Design of Artificial Short-Chained RNA Species That Are Replicated by Qβ Replicase†

Humberto Zamora,‡ Rüdiger Luce, and Christof K. Biebricher*

Max-Planck-Institute for Biophysical Chemistry, Am Fassberg, D-3400 Göttingen, Federal Republic of Germany

Received August 9, 1994; Revised Manuscript Received November 16, 1994®

ABSTRACT: Different RNA species that are replicated by Qβ replicase have related secondary structures: for both plus and minus strands, "leader" stem structures were found at their 5′ termini, while their 3′ termini were unpaired. Parallel structures in complementary strands rather than antiparallel ones require the occurrence of wobble pairs and other imperfections in the stem regions. To test whether the leader structures are required for replication, artificial RNA sequences were synthesized by transcription from synthetic oligodeoxynucleotides with T7 RNA polymerase and assayed for their ability to be replicated by Qβ replicase. A synthetic short RNA species known to be replicated was amplified, forming a stable quasi-species; i.e., its sequence was conserved during hundreds of replication rounds. A synthetic mutant of this sequence that stabilized the leader in one strand but favored a 3′-terminal stem in the other one led to the complete loss of template activity. When new RNA sequences with the described structural requirements were designed and synthesized, their template activity was too low to be directly measurable; however, incubation with replicase produced replicating RNA whose sequence was closely related to the synthesized RNA species. The most likely interpretation is that the designed sequences were in a low montainous region in the replication fitness landscape and were optimized during amplification by Qβ replicase to a nearby fitness peak. The structural features postulated to be required for replication were not only conserved but even improved in the outgrowing mutants. The presented results corroborate the postulated requirement, but indicate that additional, not yet identified features are required to fine-tune the replication efficiency.

Qβ replicase is highly specific in the choice of its templates: *in vivo*, it amplifies its natural template, the Qβ RNA and Qβ minus strands, while ignoring the vast excess of other RNA in the cell (Haruna & Spiegelman, 1965). From Qβ-infected cells late in the infection cycle, a heterogeneous replicating short-chained RNA fraction with an average sedimentation coefficient of 6 S can be isolated (Banerjee et al., 1967). *In vitro*, in template-free reactions, a large variety of short-chained RNA species have been isolated that are excellent templates of Qβ replicase (Mills et al., 1973, 1975; Schaffner et al., 1977; Biebricher & Luce, 1993). Comparing the many primary sequences known reveals little homology: A necessary, but not sufficient condition is a 5′-terminal GGG sequence corresponding to a 3′-terminal CCC sequence. The 5′-terminal GTP cannot be replaced by ITP (Feix & Sano, 1975; Mills & Kramer, 1979) or by [α-S]-GTP (Biebricher and Luce, unpublished results), while replacement of guanylate residues in the interior of the sequence readily takes place. There are two hypotheses for the requirement of the terminal pppG; they do not exclude each other: EFTu present as one subunit of the replicase (Blumenthal et al., 1972) may provide one GTP binding site for the replication initiation, which requires geminal association of two GTP molecules (Biebricher, 1983). The second is the particularly high stacking energy of two G bases.

The recognition of RNA by protein, however, is not governed by the primary structure alone, because RNA folds into defined tertiary structures (Westhof & Michel, 1992). Complex formation between the two polymer chains involves specific surface contacts (Giegé et al., 1993). While tertiary structures cannot yet be determined, minimal energy secondary folding algorithms for RNA are available (Zuker & Stiegler, 1981; Turner et al., 1988). Newer algorithms give a choice among several foldings (Zuker, 1989; McCaskill, 1990). When comparing the calculated secondary structures of short-chained RNA species replicated by Qβ replicase, we found a common structural feature: the 5′ terminus is always forming a leader stem while the 3′ end is unstructured (Biebricher & Luce, 1993). The constraints for this feature are rather high. After a replication round, a complementary replica is produced, and a cross-catalytic amplification results only if this replica is also able to serve as a template. In this paper, we show that mutational changes affecting this structure destroy the ability to replicate with Qβ replicase and that artificially devised RNA species fulfilling the postulated requirements can replicate, even though sometimes with low efficiency.

MATERIALS AND METHODS

Materials. Materials and the techniques of cloning and sequencing of replicating RNA species have been described previously (Biebricher & Luce, 1993). Because many of the templates described in this work replicate with very low efficiency, the use of Qβ replicase lacking RNA impurities was crucial for obtaining reliable results.

