

A design principle for a single-stranded RNA genome that replicates with less double-strand formation

Kimihito Usui^{1,†}, Norikazu Ichihashi^{1,2,†} and Tetsuya Yomo^{1,2,3,*}

¹Exploratory Research for Advanced Technology, Japan Science and Technology Agency, Osaka University, 1–5 Yamadaoka, Suita, Osaka 565–0871, Japan, ²Graduate School of Information Science and Technology, Osaka University, 1–5 Yamadaoka, Suita, Osaka 565–0871, Japan and ³Graduate School of Frontier Biosciences, Osaka University, 1–5 Yamadaoka, Suita, Osaka 565–0871, Japan

Received May 11, 2015; Revised July 07, 2015; Accepted July 09, 2015

ABSTRACT

Single-stranded RNA (ssRNA) is the simplest form of genetic molecule and constitutes the genome in some viruses and presumably in primitive life-forms. However, an innate and unsolved problem regarding the ssRNA genome is formation of inactive double-stranded RNA (dsRNA) during replication. Here, we addressed this problem by focusing on the secondary structure. We systematically designed RNAs with various structures and observed dsRNA formation during replication using an RNA replicase (Q β replicase). From the results, we extracted a simple rule regarding ssRNA genome replication with less dsRNA formation (less GC number in loops) and then designed an artificial RNA that encodes a domain of the β -galactosidase gene based on this rule. We also obtained evidence that this rule governs the natural genomes of all bacterial and most fungal viruses presently known. This study revealed one of the structural design principles of an ssRNA genome that replicates continuously with less dsRNA formation.

INTRODUCTION

Translation and replication of genetic information are fundamental functions of all living organisms. The primary carrier of genetic information (genome) is DNA; however, some viruses utilize single-stranded RNAs (ssRNAs) as the material basis of their genomes. Because ssRNA replicates by a simpler mechanism than DNA, it is used as the genome in primitive life-forms that evolved in the RNA world or the later RNA-protein world (1,2), and is also proposed to be useful as the basis of a genome for an artificial cell (3).

A serious problem for using ssRNA as a genome is the formation of double-stranded RNA (dsRNA), which occurs during replication. For recursive replication, an ss-

RNA genome must replicate as a single strand; however, the proximity of the newly synthesized strand to the template genome at replication would lead to formation of dsRNA through hybridization. This dsRNA formation is a problem for recursive replication in the RNA world (3) and is one of the most serious problems in the reconstitution of a replication system of an artificial ssRNA genome (4,5).

To date, the dsRNA formations that are produced during replication have been extensively investigated using Q β replicase, an RNA-dependent RNA polymerase from bacteriophage Q β , which synthesizes a complementary strand in a primer-independent manner (6). The molecular mechanism of the replication has been intensively studied recently (7–9). This replicase requires CCC at the 3'-terminus (10), stem structures at the 5'-terminus (11), an internal C/U-rich domain (12) and a long-range interaction between 5'- and 3'-termini (13) for efficient replication. However, even if all of these requirements are satisfied, RNA-encoding genes that originate from organisms other than bacteriophage Q β mainly produce dsRNA and then terminate replication (14–21). Although some factors, such as components in a cell-free translation system and ribosomal protein S1, are known to inhibit dsRNA formation (22–24), they were insufficient to render ssRNA-encoding arbitrary genes replicable without dsRNA formation (22). In addition, rigid secondary structures of a template RNA have been reported to inhibit dsRNA formation (25,26). Because quantitative requirements for the structures (i.e. the size or stability of structures) have not been determined, it is not possible to design an ssRNA that encodes an arbitrary gene at this time.

In this study, we attempted to determine a simple principle to help design an ssRNA that encodes a gene and that continuously replicates without dsRNA formation using Q β replicase. We first propose a structure-dependent dsRNA formation model and then examine the model by measuring replication of RNAs with various structures. From the results, we extracted a simple rule to design an ssRNA that replicates with less dsRNA formation and actu-

*To whom correspondence should be addressed. Tel: +81 6 6879 4171; Fax: +81 6 6879 7433; Email: yomo@ist.osaka-u.ac.jp

†These authors contributed equally to the paper as first authors.

ally designed a new replicable RNA that encodes a domain of the β -galactosidase gene.

MATERIALS AND METHODS

Plasmid construction and RNA preparation

The RNAs with 28 insertions shown in Figure 2 were constructed by PCR amplification of a plasmid encoding the MDV-1 sequence (pUC-MDV-LR (27)) with primers that consist of a common region and an insertion sequence attached to the 5'-end (sequences of the insertions are shown in Supplementary Table S1). The common regions of the primers are ACGGGCTAGCGCTTTC and CGTCACG-GTCGAACTCCC. The inserted sequences are shown in Supplementary Table S1. Each RNA was synthesized by *in vitro* transcription with T7 RNA polymerase (Takara, Japan) after SmaI (Toyobo, Osaka, Japan) digestion to determine the 3' end. The synthesized RNA was purified using a column (RNeasy mini, QIAGEN) before and after digestion of the remaining template DNA with DNase I (Takara, Japan).

Plasmids encoding the α -domain were constructed by ligating two PCR fragments prepared using pUC-MDV-LR (27) as a template and primers (CGTCACGGTC-GAACTCCC and AGTCACGGGCTAGCGCTTTC) or using pUC19 as a template and primers (AGTTC-GACCGTGACGAATTGTGAGCGGATAACAATTC and GCGCTAGCCCGTGACTCTATGC). The latter fragment contains the α -domain gene that originated from the pUC19 vector. The ligation was performed with an InFusion kit (Clontech). The SmaI site inside of the α -domain was removed by inserting a point mutation (CCCGGG to CCCGTG). The mutant m1 sequence was synthesized chemically (Gene Design, Inc., Japan) and inserted into pUC-MDV-LR (27) as described above using the synthesized DNA as a template instead of pUC19 to obtain pUC-mdv- α .mutA. The other mutants (m2–14) were constructed by successive PCR and self-ligation processes using primers to delete or change the target sequences and pUC-mdv- α .mutA as a template. The entire sequences of these RNAs are provided in the Supplemental text. These RNAs were prepared by *in vitro* transcription as described above.

RNA synthesis by Q β replicase

RNA replications of the RNA species shown in Figure 2 were performed in Q β replicase buffer (125 mM Tris-HCl, pH 7.8, 5 mM MgCl₂), 0.005% BSA, 1.25 mM NTPs, [³²P]-UTP with each RNA (100 nM) and Q β replicase (100 nM) for 20 min at 37°C. The mixture was applied to 8% polyacrylamide gels for electrophoresis, and the ssRNAs and dsRNAs were quantified by autoradiography according to a previous study (5). RNA replication of the RNAs containing an α -domain gene was performed in an amino acid-depleted translation system (28) containing [³²P]-UTP with each RNA (100 nM) and Q β replicase (100 nM) for 25 min at 37°C. The Q β replicase was purified according to a previous study (29). To obtain the time course data (Figure 7C), the reaction was encapsulated in a water-in-oil emulsion to repress the amplification of nonspecific RNA according

to a previous study (4). The quantification of ssRNA and dsRNA was performed by autoradiography after agarose gel electrophoresis according to a previous study (28).

α -Complementation assay

RNAs encoding the α -domain gene were mixed with the highly purified reconstituted translation system (30) containing ω protein (EA reagent, Clontech) and a fluorescent substrate (50 μ M Tokyo Green[®]- β Gal, Sekisui Medical, Japan) according to a previous study (31). The fluorescence at 37°C was detected using an Mx3005P QPCR system (Agilent Technologies, USA).

