PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Significance of Strand Configuration in Self-Replicating RNA Molecules

Brian K. Davis

Phil. Trans. R. Soc. Lond. B 1995 **350**, 345-352
doi: 10.1098/rstb.1995.0169

Significance of strand configuration in self-replicating RNA molecules

BRIAN K. DAVIS

Research Foundation of Southern California, 5580 La Jolla Boulevard, La Jolla, California 92037, U.S.A.

SUMMARY

The kinetic theory of replication has been extended to include dual mechanisms for conversion of self-annealed single-strand RNA to double-strand molecules, which do not replicate. An analysis of experimental results established that the replicate–template annealing reaction during transcription significantly retarded replication *in vitro* among three RNA variants copied by Qβ replicase. Annealing between complementary RNA strands free in solution had far less significance. The finding that an RNA variant can be replicated in a multiple hairpin configuration, but not as its single, long hairpin conformer, the correlation between stability of strand secondary structure and replicative fitness, and a lack of homology in the internal sequence of RNA variants copied by Qβ replicase support the conclusion that template competence depends on strand configuration, independent of most of the underlying base sequence. Occurrence of self-annealed strands in the Qβ replicase system was attributed to its reliance on RNA-driven strand separation, in the absence of enzyme catalysed strand unwinding. A 'configuration before sequence' path to self-replication exhibited a substantially lower combinatorial barrier than standard sequence-dependent evolution. RNA-dependent RNA synthesis in the Qβ system thus displays features of an RNA World and, interestingly, they reveal a rapid path for evolution of the first self-replicating molecule on Earth.

1. INTRODUCTION

The replicative form of various small RNA molecules copied by Qβ replicase consists of a single strand self-annealed into a series of hairpin loops (Schaffner *et al.* 1977; Mills *et al.* 1978; Priano *et al.* 1987; Munishkin *et al.* 1988, 1991; Biebricher & Luce 1992; Voronin 1992). A shared generic folding pattern in this heterogeneous set of RNA molecules suggests that replicative fitness depends on molecular shape, independent of most of the underlying base sequence. There is evidence that they require a minimum of one hairpin loop at the 5′-end for template competence (Schaffner *et al.* 1977; Biebricher & Luce 1993), with the preferred configuration including a second at the 3′-end (Voronin 1992), giving the transcription initiation site a tRNA-like structure (Adhin *et al.* 1990). Quasispecies-forming RNA variants are highly folded, however. They contain 3–9 hairpin loops, each typically of 8 b.p., for sequence lengths between 77 and 221 nucleotides. Thus, it remains to be explained why approximately two-thirds of their base sequence is paired in loop stems.

Annealing between complementary strands is known to be retarded by secondary structure. Biebricher *et al.* (1985) noted a lower rate of duplex formation in solution among + and − strands of RNA variants than anticipated from experiments with randomly coiled polynucleotides. An established fall-off in the (second order) rate coefficient for annealing with the square root of reciprocal polymer length (Wetmur & Davidson 1968), consequently, understates the dependence of this parameter on chain length among RNA variants. This is significant, because Qβ replicase cannot copy double-stranded RNA (Biebricher *et al.* 1982; Nishihara *et al.* 1983). The occurrence of self-complementary regions in self-replicating RNA molecules, favouring formation of a folded configurations, can be interpreted, therefore, as a defensive measure designed to minimize the loss of template competence by conversion to the double-strand form. The kinetics of replication among RNA variants, incorporating inactivation via duplex formation between strands free in solution, has been investigated previously by Biebricher *et al.* (1984). These investigators did not rule out the possibility of a more complex process, however, as they stated 'It cannot be excluded that replication may sometimes lead to the double strand.' Their conjecture was substantiated by Priano *et al.* (1987), who obtained experimental evidence of two mechanisms of double strand formation: a first order process, as well as standard, second order annealing.

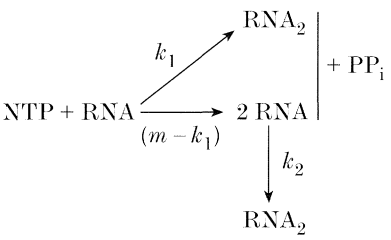
Duplex formation can be first order, when a certain fraction of nascent RNA chains fail to separate from template on completion of each transcription cycle. In the absence of replicase catalyzed unwinding of replica and template strands, formation of a double strand molecule appears to be a likely outcome of their one-to-one interaction and the issue is really: how does strand separation occur in a simple system. Consistent with involvement of strand secondary structure in separating replica and template, the helical stems of hairpin

loops contained by RNA variants form during transcription (Kramer & Mills 1981). It is not surprising then that the regio-specific copying of poly C and random poly CG, in a replicase-free system (Inoue & Orgel 1981, 1982; Joyce & Orgel 1986) should cease, once all template sites have been occupied by complementary bases. As strand separation is required for sustained replication, a highly folded configuration may be a general design feature of polymers produced in simple replicating systems.

