

# Hypothesis: Emergence of Translation as a Result of RNA Helicase Evolution

Nikolay Zenkin

Received: 4 December 2011 / Accepted: 13 April 2012 / Published online: 28 April 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** The origin of translation and the genetic code is one of the major mysteries of evolution. The advantage of templated protein synthesis could have been achieved only when the translation apparatus had already become very complex. This means that the translation machinery, as we know it today, must have evolved towards some different essential function that subsequently sub-functionalised into templated protein synthesis. The hypothesis presented here proposes that translation originated as the result of evolution of a primordial RNA helicase, which has been essential for preventing dying out of the RNA organism in sterile double-stranded form. This hypothesis emerges because modern ribosome possesses RNA helicase activity that likely dates back to the RNA world. I hypothesise that codon–anticodon interactions of tRNAs with mRNA evolved as a mechanism used by RNA helicase, the predecessor of ribosomes, to melt RNA duplexes. In this scenario, peptide bond formation emerged to drive unidirectional movement of the helicase via a molecular ratchet mechanism powered by Brownian motion. I propose that protein synthesis appeared as a side product of helicase activity. The first templates for protein synthesis were functional RNAs (ribozymes) that were unwound by the helicase, and the first synthesised proteins were of random or non-sense sequence. I further suggest that genetic code emerged to avoid this randomness. The initial genetic code thus emerged as an assignment of amino acids to codons

according to the sequences of the pre-existing RNAs to take advantage of the side products of RNA helicase function.

**Keywords** Evolution of ribosome · tRNA · RNA world · Ribozyme · Evolution

## Introduction

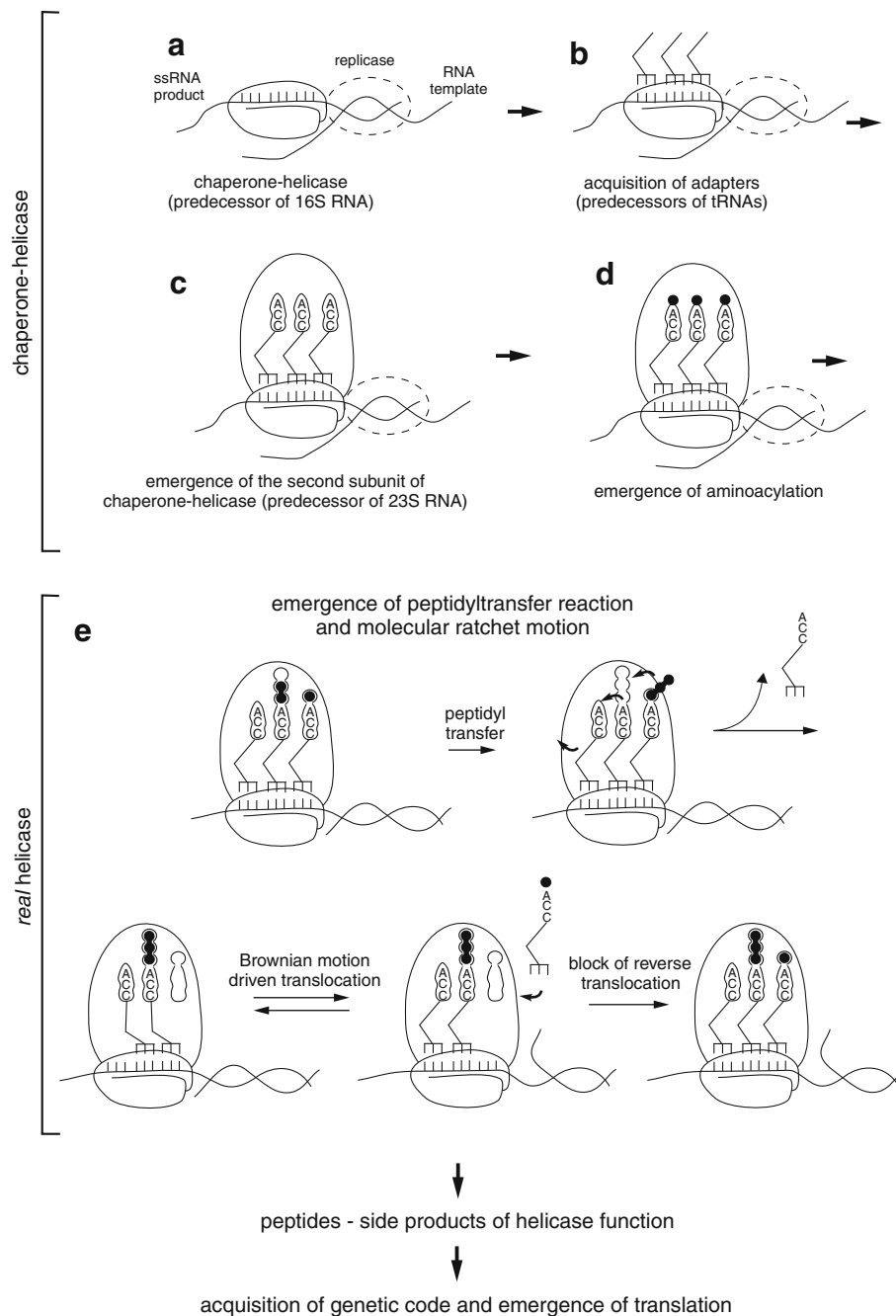
The concepts of RNA and RNA–protein worlds that preceded the modern DNA–RNA–protein world are widely accepted, and they find more and more proof with advances in biochemistry and bioinformatics (Gesteland et al. 2006). The transition between RNA and RNA–protein worlds took place when templated protein synthesis (translation) emerged. The translation machinery we know today is extremely complex. Importantly, even today, all the basic functions of the ribosome are accomplished by RNAs, 16S and 23S (bacterial nomenclature will be used throughout) ribosomal RNAs (rRNAs) and tRNAs (reviewed in Moore and Steitz 2006; Moore and Steitz 2010), meaning that the ribosome is a relic from the RNA world (Noller 2006; Poole et al. 1998; Wolf and Koonin 2007).