Secondary structure prediction and calculation of GC number in loops

The secondary structure predictions of RNAs shown in Figure 2, and mdv- α mutants were prepared by Vienna RNA web services (<http://rna.tbi.univie.ac.at/>) (32,33) using the centroid structures. For viral genome analysis, we obtained all viral genome sequences registered in the NCBI database on 8 January 2015 (the list of accession numbers is provided in the Supplemental text), and the centroid secondary structures were predicted using the RNAfold algorithm in the 2.1.8 Vienna RNA package (32,33). We applied this analysis to sequences of all types of genomes (ssRNA, dsRNA, ssDNA, dsDNA and retrovirus genomes). Based on the secondary structural predictions, we counted the average loop size and average GC number in loops. For genomes larger than 4500 nt, we separated the sequences to a set of fragments less than 4500 nt, and after secondary structure prediction of the fragments, the GC numbers in loops of all of the separated fragments were averaged. Extremely large genomes (>40 000 nt) of some bacterial dsDNA viruses were removed before analysis to reduce the calculation load. The numbers of virus genomes in each category are shown in Supplementary Table S3.

RESULTS

A schematic model of the dsRNA formation at replication

In a previous study, Axelrod et al. reported that the dsRNA ratio of the replication product decreases with insertions of sequences that form a stem structure, but not with sequences that do not form any particular structure (i.e. a loop structure) (26). In another study, dsRNA formation was reported to occur during replication primarily between a template RNA and the complementary newly synthesized RNA using the template RNA (28). Based on these results, we constructed a model of dsRNA formation during replication (Figure 1). In this model, as shown in the figure, (A) a template and a newly synthesized strand are extruded as ssRNAs from the replicase (34,35), and each strand forms intra-molecular secondary structures, (B) if both strands have loop structures at the same position, these loops are likely to associate to form a partial dsRNA (36), (C) the partial dsRNA propagates to other regions of the strand and (D) full-length dsRNA is generated after replication is complete.

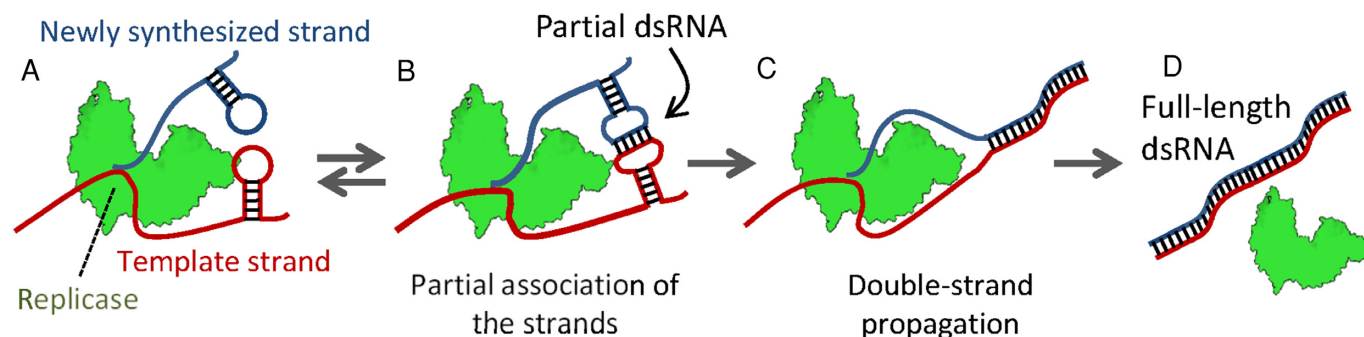


Figure 1. Schematic model of dsRNA formation during replication. (A) Both a template RNA and a newly synthesized complementary RNA exit the replicase as single strands and form intra-molecular structures. (B) If both strands contain loop structures at the same position, the loops associate to form a partial dsRNA. (C) The partial dsRNA propagates and (D) forms full-length dsRNA after replication is complete.

A key event in the formation of dsRNA in this model is the partial formation of dsRNA shown in (B); therefore, the rate of formation and the stability of the partial dsRNA should influence the formation of the full-length dsRNA. The rate of partial dsRNA formation should depend on the sizes of the associated loops in both the template and the newly synthesized strands. The stability of the partial dsRNA should depend on the GC ratio in the loop region because hybridization strength is greater for GC pairs than AU pairs. In this study, we first attempted to verify this model by examining these predictions using RNAs with various structures and nucleotide compositions.

Design of RNAs with various structures and nucleotide compositions

To investigate the effect of structures and nucleotide compositions on the dsRNA formation, we designed 28 RNAs with various secondary structures and nucleotide compositions and inserted the RNAs into position 63–64 of a 222-nt RNA (MDV-1), a replicable template for Q β replicase (Figure 2, Supplementary Table S1). These inserts were designed not to destruct the secondary structures of MDV-1 and to preferably form only a single stable structure, except for RNAs 23–28, which were designed to form two different stable structures.

Effects of the sizes and nucleotide compositions of loop structures

We investigated the effect of the insertion of loop structures. We first used RNAs with insert sequences that form 4- to 20-nt loop structures consisting of G and A (1. Loop GA4, 2. Loop GA6, 3. Loop GA8, 4. Loop GA10 and 5. Loop GA20). We incubated these RNAs (100 nM) with Q β replicase (100 nM) and nucleotides containing [32 P]-UTP for 25 min at 37°C and measured the ssRNAs and dsRNAs in the replication product by autoradiography after separation by agarose gel electrophoresis. We plotted the sum of the newly synthesized ssRNA and dsRNA amounts (total RNA synthesis) and the ratio of newly synthesized dsRNA to the newly synthesized total RNA (dsRNA ratio). Depending on the size of the inserted loop, total RNA synthesis exhibited a tendency to decrease (Figure 3A). The dsRNA ratio

decreased by the insertion of 4- to 8-nt loops but increased by the insertion more than 8-nt loops (Figure 3B). This increase of the dsRNA ratio by larger loops supports the notion that dsRNA formation is mediated by loop regions in a template RNAs and is consistent with the prediction according to the model shown in Figure 1.

According to the model, the dsRNA ratio should also depend on the stability of the partial dsRNAs that are formed during replication, and the stability should depend on the GC ratio in the hybridized region. Hence, the GC ratio in the inserted loop sequence is expected to affect the dsRNA ratio. To examine this prediction, we performed replication experiments using RNAs with insert sequences that consist of only U and A (6. Loop UA5, 7. Loop UA10, 8. Loop UA15 and 9. Loop UA20) and compared to the results with the loops that consist of half G and half A described above. In contrast to the G and A loops, the dsRNA ratio of the RNA containing U and A loops did not increase, even with the insertion of a 20-nt loop (Figure 3D). To further investigate the effect of the GC ratio, we compared the dsRNA ratios of the RNAs with 10-nt loops inserted with various nucleotide compositions, such as all G (10. Loop G10), all A (11. Loop A10), all U (12. Loop U10), G and A (4. Loop GA10) and U and A (7. Loop UA10). The dsRNA ratio exhibited larger values for two RNAs (Loop G10 and GA10) that had higher GC ratios (100% and 50%, respectively, Figure 3F). These results support the notion that the GC ratio, and hence the stability of the partial dsRNA formation between loops, is an important factor for dsRNA formation.

