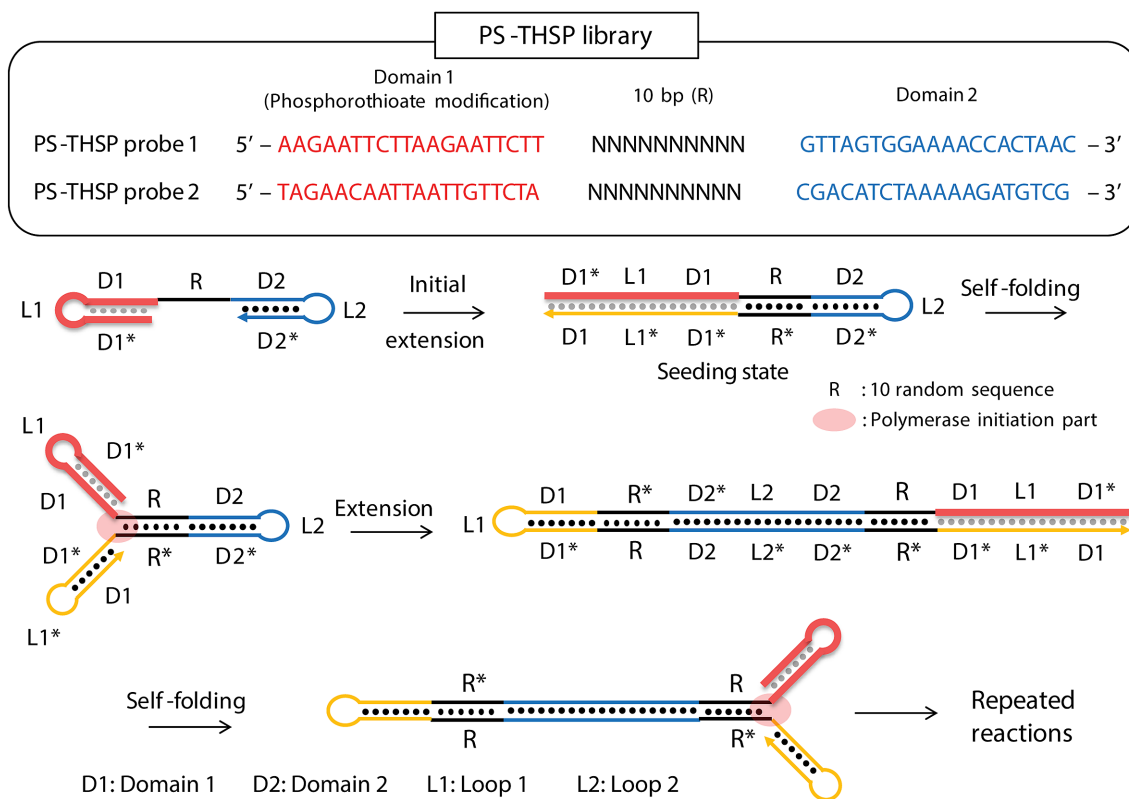


Figure 1. Scheme of a phosphorothioated-terminal hairpin formation and self-priming extension (PS-THSP) for selection of self-replicators.

hairpin, 5' D1-L1-D1* forms a duplex with its complement (5' D1-L1*-D1*), but the original inclusion of phosphorothioates in D1-L1-D1* leads to destabilization of the duplex, allowing the formation of a foldback structure that can again be extended. Multiple cycles of foldback priming and extension lead to the formation of long concatamers.

While this replication mechanism can utilize short templates and does not require a primer, different sequences at origins would likely have competed with one another in the presence of limited resources for survival. In order to recapitulate this process and determine what sorts of genotypes and phenotypes would have emerged based solely on replicability, we have attempted to extend our previous work with phosphorothiolated foldback amplification to encompass pools of random sequences. Based on the PS-THSP model, we designed a pool for the selection of more efficient PS-THSP replicators in which 10 random bases (R) were included in the original dumbbell-like structure (Figure 1). This pool represented about a million (4^{10}) variants, a population size easily encompassed within the original synthetic DNA. To take control for any impact of the hairpin sequences on replicative competition, we used two different constant sequences (PS-THSP probes 1 and 2). Importantly, in this replication mechanism, a cycle consisting of self-folding and extension is continuously repeated. Since the sequence in the random region continuously affects initiation, any sequence dependency will be accumulated in the concatamers. Therefore, while these experiments rely on evolved protein polymerases for catalysis, they nonetheless

provide an example of what might have occurred in competitions between short replicators.

MATERIALS AND METHODS

Reagents

All chemicals were of analytical grade and were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless otherwise indicated. All oligonucleotides were ordered from Integrated DNA Technology (IDT; Coralville, IA) and the sequences are summarized in Supplementary Table S1. Bst 2.0 DNA polymerase and dNTPs (10 mM each) were purchased from New England Biolabs (NEB; Ipswich, MA, USA). SYTO 82 dye was obtained from Thermo Fisher Scientific (Waltham, MA, USA).