The function of the ribosome facilitated by 16S and 23S rRNAs and tRNAs can be divided in several steps (Fig. 1e): (i) Peptidyl transfer results in energetically favourable formation of a peptide amide bond from an activated ester bond. This reaction leads to a modification of tRNAs bound to the ribosome: aminoacyl-tRNA becomes peptidylated, while peptidyl-tRNA becomes completely deacylated; (ii) The A, P and E sites of the 23S RNA have affinities to aminoacylated, peptidylated and discharged tRNAs, respectively. Therefore, the peptidyl transfer is followed by the energetically downhill unidirectional movement of tRNAs' acceptor

---

The hypothesis is dedicated to the memory of Dr Denis Izyumov.

N. Zenkin (✉)  
Centre for Bacterial Cell Biology, Institute for Cell and  
Molecular Biosciences, Newcastle University, Baddiley-Clark  
Building, Richardson Road, Newcastle upon Tyne NE2 4AX,  
UK  
e-mail: n.zenkin@ncl.ac.uk



**Fig. 1** Schematic representation of the proposed evolution of the RNA helicase that led to the emergence of translation (see text for details). **a** In one of the possible scenarios, chaperone-helicase was dragged by ribozyme-replicase (*dashed oval*), thus keeping the nascent RNA chain temporarily single-stranded, preventing its annealing to the template. The initial chaperone-helicase interacted with substrate single-stranded RNA non-specifically, and shifted the equilibrium between double-stranded and single-stranded forms to the latter one. According to the hypothesis, this chaperone-helicase was the predecessor of the 3'm domain of the 16S RNA. **b** The chaperone-helicase acquired adapters (the predecessors of tRNAs) to tighten transient interactions of chaperone-helicase with substrate RNA. **c** The second subunit of the chaperone-helicase (predecessor of modern 23S RNA) possibly emerged to increase specificity of

chaperone-helicase to adapters. **d** Emergence of adapter's aminoacylation may or may not be directly connected to evolution of the helicase (see text). As a possibility, it could have emerged to increase specificity of adapters binding to the second subunit of the chaperone-helicase. **e** Emergence of transpeptidation as a mechanism ensuring unidirectional movement of the helicase via molecular ratchet mechanism, and establishment of translation cycle seen in the modern ribosomes. Upon peptide bond formation, acceptor stems of adapters experience a favourable shift (*arrows*), while thermal motion shifts pre-anticodons with the substrate RNA. Backward shift of pre-anticodon stems is blocked by binding of a new aminoacylated adapter in the vacant A-site. From this stage, the helicase did not require assistance from the replicase for its movement, and could move and melt RNA secondary structures on its own

stems through the ribosome; (iii) Thermal motion then can drag the mRNA attached to the anticodon stems of tRNAs thus completing the act of translocation; and (iv) Backward movement of anticodon stems and mRNA is blocked by the newly emerging aminoacyl-tRNA which binds to mRNA in the vacant A site, which is followed by repetition of the cycle.