To clarify the significance of RNA-driven strand separation in a system whose replicase lacks helicase activity, the kinetic theory of replication (Eigen 1971; Thompson & McBride 1974; Tyson 1974; Davis 1978*a*, *b*; Eigen & Schuster 1979; Biebricher *et al.* 1983, 1984, 1985; Eigen *et al.* 1989; Fontana *et al.* 1991) has been extended to accommodate dual mechanisms of duplex formation. A mathematical analysis of this process is given in the following section. Subsequent examination of the experimental results of Priano *et al.* (1987) on growth and duplex levels among RNA variants in the Qβ system, described in §3 and §4, reveals that the replica-template annealing reaction is of more importance than previously considered. It will be seen that a duplex-avoidance strategy based on strand secondary structure has fundamental implications for the origin of replication.

2. REPLICATION KINETICS WITH DUAL DUPLEX-FORMATION MECHANISMS

Synthesis of the replicative form of an RNA variant concurrent with both side reactions producing double-stranded RNA is depicted by the following reaction scheme;



(1)

k_1 is a first order rate coefficient. It applies to the side reaction in which template and replica strands fail to dissociate at termination of template transcription. k_2 represents a standard annealing rate coefficient, which characterizes association of complementary strands in solution, by a concentration dependent, second order reaction. $(m-k_1)$ is the first order rate coefficient for single-strand formation. m refers to the propagation rate that would apply if first order duplex formation were suppressed. RNA chain elongation occurs by standard end-addition condensation at its 3′-terminus. Incorporation of a nucleoside triphosphate (NTP), specifically H-bonded to template, liberates a pyrophosphate molecule, PP_i. The new RNA strand is destined to be released or bound with template in a duplex, RNA_2. Whereas RNA replication is a multistep cycle of reactions that can be represented in terms of many kinetic parameters (Biebricher *et al.* 1983), polymer propagation itself is a pseudo-first order process, dependent on a single rate constant, m, during exponential kinetics. The model considered here is designed for evaluating changes over time in the concentration of polymers in replicative form and in duplexes from both sources depicted in (1), under the experimental conditions of Priano *et al.* (§3 and §4).

The rate equation for single strand formation is

$$\mathrm{d}S/\mathrm{d}t = (m-k_1)S - k_2S^2. \tag{2}$$

Solving this ordinary differential equation yields,

$$S = S_0\mathrm{e}^{(m-k1)\tau} \Big/ \left(1 + \frac{k_2}{(m-k_1)}S_0(\mathrm{e}^{(m-k1)\tau} - 1)\right) \tag{3}$$

Single strand concentration, S, is given in terms of initial concentration, S_0, and elapsed reaction time, τ. k_2 corresponds to the rate coefficient used by Priano *et al.* (1987). It equals half the standard second order coefficient k_2':

$$\mathrm{d}A/\mathrm{d}t = -k_2'AB = -k_2'A^2;\ A = B$$

$$\frac{\mathrm{d}S/\mathrm{d}t}{2} = -k_2'\left(\frac{S}{2}\right)^2;\ S = A + B,\ A = \frac{S}{2},$$

k_2 characterizes the overall depletion of single strands, including both complementary strands; A and B. When replication is initiated with RNA from a homogeneous stock of one strand type, equal strand concentrations follow the first RNA doubling cycle. In the first growth cycle, single strand concentration of the initiating type should actually decrease mainly through annealing with nascent RNA strands. Equations (2), (3) then apply only to post-first cycle synthesis.

From (1), the rate equation for replica-template annealing is

$$\mathrm{d}D_1/\mathrm{d}t = k_1S,$$

D_1 being the concentration of RNA strands in duplex form produced by this first order reaction. The single strand conversion rate is then a constant fraction of the number of replications. The rate of change in D_1 may be stated in terms of S. After recalling (2),

$$\mathrm{d}D_1/\mathrm{d}S = k_1/((m-k_1) - k_2S),$$

and following integration,

$$D_1 = -\frac{k_1}{k_2}ln\left[\frac{(m-k_1) - k_2S}{(m-k_1) - \mathrm{k}_2S_\mathrm{o}}\right]. \tag{4}$$

An equimolar preparation of complementary, non-replicating strands should retain as single strands the fraction,

$$f_{ss} = (1 + k_2S_0\tau)^{-1},$$

after the interval, τ. k_2 will be smaller for initially self-annealed strands versus randomly coiled strands (Biebricher *et al.* 1985). The concentration of single strands converted to duplex RNA in this process, D_2, is

$$D_2 = S_0(1 - f_{ss}).$$

Reliance on initial single strand levels would obviously yield an underestimate of duplex formation in a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

growing polymer population. The average single strand RNA concentration throughout an interval is used for this reason. From (3),

$$\bar{S} = \frac{1}{\tau}\int_0^{\tau} \mathrm{d}t\ S$$

$$= \frac{1}{\tau}\int_0^{\tau} \mathrm{d}t\ S_0 e^{(m-k1)t}/(1+(k_2(m-k_1)^{-1})S_0(e^{(m-k1)t}-1)).$$

Consequently,

$$\bar{S} = \alpha^{-1} + \frac{1}{k_2\tau}\ln\left[\frac{\alpha-(\alpha-S_0^{-1})e^{-(m-k1)\tau}}{\alpha-(\alpha-S_0^{-1})}\right] \quad (5)$$

