BIOPHYSICS

Optimal terminal sequences for continuous or serial isothermal amplification of dsRNA with norovirus RNA replicase

Hidenao Arai[1], Koichi Nishigaki[1], Naoto Nemoto[1], Miho Suzuki[1] and Yuzuru Husimi[2]

[1]*Graduate School of Science and Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan*
[2]*Innovative Research Organization, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan*

The norovirus RNA replicase (NV3D^pol, 56 kDa, single chain monomeric protein) can amplify double-stranded (ds) RNA isothermally. It will play an alternative role in the *in vitro* evolution against traditional Qβ RNA replicase, which cannot amplify dsRNA and consists of four subunits, three of which are borrowed from host *E.coli*. In order to identify the optimal 3′-terminal sequence of the RNA template for NV3D^pol, an *in vitro* selection using the serial transfer was performed for a random library having the 3′-terminal sequence of ---UUUUUUNNNN-3′. The population landscape on the 4-dimensional sequence space of the 17th round of transfer gave a main peak around ---CAAC-3′. In the preceding studies on the batch amplification reaction starting from a single-stranded RNA, a template with 3′-terminal C-stretch was amplified effectively. It was confirmed that in the batch amplification the ---CCC-3′ was much more effective than the ---CAAC-3′, but in the serial transfer condition in which the ----CAAC-3′ was sustained stably, the ---CCC-3′ was washed out. Based on these results we proposed the existence of the "shuttle mode" replication of dsRNA. We also proposed the optimal terminal sequences of RNA for *in vitro* evolution with NV3D^pol.

Key words: *in vitro* selection, initiation of RNA replication, breathing of dsRNA terminus, shuttle mode replication, RNA-dependent RNA polymerase

Corresponding author: Yuzuru Husimi, Innovative Research Organization, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama 338-8570, Japan.
e-mail: husimi@mail.saitama-u.ac.jp

The norovirus RNA replicase (NV3D^pol) is a comparatively small (510 a.a.) RNA-dependent RNA polymerase with a single polypeptide chain. It can amplify double-stranded (ds) RNA isothermally. On the other hand, a traditional Qβ phage RNA replicase cannot amplify dsRNA[1–3] and consists of four subunits, three of which are proteins of the translational apparatus of host *E.coli*[4,5]. Therefore it was considered that NV3D^pol will play an alternative role in the *in vitro* autonomous RNA evolution. Especially NV3D^pol is a candidate of RNA replicase displayed on an *in vitro* virus[6–9] in order to realize an *in vitro* autonomous co-evolution of RNA and protein. Thus the *in vitro* enzymatic activities of NV3D^pol synthesized using cell-free translation system have been studied extensively[10]. Summary of the results (together with that of other studies[11–14] is as follows:

1. Replication manner is various as follows.
(1) Primer–independent replication starting from ssRNA and isothermal amplification of dsRNA.

In this case one of favorable 3′-terminal sequences was ---CCC-3′ or ---CCCC-3′. On the other hand, poly(A) tail was a unfavorable 3′-terminal sequence. Champion data of amplification factor was 54-fold (single strand). The mechanism of isothermal amplification of dsRNA is a strand dissociation replication.

(2) Long hairpin synthesis starting from a small stem-loop structure at 3′-terminus.

This is the strongest activity. The optimal size is a stem of 2 bp with a loop of 3–4 nts, and a larger stem-loop is very unfavorable. This long hairpin closed with 5′-G stretch and 3′-C stretch was unfavorable for further replication.

(3) RNA primer extension replication.

This activity is detected, but very weak.

In addition to above list, NV3D^{pol} has the terminal nucleotidyl transferase (TNT) activity. Addition of 1–3 C's (mainly C, but also other nucleotides) occurred at the probability ca. 50%.

2. Replication rate is not so high.

For a template of ssRNA of 50 nts having terminal sequence 5′-GGG------CCC-3′, the apparent doubling time in the initial phase is 12 min.

3. Replication fidelity is medium.

In the sequencing data of about 90 clones, we have rarely detected base substitutions at the constant region. Thus the error rate seems to be less than 1.5×10^{-4} miss copy/base. Fidelity of NV3D^{pol} is comparable with that of Qβ replicase.

4. Replication processivity is high.

When an ssRNA template of 500 nts was replicated without primer, a sharp amplified band at 500 bp on the nondenaturing PAGE was observed[10]. Considering that the primer extension activity is very weak, the drop-off event of NV3D^{pol} from the template strand seemed to be rare during 500 nts replication.

Objectives in this study at the planning phase were as follows: (1) to demonstrate the applicability of NV3D^{pol} as an amplifier to the *in vitro* selection, and (2) to identify the optimal 3′-terminal sequence of the RNA template for NV3D^{pol} using this *in vitro* selection. We used a serial transfer as an evaluation method. We designed a random RNA template library to have the 3′-terminal sequence of ---UUUUUUNNNN-3′, in order to minimize the interference due to the hairpin synthesis. The reason of a random sequence of size four is the saturation size of C-stretch at the 3′-terminus[10].

Unexpected results led us to the discovery of the "shuttle mode" of dsRNA replication. And we realized the optimal 3′-terminal sequence was different between for a batch replication reaction and for a serial or continuous replication reaction. Note that a serial transfer is essentially equivalent to a continuous replication reaction, and in this study we used only serial transfer, but the essential conclusions are able to be applied also to a continuous reactor. Based on these results we propose the optimal terminal sequences of RNA for *in vitro* evolution with NV3D^{pol}.