Phosphorothioated-terminal hairpin formation and self-priming (PS-THSP)

All basic PS-THSP reactions were performed as detailed here. Reactions were prepared with 1X isothermal amplification buffer (20 mM Tris-HCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 50 mM KCl, 2 mM MgSO_4 , 0.1% Tween-20, pH 8.8), various concentrations of dNTPs and 100 nM of PS-THSP probe. Reactions were heated to 95°C for 5 min and cooled at 0.1°C/s to 60°C for 5 min, followed by the addition of 8 U of Bst 2.0 DNA polymerase to produce a final volume of 100 μl . Each PS-THSP reaction was then initiated at 60°C and continued for 1 h (on a T100™ thermal cycler; Bio-Rad; Hercules, CA, USA).

Real-time phosphorothioated-terminal hairpin formation and self-priming (PS-THSP)

PS-THSP reactions were prepared with 1× isothermal amplification buffer (20 mM Tris-HCl, 10 mM (NH₄)₂SO₄, 50 mM KCl, 2 mM MgSO₄, 0.1% Tween-20, pH 8.8), 2 mM of dNTPs, 10 nM of PS-THSP probe and 2 μM of SYTO 82 dye. Reactions were heated to 95°C for 5 min and cooled at 0.1°C/s to 60°C for 5 min, followed by the addition of 0.08 U/0.8 U of Bst 2.0 DNA polymerase. Each PS-THSP reaction was then initiated at 60°C and continued for 2 h (on a T100™ thermal cycler; Bio-Rad; Hercules, CA, USA).

Gel electrophoresis

A 10 μl aliquot of a PS-THSP reaction was mixed with 6× agarose gel loading buffer (NEB; Ipswich, MA, USA) and subjected to electrophoretic analysis on a 2% agarose gel containing ethidium bromide. The amplification products were visualized using a UV transilluminator.

NGS analysis

The paired-end 126 bp raw data were generated by Illumina HiSeq 2500, and the base quality scores of the fastq files were plotted by FastQC (17). The phred scores were over 30 for at least 110 base, so the data were analyzed without any trimming. Since, regardless of read directionality, R always appears before D2 whereas R* is present before D1 (Figure 1), the random 10 nucleotides were extracted as follow. For probe 1, 10 nucleotides before AAGAATTCTTAAGAAT TCTT were reverse-complemented then taken whereas 10 nucleotides before GTTAGTGGAAAACCACTAAC were just taken within a pair of R1 and R2 read. For probe 2, TAGAACAAATTAATTGTTCTA and CGACATCTAA AAAGATGTCG are for reverse-complement and forward, respectively. By allowing one mismatch for domain sequence search, additionally extracted reads were <2%. After considering false positive identification with mismatch, the increment is marginal, so we searched perfect matches. Then the random sequences were counted per pair. When the most frequent random per read pair (fragment) was only one, the reads were discarded. Also, the reads whose top frequencies are tie were filtered out for downstream analysis. The frequency heat maps of mono-/di-/tri-nucleotide were visualized by Seaborn v0.7.1 in python 3 (Seaborn can be cited using a DOI provided through Zenodo: DOI: 10.5281/zenodo.54844) Sequence logo and other sequence analysis were performed using Biopython version 1.69 (18).

RESULTS

Analyzing the selection of replicators via high-throughput sequencing

It seemed likely that one of the key variables impacting the relative replicability of template would be resource limitation. Therefore, various concentrations (from 2 μM to 2 mM) of nucleotides were used to amplify PS-THSP probe 1 for 1 h at 60°C (Figure 2A). As can be seen, at least 200 μM of nucleotides were required for efficient amplification of a substantial portion of the population, although some amplification can be seen at lower concentrations, as well.

In order to discern whether replicative improvement within the population could be observed, the amplification product was diluted by 10-fold, and amplification was again carried out for 1 h at 60°C (Figure 2A). Serial dilution followed by re-amplification was carried out three times in total. The trends originally observed held largely true, except that there was clearly selection for more efficiently replicating species that more readily formed concatemers at all but 2 μM, which did not seem to support replication. We observed a similar trend for the PS-THSP probe 2 with an alternative constant region (Supplementary Figure S1).

Therefore, we chose the 200 μM concentration evolution via serial transfer. After three 10-fold dilution serial transfers PS-THSP products were analyzed via Next Generation Sequencing (NGS). The PS-THSP products were purified using a PCR purification kit to remove short probes that were not amplified, then the remaining concatemers were sonicated to generate ~300 bp fragments. After measuring the concentration of the fragmented PS-THSP products, 500 ng (>1 ng/ul) of each sample was ligated to adaptor sequences at both ends for NGS. Paired-end reads were produced from the libraries by an Illumina HiSeq.