α is $k_2(m-k_1)^{-1}$. D_2 can be related to the average single-strand RNA concentration, $\bar{S}$, over this interval,

$$D_2 = \bar{S}(1-(1+k_2\bar{S}\tau)^{-1}). \quad (6)$$

Initiation of replication by homogeneous RNA yields unequal synthesis of complementary strands in the first growth cycle. In this event, the average single-strand concentration of a derived sequence, B, throughout the first cycle may be used to determine duplex levels, $D_2(1)$, from this source,

$$D_2(1) = 2k_2A_0\bar{B}t_D = 2k_2A_0\left(A_0\left(\int_0^{tD}\mathrm{d}te^{(m-k1)t}\right)-1\right);$$

$$A_0+B = A_0e^{(m-k1)t}$$

$$= 2k_2A_0^2\left(\frac{e^{(m-k1)tD}-1}{m-k_1}-1\right). \quad (7)$$

In the kinetic model outlined by Priano *et al.*, estimates of D_2 mole fraction, E, were obtained by assuming

$$E = k_2R\langle t_D\rangle/(1+k_2R\langle t_D\rangle)$$

$$= 1-1/(1+k_2R\langle t_D\rangle),$$

R is the total RNA concentration $(S+D)$ at τ. $\langle t_D\rangle$ refers to the expectation interval for replication per strand within the polymer population. It equals a reciprocal rate coefficient: $m^{-1} = t_D/\ln 2$, with t_D being RNA doubling time. This estimate for E is compatible with (6), on replacing R by S. k_2 characterizes the rate of depletion among all single strands, as in (2). It equals half the standard rate coefficient, k_2', as noted. However, as R is the RNA concentration at τ, it far exceeds the average single-strand concentration, S, over the interval τ. Substantial overestimates of duplex production result for annealing between initially dissociated complementary RNA strands.

Increases in duplex production rate with RNA growth were attributed by these workers to increased rates of annealing between free strands, driven by elevations in strand concentration with replication. The plausibility of this inference is supported by the following inequality,

$$\mathrm{d}D_2/\mathrm{d}\bar{S} = \mathrm{d}(\bar{S}(1-1/(1+k_2\bar{S}\tau))/\mathrm{d}\bar{S}$$

$$= 1-(1+k_2\bar{S}\tau)^{-2} \geqslant 0;\ (1+k_2\bar{S}\tau)^2 \geqslant 1.$$

It results from differentiation of (6) with respect to positive changes in single-strand RNA concentration. Another, equally significant source of time variations in duplex production rate is revealed by the proposed model. It reflects a dilution effect linked to replica-template annealing. This may appear surprising, considering that the differential form of (4) implies that first order duplex formation, D (holding the second order process constant), is a constant fraction of all replications,

$$\left(\frac{\delta D}{\delta S}\right)_{D_2} = \frac{k_1}{m-k_1}.$$

In the initial stages of synthesis, however, S_0, levels exceed δS and D, and this distorts duplex mole fractions: $\delta D/(S_0+\delta S+\delta D)$. With RNA growth, this effect is progressively offset.

3. ANALYSIS OF RNA GROWTH

Duplex formation in replication can be detected from the pattern of change in total RNA. As single strands are the replicative form of RNA variants copied by Qβ replicase, synthesis will be maximized in the first RNA doubling cycle, following initiation with homologous, single-strand RNA. Production of non-replicated double-strand molecules decreases the rate of RNA synthesis per strand in subsequent doublings. When this fall-off is constant from the first doubling cycle, duplex formation is first order. A growth dependent fall-off that becomes steeper at elevated strand concentrations indicates second order duplex formation.

Evidence confirming that duplex RNA molecules did not replicate in the experiment of Priano *et al.* (1987) can be obtained from a comparison of initial formation rate coefficients and maximum polymerization velocities. In their experiment, total RNA concentration increased between five- and ten-fold for cordycepin tolerant variant (CTV), microvariant (MCV) and midivariant (MDV) after replication for 10 min, commencing with an estimated 2.4×10^{11} single strands (R_{10}/R_0, table 1); these RNA variants have nucleotide sequence lengths of 77, 115 and 221, respectively. Based on these increases, they have effective propagation rates, m_{10}, from 2.68×10^{-3} to 3.80×10^{-3} s^{-1}. Taking into account variant single strand frequency, which ranges from 0.50 to 0.81 at 10 min, the effective formation rate coefficient can be boosted to the putative rate coefficient for a single strand preparation. We note that

$$m_0 = \bar{m}_{10}p_s^{-1},$$

m_0 being single-strand replicative fitness, or initial formation rate coefficient. p_s is the fraction of single strands over the replication interval. Omitted is the term, $(1-p_s)m_D$, related to duplex strands, as their fitness, m_D, is zero (Biebricher *et al.* 1982; Nishihara *et al.* 1983). With between one fifth and one half of all RNA strands occurring in duplexes (D/R_{10}) following replication for 10 min (see table 1), midrange single-strand frequencies, $(1-D/2R_{10})$, were between 0.750–0.905. Single strand formation rate coefficients for CTV, MCV and MDV were found to be 3.57, 3.94 and 4.20×10^{-3} s^{-1}, respectively. Comparatively, these single-strand rate coefficients for exponential growth,