As observed from the above, in the absence of factors, the energy source for translocation comes from Brownian thermal motion, while unidirectionality is ensured by chemical modification of tRNAs during peptidyl transfer reaction (Noller 2006; Spirin 1985, 2009; Woese 1970). This mode of translocation, known as ‘molecular ratchet’, can be imagined as ‘directed diffusion’ and is utilised by many other unrelated molecular machines of the modern cell, such as, for example, multi-subunit RNA polymerase (Abbondanzieri et al. 2005). Importantly, factor-independent (in the absence of EF-G and GTP) molecular ratchet translocation by the modern ribosome has been observed experimentally, supporting the idea that translocation is an intrinsic property of the ribosome: (i) Under certain in vitro conditions or after chemical modification of ribosomes, poly-U-dependent synthesis of polyphenylalanine can proceed in the absence of EF-G (Gavrilova et al. 1976; Pestka 1968). Possibility of ‘slippage’ of the ribosome on poly-U template was addressed by means of mRNA of a defined sequence (Southworth et al. 2002); (ii) Peptidyl transferase inhibitor Sparsomycin can efficiently trigger one round of translocation in the complete absence of EF-G and GTP (Fredrick and Noller 2002).

The ribosome brings together completely unrelated activities: (i) the small subunit 16S RNA binds mRNA, and the anticodon hairpins of tRNAs that are base paired with this mRNA; and (ii) the large subunit 23S RNA binds the opposite ends, the acceptor stems, of these tRNAs and performs transpeptidation. It is important that the large subunit is ‘blind’ towards the sequence of the mRNA that is read by the small subunit, indicating that the function of either subunit does not make sense for *templated* protein synthesis without the function of the other. Evolution of such complex RNA machinery is unclear: the advantage of *templated* protein synthesis could only have been achieved when the translation apparatus had already become very complex (Noller 2006; Poole et al. 1998; Wolf and Koonin 2007). In other words, the multistep processes mentioned above (Fig. 1e) must be established before encoded protein synthesis can function. This implies that the intermediate evolutionary steps towards translation would not have had any evolutionary advantage if these were solely made towards *templated* protein synthesis. Therefore, *templated* synthesis of proteins could not have evolved directly from primitive ribozyme activity(s): evolution does not have foresight. This means that the translation machinery, as we know it today, must have evolved towards some different

function (pre-translation) in the RNA world that subsequently sub-functionalised into *templated* protein synthesis (modern translation). The genetic code must have been a late addition to the pre-translation apparatus, and served as a trigger for the switch from pre-translation to modern translation. Note that impossibility of evolution of translation from a simple ribozyme does not deny that some individual functions of translation, such as *template-independent* peptide synthesis (peptidyl transferase), could have been advantageous to, and could have existed in the primordial RNA organism (Wolf and Koonin 2007, and references therein)

In an attempt to deduce the function of the pre-translation system, one must take into account that it (i) should have been essential for primordial RNA systems, even at very early stages, not to be eliminated by selection; (ii) must have evolved from a primitive system by small steps, each conferring a distinct selective advantage for the organism; (iii) led to the acquisition of basic functions that are characteristics of modern translation: coordinated binding of template RNA and tRNAs, peptide bond synthesis, and translocation; and (iv) was prone to acquire the genetic code, which led to sub-functionalization into modern translation.

A hypothesis providing a coherent scenario for emergence of translation was proposed by Poole and co-authors (Poole et al. 1998). They suggested that the predecessor of the modern ribosome was an RNA replicase, implying an activity which was clearly essential for the life of the RNA organism. According to this hypothesis, this replicase synthesised a new strand of RNA on an RNA template (the predecessor of mRNA) using trinucleotides that were cleaved from longer RNAs (predecessors of tRNAs) and ligated in a template-dependent manner. The emergence of aminoacylation of tRNAs’ predecessors led to an increased specificity of their binding by the replicase, and to an increased fidelity of replication. Formation of the peptide bond evolved to be utilised in the molecular ratchet movement of the replicase. There are, however, several aspects, which this hypothesis cannot explain. For example, remnants of activities of the intermediate evolutionary forms are expected to be found in the modern translation machinery. However, the tRNA cleavage and triplet ligation activities of the small subunit are clearly absent from modern ribosomes. It is also likely that the mechanism of regenerating cleaved tRNAs would remain in some form in today’s mechanisms of tRNA maturation. The acquisition of peptide bond formation as a driving force for molecular ratchet motion does not seem to have been under strong evolutionary pressure, given that the replicase already possessed the molecular ratchet mechanism of movement which was driven by RNA synthesis itself (as in modern multi-subunit RNA polymerases (Abbondanzieri et al.

2005)). I present an alternative hypothesis in the next section, which apparently resolves the above difficulties.