Considering that a R1-R2 read pair fragment is 250 bp long (on the 2 × 125 bp HiSeq platform), a unique sequence derived from the randomized region (R region in Figure 1) should be repeated on average six times per read pair. However, due to sequencing errors there were read pairs that contained irregular tandem sequences or multiple non-identical R region sequences. In order to filter out the most irregular read pairs, we included the majority read pair when it existed, and excluded all read pairs when there was no majority read pair (Figure 2B). For example, when the frequency of non-identical R region sequences (e.g. AGATAAAAGG and AGATAAAGGG) within one read pair was the same, the read pair was discarded.

Table 1 shows the sequence retained at each of the filtering steps, and there was no obvious bias in the sequences ultimately used for analysis. To depict the coverage of reads at each round of selection out of the total possible 4¹⁰ R region sequences (= 1 048 576) we use a four-way Venn diagram that shows the number of unique sequences and overlapped sequences from all four rounds of selection (R1, R2, R3 and R4; Figure 2C). The total number of unique sequence from all the rounds was 982, 470, which is very close to the theoretical number of 4¹⁰. This means that the diversity of the library was nearly complete (94%) and the sequencing data should therefore allow us to comprehensively observe the landscape of replicators.

Self-replicators evolve via sequence selection

In order to determine whether selection was occurring during rounds of serial dilution, we tracked the ranks of successful sequences over the four consecutive rounds (Figure 3A). Interestingly, the highly ranked sequences from the first round did not necessarily maintain their standing by the final round, indicating that there was likely competition at every round.

The evolution of probe 1 was essentially complete by the conclusion of the first round, as shown by the continuity of its sequence logo (Figure 3B). The selected distribution

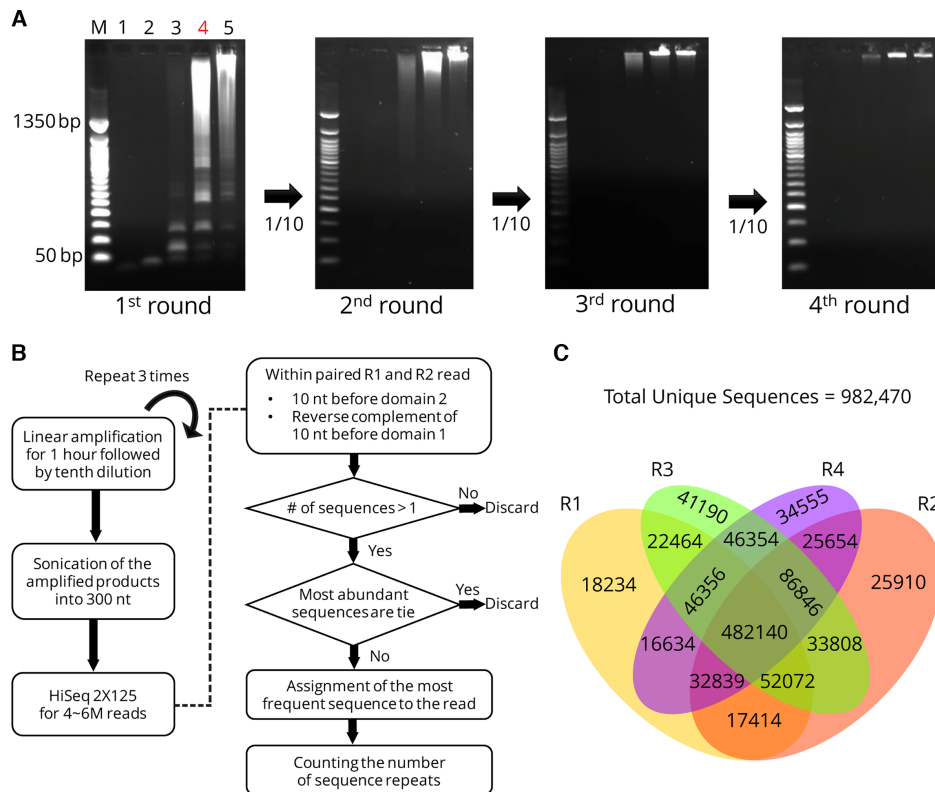


Figure 2. High-throughput sequencing for PS-THSP products of random sequences with a serial dilution. (A) PS-THSP was performed with a PS-THSP probe 1 (100 nM) and different amount of dNTPs (M: 50 bp DNA ladder, lane 1: 0 μ M, lane 2: 2 μ M, lane 3: 20 μ M, lane 4: 200 μ M, lane 5: 2 mM). The product of first round was 10 times serially diluted to fourth round and PS-THSP was performed for each round. (B) The flow chart shows that how the next generation sequencing data for PS-THSP products was analysed. (C) The four-way Venn diagram was prepared with the analysed NGS data for all the rounds.