Effects of the sizes and the nucleotide compositions of stem structures

We next investigated the effect of insertion of sequences that form stem structures. We performed replication experiments with RNAs inserted with 20- to 80-nt stem structures consisting of G and A on one side and C and U on the other side of a stem structure (13. Stem GA20, 14. Stem GA40 and 15. Stem GA80) and measured the total RNA synthesis and the dsRNA ratios as described above. Both the total RNA and the dsRNA ratios exhibited a tendency to decrease slightly as the stem size increased (Figure 4A and B), consistent with a previous study (26) and our model

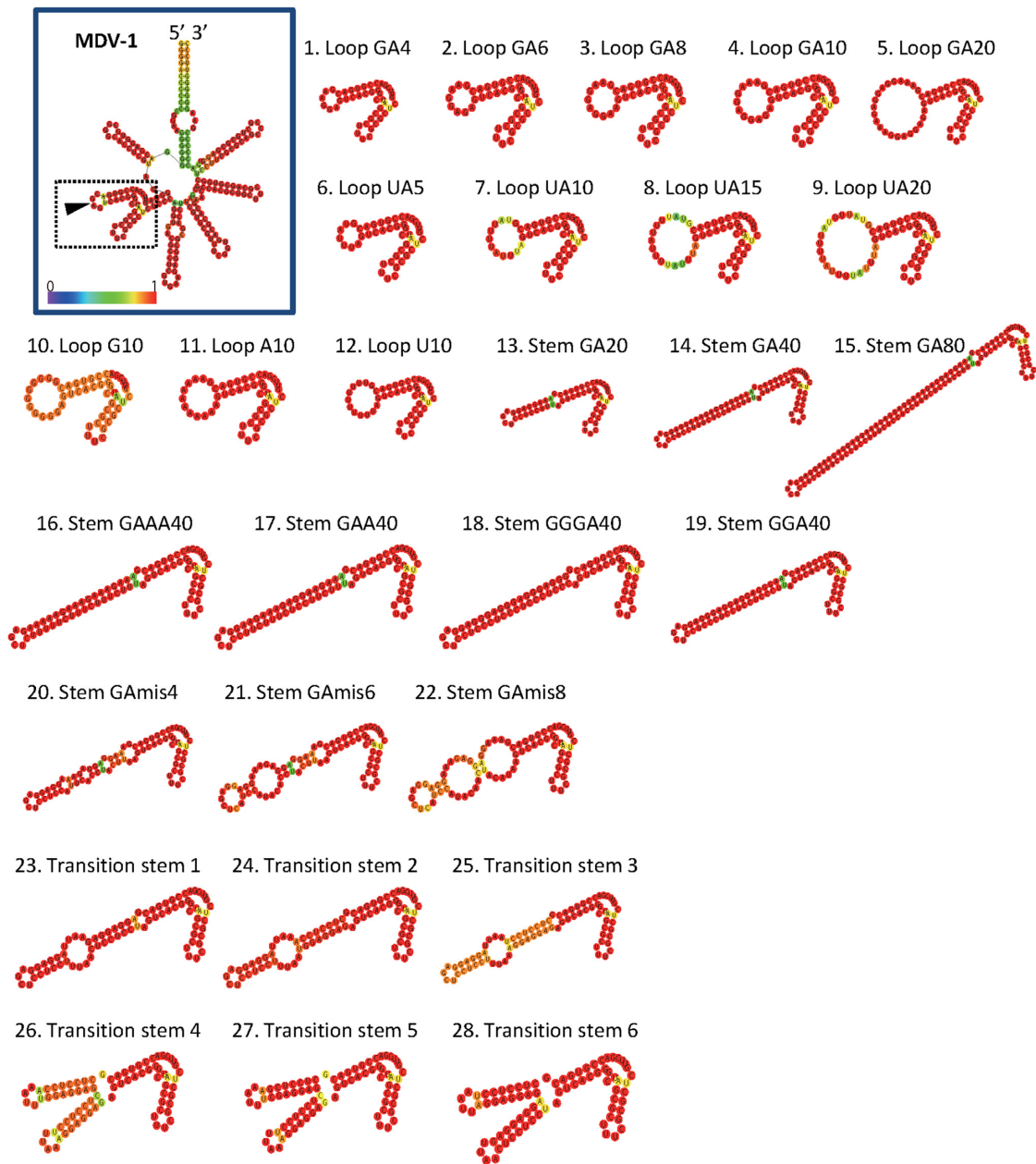


Figure 2. Predicted secondary structures of various inserted RNAs. We prepared 28 RNAs by inserting sequences that form various structures in site (63–64) of MDV-1 (arrow head in the solid box). The part of MDV-1 in the dashed box was replaced with various structures shown outside of the solid box. The color represents the assurance to form the structure according to the Centroid structure algorithm of Vienna RNA (33). The inserted sequences are shown in Supplementary Table S1.

shown in Figure 1, which predicts that the insertion of stem structures does not facilitate dsRNA formation.

We investigated the effect of the stability of stem structures by changing the GC ratio of the inserted stem structures. We performed replication experiments with RNAs inserted with 40-nt stems with 25–75% GC ratios (14. Stem GA40, 16. Stem GAAA40, 17. Stem GAA40, 18. Stem GGGA40 and 19. Stem GGA40) as described above. Both the total RNA synthesis (Figure 4C) and the dsRNA ratios (Figure 4D) did not change significantly in this range of GC

ratios, indicating that the stability of a 40-nt stem with a 25% GC ratio is sufficient to inhibit dsRNA formation.

To further investigate the effect of the stability of stem structures, we introduced 4–8 mismatch mutations into the 40-nt stem structure (20. Stem GAmis4, 21. Stem GAmis6 and 22. Stem GAmis8). The mismatch mutations did not affect the total RNA synthesis (Figure 4E) but significantly increased the dsRNA ratio when the mismatch number exceeded 4 (Figure 4F), indicating that the mismatch pair in stem structures should be at least less than 10% to inhibit dsRNA formation.