## Hypothesis

There are two observations that provide a basis for the hypothesis discussed below:

(1) One of the important properties of the translation machinery is its ability to melt almost all structures in an RNA molecule. This helicase property should be essential for the ribosome given that it functions on various RNAs that may possess very extensive secondary or tertiary structures. Helicase activity involves all RNA parts of the translation machinery and is independent of translation factors (Takayar et al. 2005). Though ribosomal proteins may assist helicase activity (Takayar et al. 2005), it seems to be a property of the rRNAs and tRNAs: 16S rRNA provides a channel for the substrate RNA binding; tRNAs interact with substrate RNA and together with 23S rRNA determine the molecular ratchet for ribosome movement. This suggests that the helicase function of the ribosome is its ancient property and most likely dates back to the RNA world.

(2) It was suggested that the self-replicating RNA systems of the RNA world required an RNA helicase. The requirement for helicase is explained by the fact that ribozymes act as single-stranded units, while their replication may result in extensive double-stranded helices which would need to be melted to produce single-stranded species (Taylor 2005a, b). The equilibrium between single-stranded and double-stranded species is always shifted towards the latter forms. If not melted, such an RNA system would die out in an inactive double-stranded form. Thermal denaturation is widely suggested to be the way of melting of extended duplexes in the RNA world. Lowering of temperature after denaturation is required to allow folding of the replicase, which would also have been melted at high temperature. It is not clear, however, how to get the separated strands far enough apart to prevent self-annealing at lower temperature (Kovac et al. 2003). Getting strands too far apart would also be detrimental because this would diminish (or eliminate, in the case of the replicase gene) chances of one of them to be further replicated. It has been suggested that ligands and/or cofactors of ribozymes could facilitate strand separation by binding and the shifting equilibrium from double-stranded to single-stranded form ('Replicator induction by metabolite' hypothesis) (Wong 2009). This mechanism also ensures that RNAs without aptameric functions (non-catalytic RNAs without ligands) would be eliminated through the lack of further replication. However, as the RNA organism became more developed, separation of longer duplexes likely required invention of more sophisticated mechanisms.

Synthesis of a complementary strand is a thermodynamically favourable process since the free energy of the system decreases. However, in order to start a new round of copying, the low-free-energy double helix needs to be unwound, and this requires energy input. Besides involvement of a dedicated helicase, formation of double-stranded product during replication can be avoided through several mechanisms: (i) Ribozyme-replicase could act on a single-stranded template and facilitate disengagement of the product RNA from the template by itself, like the modern  $Q\beta$  replicase.  $Q\beta$  replicase uses a power-stroke mechanism of translocation thus forcing the rear edge of the nascent RNA–RNA duplex against a wedge formed by its subunits, which leads to disengagement of product RNA from the template. It is however unlikely that a power-stroke mechanism could have emerged early in evolution of the RNA organism, since it involves complex structural rearrangements coupled to the energy of NTP hydrolysis. (ii) Ribozyme-replicase most likely used molecular ratchet mechanism of translocation and theoretically could function on double-stranded template in a manner used by modern multi-subunit RNA polymerase. Multi-subunit RNA polymerase, which also uses molecular ratchet translocation, forms an elongation complex, which consists of a melted region (bubble) in the double-stranded template where synthesis takes place. In such a complex, the non-template strand that anneals to the template strand behind the transcription bubble helps RNA polymerase to displace the product strand from the template strand. On single-stranded templates, in the absence of the non-template strand, RNA polymerase finds it difficult to melt the product–template duplex (Zenkin et al. 2006). Synthesis on double-stranded template, however, does not seem to be likely in the primordial RNA organism since it would have required a large and complex ribozyme-replicase that could support the structure of the elongation complex. (iii) Replication could proceed via a strand displacement mechanism, when the non-template strand of double-stranded template is released in the single-stranded form. In this case, however, the replicase would likely require an RNA helicase to unwind a double-stranded template in front of it, much similar to the modern DNA polymerases that function on double-stranded DNA genomes and require helicases for their functioning.

Taken together, it seems reasonable to suggest that the RNA replicase should have been accompanied by a helicase activity, which would either follow the replicase and melt the nascent RNA–RNA duplex, or move in front of the replicase to provide it with a single-stranded template (Fig. 1a–d). Furthermore, as in the modern cell, the problem of RNA misfolding is solved by proteinaceous RNA chaperones and helicases (Rajkowitsch et al. 2007): a ribozyme that would accomplish this function in the

primordial RNA organism seems to have been highly advantageous.

That the RNA helicase activity of the modern translation apparatus likely dates back to the RNA world, and that an RNA helicase was required in the RNA world, may have the following important implication. Taking into account that the ribosome could not have evolved from a primitive ribozyme towards templated protein synthesis, the predecessor of the modern translation machinery could have been an RNA helicase. This hypothesis implies that the increasing demand to unwind RNA in the growing complexity of the RNA organism was the driving force of evolution that has led to the modern translation machinery. The substrate RNA that was unwound by the helicase later became mRNA of modern translation. The hypothesis suggests that helicase evolution culminated in the acquisition of a molecular ratchet motion, which was directed by the chemical modification of predecessors of tRNAs during peptide-bond formation. Accordingly, protein synthesis first emerged as a side product of helicase activity, while the subsequent recruitment of the genetic code has led to the emergence of modern translation.

