Transcription and translation in an RNA world

William R. Taylor*

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

The RNA world hypothesis requires a ribozyme that was an RNA-directed RNA polymerase (ribopolymerase). If such a replicase makes a reverse complementary copy of any sequence (including itself), in a simple RNA world, there is no mechanism to prevent self-hybridization. It is proposed that this can be avoided through the synthesis of a parallel complementary copy. The logical consequences of this are pursued and developed in a computer simulation, where the behaviour of the parallel copy is compared to the conventional reverse complementary copy. It is found that the parallel copy is more efficient at higher temperatures (up to 90°C). A model for the ribopolymerase, based on the core of the large subunit (LSU) of the ribosome, is described. The geometry of a potential active site for this ribopolymerase suggests that it contained a cavity (now occupied by the aminoacyl-tRNA) and that an amino acid binding in this might have ‘poisoned’ the ribopolymerase by cross-reacting with the nucleoside-triphosphate before polymerization could occur. Based on a similarity to the active site components of the class-I tRNA synthetase enzymes, it is proposed that the amino acid could become attached to the nascent RNA transcript producing a variety of aminoacylated tRNA-like products. Using base-pairing interactions, some of these molecules might cross-link two ribopolymerases, giving rise to a precursor of the modern ribosome. A hybrid dimer, half polymerase and half proto-ribosome, could account for mRNA translocation before the advent of protein elongation factors.

Keywords: RNA world; replication; translation; ribosome; origin of life

1. REPLICATION IN AN RNA WORLD

(a) *Basic assumptions*

At some stage during the early evolution of life, there must have been a point where information began to be passed from one nucleic acid molecule to another. Given that this is the system we have today, it does not seem to be unreasonable to say that it arose in the past. Whether the information-carrying molecule was RNA or DNA (or some alternative) is more difficult to specify and whether the system was self-replicating or assisted by other molecules (or even minerals) then gives considerable scope for speculation (for reviews see Maynard Smith & Szathmáry (1995) and Gesteland *et al.* (1999) and other contributions to this issue).

There have been many ideas proposed for the nature of the first replicating system and rather than try to compare and evaluate these here, I am going to assume that both the replicating molecule and the replicator were RNA. This places my speculations in what has been called the ‘RNA world’ (Gilbert 1986). I will not consider how this world might have arisen, since there are many contributions to this issue and elsewhere that cover this aspect (e.g. Joyce 1989). I will assume that there is either a good abiotic source of raw materials or that there is, possibly in addition, an established metabolism providing these.

I will also assume that the transmission of the ‘genetic’ information is mediated through Watson/

Crick nucleotide base pairing. Again, the basis for this is the system that we have today—so it must have arisen in the past. Alternative molecular interactions have been proposed (Benner & Ellington 1991; Benner *et al.* 1993), but these are largely motivated by constraints on prebiotic chemistry rather than anything fundamental to the transmission of information.

(b) *A fundamental problem*

Making the assumptions outlined earlier gives the situation where an RNA molecule, perhaps acting as a ribozyme, must be copied into its reverse complement and then recopied again to produce an identical (and functional) copy of the original molecule. Whatever the nature of the replicating mechanism, this requires the wasteful creation of an intermediate reverse complement (unless all the sequences are palindromes) and also a doubling of the error rate as there must be two copying steps. The former problem might be alleviated if the reverse complement is also a different functional ribozyme.

There is also a more fundamental problem associated with this replication mechanism, that a single replication step is unlikely to occur in isolation. This means that there will be copies of each molecule and its reverse complement in the same space at the same time. In the modern world, this does not cause any difficulty, as the nucleic acid is either double stranded or protected from hybridization by single-strand-binding proteins. Even if the single-stranded molecule quickly adopted a secondary and a tertiary structure, it must still remain sufficiently accessible to be unfolded and copied by the

*wtaylor@nimr.mrc.ac.uk

One contribution of 19 to a Discussion Meeting Issue ‘Conditions for the emergence of life on the early Earth’.