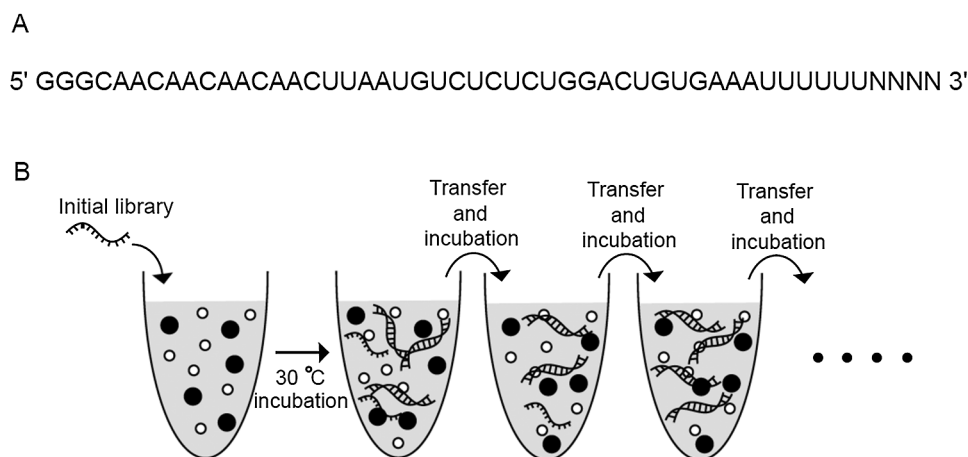## Materials and Methods

DNA

Plasmid pVL3Dwt (GenBank: AB039782[11]) harboring NV3D^{pol} gene was kindly provided from BML Inc. DNA oligomers used in this study were from Operon Inc. (Tokyo, Japan).

Preparation of Norovirus RNA replicase (NV3D^{pol})

Preparation procedure of NV3D^{pol} was same as preceding report[10] except that of adding an enrichment process with Amicon Ultra-0.5 (MWCO = 10 K; Millipore, Billerica, MA, USA) after translation reaction with cell-free protein synthesis system. Prepared NV3D^{pol} solution was stored at −80°C before use. Quantification of NV3D^{pol} was performed as described in ref. 10. Approximately 1.3 μg (= 23 pmol) NV3D^{pol} enzyme was synthesized from 1 mL cell-free protein synthesis system (Trans direct *insect cell*, Shimadzu, Kyoto, Japan). Concentration of prepared NV3D^{pol} solution was adjusted to approximately 0.3 pmol/μL.

Preparation of RNA random library

We designed the 3′-terminal sequence of RNA template random library as 5′---UUUUUUNNNN-3′ (Fig. 1A) in order to minimize the interference by hairpin synthesis. The rea-

A

5′ GGGCAACAACAACAACUUAAUGUCUCUCUGGACUGUGAAAUUUUUUNNNN 3′

B

Figure 1 Base sequence of the initial ssRNA library (A) and schematic illustration of serial transfer (B). Details of procedure of the serial transfer were described in Materials and Methods.

son of a random sequence of size four is the saturation size of C-stretch at the 3′-terminus[10]. When ---NNNN-3′ is ---NNAA-3′, a small stem-loop may be formed and an abortive long hairpin may be synthesized (Supplementary Fig. 1). Because double AU base pairs stem may be unstable, the probability of occurrence of this event may be minimized. Even if it occurs and the terminal AA sequences are omitted from the library before the serial transfer selection, the fact that poly(A) tail was a unfavorable terminal sequence[10] will allow the neglect of this event.

RNA library (GGG-U6N4 RNA) was synthesized by *in vitro* transcription. GGG-U6N4 DNA, which is a template DNA of GGG-U6N4 RNA, was prepared by PCR with cTemps(GGG-CCC) (5′-GCCAGTCGCCTGCAG<u>TAATAC GACTCACTATA</u>GGGCAACAACAACAACTTAATGTC TCTCTGAACTGTGAAATCTTATTCCC-3′, underline indicates T7Φ6.5 promoter sequence), prT7tempsGGG(+) (5′-GCCAGTCGCCTGCAGTAATACGACTCACTA-3′) and prU6N4(-) (5′-NNNNAAAAAATTTCACAGTTCAGAGA GACATTAA-3′, N; A:T:G:C ≈ 25:14:23:38, observed value for incomplete random synthesis). *In vitro* transcription was performed with RiboMAX™ Large Scale RNA Production system-T7 (Promega, Madison, WI, USA) using GGG-U6N4 DNA as a template. GGG-U6N4 RNA was purified on a denaturing polyacrylamide gel electrophoresis (PAGE) containing 8 M urea and desalted with Micro Bio-Spin column (Bio-Rad, Hercules, CA, USA). The concentration of GGG-U6N4 RNA was quantified from A_{260}. GGG-CCC and GGG-U6CAAC RNA were prepared in a same manner as GGG-U6N4 RNA using prcTempsCCC(-) (5′-GGGAATAAGAT TTCACAGTTCAGAGAGAC-3′) and prU6CAAC(-) (5′-GTTGAAAAAATTTCACAGTTCAGAGAGACATTAA-3′) in PCR, respectively.

Batch amplification reaction

In the batch reaction and the first round of serial transfer, RNA amplification was performed with purified NV3D^pol (0.06 μM) and the *in vitro* transcribed RNA template (0.25 μM) in a reaction buffer (50 mM Hepes-KOH [pH 7.0], 3 mM MnCl$_2$, 4 mM DTT, 1.25 mM rNTPs, and 1 U/μL of RNasin ribonuclease inhibitor plus (Promega)) at 30°C. Reaction solutions sampled at each scheduled time were stopped by adding 57 mM ethylenediamine-N,N,N′,N′-tetraacetic acid (EDTA) and subjected to non-denaturing or denaturing 10% PAGE followed by SYBRgreenII staining (LONZA, Rockland, ME, USA).