Table 1. Statistics at every filtering steps

	PS-THSP probe 1				PS-THSP probe 2			
	First	Second	Third	Fourth	First	Second	Third	Fourth
Read pairs (fragment)	5 227 625	4 916 639	5 436 221	4 371 625	3 996 775	5 783 549	6 206 215	4 895 878
Fragments with no random sequence	410 021	529 069	615 582	543 939	483 395	695 716	699 269	791 561
Fragments with one random sequence	501 272	643 450	780 521	645 213	559 758	949 547	978 741	831 621
Fragments with tie for the most frequent sequence	286 462	123 023	139 124	100 436	95 283	125 033	114 885	91 687
Fragment analyzed	4 026 153	3 615 249	3 888 538	3 069 541	2 858 337	4 013 244	4 413 316	3 181 003
Fragment analyzed (%)	77.02%	73.53%	71.53%	70.22%	71.52%	69.39%	71.11%	64.97%
Unique random sequence	688 153	756 683	811 230	771 378	598 334	715 368	790 995	758 091

proved to be highly skewed, with the top 20 sequences accounting for 95% of the selected population in every round. Indeed, the sequences that were eventually found >50 times total by sequencing comprised <1% of the initial population in every round (Supplementary Figure S2).

The top 100 replicators showed a strong preference for purine residues in all rounds. Given that virtually all sequences were observed in the population (Figure 2C), the purine skew is not due to library bias, but strongly suggests that there was a strong selection pressure during the evolution of PS-THSP replicators. The frequency of the first base proved to be the most variable over the rounds (Figure 3B and C). In probe 1 an evolution from preferring A to prefer-

ring G at the first position of the pool was observed, while for PS-THSP probe 2 there were a number of sequences that initially started with thymidine but these were almost completely displaced by A and G by Round 4 (Figure 3B and C).

These sequence logos also suggest that stringent selection for efficient replicators occurs quickly. To confirm this, we attempted to profile synthetic R region sequences at the initial stage (R0), we allowed the probe to extend only once, and then stopped and sequenced the reaction. These elongation reactions were performed at 37°C, as self-folding and self-replication only occur at higher temperatures (Supplementary Figure S3). Overall, A at the first position was en-