Figure 1. Replication strategies. (a) Replication via a reverse complementary strand leads to (b) a stable double-stranded duplex if the two copies meet. (c) Replication via a parallel complementary strand leads to (d) a relatively unstable double-stranded duplex if the two copies meet.

replicase. This degree of accessibility would also allow some hybridization with its complement and once completely base paired, the resulting duplex would have a lower energy than any internal bonding structure (Bartel 1999; Joyce & Orgel 1999). The duplex would therefore act as a low-energy sink, removing functional molecules from the population (figure 1).

In a more primitive world, it is difficult to imagine a mechanism that would keep the template strand and its reverse complementary transcript strand apart. The physical screening of the replicase itself, whatever its nature, would provide some protection for the current template from its transcript, but this is only a local solution as an adjacent replicase could easily be making a complementary strand. Some physical separation such as a membrane might be imagined, but would require an unrealistic degree of synchronization to keep all complementary copies apart. As in the modern world, some single-stranded binding mechanism might be imagined, based on either random oligonucleotides or peptides. While the latter is not impossible, a solution is proposed later that also opens some further possibilities.

(c) A parallel complement

The propagation of information in a nucleic acid strand from one 'generation' to the next using Watson–Crick base pairing logically does not have to involve a reverse complementary strand. Providing that there is complementary base pairing, a parallel complement would also propagate the same information. In the modern context, this possibility is ignored since the parallel DNA duplex cannot form a stable double helix and thus would make a poor information archive. In our more primitive RNA world (with no double-stranded genome), this is not a relevant constraint and it is

Figure 2. A conceptual model for a ribopolymerase. The ribopolymerase (replicase) is shown in outline as a bean shape. It is copying a template (horizontal strand) to synthesize a transcript (upright line) and the fates of all these components are depicted by arrows. Some transitions are reversible (e.g. *fold* ↔ *unfold*), while others are not (e.g. *unfold* → *hydrolyse*). The formation of a new *replicase* happens only when a plus strand folds and a *duplex* is formed only when an unfolded plus and minus strand come together.

possible to postulate that replication progressed via a parallel-stranded complement (Taylor 2005a).

In the context of the replication models discussed earlier, all that need change from the viewpoint of replicase is the direction of its progression along the template. The resulting transcript could only be expected to base pair with the template over a short region before parting, but, faced with the problem of irreversible hybridization, this would be a desirable feature of the model.

2. CONCEPTUAL MODELS

(a) A mathematical model

(i) A simple RNA world

The interactions discussed earlier can be modelled in greater detail using the calculated energies of RNA strand folding and hybridization. A mathematical model was constructed in which a replicase copies a strand with a given probability. The transcript then has a chance to fold, and if folded also has a chance to unfold, and if unfolded has a chance to hydrolyse. It also has a chance to meet its complement, and if it does, then to hybridize and once in a duplex, also to separate (figure 2). For simplicity, the model was applied to a simple RNA world consisting only of the replicase and its complement (Taylor 2005b).

There are just three free parameters in the model: the probability of the replicase finishing a copy, chance of hydrolysis and temperature. As the first is largely unknown it was set to 0.5, leaving hydrolysis and temperature to be varied. The former depends heavily on the nature of the primordial 'soup', especially the metal ion compositions, and it was varied in a range around conditions for zero growth. All other folding/unfolding and hybridization rates are a function of the calculated stability of the RNA molecule, which was made using the Vienna package (Zuker *et al.* 1998; Hofacker 2003).

Figure 3. Ribopolymerase population models. (a) The simulation of the molecular populations were run 50 times with the number of single-stranded ribopolymerase molecules plotted over 10 000 cycles. (b) The populations started with two molecules and were allowed to grow un-hindered until 100 ribopolymerases were created. After this, the ‘catabolic’ processes of duplex formation and hydrolysis were activated.

This simple model was investigated using both parallel and reverse complementary replication strategies. While the energy of the latter can be calculated using conventional methods, these do not allow the calculation of the energy of a parallel duplex. Parallel-stranded DNA duplex structures are known from X-ray studies (Rippe & Jovin 1992; Parvathy *et al.* 2002) and although no-one has synthesized an RNA equivalent, there is little to make one suspect that it could not exist. By analogy with the DNA structures, it can be expected that this would be much less stable than the reverse complementary duplex. Some reasonable limits can be placed on what this might be. If the parallel duplex is very unstable, then a lower bound will

be given by the internal folding energy of each strand separately. An upper bound can be taken as having the same energy as the reverse construct, but with the loss of one hydrogen-bond for each G : C pairing. (The G : C H-bonds are asymmetric and only two can be made between parallel strands, whereas with only two bonds, the A : T pairing must be symmetric.) This can then be estimated using conventional methods by changing each G : C pair to A : T (Taylor 2005b).