Table 1. *RNA levels and replicative fitness during exponential replication with duplex production among three RNA variants*[a]

RNA species	$\frac{R_{10}}{R_0}$[b]	$\frac{D}{R_{10}}$	$\frac{(R_{10}-\bar{D})}{R_0}$	m_{10}	m_0 (s^{-1})	$\frac{m_0}{m_0(\text{MDV})}$	$\frac{v_m}{v_m(\text{MDV})}$
				$\times 10^{-3}$			
MDV	9.8	0.19	7.94	3.80	4.20	1.0	1.0
MCV	7.1	0.34	4.69	3.27	3.94	0.94	0.87
CTV	5.0	0.50	2.50	2.68	3.57	0.85	0.77

[a] Based on experimental observations of Priano *et al.* (1987). Exponential growth was initiated by addition of 10 nM single strand MDV, MCV or CTV to a reaction mixture, 40 μl, containing 56 μg protein/ml Qβ replicase, 0.1 mM each NTP, with 250 pmol. [α-^{32}P]-GTP with 84 mM Tris, pH 7.5, and 12 mM $MgCl_2$ at 37 °C. RNA concentrations were determined from acid precipitable radioactivity at various times. Duplex formation was quantitated after noting radioactivity distribution on polymer gels after electrophoresis. Linear growth was initiated by addition of 500 nM RNA from single strand stock.

[b] R_0, R_{10} refer to number of RNA strands initially and after replication for 10 min. D denotes number of duplex strands at 10 min. m_0, m_{10} are single strand (initial) and mean fitness parameters: $m_{10} = \ln [R_{10}/R_0]/600$, and $m_0 = (1-D/2R_{10})^{-1}m_{10}$. v_m is the maximum (linear phase) polymer formation velocity.

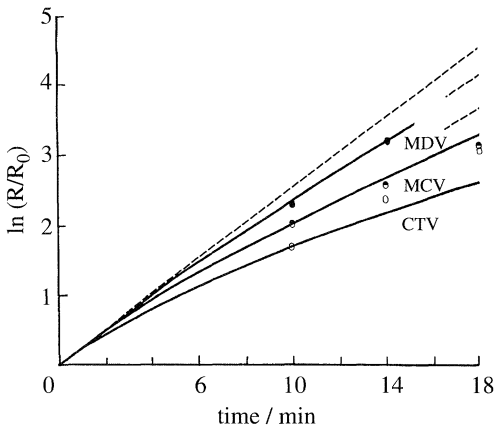


Figure 1. Pattern of growth by MDV, MCV and CTV during exponential replication in the presence of Qβ replicase. Growth unretarded by duplex formation is indicated by broken lines. They were calculated using m_0 values in table 1. Replicate-template annealing reduced RNA levels to those indicated by the continuous line. The rate of this first order reaction was inferred from duplex levels after replication for 10 min. Negative deviations by points from a continuous line indicates annealing between complementary strands free in solution. Points mark RNA levels observed by Priano *et al.* (1987).

agree with maximum polymerization rates observed (Priano *et al.* 1987) under linear phase kinetics (table 1). Parenthetically, linear phase velocities were about one tenth of those attained during exponential growth. Initiation of replication is evidently facilitated by a molar excess of replicase (Biebricher *et al.* 1981). Agreement between the magnitudes of initial formation rate coefficients deduced for these variants and maximum polymerisation velocity, at molar excess of RNA, verifies that their double-strand form does not replicate.

Growth curves for MDV, MCV and CTV deduced from RNA levels at 10 min of reaction time (table 1) are shown in figure 1. These curves are diagnostic of the kinetic mechanism underlying propagation of each variant. A linear relation for RNA growth in a ln $[R/R_0]$ versus time plot, with slope equal to replicative fitness, m, when replicase is in molar excess, for instance, indicates duplex formation is suppressed. No variant achieved this growth pattern (figure 1). They all produced nonreplicating, duplex RNA. Consequently, their growth shows a deflection to lower rates over time. Deflection from the maximum formation rate coefficient, slope m in figure 1, should increase with time owing to a dilution effect, even when duplex formation is solely first order. This effect follows initiation with RNA from a single-strand stock, as noted (§2). With continued growth the effect decreases widening the deflection from RNA synthesis at its optimal rate, m.

Consistent with first order replication and duplex formation, the fastest growing variant, MDV, grew at a virtually constant rate (figure 1): $m_0(\text{MDV}) = 4.10 \times 10^{-3}\ s^{-1}$, with $\delta D/\delta R = 0.189$. By contrast, MCV levels at 14 min and 18 min of RNA synthesis show a negative deflection from first order kinetics: $m_0(\text{MCV}) = 3.77 \times 10^{-3}\ s^{-1}$ and $\delta D/\delta R = 0.396$. Hence, second order duplex formation occurred. The deflection became more robust with growth, consistent with a direct dependence on concentration of complementary single–strand RNA. A negative action by second order duplex formation on RNA growth, already retarded by first order duplex formation, can be inferred from (2). Despite annealing rapidly (table 1), CTV strands unexpectedly displayed positive deflions in growth rate with time (figure 1): $m_{10} = 2.682 \times 10^{-3}$, $m_{14} = 2.788 \times 10^{-3}$ and $m_{18} = 2.792 \times 10^{-3}\ s^{-1}$. The spread of a faster replicating variant within the polymer population could account for this anomalous growth pattern. This variant had an intrinsic single-strand replicative fitness, $m_0(\text{CTV})$, equal to $3.35 \times 10^{-3}\ s^{-1}$ and $\delta D/\delta R$ of 0.625.