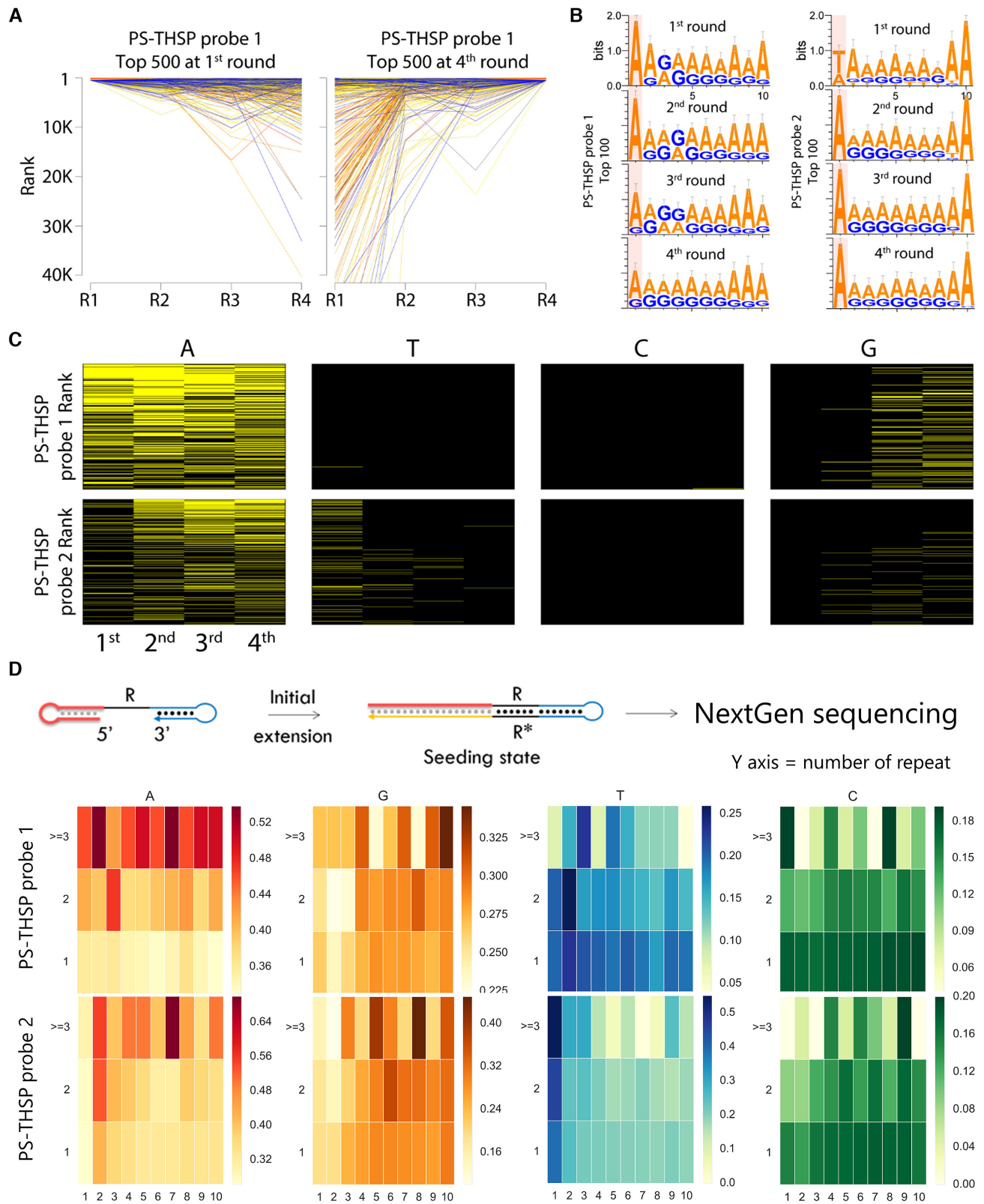


Figure 3. Evolution of selected R region sequences. **(A)** Rankings of the top 500 sequences from the first through fourth rounds of selection for PS-THSP probe 1. **(B)** Sequence logos for the top 100 PS-THSP probe 1 and probe 2 sequences over four rounds. **(C)** The yellow bars in the graph represent the identity of the initial nucleotide for the top 500 ranked sequences in each of the 4 rounds of selection, with the y axis showing the rank of the sequences (higher numbers of sequences towards the top). **(D)** Scheme for extension without concatamerization of the PS-THSP template, and corresponding analysis of the resultant R region sequences at different positions for PS-THSP probe 1 and probe 2. The x and y axes represent base position and count of sequences, respectively.

riched in individual molecules in the initial probe 1 pool, whereas T was strongly enriched in the initial probe 2 molecules, potentially indicating a skew during initial pool syntheses (Figure 3D). This skew fortuitously serves as a control for sequence bias being the rationale for our results: despite the initial bias towards thymidine at the first position in probe 2, thymidine at this position was heavily discriminated against during directed evolution (Supplementary Figure S4).

The best replicators are similar to the consensus

To further investigate the characteristics of the most fit replicators we investigated nucleotide preferences at every position; these are displayed in the form of a heat map (Figure 4A). For the heat map, the counts of particular nucleotides (A, G, C, T) at any one position for the most frequently represented variants in a given set (top 100, 1K, 10K, 50K, 100K, 200K and 500K) were divided by the total number of sequences, and then represented in terms of color. Notably, good replicators were found to contain largely A and G nucleotides from almost the very first round of selection (Supplementary Figure S5). A at the first base position and G at the third and/or fourth base positions were important for the best replicators. In contrast, T and C were actively discriminated against during selection. For example, although the first round of selection with probe 2 showed some enrichment of T at the first position, the final round had fixed A and G residues.

Beyond looking at frequencies at individual positions we examined nearest neighbor effects (Figure 4B and C). Dinucleotide (Figure 4B) and trinucleotide (Figure 4C) frequencies for the top 100 replicators were calculated at each position within a scanning window for the R region (nine possible dinucleotide positions; eight possible trinucleotide positions), and again represented as a heat map, relative to the top 500K replicators. To make the heat map, the counts of specific dinucleotide and trinucleotide sequences within the top 100 and 500K sequences were divided by the total number sequences, and the fraction was converted to color. Beyond the preference for purines, there was no particular skew towards a given di- or trinucleotide, although AAG was overall the best initial three nucleotide sequence and AAA was the best final three. As in previous analyses, the top 100 replicators showed a strong preference for purines at all positions, although the top 10 replicators favored a mixed composition of A and G (Figure 4D).

Examining the impact of polymerase concentration on initiation of replication

To explain the sequence bias that was quickly established and maintained, we hypothesized that either polymerase selectivity or purine stacking better enabled replication initiation, and subsequently competition of the replicators. To distinguish these hypotheses, we examined the replicability of individual sequences as a function of polymerase concentration. We chose two sequences that were ranked highly throughout the selection (HH) and two that were infrequent throughout (LL), and carried out PS-THSP reactions with two different concentrations of Bst DNA polymerase: the

low concentration used in the selections and a higher concentration that should be less limiting (Figure 5A and B). As expected, the reaction rates of LL sequences were slower than those of HH sequences at low concentrations of Bst DNA polymerase. In contrast, the rates of LL and HH replication were similar when the concentration of Bst was increased by 10-fold. These results strongly argue that it is not sequence selectivity of the polymerase that guides selection (since this should have been similar at both low and high concentrations of the polymerase), but rather that the initial formation of stacked base-pairs promotes initiation by the polymerase.