(ii) Simulations

Both the parallel and reverse constructs were simulated at different temperatures for different rates of single-strand hydrolysis. The success of a population was

Figure 4. Viability change with temperature. The average number of active ribopolymerase molecules remaining at the end of 50 simulations is plotted for different temperatures (x -axis, $^{\circ}\text{C}$) and different values of the hydrolysis rate (0.1–0.9, outer curves to inner curves, respectively). The curves are plotted for both antiparallel (solid lines) and parallel (dashed lines) constructs using two variants of the models (a) and (b) (see Taylor (2005b) for details).

measured by the number of active replicases at the end of 10 000 cycles (figure 3). Irrespective of the hydrolysis rate, there was a clear trend for the parallel construct to be more viable at higher temperatures, typically 20–30 $^{\circ}\text{C}$ higher than the reverse construct. If the temperatures used in the RNA ‘folding’ programs can be interpreted in absolute terms, this would give the parallel construct a peak efficiency *ca* 40–50 $^{\circ}\text{C}$, but still retaining significant activity up to 90 $^{\circ}\text{C}$ (figure 4).

The source of this benefit can be compared to the replication of strands in the polymerase chain reaction (PCR) in which strands are copied using a protein polymerase. In PCR, the DNA strands are heated to above 90 $^{\circ}\text{C}$ to cause strand separation then the temperature is lowered to allow primers to anneal and copying to proceed. The process is then repeated many times. In the RNA world, there was no one (we assume) to cyclically raise and lower the temperature to encourage replication, but with a parallel construct, the duplex is sufficiently unstable that it is not necessary to raise the temperature for separating the strands.

Many of the other contributions to this volume propose that the temperatures in the Hadean were higher than present and some studies also propose

quite an early date for the presence of liquid water. Whether on the surface or subterranean, an operational temperature of up to 90 $^{\circ}\text{C}$ would allow a replicase using a parallel complement to be viable long before it might have faced competition from a reverse complementary competitor.

(b) *A cooperative replicase*

(i) *Protecting ssRNA*

One of the poorly fixed parameters in the model described earlier was the hydrolysis rate of exposed single-stranded RNA (ssRNA). During replication, both the template and the transcript must be unfolded to some extent and even if unfolding of the template occurs only in a local region as the replicase progresses or even if the transcript folds as it is synthesized, there will always be segments that must wait for their bonding partners to be copied or synthesized, respectively. If these vulnerable situations can be protected from hydrolysis, then as shown in the simulations, the system becomes increasingly viable.

If the system is replicating a reverse complementary strand and the replicase begins at the 5' end, then the transcript will appear at the 3' end first and cannot be recopied until a new 5' end has been synthesized and it is free of the replicase. This means that for the whole duration of its replication, the copy will be vulnerable to hydrolysis. If, on the other hand, a parallel copy is being made, this will emerge led by another 5' end, which after a short interval will be available to be copied before its 3' end is complete (figure 5).

This situation is similar to the way in which ribosomes attach to the emerging transcript in bacterial protein synthesis, while the mRNA is still being transcribed. It has been argued that this is a strategy used by bacteria to minimize the exposure of single-stranded mRNA, allowing the survival of thermophilic bacteria and consistent with a thermophilic origin for all bacteria.

(ii) *A dimeric replicase*

To efficiently exploit this strategy, the replicase could function as a dimeric unit in which one transcript is immediately ‘fed’ into the second subunit as a template. In this arrangement, the transcript would not need to fold and unfold before being recopied, and if the dimer relationship was tight, then the short segment of ssRNA would be protected from hydrolysis. Taken as a functional unit, the dimer would take a strand and generate an identical copy plus a parallel complement.