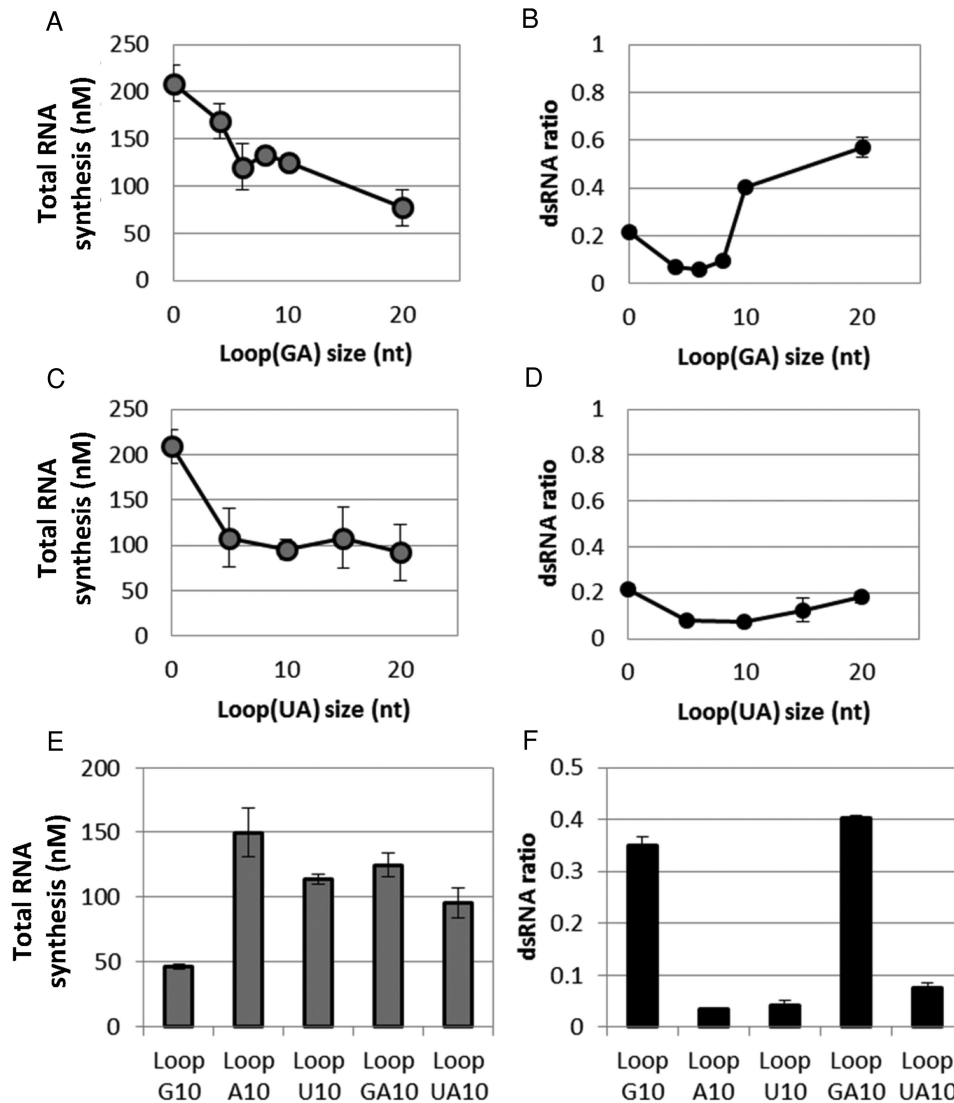


Figure 3. Effects of insertions of loop structures. RNA replication experiments were performed with 100 nM of each RNA and 100 nM of Q β replicase for 20 min at 37°C, and the total RNA synthesis, the sums of newly synthesized ssRNA and dsRNA (A, C, E), the dsRNA ratios and the ratios of newly synthesized dsRNA to the newly synthesized total RNA synthesis (B, D, E) were measured. The size effects of two types of loops composed of G and A (A, B) or U and A (C, D) were analyzed. The effects of nucleotide composition of 10-nt loop insertions were compared (E, F). Error bars indicate the standard deviations ($n = 3$).

In the experiments described above, we used RNAs that form primarily single stable structures, but generally, RNA possesses several semi-stable structures. We next investigated the effects of such multiple stable structures. We designed and inserted six sequences that form two different stable structures (one large stem or two small stems) with different ratios (23–28, transition stems 1–6). These sequences were designed to stabilize the two small stems rather than the one large stem as the number increased. The analysis of the replication products of these RNAs revealed that neither the total RNA synthesis nor the dsRNA ratios significantly differed among all of the RNAs (Figure 5A and B), indicating that the existence of two different stable structures does not affect the dsRNA ratio.

A rule for designing an RNA that replicates with less dsRNA formation

We next attempted to extract a simple index that quantitatively predicts the dsRNA ratio from all of the data obtained above. According to the model in Figure 1 and the results in Figure 3, the total numbers of G or C bases in the inserted loops were important factors for determining the dsRNA ratio. Based on this notion, we plotted the total GC numbers in loops of the inserted sequences against total RNA synthesis (Figure 6A) or the dsRNA ratio (Figure 6B). Total RNA synthesis was almost independent, but the dsRNA ratio strongly correlated with the GC number in loops of the inserted sequences (correlation coefficient is 0.81), indicating that the GC number in loops provides a useful index to predict the dsRNA ratio. According to this index,

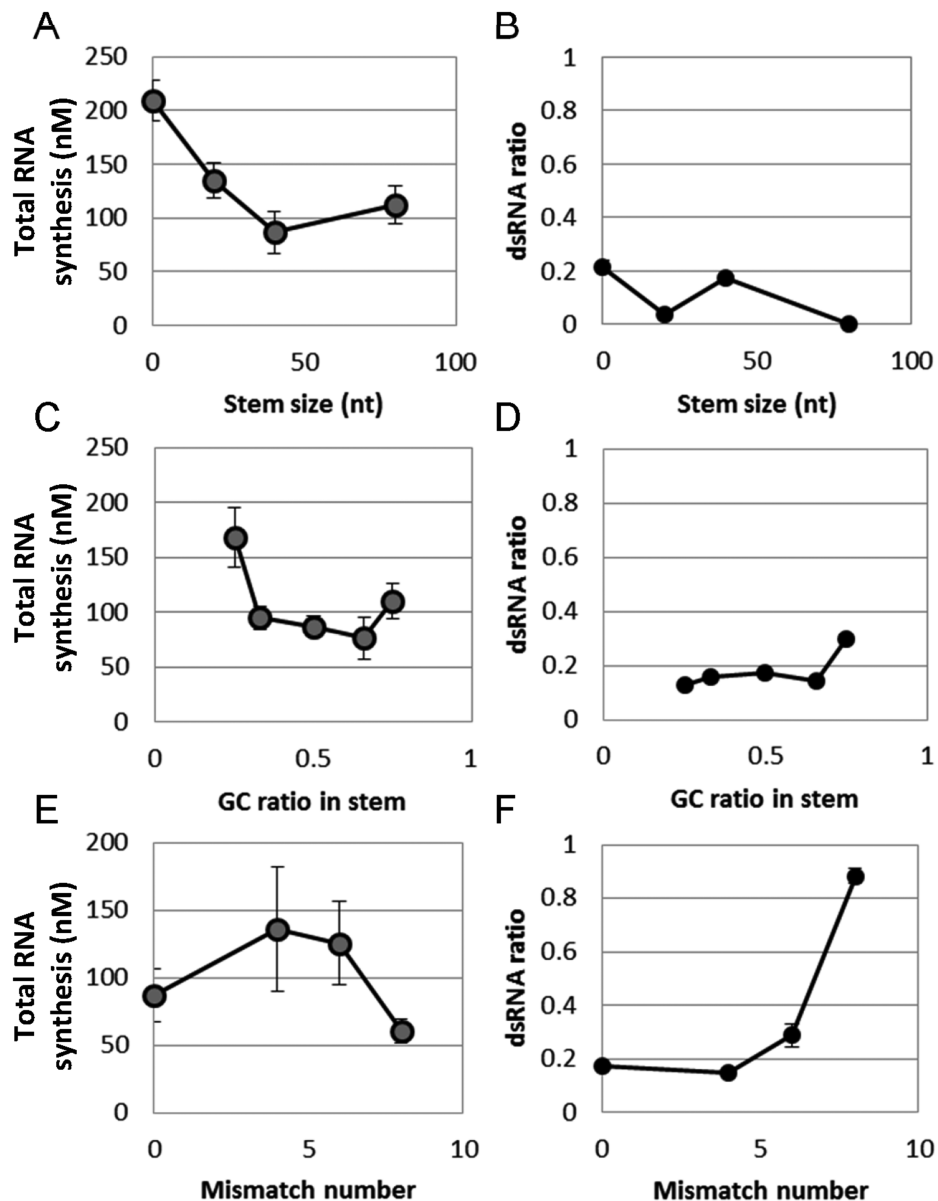


Figure 4. Effects of insertions of stem structures. RNA replication experiments were performed as described in the legend of Figure 3, and the total RNA synthesis (A, C, E) and dsRNA ratios (B, D, E) were measured. The effects of the insert sizes of stems (A, B), the GC ratios in the inserted 40-nt stem sequence (C, D) and the number of mismatch mutations in the 40-nt stem with a 50% GC ratio (E, F) were analyzed. Error bars indicate the standard deviations ($n = 3$).

we obtained a simple rule that RNA must obey to replicate as a single strand: ‘fewer (ideally <4) GC bases in loops.’ We also plotted other parameters (the sum of the loop size and the decrease of Gibbs’s free energy if the inserted loop regions form a partial dsRNA) against the dsRNA ratio, but we observed weaker correlations than the GC number in loops (Supplementary Figure S1).

There is a possibility that dsRNA is formed both during sample processing and replication. We performed RNA replication with three types of RNAs (MDV-1, 4. Loop GA10, and 22. Stem GAMis8) as described in ‘Materials and Methods’ section, to evaluate the amount of dsRNA formed during sample processing. T1 ribonuclease (Life Technologies), an ssRNA-specific ribonuclease, were added

immediately after replication, according to a previous report (23), to avoid further dsRNA formation. The amount of dsRNA was almost unchanged even after T1 ribonuclease treatment (Supplementary Figure S3), which indicates that dsRNA formation during sample processing was negligible in our experiments.