Serial transfer

The first round reaction (reaction volume = 10 μL) in the serial transfer (Fig. 1B) had various predetermined incubation times; 60 min for GGG-U6N4 RNA, and 120 min for GGG-CCC and GGG-U6CAAC RNA. 7 μL reaction aliquot was transferred to the next reaction tube containing 3 μL reaction solution (containing 50 mM Hepes-KOH [pH 7.0], 3 mM MnCl$_2$, 4 mM DTT, 4.17 mM rNTPs, 1 U/μL of

RNasin ribonuclease inhibitor plus and the same amount of fresh NV3D^pol solution), that is 1.4-fold dilution. From the second to the (N-1)^th round, the incubation time was 120 min, and for the last (N^th) one it was 240 min. All reactions were stopped by adding 57 mM EDTA and subjected to non-denaturing 10% PAGE followed by SYBRgreenII staining.

Analysis of 3′-terminal sequence of selected RNA species in the serial transfer

Sequencing of the 3′-terminal sequences of enriched dsRNA was performed as described in the preceding report[10]. In this study, U6N4Yadapter DNA (5′-CAAA<u>AAA AAATTTCACAGTTCAGAGAGAC</u>CTTAGATAATACG ACTCACTATAGGGTTAAC-3′, the underline indicates the hybridization region to RNA template) was used as a donor molecule of Y-ligation with the target RNA. The sequence data were analyzed by showing the population landscape on the 4-dimensional sequence space.

Results

Amplification from ssRNA (Batch reaction and the first round of serial transfer)

The RNA library was incubated with NV3D^pol at 30°C for 240 min and analyzed on PAGE (Fig. 2). As shown in Figure 2A, on a non-denaturing PAGE, a band of 50 bp was amplified with time, although a band of 50 bp did not exist at 0 min. This indicates that the amplified band corresponded to 50 bp dsRNA (including hairpin RNA if it existed). When the reaction solution was subjected to a denaturing PAGE containing 8 M urea, the amplification was observed for the bands of both 50 nts and 90–100 nts (Fig. 2B). The band of 50 nts corresponded to the amplified dsRNA from ssRNA having active 3′-terminal sequence, and that of 90–100 nts corresponded to the hairpin RNA elongated from ssRNA having the 3′-terminal small stem-loop. These situations in the batch amplification should occur also in the first round of the serial transfer, as shown in the next section.
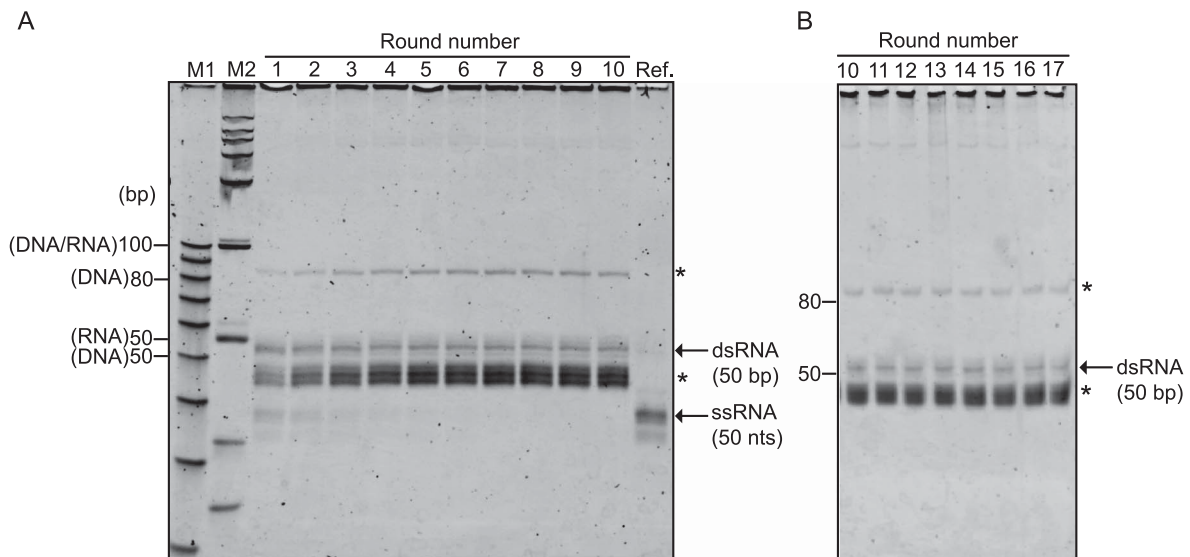
Serial transfer

The first round reaction solution was transferred to the next round reaction tube at 1.4-fold dilution and then incubated 120 min. This serial transfer procedure was repeated 9 or 16 times and analysed on a non-denaturing PAGE (Fig. 3A or B). Note that the last round incubation time was 240 min. In Figure 3, a band of 50 bp dsRNA appeared at a nearly constant intensity through the serial transfer. On the other hand, a band of 50 nts ssRNA decreased by round 6. The exogenous RNAs from a cell-free protein synthesis system (e.g. tRNAs), indicated with asterisks, were enriched naturally through serial transfer with 1.4-fold dilution.