As the emerging ends of all transcripts and templates would be in the correct polarity for immediate copying, the system need not be limited to a dimeric pair, but could form an extensive network of linked replicases (figure 5). Indeed, it is possible to imagine such a system extended to macroscopic size, taking the form of a gel (or slime) either on or in a porous rock. Although I have focused on the replicase itself, it would be expected that the RNA molecules being copied would include those involved in metabolic processes that maintain a supply of suitable building blocks.

Figure 5. Co-operative ribopolymerase networks. (a) Making a reverse complementary transcript means that re-copying cannot start until completion; with a parallel complementary transcript any transcript or template can be immediately re-copied. (b–d) This would allow extensive cooperating networks to become established.

Figure 6. A structural model for the ribopolymerase. (a) The backbone structure of the ribosome (PDB code 1JJ2) is shown as a thick trace for nucleic acid and a thin black trace for protein. The catalytic core of the ribosome (positions 2469–2651) is shaded in darker grey. The view is looking down into the tRNA-binding sites along the axis of the polypeptide exit channel. (b) The core fragment (dark grey in part (a)) is shown in isolation with its terminal tails interacting with a model template strand (2084–2134) coloured black (the 5'-terminus of the 'template' lies to the right). As a model for an emerging transcript, the 3'-tail of the tRNA molecule bound in the A-site has been extended upwards (darker grey).

3. A MOLECULAR MODEL

(a) *Ribozyme-based models*

In the conceptual models developed earlier, there is little molecular detail and almost nothing that gives any indication of the structural nature of the replicase.

This is, of course, not unexpected as it is difficult to imagine any source that could provide information at this level of detail. The best that might be done is to follow the Hutton–Lyell principle of uniformitarianism that the processes we see in action today were also

Figure 7. The catalytic core of the ribosome. (a) Part of domain V from the LSU of *Escherichia coli* (2045–2627) and yeast (1729–2397) are shown as superposed two-dimensional secondary structure plots (based on Wuyts *et al.* 2001). The structures were superposed and coloured with their degree of similarity (Red $\rightarrow$ blue = high $\rightarrow$ low conservation). The secondary structure helices branching off the central circle are, from left to right, G2 (G15), G16, G17 (G18,G19), G20 and G1 (those in parentheses are not directly attached to the circle). (b) The structure/sequence conservation values for the *E. coli* molecule have been mapped onto the known three-dimensional structure (PDB code: 1fg0). A tRNA fragment bound in the active site can be seen in atomic ball-and-stick representation between the two hairpins (G16 and G17). The orientation is close to that in part a. Both the parts were drawn with the program RASMOL.

Figure 8. A hole in the ribopolymerase active site. (a) The construction of an A : U base pair on position 2103 (red : yellow) in the ribosome structure (PDB code: 1fg0) does not cover the position of amino acid attached to the tRNA molecule (green). This leaves a cavity in the ribopolymerase active site model corresponding to the side chain of the tyrosine analogue found in the X-ray structure. The tyrosine side chain is oriented, β -carbon top and OH bottom, and is shown as green spheres. (b) The space can be filled with a smaller amino acid (threonine) along with its main chain atoms. If the base at position 2102 (pink) were a pyrimidine, then a larger amino acid could be accommodated. The faint blue lines are the backbone trace of the rest of the structure with the model template backbone shown in thicker orange. The position of the transcript (tRNA coordinates) is coloured cyan.

active in the past. For the players in an RNA world, this directs us to consider RNA enzymes (ribozymes) as a source of molecular models.

The known structures of ribozymes are not numerous, but among these there is a general tendency to find a reasonably globular shape consisting of a collection of a small number of stem-loops (short segments of double helix). For a primitive replicase ribozyme, we also have the additional constraint that at sometime in its past it would have faced the problem of poor replication fidelity. Given that it must also replicate itself, this means that it is constrained in length to be less than 200 bases.

The hammerhead ribozyme, with three short-stem loops, provides a typical example of a ribozyme of this

type on the smaller end of the size range (49 bases). The hairpin ribozyme (71) or the hepatitis delta ribozyme (111) (Lilley 2003) are more useful size, but still rather small to marshal the necessary template and transcript components. The larger group-I self-splicing intron ribozyme (Protein Data Bank structure code: 1grz), with 246 bases, exceeds our size limit, but a core substructure can always be extracted from larger molecules.