† This work has been supported in part by a grant from the Deutsche Forschungsgemeinschaft.

* To whom correspondence should be addressed.

‡ Present address: Universidad Industrial de Santander, Ciudad Universitaria, Bucaramanga, Colombia.

® Abstract published in *Advance ACS Abstracts,* January 1, 1995.

0006-2960/95/0434-1261$09.00/0

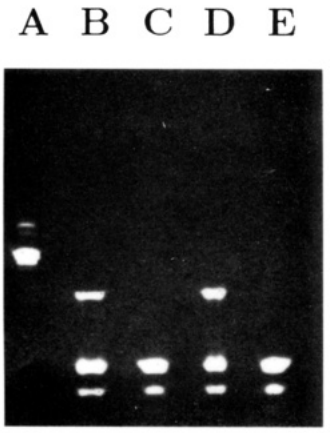

FIGURE 1: Stable and metastable folding of SV-11 RNA after heating and cooling. SV-11 plus strand was prepared by *in vitro* transcription with T7 RNA polymerase from a plasmid harboring SV-11. Gel electrophoresis separated four bands, corresponding (with increasing electrophoretic mobility) to an imperfect double strand (due to the nearly palindromic sequence of SV-11), the metastable active form, the stable hairpin structure (Biebricher & Luce, 1992), and a shorter abortive transcription product. Upon heating, all bands except the last one were converted into the stable structure. The double-stranded band was isolated by electrophoretic elution and sealed into polypropylene capillaries in water (A, B, C) or 40% aqueous formamide (D, E). Capillaries were heated for 60 s to 115 °C and quenched in dry ice (B, D) or normally cooled (C, E). Rapid cooling regenerated the metastable form, because close C_n:G_n stems form more rapidly.

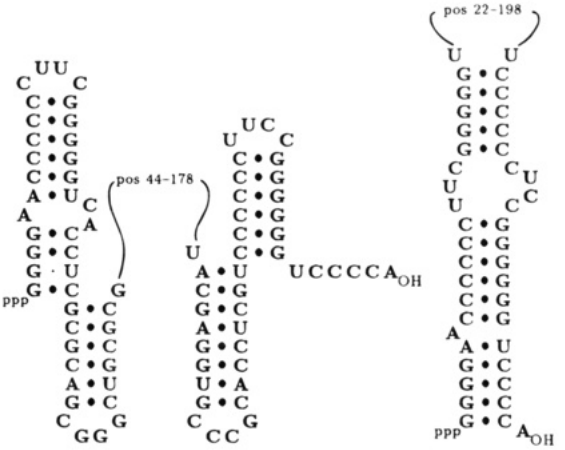

FIGURE 2: Tentative structures of the MDV-1 minus strand. On the right side, the structure favored by the Zuker algorithm is shown; on the left side, a form is depicted that fits the proposed structural pattern necessary for replication.

RESULTS

Secondary Structures of Optimized RNA Species. It was shown previously (Biebricher & Luce, 1993) that short-chained RNA species replicated by Qβ replicase share secondary structure features: in both complementary strands, the 5′ ends are engaged in a stem structure, while the 3′ ends are free. Among the many replicating RNA species with larger chain lengths (Mills et al., 1973, 1975; Schaffner et al., 1977; Priano et al., 1987; Biebricher, 1987; Munishkin et al., 1987; Biebricher & Luce, 1992, 1993), only two were found whose calculated minimum energy structures (Zuker & Stiegler, 1981; Turner et al., 1988) did not fit this pattern: for SV-11 (Biebricher & Luce, 1992) and MDV-1 (Priano et al., 1987), hairpin structures were calculated where the two termini base-paired one to the other. However, minimum energy calculations do not necessarily give the structures that are active in replication. Kinetic elements also determine the structure (Baker & Agard, 1994): Folding begins already during stepwise elongation of the strand during replication, and metastable structures may be produced. This has been shown to occur with SV-11: The most stable structure is a hairpin unable to replicate, while a metastable structure formed during replication has been shown to be the active template (Biebricher et al., 1982); the replicating metastable structure fits the postulated pattern well (Biebricher & Luce, 1992). The metastable structure is quantitatively converted to the stable structure by heating and slow cooling, but can be restored by melting and rapid cooling of SV-11, because G_n clusters close to C_n clusters form double-stranded stems more rapidly (Figure 1).