When two other serial transfer experiments was performed with the repetition of 5-fold dilution and 30 min
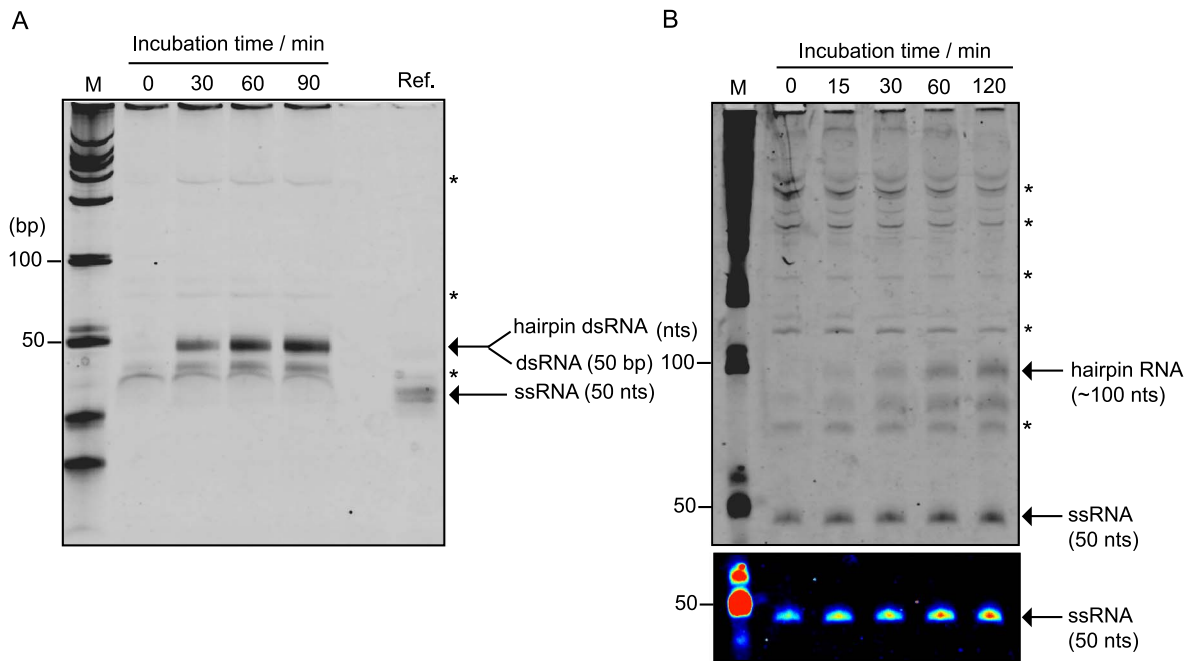
Figure 2 Batch amplification of the RNA random library with NV3D^{pol}. GGG-U6N4 RNA (5 pmol) was incubated with NV3D^{pol} (4 pmol) at 30°C. Reaction aliquots sampled at the indicated times were subjected to a non-denaturing 10% PAGE (A) or a denaturing 10% PAGE containing 8 M urea (B) followed by SYBRgreenII staining. In order to clarify the change of band intensity, the color-coded image of the bands of 50 nts ssRNA was shown at the bottom (red: the strongest, …., violet: the weakest). M; dsRNA ladder marker (BioDynamics Laboratory inc., Tokyo, Japan). Ref.; GGG-U6N4 RNA. Asterisks are exogenous RNAs from cell free protein synthesis system (e.g. tRNAs).


Figure 3 Non-denaturing PAGE analysis of the products in independent two ((A) and (B)) serial transfers. Initial random library, that is, GGG-U6N4 RNA (5 pmol) and NV3D^{pol} (1 pmol) was incubated at 30°C for 60 min in the first round reaction tube (reaction volume = 10 μL). 7 μL aliquot was transferred to the next round reaction tube (1.4-fold dilution). 57 mM EDTA was added into the first round reaction tube with the remaining 3 μL to stop the reaction. The second to the 9th (A) or 16th (B) round reaction tube was incubated for 120 min and transferred to the next test tube at 1.4-fold dilution. The 10th (A) or 17th (B) round reaction tube was incubated for 240 min. Each reaction samples were subjected to a non-denaturing 10% PAGE followed by SYBRgreenII staining. M1; 10 bp DNA step ladder (Promega). M2; dsRNA ladder marker. Ref.; GGG-U6N4 RNA. Asterisks are exogenous RNAs from cell free protein synthesis system (e.g. tRNAs).

Table 1   Sequences of clones selected from round #17

Clone number #	Sequence (3′-terminal region)	Number of clones
Initial library	ugucucucugaacugugaaauuuuuuNNNN 3′	–
1, 2	ugucucucugaacugugaaauuuuuu(A/G)CAA*C*	6
3	ugucucucugaacugugaaauuuuuuAGAA*C*	6
4	ugucucucugaacugugaaauuuuuuAACA*GUU*	6
5	ugucucucugaacugugaaauuuuuuGAAC*C*	4
6	ugucucucugaacugugaaauuuuuuCAGA*GAG*	4
7	ugucucucugaacugugaaauuuuuuGCCA*C*	3
8	ugucucucugaacugugaaauuuuuuCAAA*UU*	3
9	ugucucucugaacugugaaauuuuuuuAACU*UC*	2
10	ugucucucugaacugugaaauuuuuuuACGG*C*	2
11	ugucucucugaacugugaaauuuuuuuACCA*GA*	2
12	ugucucucugaacugugaaauuuuuuuCCCA*U*	2
13	ugucucucugaacugugaaauuuuuuuACGA*U*	2
14	ugucucucugaacugugaaauuuuuuuGGUA	2
15	ugucucucugaacugugaaauuuuuuuGCCC*CC*	1
16	ugucucucugaacugugaaauuuuuuuAACA*C*	1
17	ugucucucugaacugugaaauuuuuuuACAG*C*	1
18	ugucucucugaacugugaaauuuuuuuCGCA*G*	1
19	ugucucucugaacugugaaauuuuuuuCCAA	1