I suggest in the following a stepwise scenario that comprises a series of small advantageous changes that first served to improve functions of an RNA helicase and eventually led to the emergence of translation. The evolutionary intermediates in this scenario are proposed based on the structure and the functions of the modern translation apparatus.

(i) It is reasonable to suggest that the initial helicase function was accomplished by a simple molecule that acted as a chaperone rather than an actual helicase to melt stable duplexes. I suggest that this initial chaperone-helicase that destabilised RNA duplexes by transiently interacting with them, was an ancestor of the 16S rRNA of the modern ribosome (Fig. 1a). This idea is based on the structure of the modern 16S rRNA. A small highly conserved domain (3′m) of 16S rRNA forms a channel where mRNA is bound in a conformation that is incompatible with the formation of a double helix (Weixlbaumer et al. 2008; Yusupov et al. 2001; Yusupova et al. 2001). Curiously, the small subunit of eukaryotic ribosome is known to have the ability to melt (to some extent) RNA secondary structures without help from the rest of the translation machinery (Kozak 1978; 1989). This may possibly be a remnant from its chaperone-helicase past. However, it is to be noted that sufficiently stable structures in RNA cannot be melted by the small subunit alone. This implies that growing complexity of the RNA organism, which led to the increased sizes of genes/genome, required improvement of the chaperone-helicase, to facilitate melting of larger and more stable secondary structures.

(ii) The formation of a double helix is a cooperative process, which after nucleation (i.e. formation of 3–4 base

pairs) proceeds spontaneously (Craig et al. 1971; Porschke 1977; Zeiler and Simons 1998). Therefore, the selection would favour a chaperone-helicase that interacts with the single-stranded substrate tightly to prevent double helix formation, but does this transiently. I suggest that this property of the chaperone-helicase could have been achieved through utilisation of attachable ‘adapters’ (predecessors of modern tRNAs) that could base pair with substrate RNAs via short stretches of variable sequences (predecessors of anticodons, pre-anticodons) when substrate RNAs were bound by chaperone-helicase (Fig. 1b). While binding of adapters to the substrate RNA was transient (due to only short complementary stretches), a cooperative binding of several adapters possibly provided enough energy to counteract the annealing of a substrate RNA to its complement. Short sequences of pre-anticodons also allowed only a few variants of adapters to interact with a substrate RNA of any sequence. The existence of this step in the evolution of ribosomes is compatible with biochemical findings showing that anticodon stem-loops of modern tRNAs can specifically interact with the small ribosomal subunit, recognise codons on the bound mRNA and translocate in the absence of the rest of the tRNA molecule or the large ribosomal subunit (Joseph and Noller 1998; Ogle et al. 2001). It is possible that the lengths of the modern codon and the ribosome movement step were established, long before emergence of translation, as properties of the RNA chaperone-helicase.

(iii) I suggest that the second subunit of the chaperone-helicase, the predecessor of the 23S rRNA, has emerged to increase the specificity of binding of adapters to the chaperone-helicase. The second subunit possibly recognised some common feature of the adapters, thus, increasing local concentration of adapters and tightening their cooperative binding (Fig. 1c). This idea is compatible with the following observation. The peptidyl-transferase centre (PTC), the most ancient domain of the 23S rRNA (Bokov and Steinberg 2009), is functionally asymmetrical in the modern ribosome: A and P sites of the PTC have affinities to aminoacylated and peptidylated 3′ CCA ends of tRNAs, respectively. However, A and P domains are structurally very similar to each other, and have been formed as a result of an ancient duplication, thus, building a symmetrical PTC (Bokov and Steinberg 2009). This suggests that, at some stage in evolution, A and P domains of PTC may have recognised identical 3′ ends of ancient tRNAs (adapters). The common feature of adapters that was recognised by the second subunit of chaperone-helicase could have been some sequence, possibly the CCA trinucleotide, which is absolutely conserved in all modern tRNAs (Fig. 1c).

Note that the activity of this chaperone-helicase was still *chaperone-like*. This means that chaperone-helicase could not processively melt stable double-stranded structures.



The processive movement by chaperone-helicase could have been assisted by the replicase (Fig. 1a–c). In this scenario, the chaperone-helicase was a part of the ‘replisome’ or a domain of the replicase, and served to stabilise single-strandedness of the template or the nascent transcript near the replicase.