A quantitative model of variant replication involving both mechanisms of duplex formation follows.

4. VARIANT REPLICATION WITH DUPLEX PRODUCTION

Mole fractions of double strand RNA produced by either annealing between replicate and template molecules or complementary strands free in solution, during exponential replication of MDV, MCV and CTV, were compared with values calculated using RNA formation rate coefficients based on those in table 1, and with empirically determined second order annealing rate

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 2. *Kinetic parameters and duplex RNA levels during exponential replication of three RNA variants*

growth interval			rate coefficients						
		conc.[a] of RNA	single strand formn.	duplex[b] formation		fraction of duplex RNA[b]			
exptl (min)	model (min)	strands (nM)	$(m-k_1)$ (s^{-1})	k_1 (s^{-1})	k_2[a] (M^{-1} s^{-1})	D_1	D_2	D_1+D_2	D_1+D_2[a] (observed)
			$\times 10^{-3}$	$\times 10^{-3}$	$\times 10^{3}$				
				MDV					
10	10.0	98	3.50	0.82	0.033	0.17	< 0.01	0.17	0.19
14	14.0	232				0.18	< 0.01	0.18	0.15
—	18.0	537				0.19	< 0.01	0.19	—
				MCV					
10	10.0	71	2.62	1.61	1.70	0.33	0.01	0.34	0.34
14	13.6	126				0.35	0.01	0.36	0.38
18	17.3	226				0.37	0.02	0.39	0.42
				CTV					
10	10.0	50	1.73	2.20	11.0	0.45	0.04	0.49	0.50
14	16.9	104				0.52	0.05	0.57	0.65
18	24.6	204				0.56	0.07	0.63	0.74

[a] Experimental values are from Priano *et al.* (1987). Polymerization and detection methods are outlined under table 1. Rate coefficients for double helix formation were determined from the rate of annealing between + and − strands of each variant, using strands prelabelled with [^{32}P] and a Tris-$MgCl_2$ buffer with NTP and added ammonium sulphate, EDTA, β-mercaptoethanol and glycol at 37 °C.

[b] $(m-k_1)$, k_1, k_2, D_1, D_2 refer to rate coefficients and duplex levels of first and second order double helix formation reaction.

coefficients (Priano *et al.* 1987). Table 2 includes rate coefficients for single-strand formation, $(m-k_1)$, first order, k_1, and second order, k_2, conversion to duplex, together with replication intervals for both experiment and kinetic model (§2). Effective single-strand formation rate coefficients vary with variant chain length, $(m-k_1)$ values being 3.50×10^{-3}, 2.62×10^{-3} and 1.73×10^{-3} s^{-1} for MDV, MCV and CTV, respectively. Variant duplex production rate coefficients vary inversely with chain length. First order coefficients, k_1, exhibit a spread of nearly three-fold, from 0.82×10^{-3} to 2.20×10^{-3} s^{-1}. While second order coefficients, k_2, span more than two orders of magnitude, ranging from 33 to 11×10^{3} M^{-1} s^{-1}.

Dissociated MDV strands anneal slowly, judging from RNA growth rates (figure 1). Experimental error could account for a small discrepancy between calculated and observed duplex levels from first order conversion of single strands of MDV. This variant produced the least duplex RNA and propagated fastest.

The negative deflection from first order kinetics in MCV growth (figure 1) is shown in table 2 to result from slow second order conversion to duplex. This yields strand mole fractions, D_2, between 0.01–0.02. First order conversion to duplex is far more significant, accounting for almost all MCV duplex formation; D_1 mole fractions vary from 0.33 to 0.37. Time variations in D_1 are seen to account for most of the observed increase in duplex levels over time. This can be traced to initiation of replication using a homogeneous, single-strand stock (§2).

CTV contains only 77 nucleotides. It is the smallest variant known to form a stable quasispecies. Its propagation is retarded, however, by high duplex production (table 2). A feature of the present kinetic analysis is an underestimation of observed duplex levels for this variant. This amounted to 13% at the 14 min and 18 min. intervals. Because this shortfall occurs despite expanded incubation intervals, the effect is viewed as significant. Another interesting feature of CTV kinetics, noted earlier, is an increasing formation rate coefficient. This is illustrated by a lengthening of incubation intervals in the model, which relies on a fixed rate coefficient (table 2).