In the above-mentioned experiments, we used Q β replicase that contained a host factor, the ribosomal S1 subunit. To investigate whether the dsRNA ratio observed was dependent on the host factor S1, we performed RNA replication using core Q β replicase lacking the S1 subunit, both in the presence and the absence of the S1 subunit. We used three RNA templates (MDV-1, 4. Loop GA10, and 22. Stem GAMis8) with varying GC numbers in the inserted

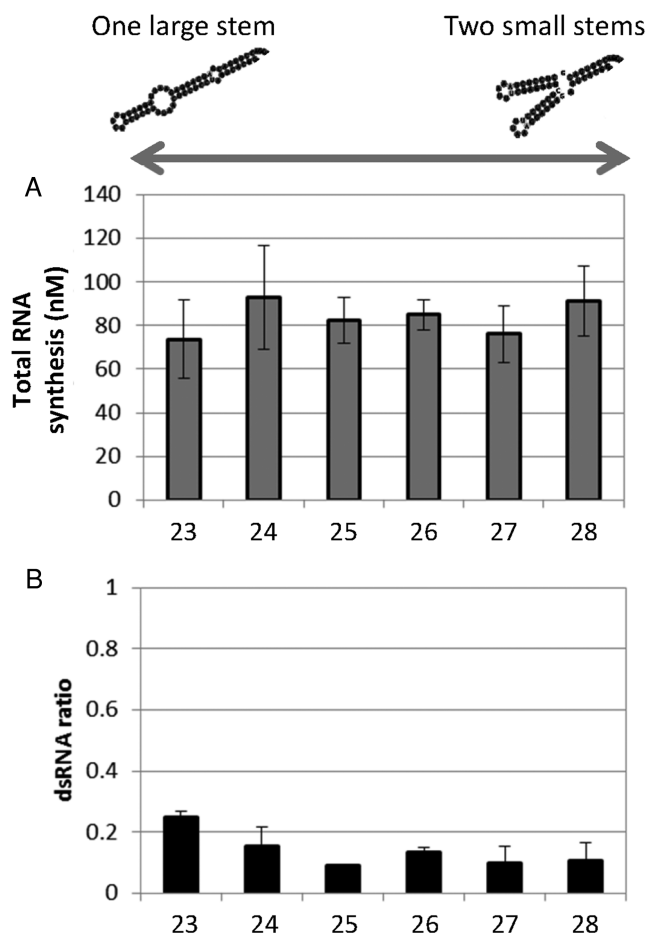


Figure 5. Effects of insertions of RNAs forming two different stable structures. RNA replication experiments were performed with 100 nM of each RNA (number 23–28) as described in the legend of Figure 3, and the total RNA synthesis (A) and dsRNA ratios (B) were measured. The inserted sequences were designed to form two different stable structures (a large stem or two small stems). The probability of forming the two small stems was designed to increase as the RNA number (23–28) increased. Error bars indicate the standard deviations ($n = 3$).

loop (3, 6 and 12, respectively). Presence of the S1 subunit decreased the dsRNA ratio for all three RNA templates, consistent with a previous report (8,23). However, the tendency of increasing dsRNA ratios, dependent on the GC number in the inserted loops, remained unchanged even in the absence of the S1 subunit (Supplementary Figure S5). This result indicates that the relationship between the dsRNA ratio and the GC number in loops is independent of the presence of the S1 subunit.

Development of an RNA that encodes an α -domain gene of β -galactosidase and continuously replicates as a single strand

Although we found a simple rule for an RNA that replicates with less dsRNA formation, it is not guaranteed that we can develop a continuously replicable RNA by following this rule. More importantly, there is also no guarantee that we can find a sequence that both satisfies the rule and maintains the activity of the encoded gene. To examine the usefulness of this rule, we attempted to design an ssRNA that is con-

tinuously replicable as a single strand and that encodes an active gene by following this rule. As a criterion of 'continuously replicable as a single strand,' we employed the condition of 'less than 0.5 dsRNA ratio,' because if the dsRNA ratio becomes >0.5 , the ssRNA only decreases as replication proceeds. As a target gene, we used the α -domain of β -galactosidase of *Escherichia coli*. The activity of this small gene (~ 300 nt) is monitored by α complementation in the presence of the omega-domain of β -galactosidase (31,37). We introduced the α -domain into the same site of MDV-1 as with the RNAs in Figure 2 to obtain mdv- α RNA (original mdv- α). Secondary structure prediction of this RNA revealed the existence of many large loops (Figure 7A, original mdv- α). We performed the replication experiment with this RNA and found that almost all synthesized RNA was dsRNA (Figure 7B, original).

To render this RNA replicable as a single strand, we first introduced 75 synonymous mutations that reduced the GC number in loops to <5 to obtain mutant m1. The predicted secondary structure of m1 exhibited reductions in size and the number of loops compared to the original (Figure 7A, m1). The replication experiment with m1 indicated that the synthesis of ssRNA increased significantly, but the dsRNA ratio was still greater than 0.5 (Figure 7B, m1), likely because the mutant m1 still had six remaining loops that contained three or four GCs (Loops 1–6 of m1, Figure 7A). Further reduction of the GC number by introducing synonymous mutations would be difficult in most of these loops because it would require massive reconstruction around the loop regions. Therefore, we first attempted to identify target loops that must be improved and then to reconstitute the structures around the loops. To identify the target loops, we deleted the remaining Loops 1–6 with various combinations to obtain the mutants m2–m12 and measured the dsRNA ratios of the replication products (Table 1). The dsRNA ratios of these deletion mutants were all lower than that of m1. Among the deletion mutants, m3, which lost Loops 1–5, showed the lowest dsRNA ratio (0.34), and m11, which lost three loops (Loops 1, 2 and 4), lost the fewest loops, yet exhibited a dsRNA ratio of <0.5 . We then attempted to modify the Loops 1, 2 and 4 regions to reduce the GC number by introducing synonymous mutations instead of deletions. For Loops 2 and 4, we reconstructed structures around the loops by introducing 10 synonymous mutations to decrease the GC number in the loops to <3 . For Loop 1, which locates in the 5'-untranslated region, we simply deleted G and C in this loop to obtain mutant m13 or substituted the sequence of the loop with all U bases to obtain mutant m14. Both mutants (m13 and m14) exhibited dsRNA ratios of <0.5 (0.47 and 0.45, respectively, Figure 7B) while maintaining the original amino acid sequence of the encoding gene.

For an ssRNA to serve as a genome, it must replicate continuously and also encode active genes. To examine whether the ssRNAs developed above satisfy these requirements, we first examined the continuity of the replication. The time course data indicated that the replication of the original mdv- α terminated abruptly at 20 min, but m13 and m14 continued linearly up to 50 min or 30 min, respectively, to amplify up to ~ 30 - or 15-fold, respectively (Figure 7C). We next examined the activity of the encoded α -domain

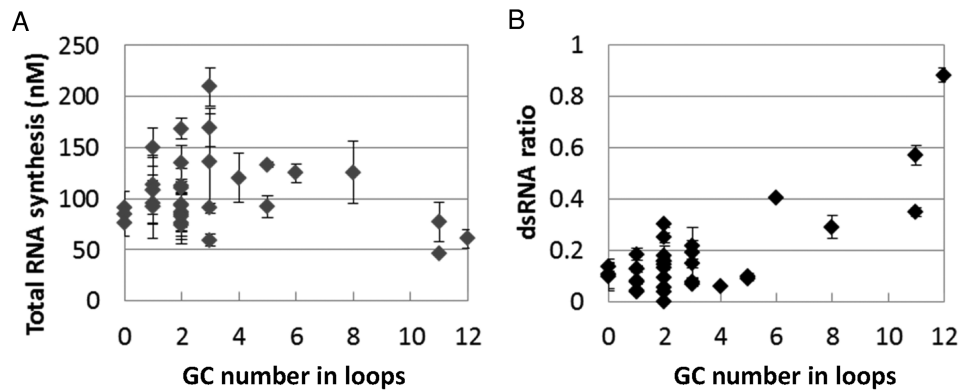


Figure 6. Effects of the GC numbers in loops of the inserted structures. The data shown in Figures 1–3 were re-plotted against the GC numbers in loops of the inserted sequences. The correlation efficient between the dsRNA ratio and the GC number in loops is 0.81. The plot against other parameters (GC ratio and dG) exhibits weaker correlations (Supplementary Figure S1).