Along with the growing complexity of the RNA organism grew the length of RNAs to be unwound, thus increasing the demands to the chaperone-helicase. The following steps of the proposed evolutionary scenario were driven by a need for a *real* helicase (rather than a chaperone) that would actively unwind RNA duplexes.

(iv) Emergence of aminoacylation of adapters may or may not have initially been bound to the development of the helicase. Few reasons have been suggested for the emergence of aminoacylation of the 3' ends of adapters. It might have emerged to increase specificity of binding to the predecessor of ribosome (Poole et al. 1998). It is also possible that aminoacylation has emerged to protect adapters from 3' end exonucleases, or as a result of misincorporation of amino acids instead of NMPs at the ends of the RNA chains during replication by RNA replicase (Taylor 2006). ‘Coding coenzyme handles’ hypothesis suggested that aminoacylated adapters acted as coenzymes of ribozymes before emergence of translation (Szathmari 1993). A somewhat similar hypothesis proposed that adapters served as handles to hold metabolites, including amino acids, for coordinated biosynthesis and metabolism (Gibson and Lamond 1990). Though the present hypothesis does not necessitate a driving force for emergence of aminoacylation, it is reasonable to suggest that aminoacylation of tRNAs preceded ribosome-dependent peptide synthesis. Ribozymes catalysing the activation of amino acids and aminoacylation were obtained by selection *in vitro*, suggesting that such a process could have occurred in the RNA world (Illangasekare and Yarus 1999).

Irrespective of the reasons behind its emergence, aminoacylation of adapters should have made a breakthrough in evolution of the chaperone-helicase. The peptidyl-transferase reaction might have proceeded spontaneously when aminoacylated ends of adapters were bound in close proximity on the second subunit of the chaperone-helicase. The chemical modifications of the adapters resulting from the peptide bond formation (discharging and peptidylation) could then be used to facilitate unidirectional molecular ratchet movement of the chaperone-helicase (see Introduction, and Fig. 1e). The development of binding sites on the second chaperone-helicase subunit with specific affinities towards three adapter forms (A for aminoacyl-, P for peptidyl-, and E for discharged adapters) resulted in a downhill unidirectional shift of acceptor ends of adapters upon peptide-bond formation (Fig. 1e). Thermal motion then shifted the pre-anticodon stems into the respective

binding sites on the first subunit of the chaperone-helicase. The reverse motion of pre-anticodon stems was blocked by binding of a new aminoacylated adapter in the vacant A-site (Fig. 1e). These motions led to the threading of the substrate RNA through the helicase, leading to the active unwinding of secondary structures of substrate RNA. As was suggested by (Poole et al. 1998), hydrolysis of the ester bond between amino acid and adapter in P site (the reaction used during termination of modern translation) could have been an intermediate evolutionary step, preceding emergence of peptide-bond formation, to be utilised for molecular ratchet motion.

(v) Initial aminoacylation of adapters could have been random, or some amino acids could have had preference to particular adapters (see ‘Discussion’). In any case, processive movement of the helicase along pre-existing sequences resulted in the formation of random/non-sense peptide chains as a side product. Advantages of some of these peptides can be envisaged for an RNA organism (such as chaperones) (Wolf and Koonin 2007). However, it is likely that, overall, these molecules of random sequences were mostly harmful for the primordial RNA organism. Therefore, the organism was forced to find a way of avoiding randomness in peptide production during helicase work, and make sense of those peptides, i.e. to make them useful. Randomness could not have been avoided through changing the sequences of the RNAs that were unwound by the helicase, since most of these RNAs were functional ribozymes. The only possible solution to this was to modify the process of random aminoacylation of adapters (which has already been in place). Particular amino acids should have been assigned to particular pre-anticodons of adapters *according* to the sequences of the existing ribozymes that were ‘read’ by the helicase in such a way, that non-random (less harmful and/or more useful) peptides were produced during helicase activity on these ribozymes.

The assignment of amino acids to pre-anticodons required emergence of ribozymes that aminoacylated adapters not randomly but according to their pre-anticodons. Such evolutionary task should not have been complicated given that aminoacylating ribozymes (that have already existed) could easily develop ability to ‘feel’ the anticodons of adapters by Watson–Crick base pairing. Taken together, I suggest that the first genetic code was assigned according to the existing sequences of ribozymes, which, therefore, served as first messenger RNAs.

(vi) I hypothesise that once the organism experienced advantages of the presence of proteins, most likely as structural components of the ribozymes (like ribosomal proteins of the modern ribosome), specialised RNA templates (mRNAs) started to emerge that were dedicated to the coding of proteins, and were no longer functional ribozymes. Novel amino acids were also recruited to the

process. At this stage of evolution, the genetic code as we know it today started to form.