The single strands of MDV-1 formed in replication are not metastable and retain their template activities and electrophoretic mobilities after heating and slow or rapid cooling. Their calculated hairpin structures have energies similar to the short-stem foldings (Figure 2). Under those conditions, the most probable structure of MDV-1 is the one shown on the left of Figure 2. Previous studies of the secondary structure (Kramer & Mills, 1981) did not decide between the two structures, and the determination of the correct structure of MDV-1 requires further experimental studies.

It seems advisable that the calculation of secondary structures should (i) show several alternative structures and (ii) give some hints regarding possible interconversions of structures. For short sequences, the method of McCaskill (1990) was found to be particularly suitable for estimating the most probable structures (Turner et al., 1988).

Synthesis of RNA by Transcription. The discovery of very short replicating species with chain lengths of 30−40 offers a simple synthetic approach for testing the structural hypothesis. RNA synthesis by transcription from synthetic oligonucleotides is a standard technique (Milligan & Uhlenbeck, 1989). It uses as a template for transcription an oligonucleotide complementary to the target sequence bearing the complement of the T7 promoter at its 3′ end; the T7 promoter sequence is annealed to this oligonucleotide, and transcription should produce the target sequence. With our RNA sequences, transcription from these artificial templates produced on average only 20−50 transcript strands per template strand, and side products were also observed (Figure 3). Completing the partially single-stranded template to the perfect double helix did not improve the transcription efficiency, and even runoff transcription of the RNA from the DNA cloned into a plasmid was much less efficient and specific than that observed with longer templates like MNV-11.

To test whether the transcription method gives reliable results, we synthesized a short-chained RNA sequence, DN3, found previously among the products of template-free RNA synthesis products of Qβ replicase (Figure 4). *In vitro* transcription with T7 RNA polymerase (Milligan & Uhlenbeck, 1989) produced RNA that was indeed readily amplified by Qβ replicase; the products showed the expected electrophoresis pattern (Figure 5). After amplification of the RNA by at least a factor of 10^6 to outdilute the input RNA, the sequence of the RNA population was analyzed and found to be identical to the plus and minus strands of the synthesized RNA species, accompanied by a few mutants

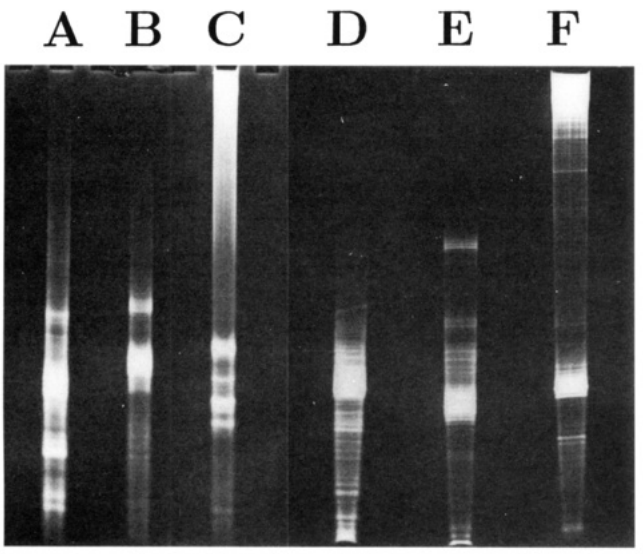

FIGURE 3: Electrophoretogram of transcription products from short-chained cDNA sequences corresponding to DN3. (A, B) Transcription from a sequence containing single-stranded synthetic DN3 cDNA followed by the reverse T7 promoter, annealed to the T7 promoter sequence DNA, according to Milligan and Uhlenbeck. (C, D) Transcription of a fully double-stranded synthetic DNA of T7 promoter followed by the DN3 strand. (E, F) Runoff transcription from a plasmid containing the DN3 sequence behind a T7 promoter, cleaved at the 3′ terminus of the DN3 sequence. Samples D, E, and F were heated in 50% formamide at 100 °C prior to loading. The yields were 20−50 strands per template strand. In comparison to the MNV-11 sequence which gave a sharp band with minor bands of $n + 1$ and $n + 2$ chain lengths, short-chained cDNA transcribed with lower yields and more side products.

differing in a few positions. This suggests that the species occupies already a local peak of the selection value landscape in the sequence space (Eigen & Biebricher, 1988); otherwise, the sequence should have been observed to drift. This is plausible, because optimization had already occurred during the isolation of the starting RNA species.