Bold letters correspond to the randomized 3′-terminal nucleotides. Italics are added nucleotide(s) by terminal nucleotidyl transferase activity of NV3D[pol]. The nucleotide of the -5 site of clone #1 and #2 are A and G, respectively, and the number of the clones were 4 and 2, respectively. The data were summarized based on the 3′-terminal four-nucleotides sequence.
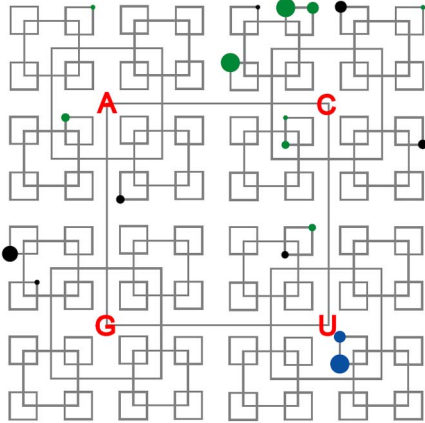
Figure 4 Population landscape of the round 17 in the serial transfer ((B) in Fig. 3) on the 4-dimensional sequence space. The quaternary 4-dimensional sequence space is displayed as a projection onto the 2-dimensional surface. Note each square is the projection of a tetrahedron, of which vertices correspond to each of four bases (A, U, G and C). The largest square corresponds to the 3′-terminal site (-1 site), the second largest to the -2 site, the third one to the -4 site, and the smallest to the -3 site. The order of sites (-1, -2, -4, -3) is arranged according to the Shannon information of the site calculated based on the base sequence of 49 clones. Thus, the largest (green) circle on the upper-most edge, for example, corresponds to (C,A,C,A), which means the terminal sequence --CAC-3′. And the upper right corner, for example, corresponds to (C,C,C,C). The area of the plotted solid circles is proportional to the number of the clones having that sequence. Green solid circles make a cluster connected with unit Hamming distance (D_H) and within $D_H \leq 2$ from the --CAAC-3′. Total population of this cluster is 24. Blue solid circles make another such cluster with total population 9. There are other isolated small populations indicated with black solid circles. Therefore the population landscape of round 17 has a main peak around --CAAC-3′ and a sub main peak around --AGUU-3′

incubation, or 2-fold dilution and 120 min incubation, the 50 bp species was washed out (Supplementary Fig. 2 shows the case of 2-fold dilution).

The sequence data of 3′-terminal region of RNA clones from round 17 in the serial transfer (Fig. 3B) were shown in Table 1. Addition of 1–3 nucleotides by the TNT activity was observed extensively. The 3′-terminal four nts sequences were plotted on the four dimensional sequence space (Fig. 4). In Figure 4, we observed the population landscape consisted with a main peak and a sub-main peak. The main peak was located around 5′-CAAC-3′ sequence and had about half of total population, shown with green solid circles.

Comparison of amplification efficiency between optimal sequences for the batch amplification and for the serial transfer
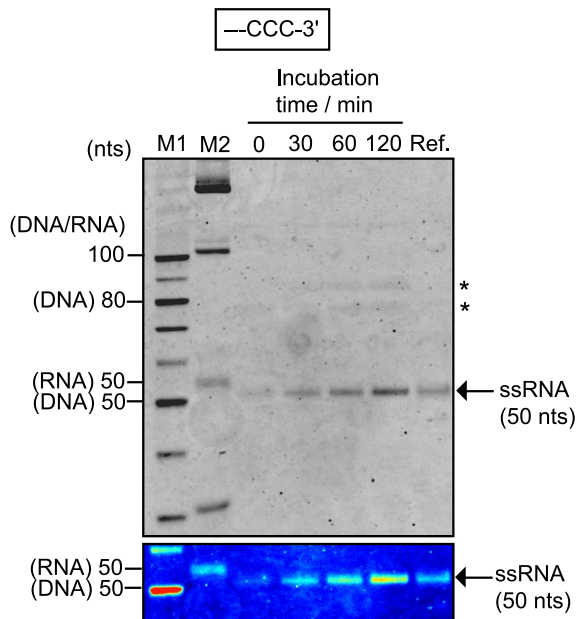
We compared the amplification efficiency of GGG-CCC and GGG-U6CAAC RNA (shown in Fig. 5A) in the batch amplification with NV3D[pol]. 5 pmol GGG-CCC or GGG-U6CAAC ssRNA was incubated with 1.2 pmol NV3D[pol] and observed the time-course on a denaturing 10% PAGE containing 8 M urea (Fig. 5B and C). As shown in Figure 5D, for the batch amplification, the GGG-CCC RNA was more optimal than GGG-U6CAAC RNA.

Next, we performed the serial transfer experiments using GGG-CCC or GGG-U6CAAC RNA, respectively. Each round in the serial transfer consisted of 1.4-fold dilution and 120 min incubation. Aliquots from each round were analyzed on a non-denaturing 10% PAGE (Fig. 6A and B). Quantification of amplification of dsRNA was shown in Figure 6C.
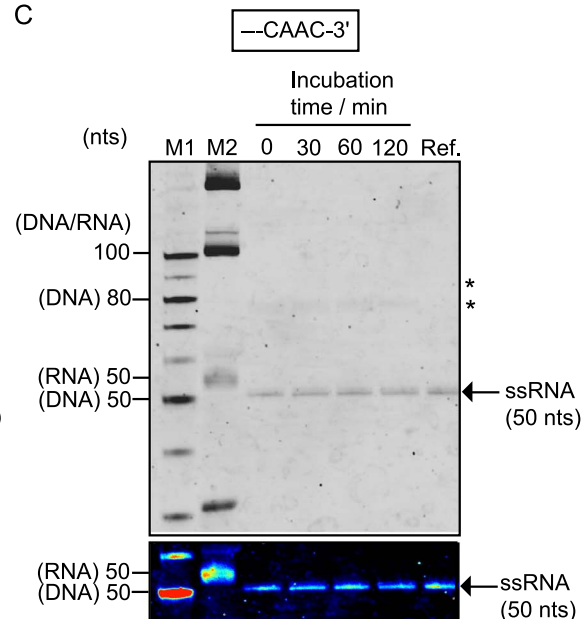
A

GGG-CCC :   5' GGGCAACAACAACAACUUAAUGUCUCUCUGAACUGUGAAAUCUUAUUCCC 3'
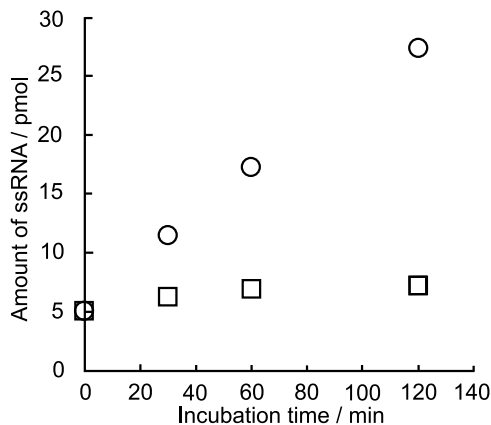GGG-U6CAAC :   5' GGGCAACAACAACAACUUAAUGUCUCUCUGAACUGUGAAAUUUUUUCAAC 3'