It was possible to duplicate CTV growth and duplex production by increasing k_2(CTV) from 11×10^{3} to 18×10^{3} M^{-1} s^{-1}, and by assuming exponential growth in all CTV RNA, single- and double-strand forms, with a rate coefficient of 2.77×10^{-3} s^{-1}. k_1 remaining unchanged. Whereas it would be interesting to re-examine replicase interactions with duplex CTV, this alternative model is not tenable. Specifically, Qβ replicase cannot copy duplex RNA (Biebricher *et al.* 1982; Nishihara *et al.* 1983). More likely, CTV preparations showed faster growth than anticipated in a cordycepin-free polymerization mixture, because they contained variants better adapted than CTV for replication, in the absence of this inhibitor. It remains intriguing, nonetheless, that 1.6×10^{12} single strands of CTV saturated 2.24 μg Qβ replicase, whereas saturation by MDV and MCV required 3.3×10^{12} and 4.1×10^{12} single strands, respectively (Priano *et al.* 1987). CTV saturates Qβ replicase at 4.6×10^{12} strands, on adding the number of duplex, 2.9×10^{12}, and single CTV strands; a value not far above saturating levels for MDV and MCV single strands.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

5. DISCUSSION

Replicate annealing with template emerged as the most significant impediment to propagation by the self-replicating RNA molecules examined. Duplex formation between complementary strands free in solution was demonstrated to have far less impact. Because Qβ replicase lacks helicase activity, unlike some RNA-dependent RNA replicases (Cho *et al.* 1993), it is understandable that the three RNA variants examined by Priano *et al.* (1987) and several other variants copied by Qβ replicase (Schaffner *et al.* 1977; Munishkin *et al.* 1988, 1991; Biebricher & Luce 1992, 1993) have folded strand configurations. In the absence of catalysed unwinding, replication relies on RNA-driven strand separation. It is spontaneous separation of a nascent strand from template, through self-annealing in competition with the replicate-template annealing reaction, that makes possible multiple template-transcription cycles.

Considering that replicate-template annealing accounts for most duplexes formed by small RNA molecules undergoing replication in the Qβ replicase system, a kinetic analysis of duplex formation during replication in this system that focused exclusively on standard, second order annealing is not appropriate. Nearly all the duplex molecules found following replication of minivariant-11 (sequence length, 86 bases), salt resistant variant-11 (length, 115 bases) and MDV in the investigations of Biebricher *et al.* (1984, 1985) assuredly originated from replicate-template annealing, rather than annealing between free strands in solution, as suggested. The level of second order duplex formation was also overestimated by Priano and her coworkers, as they based their assessment on the endpoint RNA strand concentration for a replication interval, rather than its mean concentration (§2).

Folding by nascent RNA strands during replication (Mills *et al.* 1978) actually retards transcription, as it causes Qβ replicase to pause at hairpin stems. Variants with highly folded strand configurations are consequently transcribed only slowly. From transcription speeds observed by Priano *et al.* at reduced temperature, MDV strands, with nine hairpin loops, were assembled at a rate of around three NTP s^{-1} at 37 °C, whereas MCV and CTV strands, with four and three hairpin loops, respectively, were assembled at nearly twice this rate. Not withstanding its slower rate of replicase transcription and a longer nucleotide sequence, MDV had a higher propagation rate than either CTV or MCV (see figure 1). This correlates with the higher thermodynamic stability of its replicative form and reduced rate of conversion to duplex (Davis 1995).

Compelling evidence for the significance of strand configuration was reported recently by Biebricher & Luce (1992). A salt-resistant variant, SV-11, containing 115 nucleotides arranged in a palendromic sequence, was shown to have two single strand configurations. It folds into a metastable form, with four-hairpin loops of 38 b.p. overall, or a stable form, which is a single, long hairpin with 46 b.p. Pulse-chase experiments showed the multi-hairpin configuration is the replicative form; the single loop conformer is not copied. An RNA molecule will replicate, or not, therefore, depending on its strand folding pattern.

A long hairpin, such as that in SV-11, does not exert an absolute effect on replication efficiency, however. Its action depends on structural context. Thus, a hairpin 1.7-fold longer than that formed by stable SV-11 reduced replicative fitness by only 26–39 % when spliced into MDV (Axelrod *et al.* 1991; Davis 1995). In this genetically engineered variant, the long helical stem of its hairpin seems to be destabilized by the already self-annealed 5′-tail of nascent strand. Replicase should advance as the hairpin stem unwinds, driven initially by thermal fluctuations increasing entropy of both self–annealing replica and template strands.

RNA variants copied by Qβ replicase lack sequence homology, except at each terminus. Small sub-quasispecies RNA variants characterized recently by Biebricher & Luce (1993), with sequences of 26–49 nucleotides, differ at all sites apart from a 5′ GG(G/U) and 3′ CCC(A). Although some variant sequences had common motifs, such as CUUCGG and CCGAAG, they were not a constant feature. Likewise, quasi-species-forming variants display variable internal sequences (Schaffner *et al.* 1977; Priano *et al.* 1987; Munishkin *et al.* 1988, 1991; Biebricher & Luce 1992).