The species remained evolutionarily stable for many serial transfers even though its overall replication rate (in the linear growth phase) was only 1.2×10^{-3}, less than one-fifth of the rate of MNV-11, SV-11, or other RNA species described to be found as impurities in some Qβ replicase preparations (Chetverin et al., 1992). Therefore, during a hundredfold exponential amplification of DN3, such an impurity if present would have been amplified by a factor of $>10^{10}$, sufficient to amplify the impurity to macroscopic appearance.

Site-Directed Mutagenesis Altering the Secondary Structure. The synthesis was repeated with a mutant RNA species where the 5′-terminal stem region of the resulting RNA was stabilized by four base exchanges (Figure 6). The corresponding complementary strand formed in replication now has two possible structures: a stable one containing a corresponding 3′ stem region and a much less stable alternative where the presumably required 5′ end is present. Formation of the more stable structure should thus abolish the ability to replicate. This was indeed the case: an appropriate dilution of the transcripts gave no synthesis after 4 h of incorporation. After incorporation for 40 h, RNA synthesis was eventually observed, but cloning and sequencing of the products revealed no sequence relationship to the ancestor used in the reaction. The conclusion is that the mutated sequence shown in Figure 6 is no longer a template for Qβ replicase due to its altered structure.

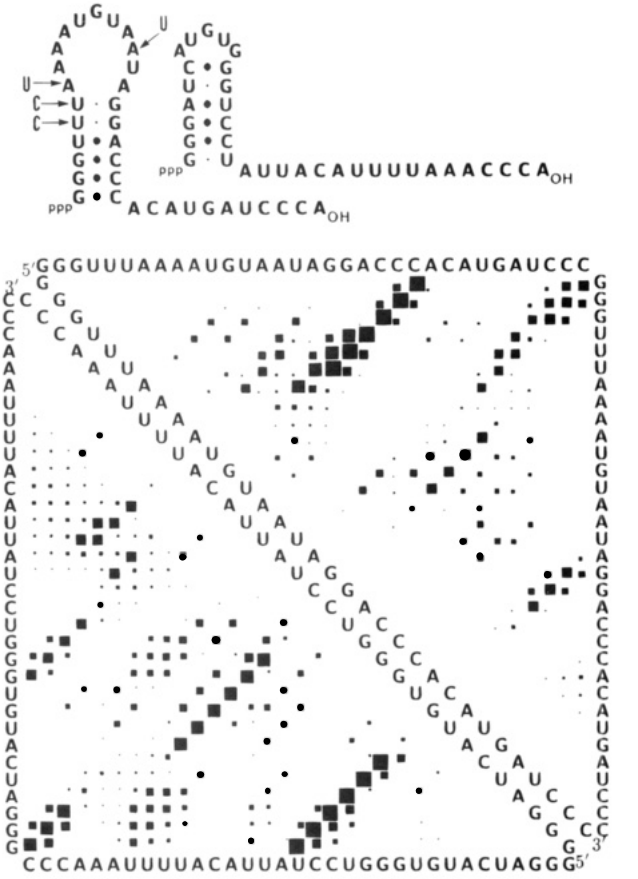

FIGURE 4: Structure of DN3 RNA. Calculated minimum energy structures (top) and the probabilities of intramolecular base-pairing (bottom) are shown for both plus and minus strands. Note the characteristics of a replicating structure: (i) a preference for a parallel structure where the 5′ termini of both plus and minus strands are involved in base-pairing and (ii) several structural alternatives with similar energies. The mutations leading to the sequence shown in Figure 6 are also shown.

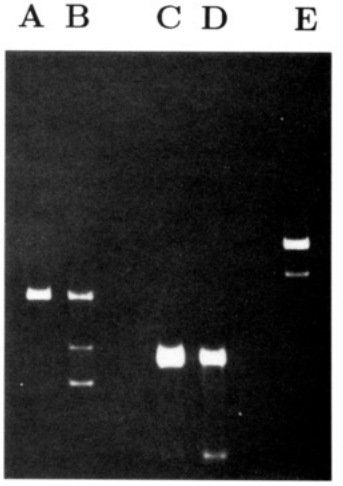

FIGURE 5: Electrophoretogram of synthetic RNA amplified by Qβ replicase. (A, B) SN5042 RNA, sequence shown in Figure 7; (C, D) DN3 RNA, sequence shown in Figure 5; (E) MNV-11 RNA as reference. Samples B and D were heated in 50% formamide at 100 °C prior to loading.