Figure 5 Comparison of RNA amplification efficiency between GGG-CCC and GGG-U6CAAC RNA. (A) Sequences of GGG-CCC and GGG-U6CAAC RNA. (B and C) GGG-CCC or GGG-U6CAAC RNA was incubated with NV3D^{pol} at 30°C and sampled at 0, 30, 60 and 120 min followed by adding 57 mM EDTA to stop the reaction, respectively. Each reaction aliquots were subjected to denaturing PAGE (10%) containing 8 M urea followed by SYBRgreenII staining ((B) for GGG-CCC (---CCC-3′) RNA and (C) for GGG-U6CAAC (---CAAC-3′) RNA). Just as in the caption of Figure 2, the color-coded images of the bands of 50 nts ssRNA were shown at the bottom. M1; DNA 10 bp step ladder. M2; dsRNA ladder marker. (D) Quantification of the amount of RNA strands from the band intensity of the denaturing PAGE (B and C) as a function of incubation time. Asterisks are exogenous RNAs from cell free protein synthesis system (e.g. tRNAs). Quantification of each band intensity was performed by a calibration curve of each RNA on the denaturing PAGE. Open circle; GGG-CCC RNA, Open square; GGG-U6CAAC RNA.

In the case of GGG-CCC RNA, the synthesis of dsRNA from ssRNA had high efficiency (Fig. 5B and D), but the amplification efficiency of dsRNA was low, as judged from Figure 6, where 50 bp dsRNA from GGG-CCC RNA was washed out gradually.

In the case of GGG-U6CAAC, the synthesis of dsRNA from ssRNA was slight (Fig. 5C and D), but the amplification efficiency of dsRNA was higher than GGG-CCC RNA, as judged from Figure 6, where GGG-U6CAAC RNA was not washed out in the serial transfer (Fig. 6B).

Discussion

Removal of hairpin species

We designed the random library of RNA template as --UUUUUUNNNN-3′ in order to minimize the interference
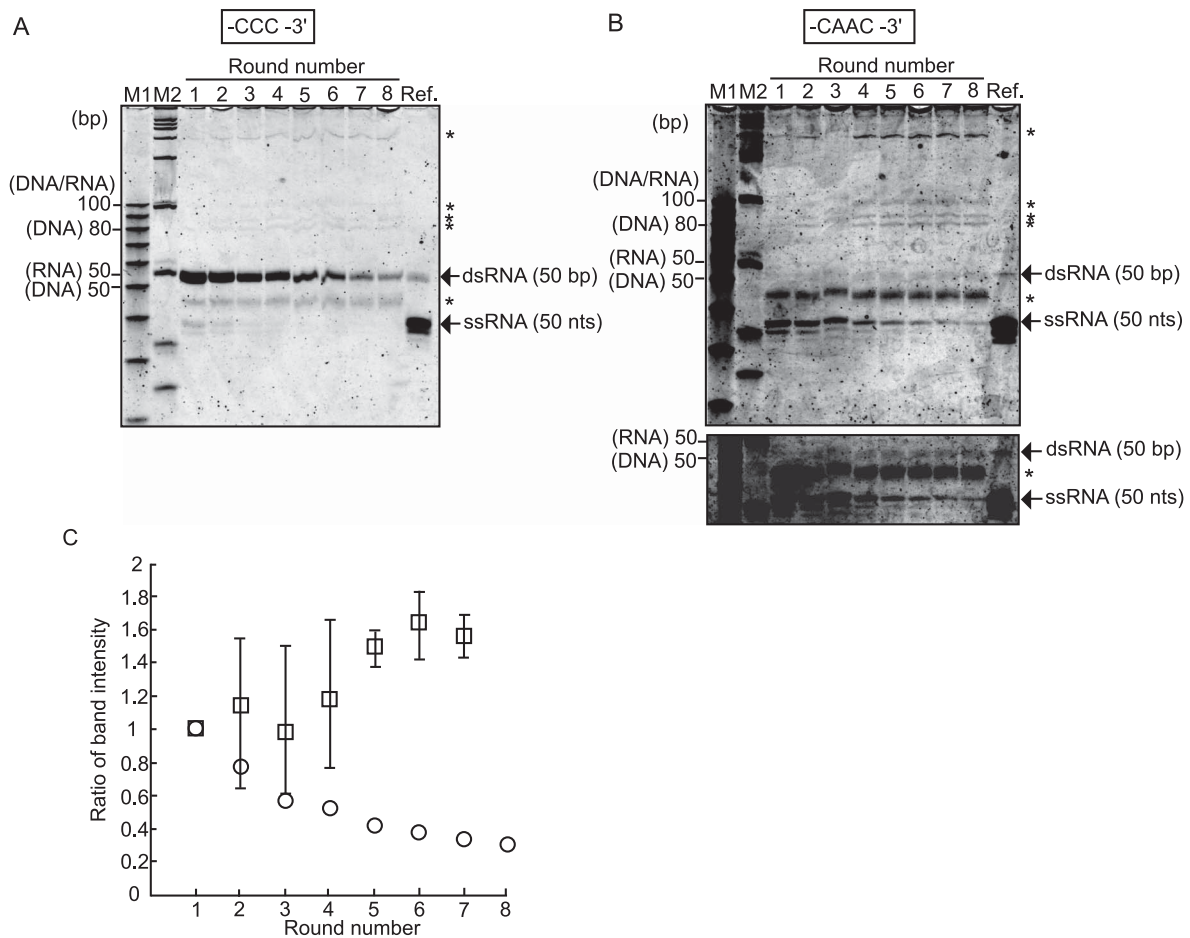
Figure 6 Non-denaturing PAGE (10%) analysis of the serial transfer (1.4-fold dilution and 120 min incubation) using GGG-CCC RNA (A) or GGG-U6CAAC RNA (B). At the bottom of (B), the same but intensified image around the asterisk region is shown in order to clarify the dsRNA band. We made an exception in the incubation time of 240 min for the final round. M1; DNA 10 bp step ladder. M2; dsRNA ladder marker. Ref.; GGG-CCC or GGG-U6CAAC RNA. Asterisks are exogenous RNAs from cell free protein synthesis system (e.g. tRNAs). (C) Quantification of 50 bp dsRNA band as a function of round number. Ordinate is the ratio of band intensity of round r against that of round 1. Open circle; GGG-CCC RNA, Open square; GGG-U6CAAC RNA.