The substrate for most ribozymes is RNA, which fits well with using them as a source of replicase models. However, most of them act upon themselves as a part of the maturation process, such as the self-splicing introns. By contrast, the function of a replicase requires it to bind both a template and a transcript along with a

Figure 9. Interfering amino-acyl RNA by-products. (a) An amino acid (black circle) 'invades' the site, becomes 'charged' by the NTP and links to the transcript. Transcription is terminated resulting in aaRNA by-products (detached '!' shapes). (b) A dimer in which the first subunit has lost its transcript from the active site but this is still held by the second subunit. The 3' tail of this has base paired (grey dots) with a visiting RNA allowing its 3' tail to reach the vacated active site of the first subunit.

Figure 10. A 'cross-linked' ribopolymerase dimer. In the dimeric model of the ribopolymerase, a tRNA-like molecule (white) can be bound in the P-site and make base-pair interactions with the transcript (light blue) as it emerges from the second ribopolymerase copy (yellow). (a) The anticodon bases from the tRNA are paired with the three magenta spheres, which are the codon bases of mRNA bound in the ribosome small subunit. These have been superposed onto the emerging template. Both the sets of coordinates were taken from the structure of the ribosomal small subunit (PDB code: 1gix). The tRNA molecule has been 'simplified' by the removal of the short hairpins (the right and the left parts of the clover-leaf in the standard secondary structure representation), so that it corresponds better to the single hairpin aaRNA discussed in the text. (b) The corresponding position of the tRNA in the large subunit is shown in dark blue. The parts are a stereo pair.

nucleotide monomer. None of the known ribozyme structures provides anything like this degree of complexity, but if a search is extended to include nucleo-protein particles, then a suitable structure can be found in the ribosome (figure 6).

(b) A ribosome-based model

Compared to the smaller ribozymes considered earlier, the ribosome is enormous. My motivation for even considering it came not from its overall structure, but from an examination of the active site of peptidyl

transfer, which lies in the core of the LSU. The active site does not involve any protein component (except the chain being synthesized); therefore, from our viewpoint, the ribosome can be viewed as a ribozyme (Nissen *et al.* 2000; Lilley 2003).

At the catalytic core of the ribosome, three polymers are brought together: two tRNA molecules and one growing polypeptide chain. The tRNA molecules carry amino acids attached to their 3' tails through a phosphate linkage, which is the same as the linkage used between nucleotide bases. By neglecting the

amino acid group, this makes it easy to view the aminoacyl-tRNA (aa-tRNA) simply as an unmodified polynucleotide chain. By replacing the amino acid by a nucleotide (extending the tRNA chain by one base), the ends of the tRNA molecules extend towards the base of the ribosome active site, where there are unpaired bases with which they can pair through normal Watson–Crick interactions (Taylor 2004).

Picturing the two tRNA molecules as a template and a transcript in a replicase model does not suggest any copying mechanism, but their base pairing with the strand at the bottom of the site places two nucleotides from different sources side-by-side—exactly what is required in a replicase. Just as the amino acids become linked, so their nucleotide equivalents might also link. Clearly, a nucleotide does not need a polynucleotide to deliver it to the active site and so one tRNA can be neglected, keeping only its terminus as a nucleotide triphosphate (NTP) position. The space occupied by the lost tRNA conveniently leaves a tunnel through which the NTP could approach the active site.

This arrangement forces the strand at the base of the active site into the role of template. The structural model can accommodate this requirement without difficulty as the strand at the base of the active site is sequentially separated from the part of the LSU chain that constitutes the rest of the LSU core (figure 7*a*). If this region is taken as a replicase model, it incorporates *ca* 180 bases in the form of two main stem-loops followed by a short third one. Each terminus of the segment also partly base pairs with the template strand, which it encircles completely (figure 7*b*).

Whether by fortune or because it is a relic of an ancient ribozyme, the core of the ribosome provides a framework that lets us visualize a possible structure for the replicase discussed in the opening sections. In addition, the transcript (tRNA) is also in the orientation described earlier for the synthesis of a parallel complement.