Artificial Templates. The second approach was designing RNA molecules with the postulated structure requirements and determining their ability to replicate. If they do replicate—no matter how slowly—their replication efficiency will increase during amplification, because it is unlikely to hit a local maximum in the "fitness landscape" by design. It

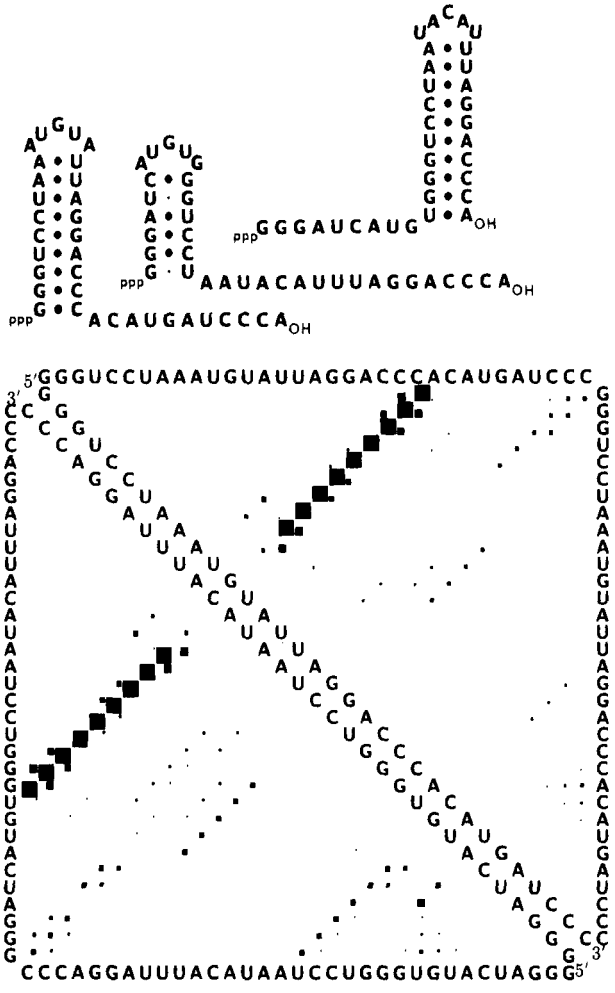

FIGURE 6: Structure of a synthetic mutant of DN3 RNA. The structure has a strong preference for the antiparallel structure and is incapable of replication.

suffices to find a sequence in a mountainous region in the fitness landscape to be led to a local fitness optimum by evolution. On the other hand, if the RNA molecule is totally unable to reproduce, it will have no opportunity to improve by mutation, as is illustrated by the experiment shown in Figure 6 where no reversion to the replicating molecule was found. A comparison of the optimized product and its precursor provides a hint about what may be important for replication.

Two sequences in the range of 35−45 having the postulated structural properties were designed by trial and error, and their overall replication rates were measured, using amounts of RNA transcript stoichiometric to the replicase. The replication rates were too low for direct detection in the standard assay ($<10^{-4}$ s^{-1}); however, when dilutions of the RNA were incubated with Qβ replicase, growth was observed within 2 h. The electrophoretograms showed several bands corresponding to different sequence lengths. The obtained RNA populations were sampled by cloning and sequencing. The sequences of the resulting RNA species were found to be closely related but not identical to those of their synthesized ancestors. Some calculated structures of the synthesized RNA species and their selected RNA mutants are shown in Figures 7 and 8. The mutational changes are highly instructive: the observed mutations improved the stability of the 5′-terminal stems. One of the mutations shown in Figure 7 had no apparent influence on