DISCUSSION

The nature of the earliest replicators is an open question. However, it is unlikely that many replication mechanisms that are currently observed would have been plausible at or near origins, as both templates and putative primers would have been present at small concentrations. In particular, terminal hairpin self-priming (THSP) allows even very short oligonucleotides with foldback structures to elongate into concatamers, thus amplifying the initial oligonucleotide sequence. We have previously determined that the introduction of destabilizing chemistries (such as phosphorothioate linkages) into THSP templates can further improve self-priming, especially at lower temperatures, which in turn may have been an important consideration for the evolution of replicators. To the extent that THSP could have played a role in the evolution of nucleic acid replicators, competition between individual replicators may have occurred, with some sequences favored over others. To determine what the outcome of such a competition may have been, we employed serial dilution to select for PS-THSP replicators from a small (10 residue) random sequence pool, where the randomized region was placed just after the foldback, self-priming structure. Selections were carried out with two different templates to determine the generality of results.

While virtually all sequences were present in initial pools (Figure 2C), both pools quickly collapsed to be exceedingly purine rich (Figure 3). This skew in replicators occurred irrespective of template, and irrespective of initial sequence biases; for example, the template 2 pool initially had a preponderance of thymidines at the first position of the random sequence region, but this bias was quickly overcome by purine-rich replicators. The dominance of purine-rich replicators was observed not only in the aggregate, but also in terms of the replicability and relative rankings of individual sequences within the population (Figure 4).

We chose the Bst DNA polymerase for the directed evolution of replicators because we needed a polymerase with a strong strand displacement activity, since we are not relying on thermal cycling to generate single-stranded templates. It was also useful that this polymerase could work over a variety of temperatures, as this allowed us to better understand how the temperature-dependent folding of the terminal hairpin structure impacted replication.

While it might be anticipated that the homopurine runs were due to polymerase preference, no such preference is observable in the *B. stearothermophilus* genome (the source of the Bst polymerase), which has no particular nucleotide