4. FROM REPLICASE TO RIBOSOME

It is tempting to speculate that the modelling of a dimeric replicase from a ribosome described earlier might have been the path followed by nature in reverse. However, just as the amino acid components were discarded in the modelling process, there must be some reason to justify their reintroduction in the path towards the ribosome. It does not seem reasonable to suppose that aminoacylated RNA molecules spontaneously arose and carried their cargo to an active site for peptide synthesis.

(a) *Amino acids as replicase poison*

In the modelling of the replicase active site, the extension of the tRNA molecule by an additional nucleotide did not completely fill the space originally occupied by the attached amino acid. Indeed, it can be argued that this space must be maintained to allow translocation to occur (Taylor 2004). If a free amino acid were by chance to enter this cavity, it would lie adjacent both to the 3' tail of the RNA transcript

and, at some point in the polymerization cycle, to a NTP (figure 8).

This juxtaposition of components is similar to that found in the modern (protein) tRNA synthases and it does not seem unreasonable to imagine that, on some occasions, the visiting amino acid would become attached to the transcript, generating an aminoacylated RNA. Since no further nucleotides could subsequently attach to the transcript, the attachment of an amino acid would prematurely terminate transcription, leaving an incomplete aminoacylated copy, which would eventually detach from the replicase. In effect, the amino acid would have acted as a catalytic poison on the replicase.

This amino acid side reaction may not have been sufficiently productive to compromise the efficiency of the replicase, but the aa-RNA products, over time, may have built-up in the environment of the replicase. Since they would be the incomplete copies of ribozymes, these by-products would be expected to have secondary structure. Being, on an average, roughly half the size of a typical ribozyme, this would probably constitute a single-stem loop structure. Folded as a secondary structure would greatly help to reduce spontaneous hydrolysis, increasing the lifetime of the by-products, and if the replicase functioned as an extended network, their escape through diffusion would be reduced.

(b) *Interfering aminoacylated by-products*

In the form of a hairpin, the aa-RNA by-products would have a loop, like the anticodon loop in modern tRNA that would be available to hybridize with any unpaired complementary strand it might encounter. Those available in its immediate vicinity would be the unfolded segments of the templates and transcripts as they either entered or exited the replicase. Being attached to either of these leaves its amino acylated tail tethered close to the active site of the adjacent replicase subunit. Through this forced proximity, the tail might either join or displace the current transcript from the adjacent subunit (figure 9).

Without being over-contrived, we now have a situation in which an amino acid component has been introduced to a system with two subunits, one of which is binding the aminoacylated tail of a small RNA, while the other is holding a template to which the other end of the aa-RNA is attached through a limited extent of base pairing. In structural terms, only one further step is needed to imagine how such a model might be developed towards a proto-ribosome. This would require that a second aa-RNA bind in the active site, bringing the two amino acids together in the active site, where spontaneous polymerization would occur, as in the modern ribosome.

It is not unexpected that this proto-ribosome model is structurally consistent with the modern ribosome, since its development has simply reversed the modelling of the replicase from the ribosome. However, it does not embody the essential feature of the ribosome since the model contains no message (encoding a protein sequence) and no genetic code through which it might be translated. At this point, these aspects might be left to chance; following Crick, the genetic code could arise as a series of accidents and the protein encoding message as segments of RNA template

Figure 11. Structure cartoons. (a) A visiting aaRNA base pairs (grey dots) to the template emerging from the second subunit and blocks the active site of the first subunit. In this blocked state, the first subunit can lose its transcript. (b) As transcription still proceeds in the second subunit, translocation of the template (in the direction of the arrow) makes space for a second aaRNA to bind and also access the active site of the first subunit. The first aaRNA (in the P-site) cannot escape until peptidyl transfer attaches the growing polypeptide chain (string of beads) to the new amino acid in the A-site. As the two halves of the dimer diversify, the template in the lower half can become inert, while the template for the second half can become any available RNA message.

(passing through the replicase subunit) that would give rise to useful protein products.

It seems unsatisfactory to leave the proto-ribosome model to evolve through a series of accidents without speculating whether there might be any mechanism or selective pressure to favour such a development. The following section will consider this further.