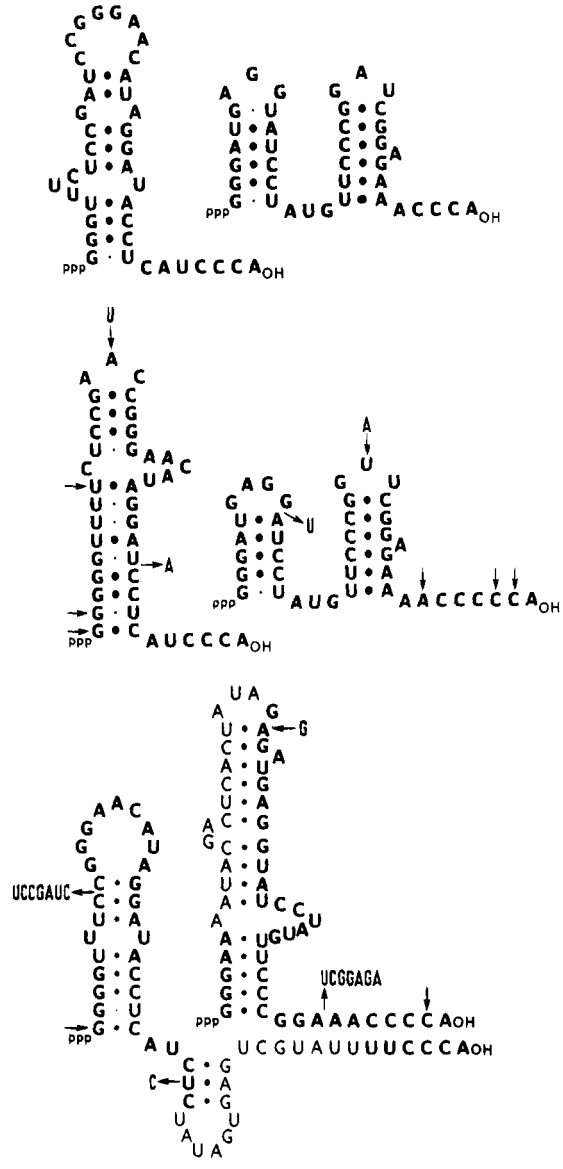

FIGURE 7: Tentative secondary structures of a designed artificial replicating RNA (top) and two of its amplification products by Qβ replicase (bottom). The mutations are indicated. The bottom sequence contains the T7 promoter sequence (indicated by thin symbols).

the stem structure but contributed to the stability by formation of the tetraloop UUCG (Tuerk et al., 1988) found often in replicating RNA species (Biebricher & Luce, 1993).

One of the replicating derivatives of the designed RNA contained the T7 promoter sequence (Figure 7, bottom). It cannot be ruled out but seems unlikely that the synthesized oligodeoxynucleotide was directly used to form this molecule, because some full-length RNA (sequence length plus promoter length) was found among the transcripts. The RNA species containing the T7 promoter sequence is readily amplified by Qβ replicase (Figure 5). Another replicating derivative contained a sequence duplication due to recombination processes during amplification (Figure 8; Biebricher & Luce, 1992).

DISCUSSION

How does an RNA template instruct the replicase to replicate it? Obviously it is not the primary structure that is responsible for the instruction, because some RNA species

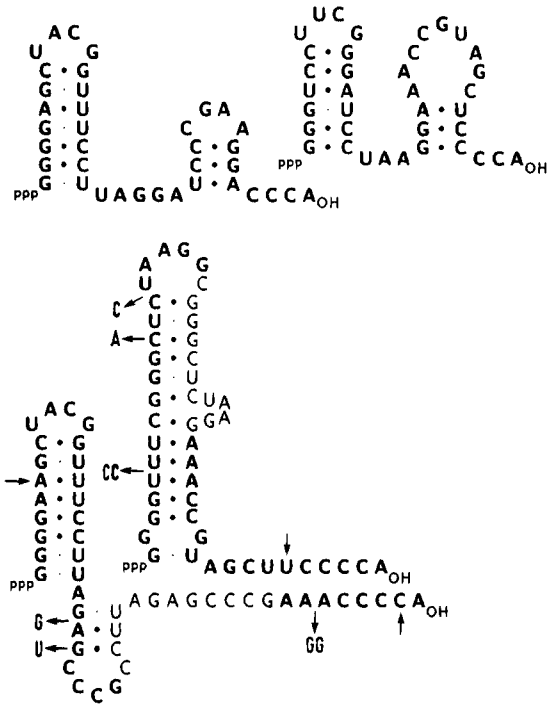

FIGURE 8: Tentative secondary structures of another deviced replicating RNA (top) and one of its amplification products (bottom). The bottom sequence contains a duplicated piece (indicated by thin symbols).