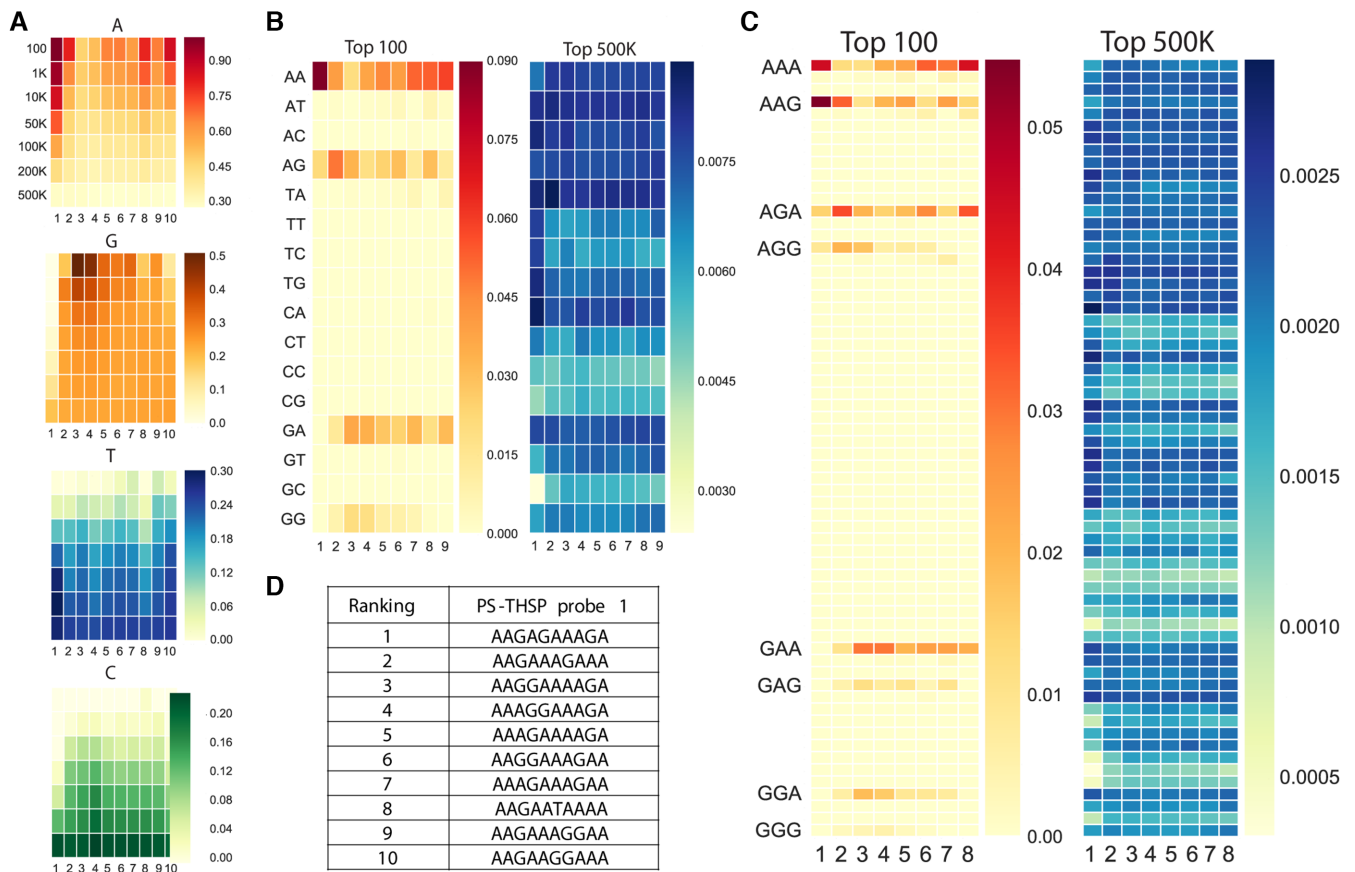
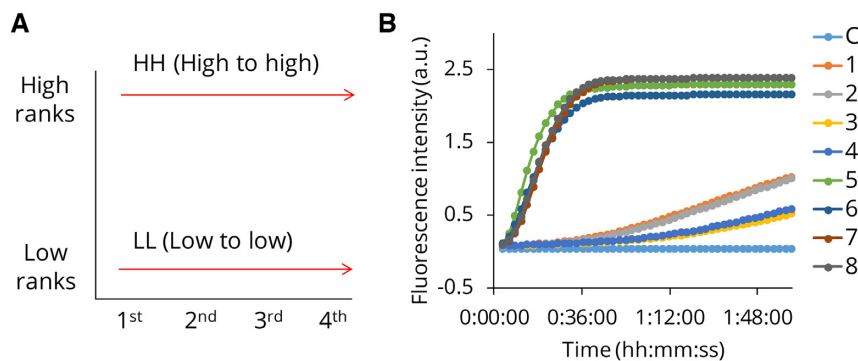


Figure 4. Analysis of different probe 1 replicators at individual positions in the R region during the first round of selection. (A) A heat map was prepared to show the relative frequency of each base (A, G, T, C) at each position within top ranking sequences (top 100, 1K, 10K, 50K, 100K, 200K, and 500K). Frequency trends for dinucleotide (B) and trinucleotide (C) sequences within top ranking 100 and 500K sequences were again represented using heat maps. (D) The top 10 ranked sequences from the first round for probe 1.



Sample	Sequence	Bst 2.0 DNA polymerase	
		0.08 U	0.8 U
HH 1	AAGGAAGAAA	1	5
HH 2	AAAAGGGAAA	2	6
LL 1	CTCCTTCCTT	3	7
LL 2	TCACCTTTTG	4	8

C: No Bst 2.0 DNA polymerase

Figure 5. Real-time data of PS-THSP for high and low ranked R region sequences. (A) Two HH and two LL sequences whose ranks remained high and low over four rounds were selected, respectively. (B) PS-THSP reactions were monitored in a real-time using a SYTO 82 dye with different concentrations of Bst DNA polymerase.

skew (0.23 A, 0.27 C, 0.26 G, and 0.24 T) and no preponderance of purine runs based on this skew (Supplementary Figure S6A). For example, the prevalence of 8 types of purine trinucleotides (AAA, AAG, AGA, AGG, GAA, GAG, GGA and GGG) is almost identical to what would be predicted based on the overall mononucleotide compositions (Supplementary Figure S6B), suggesting that the Bst polymerase has no natural sequence bias. However, to more explicitly test the possibility that polymerase bias led to sequence bias, we performed PS-THSP reactions with two sequences that had very different replicabilities (LL with low purine content and HH with high purine content) and altered the amount of Bst DNA polymerase in the reaction, using both a smaller amount (0.2 U/pmol DNA) than in the selection experiments (0.8 U/pmol DNA) and a larger amount (2 U/pmol DNA). Larger amounts of enzyme allowed the slower (LL) replicators to ‘catch up’ to their selected counterparts, indicating that it was not sequence bias of the polymerase but rather intrinsic differences in the replicability of the templates that led to competition and selection. For the selection experiments with the randomized PS-THSP libraries, the amount of Bst DNA polymerase used was 0.8 U/pmol DNA and the results were akin to those seen at the lower concentration, indicating that selection could occur due to differential foldback.