Table 1. List of the deleted loops of MDV- α mutants

	Loop 1	Loop 2	Loop 3	Loop 4	Loop 5	Loop 6	DS ratio*	RNA synthesis (nM) ^a
original	+	+	+	+	+	+	0.76	24
m1	+	+	+	+	+	+	0.74	68
m2	+	–	–	–	–	+	0.65	107
m3	–	–	–	–	–	+	0.34	170
m4	+	–	–	–	–	–	0.67	90
m5	+	–	–	–	–	+	0.70	90
m6	–	+	–	–	–	+	0.45	155
m7	–	–	+	–	–	+	0.42	122
m8	–	–	–	+	–	+	0.49	117
m9	–	–	–	–	+	+	0.42	114
m10	–	+	–	–	+	+	0.50	108
m11	–	–	+	–	+	+	0.47	109
m12	–	–	–	+	+	+	0.55	99
m13	– ^b	– ^b	+	– ^b	+	+	0.47	124
m14	– ^b	– ^b	+	– ^b	+	+	0.45	110

^aRNA replication was performed with each mdv- α mutant (100 nM) and Q β replicase (100 nM) for 20 min at 37°C, and the total RNA synthesis and the dsRNA ratios were measured as described in ‘Materials and Methods’ section. The mutants that exhibit <0.5 dsRNA ratios are shadowed.

^bThe loops were reduced by synonymous substitution rather than deletion.

gene through *in vitro* translation and the α complementation assay. The RNAs (mdv- α , m1, m3, m13 and m14) were mixed with a reconstituted translation system containing a fluorescent substrate and the omega protein, and the fluorescence was monitored during the incubation (Figure 7D). For mutant m3, which lost Loops 1–5, the fluorescence was at the same level as that with no RNA, while for the original mdv- α , mutants m14, m1 and m13 showed higher fluorescence than that with no RNA, indicating that these RNAs, except for m3, encoded active α -domain genes. Taken together, these results demonstrate that based on the design rule, we succeeded in developing RNAs (m13 and m14) that possess the capacity to replicate continuously as a single strand and encode an active gene.

Generality of the rule among natural ssRNA viral genomes

Finally, we examined whether natural viral ssRNA genomes obey the same rule that we developed. We first investigated the predicted secondary structure of the ssRNA genome of bacteriophage Q β , which has been reported to replicate with a low dsRNA ratio during *in vitro* replication (38), and

found that the GC number in loops is at the same level as that of mutant m14 and is significantly lower than that of the random sequences of the same size and GC ratio, supporting the idea that the ‘less GC number in loop’ rule is also valid in this genome (Supplementary Figure S2 and Table S2). To further examine the generality of the rule, we investigated all sequence data of ssRNA viral genomes that have been registered in the NCBI database and analyzed the secondary structures to count the GC numbers in the loops. A plot of the GC number against the genome size revealed that the GC numbers in loops of the ssRNA genomes of bacteria and fungi are specifically localized to a region of <1 GC in the loops (Figure 8A), suggesting that the ssRNA viral genomes of these organisms obey the rule. Next, we applied the same analysis to bacterial and fungal viruses possessing other types of genomes (double-stranded RNA, single- and double-stranded DNA) (Figure 8B). The GC numbers in loops of other types of genomes are higher than those of ssRNA genomes, indicating that the rule of less GC number in loops is specific to ssRNA genomes. These results suggest that the rule obtained in the *in vitro* experiments using artificial RNAs is a general rule that governs

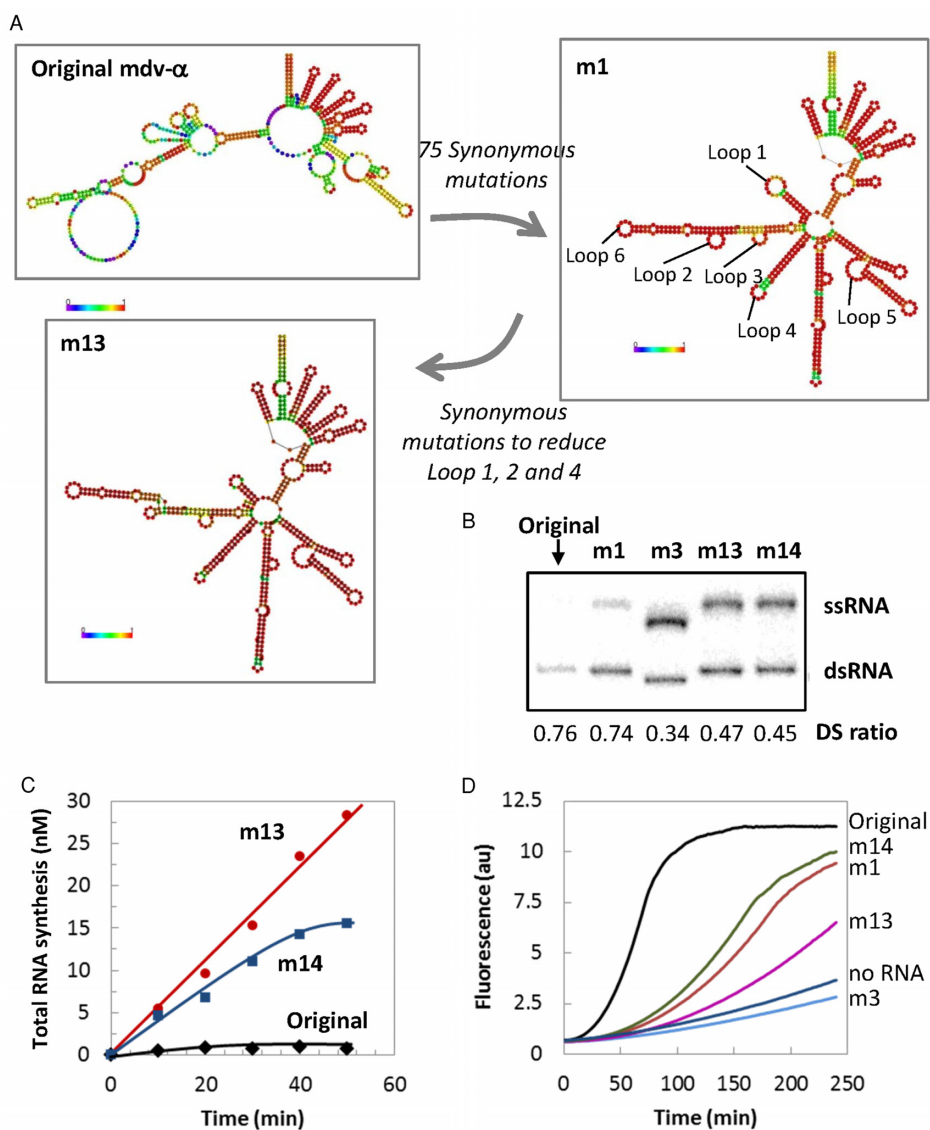


Figure 7. Development of continuously replicable ssRNAs encoding the α -domain of β -galactosidase. (A) Predicted secondary structures of some ssRNAs. The original mdv- α RNA was constructed by inserting the α -domain gene of β -galactosidase into MDV-1 RNA. First, we introduced 75 synonymous mutations to reduce the GC number in loops to less than 5 and to obtain mutant m1. Second, after selecting target loops to be improved, we further introduced several synonymous mutations and deletions to obtain mutants m13 and m14 (the structure of m14 is almost the same as that of m13). (B) Replication of the mdv- α mutants. Some mdv- α mutants (100 nM) were incubated with Q β replicase (100 nM) in the presence of [32 P]-UTP for 20 min at 37°C, and the replicated ssRNA and dsRNA were detected by autoradiography after polyacrylamide gel electrophoresis, as described in ‘Materials and Methods’ section. The dsRNA ratios are shown at the bottom. (C) Time-course data of replication. Replication was performed with each RNA (1 nM) and Q β replicase (100 nM) as described in ‘Materials and Methods’ section, and the total synthesized RNA amounts (the sum of ssRNA and dsRNA) were measured. (D) α -Complementation activities. Each RNA was mixed with the reconstituted translation system in the presence of the omega protein of β -galactosidase and fluorescence substrate, as described in ‘Materials and Methods’ section. The fluorescence was measured every minute.

natural viral ssRNA genomes of all bacteria and of most fungi presently known.