can exist in more than one conformation only one of which is an active template (Biebricher et al., 1982; Biebricher & Luce, 1992). Measurements of replicase binding to its templates, in particular to Qβ RNA itself (Vollenweider et al., 1976; Meyer et al., 1981; Barrera et al., 1993), to mRNA, rRNA, and tRNA (Silverman, 1973), or to short-chained replicating RNA (Mills et al., 1977; Werner, 1991; Biebricher & Luce, unpublished) have shown that binding is a necessary but not sufficient condition for replication. While several nonreplicating RNAs, e.g., rRNA, bind quite efficiently, some templates that do replicate bind only weakly (Luce & Biebricher, 1993), and their replication can be inhibited by addition of salt (Prives & Silverman, 1972; Biebricher, 1983). The invariant 5′- and 3′-terminal sequences are also important for correct initiation of replication. While removal of the terminal A (Kamen, 1969) or addition of up to four A residues to the 3′-terminus (Biebricher & Luce, 1993) does not affect RNA replication by Qβ replicase, removal or deamination of one or more C residues at the 3′-terminus or addition of a long A-tail destroys the template activity (Kamen, 1969; Rensing & August, 1969; Weber & Weissmann, 1970; Gilvarg et al., 1975; Mills et al., 1980). On the part of the enzyme, the viral subunit seems mainly responsible for recognition and catalysis (Hartmann et al., 1988; Lindner et al., 1991).

The structural consensus between replicating RNA species described in this paper is corroborated by the design of synthetic templates. The 5′ leader stem found to be required for replication has been detected also in eukaryotic mRNA sequences (Konings et al., 1987), possibly for similar reasons. During RNA synthesis, both by transcription and by RNA replication, the newly synthesized replica strand has to be peeled off of the template, and the 5′ leader sequence might reduce the energy required for strand separation. Furthermore, replicating RNA has been reported to contain much intramolecular base-pairing to prevent template loss by double-strand formation (Biebricher et al., 1982, 1984; Axelrod et al., 1991). Often, alternative structures with smaller stabilities are possible where the ends base-pair one to the other (Voronin, 1992). One might speculate that a dynamic switch from one to the other structure may occur during replication. However, we have sequences of efficiently replicating RNA species whose termini cannot base-pair one to the other because they have only two G residues at their 5′ termini. Küppers and Sumper (1975) described another requirement for template activity: the presence of two C-clusters in a specific geometry. As a matter of fact, the requirement of a leader stem structure, not base-paired to the 3′ terminus, necessarily demands a second C—or at least a pyrimidine—cluster. The presence of a second C-cluster cannot be responsible alone for template activity, because the shown mutation that destroyed the template activity of DN3 (Figures 4 and 6) left the C-clusters intact.

We have presented evidence that a 5′ leader stem structure is required for replication. Yet, despite the presence of the postulated requirements in the designed molecules, they were not replicated efficiently by Qβ replicase. The replication efficiency was improved dramatically by a few mutations. Even though the required structures were conserved and improved by the mutations, it seems that other, not yet identified features help to improve the replication efficiency.

ACKNOWLEDGMENT

We are indebted to J. S. McCaskill for giving us his partition function computer program.