Instead, we believe that the predominance of purine runs resulted from the fact that purines (A and G) stack better with one another than do pyrimidines (T and C) (19), so purines (A and G) can be more efficiently incorporated into the growing strand, which in turn promotes more facile strand separation from the opposing, non-template strand (Supplementary Figure S7). Thus, the choice of purines was predominantly due to the replicator itself, not to the polymerase that catalysed the replication. This conclusion has generic implications for self-priming at near origins, perhaps providing some credence to Wachterhauser’s hypothesis that initial replicators were composed only of purines (20). The simplicity of the replication mechanism also suggests possible applications in ‘self-priming’ library preparations for Next Generation Sequencing, especially given that the resultant concatemers would allow for error correction during analysis (19).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Ajou University; National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [2018R1C1B6001112]; Welch Foundation [F-1654]; John Templeton Foundation [54466]. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation. Funding for open access charge: NRF [2018R1C1B6001112].

Conflict of interest statement. None declared.

REFERENCES

- Ganai, R.A. and Johansson, E. (2016) DNA replication—a matter of fidelity. *Mol. Cell*, **62**, 745–755.
- Doudna, J.A. and Szostak, J.W. (1989) RNA-catalysed synthesis of complementary-strand RNA. *Nature*, **339**, 519–522.
- Lincoln, T.A. and Joyce, G.F. (2009) Self-Sustained replication of an RNA enzyme. *Science*, **323**, 1229–1232.
- Rohatgi, R., Bartel, D.P. and Szostak, J.W. (1996) Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J. Am. Chem. Soc.*, **118**, 3332–3339.
- Tkachenko, A.V. and Maslov, S. (2015) Spontaneous emergence of autocatalytic information-coding polymers. *J. Chem. Phys.*, **143**, 045102.
- Ertem, G. (2004) Montmorillonite, oligonucleotides, RNA and origin of life. *Orig. Life Evol. Biosphere J. Int. Soc. Study Orig. Life*, **34**, 549–570.
- Orgel, L.E. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.*, **39**, 99–123.
- Leu, K., Kervio, E., Obermayer, B., Turk-MacLeod, R.M., Yuan, C., Luevano, J.-M., Chen, E., Gerland, U., Richert, C. and Chen, J.A. (2013) Cascade of reduced speed and accuracy after errors in Enzyme-Free copying of nucleic acid sequences. *J. Am. Chem. Soc.*, **135**, 354–366.
- O’Flaherty, D.K., Kamat, N.P., Mirza, F.N., Li, L., Prywes, N. and Szostak, J.W. (2018) Copying of mixed-sequence RNA templates inside model protocells. *J. Am. Chem. Soc.*, **140**, 5171–5178.
- Prywes, N., Blain, J.C., Del Frate, F. and Szostak, J.W. (2016) Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife*, **5**, e17756.
- Kato, T., Liang, X. and Asanuma, H. (2012) Model of elongation of short DNA sequence by thermophilic DNA polymerase under isothermal conditions. *Biochemistry*, **51**, 7846–7853.
- Boczowska, M., Guga, P. and Stec, W.J. (2002) Stereodefined phosphorothioate analogues of DNA: relative thermodynamic stability of the model PS-DNA/DNA and PS-DNA/RNA complexes†. *Biochemistry*, **41**, 12483–12487.
- LaPlanche, L.A., James, T.L., Powell, C., Wilson, W.D., Uznanski, B., Stec, W.J., Summers, M.F. and Zon, G. (1986) Phosphorothioate-modified oligodeoxyribonucleotides. III. NMR and UV spectroscopic studies of the Rp-Rp, Sp-Sp, and Rp-Sp duplexes, [d(GGSAATCC)]₂, derived from diastereomeric O-ethyl phosphorothioates. *Nucleic Acids Res.*, **14**, 9081–9093.
- Jung, C. and Ellington, A.D. (2016) A primerless molecular diagnostic: phosphorothioated-terminal hairpin formation and self-priming extension (PS-THSP). *Anal. Bioanal. Chem.*, **408**, 8583–8591.
- von Kiedrowski, G. (1986) A Self-Replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed. Engl.*, **25**, 932–935.
- Scheuring, I. and Szathmáry, E. (2001) Survival of replicators with parabolic growth tendency and exponential decay. *J. Theor. Biol.*, **212**, 99–105.
- Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.
- Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinforma. Oxf. Engl.*, **25**, 1422–1423.
- Friedman, R.A. and Honig, B. (1995) A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.*, **69**, 1528–1535.
- Wächtershäuser, G. (1988) Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.*, **52**, 452–484.