The emergence of proteins allowed the helicase to further improve its efficiency by recruiting proteinaceous elongation factors, EF-Tu and EF-G, which additionally powered the existing molecular ratchet via GTP hydrolysis (Spirin 2002). I suggest that the helicase/ribosome has been working as a helicase until the specialised proteinaceous helicases replaced it, leaving protein synthesis as the only ribosomal function. However, as mentioned above, even now the ribosome is known to be able to melt almost any secondary structure in RNA, perhaps as a remnant of its ‘helicase past’. The helicase abilities of ribosomes are required for efficient translation of mRNAs with extensive secondary structures, and used by bacteria in the regulation of gene expression in phenomenon known as attenuation (Yanofsky 1981).

## Discussion

The hypothesis presented here proposes the emergence of translation as a result of the evolution of an RNA helicase that existed in the RNA world. This hypothesis provides a stepwise scenario for the evolution of the translation machinery that comprises a series of small advantageous changes that improved the fitness of the primordial organism, and which is consistent with Darwinian principles of evolution. The principal idea of this hypothesis is that all basic features of modern translation could have emerged far before the emergence of translation per se but to improve properties of the RNA helicase.

If the initial aminoacylation of adapters was random, then the present hypothesis can be unified with the ‘coevolution theory’ (Wong 1975), which suggests that the organisation of the genetic code lies in the biosynthetic relationship between amino acids, and has substantial evidence in its support (reviewed in (Di Giulio 2005)). According to this theory, a few (four or five) amino acids, which were coded initially, subsequently shared their codons with amino acids originating from them biosynthetically (Wong 1975). This theory, however, does not specify how initial assignment took place. The joint theory therefore would suggest that the first few amino acids were assigned to 64 pre-anticodons of adapters according to the pre-existing templates (helicase substrates) to take advantage of peptides that were a side effect of helicase function. This was the starting point for the sharing of codons with newly recruited amino acids, based on biosynthetic relations with initial ones.

‘Stereochemical theory’ (Crick 1968) suggests that aminoacylation of adapters has not been random before the emergence of peptide synthesis (reviewed in (Di Giulio

2005)). It proposes a stereochemical relationship between amino acids and corresponding codons and/or anticodons. This theory is somewhat supported by the finding that the sequences of RNA aptamers to three (not the simplest) amino acids (arginine, tyrosine and isoleucine) are enriched with codons and anticodons corresponding to these amino acids in the modern genetic code (Yarus et al. 2009). The apparent contradiction of the ‘stereochemical theory’ to the present hypothesis, which suggests assignment of amino acids to codons according to the pre-existing RNA sequences, can be resolved by the following consideration. If some of the amino acids had preference to particular adapters because of stereochemical relations with their anticodons, then this still would have resulted in the synthesis of non-sense peptides on the pre-existing sequences. In order to make sense out of these peptides, the RNA organism could recruit new amino acids, which had no stereochemical relations with adapters’ anticodons, and thus could be attached to adapters of the organism’s choice. This latter assignment would have shaped the rest of the code around the pre-assigned amino acids.

As suggested by the present hypothesis, the genetic code was formed (or shaped) according to the sequences of the ribozymes existing at that time. Therefore, the hypothesis predicts that sequences of these ribozyme should code for non-random amino acid sequences. Though most of the ribozymes were presumably lost during the course of evolution, some still do function in modern cells, and the obvious remnant from the RNA world is the ribosome itself. The hypothesis thus predicts that decoding of rRNAs (conserved sequences) using the current genetic code should result in non-random protein sequences, such as simple protein motifs (coiled coils,  $\beta'$  sheets, turns between elements, etc.)

**Acknowledgments** The author thanks Rob van Nues, Martin Embley, Vasisht Tadigotla, and Kathryn Doris for critical reading of the manuscript. This study was supported by the UK Biotechnology and Biological Sciences Research Council, and the European Research Council [ERC-2007-StG 202994-MTP].

**Conflict of interest** The author declares that he has no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* 438:460