of hairpin synthesis. When ---NNNN-3′ is ---NNAA-3′, a small stem-loop structure, which is comparatively unstable, may be formed and these terminal sequences may be removed from the library searching for the optimal terminal sequence (Supplementary Fig. 1). In fact, the 100 nts band in Figure 2B and the 50 bp band in Figure 2A indicated the formation of such a hairpin, because there was no 100 bp band in Figure 2A. The hairpin was hardly replicated and washed out in the serial transfer, as indicated by the sample from round 10 and round 17 (Fig. 3). On the other hand, the initial ---NNAA-3′ sequences were observed in selected sequences (clone #1, #2, #3, #8 and #19 in Table 1). It means the small stem-loop formed in the ---NNAA-3′ was in fact fairly unstable and some fraction of the ---NNAA-3′ sequences served as a template for the primer-independent replication.

According to the RNA library design, U6N4Yadapter was ligated with 3′-terminus of dsRNA, and not with hairpin

dsRNA in the sequencing process using Y-ligation method[15]. Thus we could sequence only selected dsRNAs with preferable 3′-terminal sequence for NV3D[pol].

Other selection biases

As shown in Table 1, the addition of nucleotides by TNT activity of NV3D[pol] was extensively observed. Which is the selected sequence, original quadruplet site or final 3′-terminal sequence? At least, the addition of 3′-terminal nucleotide at the final round of transfer must be omitted. Any sequence of a clone, number of which is one (clone #15–#19, in Table 1), is a candidate of this disregard. A sequence of a clone, number of which is two or more is not such a candidate, but it must be a preferably amplified sequence, considering the sampling probability. Therefore we regarded the final 3′-terminal sequence of the majority of sequence in Table 1 as the selected sequence by *in vitro* selection using the serial transfer.

In the initial random library the composition of U was about three-times smaller than other nucleotides. The molecular diversity, however, is only 256 and the number of RNA molecule in the first round test tube is more than 10^{12}. Thus deterministic behavior was expected for all molecular species having U bases.

As performed in the preceding study[10], we also applied Y-ligation method[15] to sequence the 3′-terminal random region of selected RNAs, which could not be sequenced by a conventional sequencing method using an RT primer. In Y-ligation method, 3′-hydroxyl group (acceptor) of the selected RNAs and 5′-phosphate (donor) of U6N4Yadapter were ligated with T4 RNA ligase (TaKaRa, Shiga, Japan). It was reported that the presence of U bases in the acceptor (that is, 3′-terminal region of size of five bases) reduces the ligation efficiency by T4 RNA ligase[16]. Thus, the sequence ---UUNNNN-3′ of the initial library might give an overestimated TNT activity. In spite of the reported bias in Y-ligation, about 35% of selected clones had U base at 3′-terminal region. As RNA ligation in the reference 16 is a normal one, the reported result might not be applied to Y-ligation. But if it could be applied, the sequences having U bases might be underestimated.

Starting reaction

As mentioned above, the initial ---UUUNNAA-3′ sequences were observed in the selected sequences (clone #1, #2, #3, #8 and #19 in Table 1). The fraction is 33% of all clones sequenced. If we include the initial ---UUUNNGA-3′ (clone #6 and #13) considering GU base pair, the fraction is 45%. This large value suggested that ---UUUNNAA-3′ was selected at some stage in this *in vitro* selection process. The stage must be the starting reaction, that is, the binding reaction of NV3D^{pol} to ssRNA template in the initial library, followed by the initiation of polymerization. As mentioned in Introduction, long hairpin synthesis starting from a small stem-loop structure at 3′-terminus is the strongest activity of NV3D^{pol}. Therefore NV3D^{pol} should bind rapidly to a 3′-terminal small stem-loop, which was made of the -----UUUNNAA-3′ in this study. The -----UUUNNAA-3′, however, folds to a weak small stem-loop, and easily unfolds to make free 3′-terminus. Thus the next step by the bound enzyme goes to one of two ways: the hairpin synthesis pathway or the pathway of primer-free replication on ssRNA template. It is strongly suggested that the latter pathway was realized in the clone #1, #2, #3, #8 and #19. For the clone #1, #2, #3 and #8, the TNT activity of NV3D^{pol} added additional nucleotide(s) in the following transfers and, as a result, the fitness for serial transfer increased.The initial ---UUUNNCA-3′ sequences were also observed in the selected sequences in large fraction (31%).

The reason of this phenomenon is open question.

Two modes in the replication of dsRNA:
the shuttle mode and the bimolecular reaction mode

The population landscape on the 4-dimensional sequence space of the 17th round of transfer gave a main peak around ---CAAC-3′ (Fig. 4). In the preceding studies on the batch amplification reaction starting from a ssRNA, a template with 3′-terminal C-stretch (---CCC-3′ or ---CCCC-3′) was amplified effectively. Which sequence is the optimal, ---CCC-3′ or ---CAAC-3′? This apparent contradiction was resolved as follows. It was confirmed that in the batch amplification the ---CCC-3′ was much more effective than the ---CAAC-3′ (Fig. 5), but in the serial transfer condition in which the ----CAAC-3′ was sustained (or grown gradually) the ---CCC-3′ was washed out (Fig. 6).