The dependence of template competence on a folded strand configuration, to the exclusion of its nonfolded conformer, and a manifest lack of internal sequence homology among variants copied by Qβ replicase both suggest that the generic pattern of strand folding among RNA variants determines replication efficiency. Most of the underlying base sequence is nonspecific. Giving higher priority to RNA configuration than to base sequence, in this manner, has profound implications for early evolution. This may be illustrated by considering the probability of randomly forming the simplest RNA molecule replicated by Qβ replicase. It contains 26 nucleotides, and both complementary strands fold into a least energy configuration with a single, 5′-end hairpin of three and six b.p. (Biebricher & Luce 1993). The total number of sequences which could be formed through randomly arranging nucleotides from dimer to a 26-mer RNA sequence can be estimated (Schelkunoff 1965; Davis 1996),

$$N(2,n;z) \approx \frac{z^{n+1}}{z-1}.$$

Here, z (monomer variety) and n (sequence length) are four and 26, respectively. Over 10^{16} possible sequences form the sample set from which this self-replicating RNA molecules is randomly selected.

A ‘configuration before sequence’ model substantially increases this probability. When template competence exists in any RNA molecule of around 26 nucleotides, containing a hairpin of at least three b.p., internal G–U pairs included, and starting at, or next to its 5′-end, the probability of forming a self-replicating strand through random copolymerization is estimated to be only 2.4×10^{-4}. It is not surprising that selection of a suitable configuration is twelve orders of mag-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

nitude more probable than the underlying base sequence, because many sequences can be found that will form a specified configuration. On specifying five bases (at strand ends), this probability reduces by three orders of magnitude; self-replication remains nine orders more likely than selection of the given variant sequence. Strand separation also underlies the success of von Kiedrowski and associates (von Kiedrowski 1986; von Kiedrowski *et al.* 1991; Matzen *et al.* 1994) in achieving sustained self-replication in a replicase-free system by ligation of trinucleotides on a hexadeoxynucleic acid template. In the replicase-free system developed by Orgel *et al.*, replication of a homopolymer or random copolymer ceases at the first transcription cycle, when all template bases are paired to complementary bases (Joyce & Orgel 1986). This length restriction on replication of sequences unable to self-anneal might be circumvented, in principle, through some procedure combining transcription in short segments with ligation of the resulting oligomers.

A folded strand configuration seems likely to be a general design feature of replicating molecules, which lack a mechanism for catalysed replicate-template unwinding, since it represents, as indicated, a readily accessible duplex avoidance strategy. In the Qβ replicase system, strand secondary structure was seen to underlie RNA-driven strand separation during replication. With self-annealing in competition with the replicate-template annealing reaction (equation (1)), sustained replication is possible. Strand secondary structure in Qβ and other coliphages has been implicated also in regulation of the viral genome (Skripkin *et al.* 1990; Skripkin & Jacobson 1993). As these findings expand the spectrum of RNA functions, they add credence to an RNA phase in early evolution. Combining this with the evidence that early evolution followed a rapid, 'configuration before sequence' path, suggests that the first replicating polymer on Earth was a folded preRNA strand, apparently of prochiral, acyclic monomers (Joyce *et al.* 1987; Orgel 1992; Joyce & Orgel 1993).

My thanks to K. Rousell for his assistance.

REFERENCES

Adhin, M. R., Alblas, J. & van Duin, J. 1990 Secondary structure at the 3′ terminal region of RNA coliphages: comparison with tRNA. *Biochem. Biophys. Acta* **1050**, 110–118.

Axelrod, V. D., Brown, E., Priano, C. & Mills, D. R. 1991 Coliphage Qβ RNA replication: RNA catalytic for single-strand release. *Virology* **184**, 595–608.

Biebricher, C. K., Dickmann, S. & Luce, R. 1982 Structural analysis of self-replicating RNA synthesized by Qβ replicase. *J. molec. Biol.* **154**, 629–648.

Biebricher, C. K., Eigen, M. & Gardiner, W. C. 1983 Kinetics of RNA replication. *Biochemistry* **22**, 2544–2559.

Biebricher, C. K., Eigen, M. & Gardiner, W. C. 1984 Kinetics of RNA replication: Plus–minus assymetry and double strand formation. *Biochemistry* **23**, 3186–3194.

Biebricher, C. K., Eigen, M. & Gardiner, W. C. 1985 Kinetics of RNA replication: competition and selection among self-replicating RNA species. *Biochemistry* **24**, 6550–6560.

Biebricher, C. K. & Luce, R. 1992 *In vitro* recombination and terminal elongation of RNA by Qβ replicase. *EMBO J.* **11**, 5129–5135.

Biebricher, C. K. & Luce, R. 1993 Sequence analysis of RNA species synthesized by Qβ replicase without template. *Biochemistry* **32**, 4848–4854.

Cho, M. W., Richards, O. C., Dmitrieva, T. M., Agol, V. & Ehrenfeld, E. 1993 RNA duplex unwinding activity of Poliovirus RNA–dependent RNA polymerase $3D^{pol}$. *J. Virol.* **67**, 3010–3018.

Davis, B. K. 1978*a* Rate of polymer formation and entropy production during competitive replication. *J. molec. Evol.* **10**, 325–338.

Davis, B. K. 1978*b* Change in polymerization rate and entropy production during competitive replication. In *Biomolecular structure and function* (ed. P. F. Agris *et al.*) pp. 607–610. New York: Academic Press.