5. TOWARDS THE RIBOSOME

(a) *A hybrid replicase/ribosome*

The current model proposes a ribopolymerase functioning in its core form in a dimeric state with closely linked cycles of template and transcript. If this is disrupted by amino acids, as postulated earlier, we could have one or more aaRNA molecules cross-linking the dimer from template to active site (figure 10). In this form, one half of the dimer will be inactive, but the second subunit could still function and transcribe its message. Were this to happen, then any visiting aaRNA attached to the template would be translocated as transcription progressed.

What happens next depends on a balance of forces that are difficult to quantify, but if it is assumed that the replicase subunits themselves are fixed in an extended network, then one of the following could happen:

- (i) the template 'piles-up' at the exit site;
- (ii) the aaRNA 'codon' pairing to the template is broken;
- (iii) the amino acylated end of the aaRNA is pulled out of the active site and
- (iv) translocation is halted in the active subunit.

I would argue that option (i) is unlikely if the template 5' terminus has already become attached to another replicase in the network (or is actively folding). The strength of the link considered in (ii) depends on the degree of base pairing involved, but as the aaRNA 'codon' loop cannot entangle the transcript, not more than half a turn of base pairing would be possible giving, at most, between 3 and 5 bp. This is quite a strong interaction relative to the energy of NTP hydrolysis, which makes it more likely that the other end should detach first (option (iii)), pulling the attached amino acid out of the inactive site. However, if the amino acids had already polymerized, then extraction would become increasingly difficult with longer polypeptide chains. I will therefore continue to develop the model in which option (iv) happens first, not only for the reasons advanced earlier, but also as it is the only route that allows some further deductions.

The point at which the replicase stalls, the aaRNA (possibly attached to a growing polypeptide) will be in a position corresponding to the 'P' site of the modern ribosome (the NTP entry channel in the current replicase model). This leaves the modern 'A' site free to receive a second aaRNA molecule. As amino acid polymerization is not energetically driven, it will be able to occur just as in the modern ribosome. The consequence of this is that the first aaRNA is now released from its tether, and translocation in the active subunit can progress a few steps more, pulling the second aaRNA from the 'A' site to the 'P' site until translocation stalls again (figure 11b).

In this way, a regular ratchet mechanism has been established which advances only when two aaRNAs are present in the (in)active site. The active subunit provides the motive force, avoiding the need to postulate

a complex protein elongation mechanism before any of its components could reasonably be expected to exist. As mentioned earlier, the aaRNA 'codon' interaction cannot be extensive (for steric reasons) and is likely to be less than five bases (under half a turn). It is not unreasonable to suppose that three bases might have been an optimal step-size, thus laying a structural basis for the size of the modern codon.

6. CONCLUSIONS

The arguments presented earlier have developed a model from a simple logical constraint on replication through to quite a detailed molecular model. The reason for proposing such a level of detail is not that it makes the model inherently any more testable than the conceptual model, but that it provides a visualization for those, like myself, who think in terms of molecular mechanics. It might then be argued that if a credible mechanism can be proposed, then even if it is wrong, it shows that something similar could have occurred. Given that nature probably spent 499 999 995 more years on the problem than I have, one would hope that something more complete and efficient could have been 'devised'. As stated at the beginning of this paper, because we are here, we know this did indeed occur.

(a) *No genetic code*

The final model incorporates the synthesis of a polypeptide chain regulated by tRNA-like molecules that impose a ratchet mechanism in steps of three bases along a template. If some of the, initially random, protein sequences provided an advantage to the metabolic system that contains the sequence of the ribozyme from which they were copied, then it is possible to imagine selection becoming active, but this can only occur after some link has been established between the 'anticodon' end of the tRNA-like molecules and the amino acid they carry.

Unfortunately, I see no detailed mechanism that could provide this link other than the broad similarity of the two replicase subunits to the two binding sites found in modern (protein) aa-tRNA synthases. Some basis for a link might be imagined in the structure of these binding sites, but this must await a more detailed consideration (Taylor in press).

(b) *Ribodammerung*

A final consequence of the adoption of a parallel-stranded complement in a replication mechanism is that such a system contains seeds of its own downfall. If any competing system were able to synthesize a reverse complement, then the problem of double-strand stability would return and functional ribozymes would become lost as double-stranded molecules. Collapse would be expected to be rapid as the ribopolymerases themselves would be removed in the process. It is unlikely that an RNA-based system would be able to create competition of this kind, as it would also destroy itself in the process. However, if any of the semi-random polypeptide by-products ever found the trick of making a reverse complement, then nothing in the RNA world could hold them back. The sequences of the ribozymes

would become double-stranded genomes, controlled and replicated by their new protein masters.

REFERENCES

- Bartel, D. P. 1999 Re-creating an RNA replicase. In *The RNA world: the nature of modern RNA suggests a prebiotic RNA world* (ed. R. F. Gesteland, T. R. Cech & J. A. Atkins), ch. 5, 2nd edn. New York, NY: Cold Spring Harbor Laboratory.
- Benner, S. A. & Ellington, A. D. 1991 RNA world. *Science* **252**, 1232.
- Benner, S. A., Cohen, M. A., Gonnet, G. H., Berkowitz, D. B. & Johnsson, K. P. 1993 Reading the palimpsest: contemporary biochemical data and the RNA world. In *The RNA world: the nature of modern RNA suggests a prebiotic RNA world* (ed. R. F. Gesteland & J. A. Atkins), pp. 27–70. New York, NY: Cold Spring Harbor Laboratory.
- Gesteland, R. F., Cech, T. R. & Atkins, J. A. (eds) 1999 *The RNA world: the nature of modern RNA suggests a prebiotic RNA world*. 2nd edn. New York, NY: Cold Spring Harbor Laboratory.
- Gilbert, W. 1986 Origin of life: the RNA world. *Nature* **319**, 618. (doi:10.1038/319618a0)
- Hofacker, L. 2003 Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431. (doi:10.1093/nar/gkg599)
- Joyce, G. F. 1989 RNA evolution the origins of life. *Nature* **338**, 217–224. (doi:10.1038/338217a0)
- Joyce, G. F. & Orgel, L. E. 1999 Prospects for understanding the origin of the RNA world. In *The RNA world: the nature of modern RNA suggests a prebiotic RNA world* (ed. R. F. Gesteland, T. R. Cech & J. A. Atkins), ch. 2, 2nd edn. New York, NY: Cold Spring Harbor Laboratory.
- Lilley, D. M. J. 2003 The origins of RNA catalysis in ribozymes. *TIBS* **28**, 495–501.
- Maynard Smith, J. & Szathmáry, E. 1995 *The major transitions in evolution*. Oxford, UK: Oxford University press.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. 2000 The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930. (doi:10.1126/science.289.5481.920)
- Parvathy, V. R., Bhaumik, S. R., Chary, K. V. R., Govil, G., Liu, K., Howard, F. B. & Miles, H. T. 2002 NMR structure of a parallel-stranded DNA duplex at atomic resolution. *Nucleic Acids Res.* **30**, 1500–1511. (doi:10.1093/nar/30.7.1500)
- Rippe, K. & Jovin, T. M. 1992 Parallel-stranded duplex DNA. *Methods Enzymol.* **211**, 199–220.
- Taylor, W. R. 2004 A molecular model for transcription in the RNA world based on the ribosome large subunit. *Comp. Biol. Chem.* **28**, 313–319. (doi:10.1016/j.compbiolchem.2004.09.008)
- Taylor, W. R. 2005a Stirring the primordial soup. *Nature* **434**, 705. (doi:10.1038/434705a)
- Taylor, W. R. 2005b Modelling molecular stability in the RNA world. *Comput. Biol. Chem.* **29**, 259–272. (doi:10.1016/j.compbiolchem.2005.04.009)
- Taylor, W. R. In press. A molecular model for the origin of translation in an RNA world. *J. Theor. Biol.*
- Wuyts, J., Van de Peer, Y. & De Wachter, R. 2001 Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. *Nucleic Acid Res.* **29**, 5017–5028.
- Zuker, M., Matthews, D. H. & Turner, D. H. 1998 Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. Technical report, Institute for Biomedical Computing, Washington University, St. Louis, USA, 1998. User guide for the MFOLD program version 3.1.