DISCUSSION

In this study, we attempted to understand design principles of an ssRNA genome that replicates without dsRNA formation. We first proposed a structure-dependent dsRNA formation model at replication (Figure 1) and then obtained a set of supportive evidence for the model using RNAs with various structures along with Q β replicase. From these results, we extracted a critical rule for RNA sequences (less

GC number in loops) to continuously replicate as a single strand, and then we designed an ssRNA that encoded an α -domain of β -galactosidase and was continuously replicable, mainly as a single strand. Furthermore, we also found evidence that the ssRNA genomes of all bacterial and of most fungal viruses presently known obey this rule. In summary, this study revealed one of the structural design principles of an ssRNA genome, which is useful for developing an artificial ssRNA genome and contributing to our understanding of the structural constraints governing the ssRNA genome of bacterial and fungal viruses.

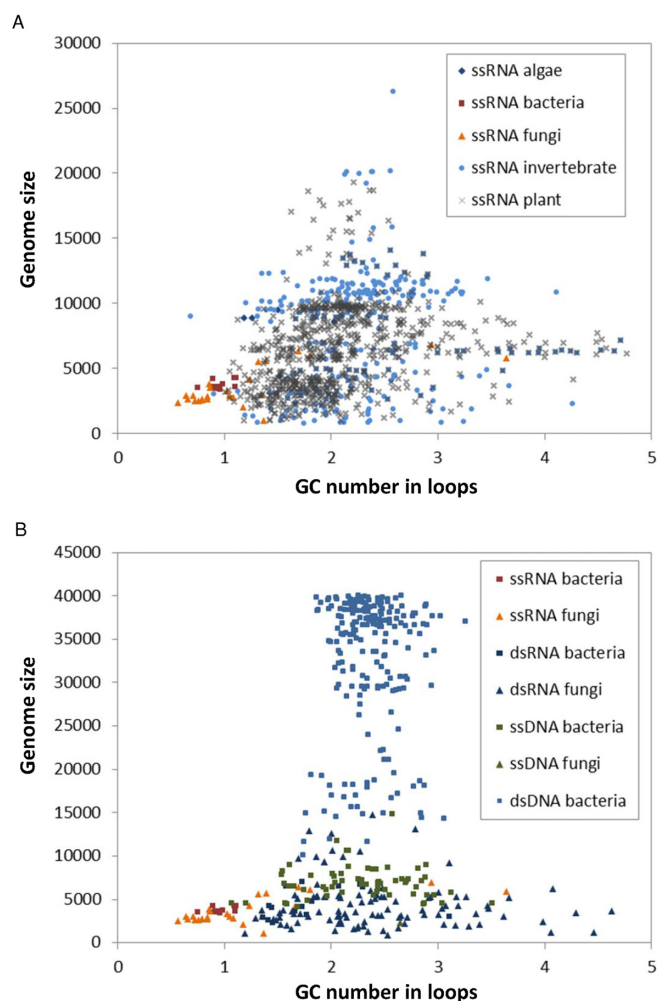


Figure 8. The GC numbers in loops of the predicted secondary structures of various viral genomes. (A) The genome sequences of single-stranded RNA viruses of various host organisms (6 algae, 11 bacteria, 26 fungi, 322 invertebrate, 787 plant and 656 vertebrate) were subjected to secondary structure prediction, and the GC numbers in loops were counted. (B) The sequences of other genome types of bacterial and fungal viruses (6 ssRNA, 15 dsRNA, 63 ssDNA, 1325 dsDNA virus of bacteria, 26 ssRNA, 108 dsRNA and 1 ssDNA virus of fungi) were analyzed by the same method as described in (A).

A plausible mechanism for the dsRNA reduction by the ‘less GC number in loops’ rule is that there is a decrease in partial dsRNA formation between loops on the template and the newly synthesized RNA, as shown in the model (Figure 1). We do not deny the possibility that the mechanism of dsRNA formation depends on the host factors of Q β replicase (EF-Tu and Ts), one of which is known to mediate dsRNA separation during replication (35). Nevertheless, if this dsRNA formation mechanism is independent of the types of replication enzymes, this rule might be applicable to other RNA replications catalyzed by different replication enzymes, including ribozymes. The formation of dsRNA is considered to be a serious problem in RNA replication by ribozymes in the hypothetical RNA world (3) and in *in vitro* RNA replication systems (39). The results obtained in this study suggest that the dsRNA problem of

primitive genomic RNAs can be solved by reducing the GC number in loops.

The design rule described in this study is also useful in the development of an artificial cell-like system that is composed of an artificial ssRNA genome. In previous studies, we constructed an evolvable artificial cell-like system harboring an ssRNA genome (4,5) that encoded only a single gene. The design method developed in this study offers a way to introduce new genes into the ssRNA genome and to maintain the activities of the encoded genes.

There is a limitation in the design rule derived in this study. The design method according to the ‘less GC in loops’ rule presently depends on secondary structure prediction, which is reported as valid for only small RNAs of typically <500 nt (40). For the design of longer ssRNAs, the development of more reliable prediction algorithms or a combination of experimental methods to detect loop structures, such as Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), (41) would be required.

A remaining unanswered question is why the ‘less GC in loops’ rule is valid for bacterial and fungal virus genomes, but not for other higher organisms, such as vertebrates and plants. Some vertebrate or plant ssRNA viruses replicate with an intermediate dsRNA stage, known as replicative intermediate forms (RF). Examples of these are poliovirus (42), enterovirus (43), corona virus (44), tobacco mosaic virus (45) and cowpea mosaic virus (46). During the replication of these viruses, a single-stranded RNA genome is synthesized from the double-stranded intermediate form. Therefore, the nascent strand can be synthesized as a single strand, even if it possesses a high GC content in loop. Another possibility is that RNA-binding proteins might bind and inhibit the hybridization between a template and a newly synthesized strand. SsRNA viruses of vertebrates include clinically important viruses, such as coronaviruses and hepatitis C virus. Although further investigations are required, the discovery of the mechanisms that inhibit dsRNA formation in vertebrates might provide a new drug target for ssRNA virus infection.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Funding for open access charge: Exploratory Research for Advanced Technology, Japan Science and Technology Agency.

Conflict of interest statement. None declared.

REFERENCES

- Gilbert, W. (1986) Origin of life – the RNA World. *Nature*, **319**, 618–618.
- Joyce, G.F. (1989) RNA evolution and the origins of life. *Nature*, **338**, 217–224.
- Szostak, J.W., Bartel, D.P. and Luisi, P.L. (2001) Synthesizing life. *Nature*, **409**, 387–390.
- Bansho, Y., Ichihashi, N., Kazuta, Y., Matsuura, T., Suzuki, H. and Yomo, T. (2012) Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. *Chem. Biol.*, **19**, 478–487.