- Bokov K, Steinberg SV (2009) A hierarchical model for evolution of 23S ribosomal RNA. *Nature* 457:977
- Craig ME, Crothers DM, Doty P (1971) Relaxation kinetics of dimer formation by self-complementary oligonucleotides. *J Mol Biol* 62:383
- Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38:367
- Di Giulio M (2005) The origin of the genetic code: theories and their relationships, a review. *Biosystems* 80:175
- Fredrick K, Noller HF (2002) Accurate translocation of mRNA by the ribosome requires a peptidyl group or its analog on the tRNA moving into the 30S P site. *Mol Cell* 9:1125
- Gavrilova LP, Kostiyashkina OE, Koteliansky VE, Rutkevitch NM, Spirin AS (1976) Factor-free (“non-enzymic”) and factor-dependent systems of translation of polyuridylic acid by *Escherichia coli* ribosomes. *J Mol Biol* 101:537
- Gesteland RF, Cech T, Atkins JF (2006) The RNA world: the nature of modern RNA suggests a prebiotic RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
- Gibson TJ, Lamond AI (1990) Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J Mol Evol* 30:7
- Illangasekare M, Yarus M (1999) Specific, rapid synthesis of Phe-RNA by RNA. *Proc Natl Acad Sci USA* 96:5470
- Joseph S, Noller HF (1998) EF-G-catalyzed translocation of anticodon stem-loop analogs of transfer RNA in the ribosome. *EMBO J* 17:3478
- Kovac L, Nosek J, Tomaska L (2003) An overlooked riddle of life’s origins: energy-dependent nucleic acid unzipping. *J Mol Evol* 57(Suppl 1):S182
- Kozak M (1978) How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* 15:1109
- Kozak M (1989) Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol* 9:5134
- Moore PB, Steitz TA (2006) The roles of RNA in the synthesis of protein. In: Gesteland RF, Cech T, Atkins JF (eds) The RNA world: the nature of modern RNA suggests a prebiotic RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 257–285
- Moore PB, Steitz TA (2010) The roles of RNA in the synthesis of protein. *Cold Spring Harb Perspect Biol*
- Noller HF (2006) Evolution of ribosomes and translation from an RNA World. In: Gesteland RF, Cech T, Atkins JF (eds) The RNA world: the nature of modern RNA suggests a prebiotic RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 287–307
- Ogle JM, Brodersen DE, Clemons WM Jr, Tarry MJ, Carter AP, Ramakrishnan V (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292:897
- Pestka S (1968) Studies on the formation of transfer ribonucleic acid-ribosome complexes. 3. The formation of peptide bonds by ribosomes in the absence of supernatant enzymes. *J Biol Chem* 243:2810
- Poole AM, Jeffares DC, Penny D (1998) The path from the RNA world. *J Mol Evol* 46:1
- Porschke D (1977) Elementary steps of base recognition and helix-coil transitions in nucleic acids. *Mol Biol Biochem Biophys* 24:191
- Rajkowitz L, Chen D, Stampfl S, Semrad K, Waldsich C, Mayer O, Jantsch MF, Konrat R, Blasi U, Schroeder R (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* 4:118
- Southworth DR, Brunelle JL, Green R (2002) EFG-independent translocation of the mRNA:tRNA complex is promoted by modification of the ribosome with thiol-specific reagents. *J Mol Biol* 324:611
- Spirin AS (1985) Ribosomal translocation: facts and models. *Prog Nucleic Acid Res Mol Biol* 32:75
- Spirin AS (2002) Ribosome as a molecular machine. *FEBS Lett* 514:2
- Spirin AS (2009) The ribosome as a conveying thermal ratchet machine. *J Biol Chem* 284:21103
- Szathmary E (1993) Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proc Natl Acad Sci USA* 90:9916
- Takyar S, Hickerson RP, Noller HF (2005) mRNA helicase activity of the ribosome. *Cell* 120:49
- Taylor WR (2005a) Modelling molecular stability in the RNA world. *Comput Biol Chem* 29:259
- Taylor WR (2005b) Stirring the primordial soup. *Nature* 434:705
- Taylor WR (2006) A molecular model for the origin of protein translation in an RNA world. *J Theor Biol* 243:393
- Weixlbaumer A, Jin H, Neubauer C, Voorhees RM, Petry S, Kelley AC, Ramakrishnan V (2008) Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* 322:953
- Woese C (1970) Molecular mechanics of translation: a reciprocating ratchet mechanism. *Nature* 226:817
- Wolf YI, Koonin EV (2007) On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol Direct* 2:14
- Wong JT (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* 72:1909
- Wong JT (2009) Chapter 1. Introduction. In: Wong JT, Lazcano A (eds) Prebiotic evolution and astrobiology. Landes Bioscience, Austin, Tex, pp 4–5
- Yanofsky C (1981) Attenuation in the control of expression of bacterial operons. *Nature* 289:751
- Yarus M, Widmann JJ, Knight R (2009) RNA-amino acid binding: a stereochemical era for the genetic code. *J Mol Evol* 69:406
- Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292:883
- Yusupova GZ, Yusupov MM, Cate JH, Noller HF (2001) The path of messenger RNA through the ribosome. *Cell* 106:233
- Zeiler BN, Simons RW (1998) Antisense RNA structure and function. In: Simons RW, Grunberg-Manago M (eds) RNA structure and function. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/New York, pp 437–464
- Zenkin N, Naryshkina T, Kuznedelov K, Severinov K (2006) The mechanism of DNA replication primer synthesis by RNA polymerase. *Nature* 439:617