In fact as reported in Introduction, for a template of ssRNA of 50 nts having terminal sequence 5′-GGG------CCC-3′, the apparent doubling time in the initial phase of batch amplification is 12 min. On the other hand, as shown in Figure 6B, the same 5′-GGG------CCC-3′ was washed out at 1.4-fold dilution and 120 min incubation.

Based on these results we propose the existence of the "shuttle mode" replication of dsRNA as one of the possibilities. "Shuttle mode" replication means the successive strand dissociation replication of dsRNA without dissociation of the enzyme from plus/minus template set, explained as follows. At the end of (–)strand polymerization on the (+)strand template, the enzyme translocates from 5′-terminus of (+)strand template to 3′-terminus of fresh (–)strand and begin to polymerize the next generation (+)strand, dissociating the parent (+)strand. And the end of this time polymerization, the similar event occurs to begin polymerization of the next generation (–)strand, and so on. The enzyme translocation is not fully successful, and all the enzymes dissociate from the plus/minus template sets after many (say 10) shuttling cycle. The phenomena that in the batch replication reaction a growth curve became nearly flat after 240 min [Fig. 3 in Ref. 10] showed probably the life-time of shuttle mode replication. The cause of this lying flat may not be inactivation of the enzyme nor consumption of NTPs. If this explanation is correct, the probability of successful translocation is estimated to be 80%.

The selected sequences in our serial transfer have passed two criteria; the primer-independent replication of ssRNA (for round 1) and the bi-molecular reaction mode replication of dsRNA (for round 2–17). On the other hand, the C-stretch found in the batch amplification, in which the amplification proceeds based on the primer-independent replication plus shuttle mode, seems to be not favorable for the bi-molecular reaction mode replication of dsRNA, because the terminal sequence having C-stretch cannot breathe easily in comparison with the sequence including weak AU base pair(s). Breathing of the terminal region of dsRNA allows NV3D^{pol} to bind to the 3′-terminal region and make the initiation complex for the next replication.

Conclusion

NV3D[pol] was successfully used for the natural selection type (that is, the fitness = the specific growth rate, as realized in the serial transfer) *in vitro* selection experiment. Analysis of selected sequences suggested a novel RNA replication mode, that is, the shuttle replication mode. A tentatively recommended terminal sequences for the *in vitro* evolution with NV3D[pol] is 5′-GGGCAAAAA------------------UUUUACAAC-3′ or 5′-GGGCAAAAA ------------------UUUUACAA-3′. The central region indicated by "--------" is able to have more than 500 nts size encoding a protein of more than 150 a.a. We hope even NV3D[pol] itself (510 a.a.) will be coded in this region.

Acknowledgement

References

1. Morozov, I. Y., Ugarov, V. I., Chetverin, A. B. & Spirin, A. S. Synergism in replication and translation of messenger RNA in a cell-free system. *Proc. Natl. Acad. Sci. USA* **90**, 9325–9329 (1993).

2. Bansho, Y., Ichihashi, N., Kazuta, Y., Matsuura, T., Suzuki, H. & Yomo, T. Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. *Chem. Biol.* **19**, 478–487 (2012).

3. Usui, K., Ichihashi, N., Kazuta, Y., Matsuura, T. & Yomo, T. Kinetic model of double-stranded RNA formation during long RNA replication by Qβ replicase. *FEBS Lett.* **587**, 2565–2571 (2013).

4. Blumenthal, T. & Carmichael, G. G. RNA replication: Function and structure of Qβ-replicase. *Annu. Rev. Biochem.* **48**, 525–548 (1979).

5. Brown, D. & Gold, L. RNA replication by Qβ replicase: A working model. *Proc. Natl. Acad. Sci. USA* **93**, 11558–11562 (1996).

6. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. In vitro virus: Bonding of mRNA bearing puromycin at the 3′-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* **414**, 405–408 (1997).

7. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* **94**, 12297–12302 (1997).

8. Tabuchi, I., Soramoto, S., Nemoto, N. & Husimi, Y. An *in vitro* DNA virus for *in vitro* protein evolution. *FEBS Lett.* **508**, 309–312 (2001).

9. Ueno, S., Arai, H., Suzuki, M. & Husimi, Y. An mRNA-protein fusion at N-terminus for evolutionary protein engineering. *Int. J. Biol. Sci.* **3**, 365–374 (2007).

10. Arai, H., Nishigaki, K., Nemoto, N., Suzuki, M. & Husimi, Y. Characterization of norovirus RNA replicase for *in vitro* amplification of RNA. *BMC Biotechnol.* **13**, 85 (2013).

11. Fukushi, S., Kojima, S., Takai, R., Hishino, B. F., Oka, T., Takeda, N., Katayama, K. & Kageyama, T. Poly(A)- and primer-independent RNA polymerase of *norovirus*. *J. Virol.* **78**, 3889–3896 (2004).

12. Belliot, G., Sosnovtsev, S. V., Chang, K., Babu, V., Uche, U., Arnold, J. J., Cameron, C. E. & Green, K. Y. Norovirus proteinase-polymerase and polymerase are both active forms of RNA-dependent RNA polymerase. *J. Virol.* **79**, 2393–2403 (2005).

13. Rohayem, J., Robel, I., Jäger, K., Scheffler, U. & Rudolph, W. Protein-primed and de novo initiation of RNA synthesis by norovirus 3D[pol]. *J. Virol.* **80**, 7060–7069 (2006).

14. Rohayem, J., Jäger, K., Robel, I., Scheffler, U., Temme, A. & Rudolph, W. Characterization of norovirus 3D[pol] RNA-dependent RNA polymerase activity and initiation of RNA synthesis. *J. Gen. Virol.* **87**, 2621–2630 (2006).

15. Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T. & Husimi, Y. Y-ligation: An efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. Mol. Divers. **4**, 187–190 (1998).

16. England, T. E. & Uhlenbeck, O. C. Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry* **17**, 2069–2076 (1978).