5. Ichihashi, N., Usui, K., Kazuta, Y., Sunami, T., Matsuura, T. and Yomo, T. (2013) Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nat. Commun.*, **4**, 2494.
6. Haruna, I. and Spiegelman, S. (1965) Autocatalytic synthesis of a viral RNA in vitro. *Science*, **150**, 884–886.
7. Takeshita, D. and Tomita, K. (2012) Molecular basis for RNA polymerization by Qbeta replicase. *Nat. Struct. Mol. Biol.*, **19**, 229–237.
8. Takeshita, D., Yamashita, S. and Tomita, K. (2014) Molecular insights into replication initiation by Qbeta replicase using ribosomal protein S1. *Nucleic Acids Res.*, **42**, 10809–10822.
9. Tomita, K. (2014) Structures and functions of Qbeta replicase: translation factors beyond protein synthesis. *Int. J. Mol. Sci.*, **15**, 15552–15570.
10. Koppers, B. and Sumper, M. (1975) Minimal requirements for template recognition by bacteriophage Qbeta replicase: approach to general RNA-dependent RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 2640–2643.
11. Biebricher, C.K. and Luce, R. (1993) Sequence analysis of RNA species synthesized by Q beta replicase without template. *Biochemistry*, **32**, 4848–4854.
12. Brown, D. and Gold, L. (1995) Selection and characterization of RNAs replicated by Q beta replicase. *Biochemistry*, **34**, 14775–14782.
13. Ugarov, V.I. and Chetverin, A.B. (2008) Functional circularity of legitimate Qbeta replicase templates. *J. Mol. Biol.*, **379**, 414–427.
14. Palmenberg, A. and Kaesberg, P. (1974) Synthesis of complementary strands of heterologous RNAs with Qbeta replicase. *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 1371–1375.
15. Obinata, M., Nasser, D.S. and McCarthy, B.J. (1975) Synthesis of probes for RNA using Qbeta-replicase. *Biochem. Biophys. Res. Commun.*, **64**, 640–647.
16. Feix, G. and Hake, H. (1975) Primer directed initiation of RNA complementary to potato spindle tuber viroid using Q beta replicase. *Biochem. Biophys. Res. Commun.*, **65**, 503–509.
17. Vournakis, J.N., Carmichael, G.G. and Efstratiadis, A. (1976) Synthesis of RNA complementary of rabbit globin mRNA by Qbeta replicase. *Biochem. Biophys. Res. Commun.*, **70**, 774–782.
18. Feix, G. (1976) Primer-dependent copying of rabbit globin mRNA with Qbeta replicase. *Nature*, **259**, 593–594.
19. Owens, R.A. and Diener, T.O. (1977) Synthesis of RNA complementary to potato spindle tuber viroid using Q beta replicase. *Virology*, **79**, 109–120.
20. Weissmann, C., Feix, G., Slor, H. and Pollet, R. (1967) Replication of viral RNA. XIV. Single-stranded minus strands as template for the synthesis of viral plus strands in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **57**, 1870–1877.
21. Biebricher, C.K., Diekmann, S. and Luce, R. (1982) Structural analysis of self-replicating RNA synthesized by Qbeta replicase. *J. Mol. Biol.*, **154**, 629–648.
22. Usui, K., Ichihashi, N., Kazuta, Y., Matsuura, T. and Yomo, T. (2014) Effects of ribosomes on the kinetics of Qbeta replication. *FEBS Lett.*, **588**, 117–123.
23. Vasilyev, N.N., Kutlubaeva, Z.S., Ugarov, V.I., Chetverina, H.V. and Chetverin, A.B. (2013) Ribosomal protein S1 functions as a termination factor in RNA synthesis by Qbeta phage replicase. *Nat. Commun.*, **4**, 1781.
24. Morozov, I.Y., Ugarov, V.I., Chetverin, A.B. and Spirin, A.S. (1993) Synergism in replication and translation of messenger RNA in a cell-free system. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 9325–9329.
25. Mills, D.R., Dobkin, C. and Kramer, F.R. (1978) Template-determined, variable rate of RNA chain elongation. *Cell*, **15**, 541–550.
26. Axelrod, V.D., Brown, E., Priano, C. and Mills, D.R. (1991) Coliphage Q beta RNA replication: RNA catalytic for single-strand release. *Virology*, **184**, 595–608.
27. Inokuchi, Y., Kajitani, M. and Hirashima, A. (1994) A study on the function of the glycine residue in the YGDD motif of the RNA-dependent RNA polymerase beta-subunit from RNA coliphage Q beta 1. *J. Biochem.*, **116**, 1275–1280.
28. Usui, K., Ichihashi, N., Kazuta, Y., Matsuura, T. and Yomo, T. (2013) Kinetic model of double-stranded RNA formation during long RNA replication by Qbeta replicase. *FEBS Lett.*, **587**, 2565–2571.
29. Ichihashi, N., Matsuura, T., Hosoda, K. and Yomo, T. (2010) Identification of two forms of Qbeta replicase with different thermal stabilities but identical RNA replication activity. *J. Biol. Chem.*, **285**, 37210–37217.
30. Kazuta, Y., Matsuura, T., Ichihashi, N. and Yomo, T. (2014) Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. *J. Biosci. Bioeng.*, **118**, 554–557.
31. Nishiyama, K., Ichihashi, N., Matsuura, T., Kazuta, Y. and Yomo, T. (2012) alpha-complementation in an artificial genome replication system in liposomes. *Chembiochem*, **13**, 2701–2706.
32. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.
33. Lorenz, R., Bernhart, S.H., Honer Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
34. Kidmose, R.T., Vasiliev, N.N., Chetverin, A.B., Andersen, G.R. and Knudsen, C.R. (2010) Structure of the Qbeta replicase, an RNA-dependent RNA polymerase consisting of viral and host proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 10884–10889.
35. Takeshita, D. and Tomita, K. (2010) Assembly of Q{beta} viral RNA polymerase with host translational elongation factors EF-Tu and -Ts. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15733–15738.
36. Lima, W.F., Monia, B.P., Ecker, D.J. and Freier, S.M. (1992) Implication of RNA structure on antisense oligonucleotide hybridization kinetics. *Biochemistry*, **31**, 12055–12061.
37. Ullmann, A., Jacob, F. and Monod, J. (1967) Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.*, **24**, 339–343.
38. Urabe, H., Ichihashi, N., Matsuura, T., Hosoda, K., Kazuta, Y., Kita, H. and Yomo, T. (2010) Compartmentalization in a water-in-oil emulsion repressed the spontaneous amplification of RNA by Q beta replicase. *Biochemistry*, **49**, 1809–1813.
39. Kim, D.E. and Joyce, G.F. (2004) Cross-catalytic replication of an RNA ligase ribozyme. *Chem. Biol.*, **11**, 1505–1512.
40. Mathews, D.H., Turner, D.H. and Zuker, M. (2007) RNA secondary structure prediction. *Curr. Protoc. Nucleic Acids Chem.*, Chapter 11, Unit 11.12.
41. Wilkinson, K.A., Merino, E.J. and Weeks, K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
42. Oberg, B. and Philipson, L. (1969) Replication of poliovirus RNA studied by gel filtration and electrophoresis. *Eur. J. Biochem.*, **11**, 305–315.
43. Triantafyllou, K., Vakakis, E., Kar, S., Richer, E., Evans, G.L. and Triantafyllou, M. (2012) Visualisation of direct interaction of MDA5 and the dsRNA replicative intermediate form of positive strand RNA viruses. *J. Cell Sci.*, **125**, 4761–4769.
44. Hagemeyer, M.C., Vonk, A.M., Monastyrskaya, I., Rottier, P.J. and de Haan, C.A. (2012) Visualizing coronavirus RNA synthesis in time by using click chemistry. *J. Virol.*, **86**, 5808–5816.
45. Buck, K.W. (1999) Replication of tobacco mosaic virus RNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **354**, 613–627.
46. Shanks, M., Lomonosoff, G.P. and Evans, D. (1985) Double-stranded, replicative form RNA molecules of cowpea mosaic-virus are not infectious. *J. Gen. Virol.*, **66**, 925–930.