REFERENCES

Axelrod, V. D., Brown, E., Priano, C., & Mills, D. R. (1991) *Virology 184*, 595−608.
Baker, D., & Agard, D. A. (1994) *Biochemistry 33*, 7505−7509.
Banerjee, A. K., Eoyang, L., Hori, K., & August, J. T. (1967) *J. Mol. Biol. 45*, 181−193.
Barrera, I., Schuppli, D., Sogo, J. M., & Weber, H. (1993) *J. Mol. Biol. 232*, 512−521.
Biebricher, C. K. (1983) *Evol. Biol. 16*, 1−52.
Biebricher, C. K. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 299−306.
Biebricher, C. K., & Luce, R. (1992) *EMBO J. 11*, 5129−5135.
Biebricher, C. K., & Luce, R. (1993) *Biochemistry 32*, 4848−4854.
Biebricher, C. K., Diekmann, S., & Luce, R. (1982) *J. Mol. Biol. 154*, 629−648.
Biebricher, C. K., Eigen, M., & Gardiner, W. C. (1984) *Biochemistry 23*, 3186−3194.
Blumenthal, T., Landers, T. A., & Weber, K. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1313−1317.
Chetverin, A. B., Chetverina, H. V., & Munishkin, A. V. (1991) *J. Mol. Biol. 222*, 3−9.
Eigen, M., & Biebricher, C. K. (1988) in *RNA Genetics, Volume III: Variability of RNA Genomes* (Domingo, E., Ahlquist, P., & Holland, J. J., Eds.) pp 211−245, CRC Press, Boca Raton, FL.
Feix, G., & Sano, H. (1973) *Hoppe-Seyler's Z. Physiol. Chem. 354*, 1186−1187.
Giegé, R., Puglisi, J. D., & Florentz, C. (1993) *Prog. Nucleic Acid Res. Mol. Biol. 45*, 129−205.
Gilvarg, C., Bollum, F. J., & Weissmann, C. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 428−432.
Hartmann, G. R., Biebricher, C., Glaser, S. J., Grosse, F., Katzameyer, M. J., Lindner, A. J., Mosig, H., Nasheuer, H.-P., Rothman-Denes, L. B., Schäffner, A. R., Schneider, G. J., Stetter, K.-O., & Thomm, M. (1988) *Biol. Chem. Hoppe-Seyler 369*, 775−788.
Haruna, I., & Spiegelman, S. (1965) *Proc. Natl. Acad. Sci. U.S.A. 54*, 1189−1193.
Kamen, K. (1969) *Nature (London) 221*, 321−325.

Konings, D. A. M., van Duijn, L. P., Voorma, H. O., & Hogeweg, P. (1987) *J. Theor. Biol. 127*, 63−78.
Kramer, F. R., & Mills, D. R. (1981) *Nucleic Acids Res. 9*, 5109−5124.
Küppers, B.-O., & Sumper, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2640−2644.
Lindner, A. J., Glaser, S. J., Biebricher, C. K., & Hartmann, G. R. (1991) *Eur. J. Biochem. 202*, 249−254.
McCaskill, J. S. (1990) *Biopolymers 29*, 1105−1119.
Meyer, F., Weber, H., & Weissmann, C. (1981) *J. Mol. Biol. 153*, 631−660.
Milligan, J. F., & Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51−62.
Mills, D. R., & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2232−2235.
Mills, D. R., Kramer, F. R., & Spiegelman, S. (1973) *Science 180*, 916−927.
Mills, D. R., Kramer, F. R., Dobkin, C., Nishihara, T., & Spiegelman, S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4252−4256.
Mills, D. R., Nishihara, T., Dobkin, C., Kramer, F. R., Cole, P. E., & Spiegelman, S. (1977) in *Nucleic Acid-Protein Recognition* (Vogel, H. J., Ed.) pp 533−547, Academic Press, New York.
Mills, D. R., Kramer, F. R., Dobkin, C., Nishihara, T., & Cole, P. E. (1980) *Biochemistry 19*, 228−236.
Munishkin, A. V., Voronin, L. A., & Chetverin, A. B. (1988) *Nature (London) 333*, 473−475.
Priano, C., Kramer, F. R., & Mills, D. R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 321−330.
Prives, L. C., & Silverman, P. (1972) *J. Mol. Biol. 71*, 657−670.
Rensing, U., & August, J. T. (1969) *Nature (London) 224*, 853−856.
Schaffner, W., Rüegg, K. J., & Weissmann, C. (1977) *J. Mol. Biol. 117*, 877−907.
Silverman, P. M. (1973) *Arch. Biochem. Biophys. 157*, 234−242.
Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Brody, E. N., & Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364−1368.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Chem. 17*, 167−192.
Vollenweider, H. J., Koller, T., Weber, H., & Weissmann, C. (1976) *J. Mol. Biol. 101*, 367−377.
Voronin, L. A. (1992) *Biochimie 74*, 491−494.
Weber, H., & Weissmann, C. (1970) *J. Mol. Biol. 51*, 215−224.
Werner, M. (1991) *Biochemistry 30*, 5832−5838.
Westhof, E., & Michel, F. (1992) in *Structural tools for the analysis of protein−nucleic acid complexes* (Lilley, D. M. J., Heumann, H., & Suck, D., Eds.) pp 255−267, Birkhäusser, Basel.
Zuker, M. (1989) *Science 244*, 48−52.
Zuker, M., & Stiegler, P. (1981) *Nucleic Acids Res. 9*, 133−148.

BI9418277