Davis, B. K. 1995 Stability of replicative form and fitness among RNA variants transcribed by Qβ replicase. *Proc. R. Soc. Lond.* B **260**, 39–43.

Davis, B. K. 1996 A theory of evolution that includes prebiotic self-organization and episodic species formation. *Bull. Math. Biol.* **58** (In the press.)

Eigen, M. 1971 Self-organization of matter and the evolution of biological molecules. *Naturwissenschaften* **58**, 465–523.

Eigen, M. & Schuster, P. 1979 *The hypercycle. A principle of natural self-organization.* Berlin: Springer-Verlag.

Eigen, M., McCaskill, J. & Schuster, P. 1989 The molecular quasispecies. *Adv. phys. Chem.* **75**, 149–263.

Fontana, W., Griesmacher, T., Schnabl, W., Stadler, P. F. & Schuster, P. F. 1991 Statistics of landscapes based on free energies, replication and degradation rate constants of RNA secondary structure. *Monat. Chem.* **122**, 795–819.

Inoue, T. & Orgel, L. E. 1981 Substituent control of the poly(C)-directed oligomerization of guanosine 5′-phosphoimidazolide. *J. Am. Chem. Soc.* **103**, 7666–7667.

Inoue, T. & Orgel, L. E. 1982 Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C). *J. molec. Biol.* **162**, 204–217.

Joyce, G. F. & Orgel, L. E. 1986 Non-enzymatic template-directed synthesis on RNA random copolymers. *J. molec. Biol.* **188**, 433–441.

Joyce, G. F. & Orgel, L. E. 1993 Prospects for understanding the origin of the RNA world. In *The RNA world* (ed. R. F. Gesteland & J. F. Atkins) pp. 1–25. Plainview, New York: Cold Spring Harbor Laboratory Press.

Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. 1987 The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4398–4402.

Kramer, F. R. & Mills, D. R. 1981 Secondary structure formation during RNA synthesis. *Nucleic Acid Res.* **9**, 5109–5124.

Matzen, M., Sievers, D., von Kiedrowski, G. *et al.* 1994 Self-replicating oligonucleotides. *Origins Life* **24**, 203–204.

Mills, D. R., Dobkin, C. & Kramer, F. R. 1978 Template-determined variable rate of RNA chain elongation by Qβ replicase. *Cell* **15**, 541–550.

Munishkin, A. V., Voronin, L. A. & Chetverin, A. B. 1988 An *in vitro* recombinant RNA capable of autocatalytic synthesis by Qβ replicase. *Nature, Lond.* **333**, 473–475.

Munishkin, A. V., Voronin, L. A., Ugarov, V. I., Bondareva, L. A., Chetverina, H. V. & Chetverin, A. B. 1991 Efficient templates for Qβ replicase are formed by recombination from heterologous sequences. *J. molec. Biol.* **221**, 463–472.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Nishihara, T., Mills, D. R. & Kramer, F. R. 1983 Localization of the Qβ replicase recognition site in MDV-1 RNA. *J. Biochem.* **93**, 669–674.

Orgel, L. E. 1992 Molecular replication. *Nature, Lond.* **358**, 203–209.

Priano, C., Kramer, F. R. & Mills, D. R. 1987 Evolution of the RNA coliphages: the role of secondary structures during RNA replication. *Cold Spring Harb. Symp. quant. Biol.* **52**, 321–330.

Schaffner, W., Ruegg, K. J. & Weissmann, C. 1977 Nanovariant RNAs: nucleotide sequences and interaction with bacteriophage Qβ replicase. *J. molec. Biol.* **117**, 877–907.

Schelkunoff, S. A. 1965 *Applied mathematics for engineers and scientists*. Princeton: D. van Nostrand.

Skripkin, E. A. Adhin. M. R. de Smit, M. H. & van Duin, J. 1990 Secondary structure of bacteriophage MS2. Conservation and biological significance. *J. molec. Biol.* **211**, 447–463.

Skripkin, E. A. & Jacobson, A. B. 1993 A two-dimensional model at the nucleotide level for the central hairpin of coliphage Qβ. *J. molec. Biol.* **233**, 245–260.

Thompson, C. J. & McBride, J. L. 1974 On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Math. Bioscience* **21**, 127–142.

Tyson, J. J. 1974 Competition, selection and evolution in chemical networks. *Lect. Math. Life Sci.* **7**, 43–67.

von Kiedrowski, G. 1986 A self-replicating hexadeoxynucleotide. *Angew. Chem. int. edn. Eng.* **25**, 932–935.

von Kiedrowski, G., Wlotzka, B., Helbing, J., Matzen, M. & Jordan, S. 1991 Parabolic growth of a self-replicating-hexadeoxynucleotide bearing a 3′-5′-phosphodiamidate linkage. *Angew. Chem. int. edn. Eng.* **30**, 423–426.

Voronin, L. A. 1992 Structural consensus of RNA templates replicated by Qβ replicase. *Biochemie* **74**, 491–494.

Wetmur, J. G. & Davidson, N. 1968 Kinetics of renaturation of DNA. *J. molec. Biol.* **31**, 349–370.

Received 21 March 1995; *accepted 10 August 1995*

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES