Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/18950 holds various files of this Leiden University dissertation.

Author: Velthuis, Arend Jan Wouter te
Title: A biochemical portrait of the nidovirus RNA polymerases and helicase
Date: 2012-05-16

Chapter 1

A more general introduction

ABSTRACT

Ribonucleic acid (RNA) is a versatile biopolymer. It is used as one of the main carriers of information in cells and a common form of information storage among RNA viruses. But it is much more than that. RNA can also drive catalytic reactions, ferry precursor molecules for proteins through the cell, regulate gene expression or it can be employed as a tool to fight pathogens. This first chapter will outline the basics of this intriguing biological molecule, by introducing the concept of RNA virus replication and providing a narrative primer for the subsequent experimental chapters, which will pursue a more in-depth biochemical exploration of the RNA replication machinery of the human pathogen SARS-coronavirus (SARS-CoV) and its distant relative, equine arteritis virus (EAV).

"Anything that happens, happens. Anything that causes something else to happen, causes something else to happen. Anything that, in happening, causes itself to happen, happens again."

Douglas N. Adams, Mostly Harmless, 1992.

INTRODUCTION

It takes little imagination to see that this ostensible tautology refers to recurring events. And, though taken from a work of humorous science fiction, it also describes something quintessentially biological: that so-called self-replicating entities (or selfish entities) multiply themselves.

On a large scale, bacteria and eukaryotic cells (ranging from meters in nerve cells to a few micrometers in white blood cells) can indeed be said to be replicating themselves. But as they do so, they need rely upon a dedicated set of proteins such as polymerases[12] to copy their genetic makeup before they can successfully split up into two daughter cells. Even the much smaller viruses (hundreds of nanometers, or less, see BOX I), which replicate without a division stage and can assemble *de novo* from viral proteins, require protein complexes for their replication.

Depending on the virus, these replication complexes usually consist of large numbers of viral and host proteins. But even if we would strip these pathogens down to their bare essentials and look at the smallest and simplest viruses (or viroids), such as the phytopathogenic viroids and the hepatitis delta virus, we find that even their genetic material needs to resort to at least some components of the molecular facilities of the host to drive its replication [1,2]. In essence this is also true for the viroids that use the enzymatic abilities encoded in other viruses, such as in the case of sputnik [3], which invades the 'replication factories' of a significantly larger virus [4].

If we look to even smaller entities, however, we can find that on the scale of small polymers (tens of nanometers), the parental molecule can be directly responsible for the formation of its descendants and the mutations that can arise in these descendants. Interestingly, it is believed that exactly these two characteristics were also the two crucial bottlenecks in the process of evolution of life on Earth.

An information-bearing and catalytically active polymer to which these characteristics are currently ascribed - and which could have made the transition to become the first self-replicating molecule given existing *in vitro* evidence and the supportive icy conditions of the early Earth [5,6,7] -, is ribonucleic acid (RNA). The hypothesis that incorporates it as the most parsimonious precursor of the last universal common ancestor (LUCA) is called the 'RNA world' [8,9]. Of course, alternative origin of life hypotheses have been proposed and exhaustively discussed elsewhere (see, *e.g.*, [10,11,12,13,14]), but, for the illustrative

12 **Polymerase**: The term polymerase commonly refers to the enzyme or enzyme complex that copies the genomic content of an organism (*i.e.*, it uses a template). The term can, however, also be used in a more generic context, and refer to enzymes that simply make polymers from monomers in a template independent fashion. Such enzymes can make, for example, starch, cellulose, or polyphosphate.

purposes of this dissertation and the use of RNA by its viral 'protagonists', this chapter will focus on the versatility and limits of RNA in its subsequent paragraphs.

BOX I: Viruses - Viruses are organic agents incapable of reproduction without the infrastructure and metabolism of a host cell. This characteristic makes them therefore obligate symbionts. However, since their discovery in 1898 [4], the definition of viruses has often been amended due to new discoveries and should certainly not be regarded as 'evil' pathogens. Reoviruses for instance, have been shown to be crucial in the maturation of eggs of parasitoid wasps [5], whereas retroviruses have been linked to the evolution of placental mammals [8,9] and the life cycle and possibly the evolution of the chloroplast using sea slug *Elysiachlorotica*[12,13].

Presently, viruses are defined as intracellular entities with nucleic acid genomes that are i) capable of directing their own replication and ii) are not cells themselves [15]. Viral infections can be host cell-specific, but may in general range from bacteria/archea to large multicellular organisms such as higher plants or humans. As part of an effort to categorise and better understand viruses, the International Committee on the Taxonomy of Viruses (ICTV) currently recognises six groups of viruses under the classical Baltimore Classification, including four groups (group III-VI) that have RNA genomes. These four are: III, dsRNA viruses; IV, positive-strand ssRNA (+RNA) viruses; V, negative-strand ssRNA (-RNA) viruses; and VI, ssRNA reverse-transcribing viruses.

Most notable human and economically important pathogens are found among the ssRNA groups, where group IV includes severe acute respiratory syndrome coronavirus (SARS-CoV, *coronaviridae*), hepatitis C virus (HCV, *flaviviridae*), poliovirus (PV, *picornaviridae*), yellow fever virus (*flaviviridae*), chikungunya virus (*togaviridae*), and porcine reproductive and respiratory syndrome virus (PRRSV, *arteriviridae*); and group V is known for lethal pathogens such as influenza A virus (*orthomyxoviridae*), ebola virus (*filoviridae*), hantavirus (*bunyaviridae*), measles virus (*paramyxoviridae*), and lassa virus (*arenaviridae*). The best-described member of group VI is the human immunodeficiency virus (HIV, *lentiviridae*). Interestingly, this last group, together with some dsRNA viruses, also includes various viruses that have mutualistic relationships with their host.

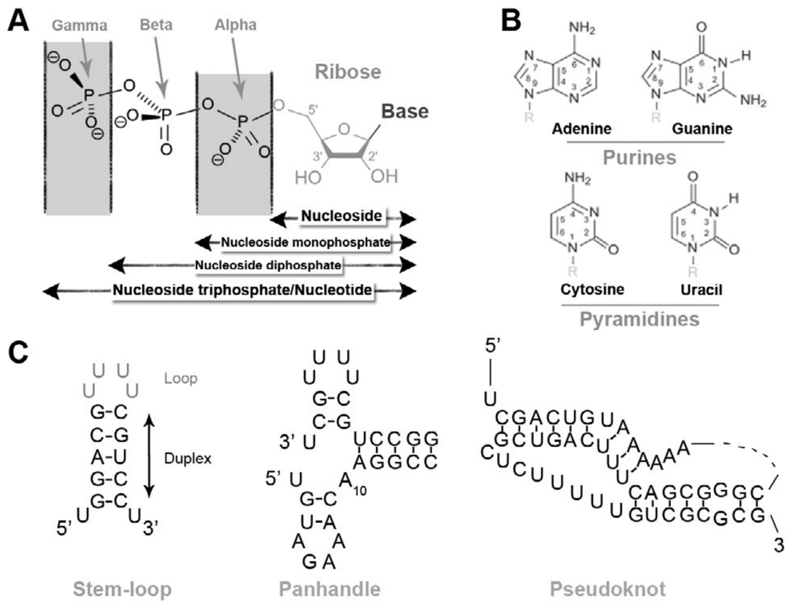

Figure 1: The basic principles of nucleotides and RNA structures. (**A**) The core structure of a nucleotide triphosphate that is the principle component of RNA. Indicated with arrows are the three phosphate groups (alpha, beta and gamma) that are attached via ester bonds to the 5′ carbon of the central ribose moiety. The position of the base is indicated with 'Base'. (**B**) The base component of a typical nucleoside triphosphate (nucleotide) can either be adenine, guanine, cytosine or uracil. Of these four, adenine and uracil can form two hydrogen bonds when brought in close proximity of each other, while guanine and cytosine can form three. (**C**) The basic secondary structure of RNA consists of a double stranded (duplex) region and a single stranded loop. Together these give rise to various structures, such as, in order by increasing complexity: stem-loop structures (or hairpins), panhandles and pseudoknots.

The versatility of RNA

RNA is - apart from being much less convenient to work with than deoxyribonucleic acid (DNA) - a flexible polymer with a very low persistence length[13] in single-stranded form (ssRNA). It consists of covalently linked subunits called nucleotides (nts), which are each composed of a phosphate group, a pentose sugar ribose and a nitrogenous base[14] (see Fig. 1A and B). Importantly, the ribose-base component of the nucleotide is the only part that can freely diffuse across a membrane, since the 5′ phosphates of the nts confer too much negative charge. On its own, the ribose-base is also referred to the nucleoside (Fig. 1A).

13 **Persistence length**: A basic mechanical property of a rod, string or polymer that defines its stiffness.

14 **Nucleotide bases**: adenine (A), guanosine (G), uracil (U) or cytocine (C). Alternative bases are, for instance, ribavirin (R), thymidine (T), ionosine (I), and urazole.

Crucially, to facilitate the replication of an RNA molecule, the order of the nucleotides in the parent molecule, the template, must be faithfully copied to the product molecule. This information is handed down through sequence-specific hydrogen bonds that can form between the pyrimidines (U or C) and purines (A or G) of the template RNA molecule and the new, nascent RNA[15]. These interactions are called the Watson-Crick base pairs. In addition, so-called stacking energies in the RNA structure provide fine tuning to these interactions [15].

However, at any time that ssRNA exists in the parent molecule, the base pairs and the helical structure of the RNA can also favour associations between ssRNA parts within the RNA molecule that have sequence complementarity. In turn, such intramolecular interactions can essentially prevent the complete duplication of the then partially double stranded RNA (dsRNA), but they may also induce the formation of more elaborate secondary structures, such as hairpins, pan-handles and pseudoknots (Fig. 1C). In addition to the Watson-Crick base pairs, hydrogen bond-derived interactions can also be established among three or even four nucleotides, thereby allowing for even more plasticity in the ssRNA. Thus, having all these interactions at its disposal, a ssRNA may quite easily acquire such an intricate secondary structure that it can start to specifically bind metal ions and facilitate the catalysis of various chemical reactions such as RNA-based RNA cleavage [16,17], peptide bond formation [18,19], and, seemingly paradoxically in respect of the duplexes in the structure, self-replication *in vitro* [20,21,22].

Polymerising RNA with RNA

Inside contemporary cells, the non-catalytic form of RNA serves predominantly as messenger molecule. In such functions it is mainly involved in providing ribosomes poised for protein synthesis with a genetic code (as mRNAs) or amino acids (as aminoacyl-tRNAs). In addition, it may play a vital role in other processes, such as the initiation of DNA synthesis [23], and countless regulatory or host-defence mechanisms in roles like riboswitches, siRNAs, miRNAs, piRNAs, snoRNAs and l-ncRNAs. Catalytic RNA, on the other hand, can be found in the functional centre of the ribosome and appears here to be actively involved in peptide bond formation [18,19]. In addition, it can be found in the active sites of the RNA cleavage enzyme RNaseP [24].

Note though that in the above two examples of catalytically active RNAs, the RNA always needs to be associated with protein in order to function. It contributes activity to a complex, but it is not enzymatically active in its own right. However, the catalytically active form of completely bare RNA, as presented in the RNA world hypothesis, is by no means an imaginary relic of the past. Even contemporary RNA molecules can be fully

---

15 **Nucleotide hydrogen bonds**: a double bond is formed between U::A; a triple bond is formed between C:::G.

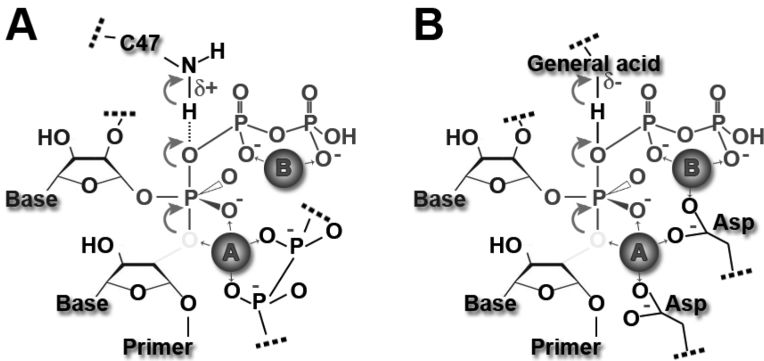

Figure 2: Ribozyme and polymerase active sites. (**A**) The present model for the active site and transition states of the ribozyme polymerase and ligases. Central to the activity are two aspartic-bound magnesium ions (grey spheres labelled A and B), of which metal A activates the 3′-hydroxyl group of the primer (highlighted in light grey) and metal B stabilises the triphosphate backbone of the incoming nucleotide (shaded in dark grey). The N4 of the cytosine base (attached ribose not shown) in the top left corner forms a hydrogen bond with the oxygen of the leaving pyrophosphate. The transition of covalent bonds is indicated with arrows. (**B**) The interactions within the active site of an RNA-dependent RNA polymerase during the catalysis of NTP incorporation. As indicated in figure 2A, two aspartic acid (Asp)-bound magnesium ions are required for activity. The general acid depicted at the top of the figures helps stabilise and protonate the pyrophosphate group as it is released. Figures adapted from reference [34].

catalytically active on their own [25]. These molecules, typically referred to as ribonucleic acid enzymes or ribozymes, can be found in cells or viral genomes as part of self-splicing introns [26] or the self-cleaving hammerhead ribozyme [16,17]. But, as will be discussed in more detail below, compared to protein enzymes that usually catalyse many chemical reactions per second, RNA self-cleavage is relatively slow and only capable of achieving cleavage rates of approximately 1-100 reactions per minute (min^{-1}).

So far, no bare, metabolic RNAs[16] have been identified *in vivo*. However, the RNA-based RNA ligases that have been presented as support for the principle, were first identified from a large set of quasispecies[17] that had been selected from random oligonucleotides through *in vitro* evolution [20,27]. *In vitro*, these ribozymes can catalyse the formation of a phosphodiester bond between the 3′-hydroxyl (3′-OH) group of the incoming

16 **Metabolic RNA**: An RNA that creates rather than destroys a biopolymer and can thus the basis the basis of self-replicating RNAs.

17 **Quasispecies**: a group or 'cloud' of related, amplifying nucleic acid sequences that are expected to contain mutations between parent and off-spring molecule. This is opposed to a species, which maintains, on average, a stable genotype.

oligonucleotide substrate and the α-phosphate of their own 5′ triphosphate, requiring initially about 1 min to complete a single reaction cycle.

Fundamentally, this 'ligase' reaction is chemically identical to the metal-dependent condensation reaction that is catalysed by the proteinaceous viral and cellular polymerases (compare Fig. 2A and 2B). However, it is, enzymatically different because it likely does not require a general acid to facilitate the proton transfer between the two substrates [22,28]. Crucially, it is on both aspects different from the ATP-dependent ligase reactions performed by cellular or viral enzymes, such as T4 ligase, even though the name of the reaction (*i.e.*, ligase) may suggest otherwise.

Overall, the ribozyme reaction proceeds in a series of discrete steps: (first) the reactants are aligned to an RNA template by Watson-Crick pairing, (second) the first metal ion activates the 3′-OH of the primer substrate, (third) the 3′-OH of the primer substrate (P) attacks the 5′-triphosphate of the ribozyme (^{ppp}R), (fourth) the second metal ion stabilises the developing negative charge on the β- and γ-phosphates, (fifth) a new internucleotide linkage is created which thereby extends the RNA primer by the length of the ribozyme to P^{+R} and releases pyrophosphate (pp_i) [22,29]. The direct reverse reaction, *i.e.*, pyrophosphorolysis, has also been detected, but this process is deemed to be too inefficient compared to the condensation reaction to significantly compete with it [29,30]. Practically, the ribozyme reaction thus follows the equation:

$$P^{OH} + {}^{ppp}R \rightarrow P^{+R} + pp_i + H_2O \qquad [1]$$

The 'true' RNA-based polymerases that incorporate NTPs onto the 3′-OH of a primer[18] and thus better resemble polymerases are currently around 189 nt long and were identified after further *in vitro* evolution of the 'ligase' ribozyme RNA sequences [30,31]. Still, the majority of their reactions does not proceed past an extension of 4-6 nt of the primer, primarily due the low stability of the primer-template complex (millimolar affinity) and the rate at which the ribozyme itself is hydrolysed in the presence of the high Mg^{2+}-containing reaction buffer (≥200 mM Mg^{2+}) [32]. Currently, the most processive RNA-based polymerase is called B6.61 - derived from *in vitro* compartmentalisation (IVC) procedures to mimic Darwinian selection in cells - and capable of incorporating up to 20 nts in a template-dependent manner over the course of 24 hours [32].

---

18 Note that this thus replaces ^{ppp}R with NTP and P^{+R} with P^{+N} in Eq. 1

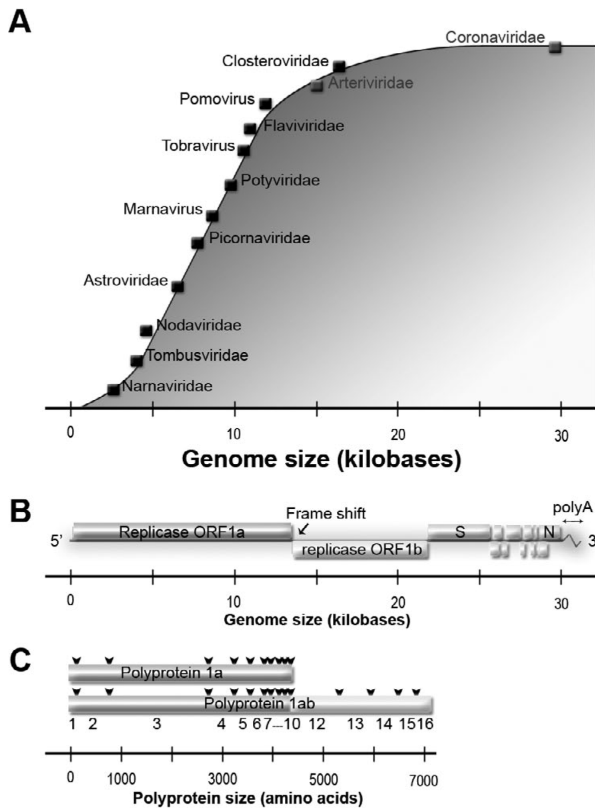

Figure 3: The genomes of selected ssRNA viruses and the genome organisation of the largest RNA genome know to date. (**A**) The mean genome sizes of various RNA viruses arranged according to genome size reveals that the majority of viruses are constraint within an upper-limit of ~13 kb. The nidoviruses studies in this thesis are highlighted in grey. Figure adapted from reference [47]. (**B**) The genome of SARS-CoV is polycistronic, and protected at the 5′ and 3′ ends with a cap and polyA tail, respectively. The first two open reading frames (ORF1a, grey; and ORF1b, light grey) are connected via a ribosomal frame shift that is triggered at a 1:4 ratio by an encoded pseudoknot structure in the coding region of nsp12. The open reading frames that encode the spike and nucleocapsid structural proteins are marked with S and N, respectively. (C) Expression of the first two open reading frames results in the synthesis of two large polyproteins. These proteins are subsequently cleaved by intrinsic protease activity (indicated by black triangles) to give rise to 16 mature non-structural proteins that concomitantly assemble into a functional RTC.

Limits of ribozymes

Polymerase properties, such as fidelity[19], reaction rate[20], processivity[21], and primer-

---

19 **Fidelity**: a measure for the consistency with which a polymerase forms Watson-Crick base pairs when it incorporates a new nucleotide.

20 **Rate**: the number of incorporated nts, usually expressed in molarity, per unit of time.

21 **Processivity**: the extension of the initial substrate in a template-dependent manner, but

template complex (PTC) recognition (*e.g.*, expressed as a binding constant, K_d) are all highly interdependent characteristics. Consequently, it is increasingly difficult, if not impossible, to fully optimise these properties simultaneously for a single ribozyme polymerase. And this even holds true when one takes the wide-array of chemical modifications into account that cellular enzymes can make on RNA's four chemical components [33].

It is nevertheless likely that, akin to the current biosphere, primordial self-replicating entities constantly competed with other genotypes for resources. Such a situation would have strongly favoured faster replication cycles and thus more energetically desirable chemical constants to allow nucleotide incorporations within seconds rather than minutes. Additionally, to counter parasitism, specific recognition of the 'own' genome would have evolved as well (*e.g.*, via RNA sequence or structure specificity, although not necessarily through *in cis* activity). If we also take into account the rapid and spontaneous degradation of RNA molecules in solution, which imposes a significant upper limit on the life-span of each ribozyme and thereby its maximum size, we arrive at a set of compelling reasons for why contemporary cells and viruses evolved the more chemically diverse and stable proteinaceous enzymes: to ensure faithful and rapid duplication of a large amount of genetic material.

Indeed, where eutectic freezing[22] can result in the template-dependent formation of RNA molecules of up to 420 nts in length [7,34] and self-replication only sustains RNAs that are even shorter [20], the RNA molecules that constitute modern RNA genomes reach up to 32 kb in length [35]. Such large molecules are found in RNA viruses (Fig. 3A) and they encode so-called RNA-dependent RNA polymerases (RdRps) to catalyse their replication [35]. Intriguing exceptions are some viroid genomes, however, which can 'lure' the cellular DNA-dependent RNA polymerase (DdRp) Pol II, an enzyme that is generally involved in cellular transcription[23], into replicating their viral RNA[36].

Replication and transcription complexes

The RdRp is essentially the sole determinant of genome size. However, where optimisation is no longer possible, some of its functions can be enhanced or supported by other protein factors. For instance, all genomes above 6 kb encode a helicase[24] co-factor, likely to support the polymerase in unwinding the large stretches of dsRNA that - as discussed

expressed as function of the incorporation rate and the dissociation constant.

22 **Eutectic freezing**: the process of molecular crowding due to physical exclusion from surrounding ice-crystals.

23 **Transcription**: the process of mRNA synthesis as it happens on, but is not exclusive to, chromosomal DNA templates, and which results in RNA molecules that can be translated by the ribosome to yield protein.

24 **Helicase**: An enzyme that can use the hydrolysis of ATP to unwind a dsRNA

above - can be found in ssRNA secondary structures and may represent the replicative intermediate of ssRNA replication [35,37]. Furthermore, although the majority of viral RNA genomes is shorter than 13 kb, an upper size-limit that is believed to be defined by the number of mutations that are made by the replication and transcription complex (RTC)[25] and the number of mutations that the virus can tolerate [38,39], some have RNA genomes that are close to three times larger, *i.e.*, the coronaviruses (CoVs). How can they achieve this, while others do not?

To get to an answer, one first has to appreciate the evolutionarily importance of RdRp fidelity and how it affects multiple crucial factors: i) all enzymes evolved as highly interconnected networks that are easily disturbed by amino acid substitutions; ii) the RNA sequence may fold into secondary structures that are vital recognition signals for enzymes or genome packaging [40,41,42]; iii) the codon bias of the host determines the pause rate of translating ribosomes and thus tunes the rate and reliability with which the polypeptide folds into a functional protein [43]; iv) viral enzymes may initially be expressed as polyproteins, in which the enzymes subunits are separated by conserved and highly coordinated cleavage sites to regulate their activity and ensure proper folding [44]; v) the RNA sequence may force initiating ribosomes to perform leaky scanning or translating ribosomes to shift reading frame, which together may regulate the molecular ratios between components of the viral RTC [45].

If one now puts these important factors - and particularly how prone they are to disruption - next to the observation that the estimated fidelity of most RNA polymerases is one mutation per every 10,000 copied nucleotides, it is not that remarkable that the average size of the majority of viral RNA genomes is roughly coinciding with this value. A substantial body of work also supports this ‘phenomenology’ quantitatively [35,46]: the RdRp fidelity limits the RNA genome size. Interestingly, additional mutations may derive from spontaneous mutation, the activity of cellular enzymes, such as the base deaminating enzyme APOBEC [47,48,49], or homologous recombination events, which either take place at random [50] or at well-defined sites in the genome such as during nidovirus discontinuous transcription (see BOX II) [51].

But why then, to better preserve the *status quo* or enlarge the coding capacity, does the average virus not evolve a less error-prone RNA polymerase or an additional subunit to improve it? Indeed, most DNA-dependent DNA polymerases (DdDp) use a proofreading mechanism. This is usually an additional enzymatic activity, a 3′-to-5′ exonuclease residing in their N-terminal domain, is able to detect misincorporations and perform metal-dependent hydrolysis, starting with the last incorporated nucleotide [52,53,54,55][26].

---

25 **RTC**: Essentially the RdRp and the collection of associated enzymes and modulating co-factors that together regulate the RdRp in its function

26 Note, that this does not necessarily imply that this process of error-recognition-and-repair will result in a correct Watson-Crick base pair; the second incorporation cycle is equally subject to

However, in the RNA world, things are seemingly never that simple, because the pressures that derive from host-pathogen relations, *e.g.*, host immune responses and other antiviral phenomena, favour a strategy that allows the virus to achieve a greater genetic variability and thereby the potential to evade any antiviral attack [39,56,57].

BOX II: Nidovirus discontinuous RNA synthesis - One of the unique features of the Nidovirales, the order to which the coronaviruses and arteriviruses belong, is the production of a nested set of subgenomic RNAs via a mechanism involving a step of discontinuous RNA synthesis. This mechanism involves the synthesis of five to nine subgenomic mRNAs (sg RNAs), whose sizes differ and which have been found to contain both identical 5′- and 3′-termini [1,2,3].

To explain the biochemistry of the underlying mechanism, several models have been proposed over the years [1,6,7]. The most widely supported model proposes that these differently sized sg RNAs are formed without splicing events in larger, precursor RNAs, as is typical of cellular mRNA maturation. The exact details of the mechanism are still a mystery, however, and the contribution of the viral and host proteins mostly unknown.

In the prevalent model, RNA synthesis initiates with the generation of -RNA using the +RNA genome as template [1,7]. Although the +RNA genome has a well-defined length, -RNA synthesis is halted at a one of the many body-TRS (transcription-regulating sequence) elements present in the genome and continued at a downstream leader-TRS. Consequently, this results in a fusion of a body elements with a common 3′ anti-leader via complementary sequences just upstream (on the -RNA) of the leader and downstream of the body elements [see for review reference 7]. Alternatively, -RNA synthesis progresses through the TRS elements until the end of the +RNA is reached.

Overall, this produces -RNAs of several sizes: full-length genomic antisense RNA and various subgenomic antisense RNA species [1,7]. Together, these -RNAs thus form, like the positive-stranded sgRNAs that the virus eventually produces, a co-terminal nested set. The full-length antisense RNA is used as template for viral genome duplication - this will either be translated and used in new RTCs or be inserted into assembling virus particles - while the shorter subgenomic antisense RNA species will serve as template for positive-stranded sg RNAs synthesis ([see for review reference 1]). In turn, these sg RNAs are translated by cellular ribosomes into the viral structural proteins.

---

the limit of polymerase fidelity as the initial round.

Going beyond limits, nidoviruses and SARS-coronavirus

It is theorised that, given the constraints summarised above, no viral RNA genome larger than ~10 kb can be stably maintained in the absence of additional "error prevention" or proofreading mechanisms [35,58]. However, the mean size of the CoV genome is ~30 kb of positive-strand RNA (+RNA)[27], and that is significantly larger than the average RNA genome (Fig. 3A). Additionally, phylogenetic analysis has suggested that the mutation rate among CoV genomes is moderate to low compared to other RNA viruses [59,60]. These observations thus suggest that there must be something special about the coronavirus replication machinery, the immune responses (or the lack thereof) the virus 'feels' during the infection of a host, or the CoV replicative cycle in general that enabled them to expand their genome beyond the limit imposed on other RNA viruses. A direct answer is presently unavailable.

So, then, what do we know about these viruses? Based on a similar genomic organisation and conserved protein domains [35,58], the coronavirus subfamily (*Coronavirinae*)[28] is unified with the *Torovirinae* in the *Coronaviridea* family and the order of Nidovirales. The viruses that are classified under this order all utilise a unique discontinuous RNA synthesis mechanism to express their structural genes (see BOX II). Strikingly, although CoVs have long been recognised to cause common and more severe respiratory diseases, including various human diseases, it was not until the sudden emergence of the severe acute respiratory syndrome CoV (SARS-CoV)[29] via zoonotic transfer from bats to humans that they became interesting for worldwide research programs [61,62,63,64]. These programs now also include their distant relatives the arteriviruses and their prototype equine arteritis virus (EAV) [62,65,66].

The nidovirus RTC

Similar to other +RNA viruses, the CoV RTC is targeted to cellular membranes [67,68]. It has been proposed that RNA viruses may benefit from this association, because these membranes may provide: i) physical support for organization of the RTC; ii) compartmentalization and an increase of the local components/products concentration; and iii) protection of the viral RNA and dsRNA intermediates from host defences [69]. The membranes that are used to support the RTC vary among viruses, however. Structures induced by nidoviruses were initially linked to endosome/autophagosome pathways,

---

27 **Positive-strand**: upon entry into the host cell, a typical +RNA virus genome serves as mRNA, thereby allowing the cellular ribosome to engage in translation. -RNA viruses first go through at least one round of transcription before translation of the viral code can take place. Interestingly, ambisense viruses can encode reading frames in both the minus and plus strand.

28 ***Coronavirinae**:* these viruses are since 2009 subdivided in the three genera of Alpha-, Beta-, and Gammacoronaviruses by the ICTV.

29 By July 2003, when the SARS epidemic had officially disappeared due to increased health and safety measures, 813 patients had died from SARS. Presently, still no effective drugs exist.

but are now believed to be derived from the ER [70,71,72,73,74]. Other +RNA viruses have been shown to recruit their membranes from the ER as well, such as in the case of PV and Dengue virus, whereas other prefer mitochondria, such brome mosaic virus (BMV) and carnation Italian ringspot virus (CIRV) [75,76], or late endosomes, as was shown for Semliki Forest virus (SFV) [77,78].

Presently, the mechanism behind the formation of the CoV vesicular structures is mostly elusive, although one step at least appears to be vital. This step involved the viral transmembrane proteins, which are believed to target the viral RTC to the cellular membranes [58]. In nidoviruses, these proteins are part of the large set of replicase or nonstructural proteins (nsps) (*e.g.*, SARS-CoV encodes sixteen mature nsps) that are encoded as two large, partially overlapping open reading frames (ORFs) [58,61,79], called ORF1a and ORF1b. Expression of these two ORFs results in two large polyproteins (Fig. 3B and C), namely polyprotein pp1a and the pp1ab fusion polyprotein [58,80] (Fig, 3B and C). The intricate mechanism behind this elegant way of expression is an internal ribosomal frameshift signal that resides in the nsp12-coding region and it ensures an asymmetrical production of the nsps encoded in ORF1a and ORF1b (Fig. 3B): a 3-4 fold overproduction of the proteins upstream of the frameshift (nsp1-11) compared to the ones encoded downstream of the frameshift (nsp12-16) [45,74,81,82].

To derive the mature nsps, the two polyproteins are processed by at least one papain-like protease (PLpro) and a 3C-like protease (3CLpro) activity (some nidoviruses contain 2 PLpro or additional protease domains) [79,80,83]. This step also releases the core enzymes around which the CoV RTC will form: the putative main, primer-dependent RdRp nsp12 and a second polymerase activity resident in nsp8 [58,84,85]. Interestingly, in addition to the processing of viral polyproteins, the PLpro enzymes have also been implicated in negating antiviral responses [86,87,88].

The nidovirus polymerases

The nsp12 structure has not yet been solved, but a model of this SARS-CoV RdRp has been proposed and used to infer that the CoV polymerase is fundamentally different from RNA polymerases encoded by viruses like poliovirus (PV) and hepatitis C virus (HCV) [89]. First of all, the amino acid conservation between CoV polymerases and those of known structure (*e.g.*, HCV and PV) is less than 10%, whereas the overall sequence conservation within the CoV subfamily is over 60% and even complete among different SARS-CoV isolates [89].

The most significant differences with known RdRp structures reside in the N-terminal domain (amino acid residues 1-375 of the total 932 in SARS-CoV) and it even appears that the tertiary structure of this domain has no structural equivalence in any other viral RNA polymerase. In addition, large differences reside in the nucleotide selection pocket, which may explain the ineffectiveness of wide-spectrum viral replication inhibitors

against SARS symptoms [89,90,91]. For instance, a resistance to ribavirin[30] - a widely used purine analogue in the treatment of, *e.g.*, Lassa virus, respiratory syncytial virus and HCV infections [92,93,94,95] and an effective inhibitor of the replication of picornaviruses, orthomyxoviruses [96,97], hantaviruses [98,99], vaccinia virus [100], and reoviruses [101] in cell culture - was observed in several studies [102,103].

The above features may somehow play a role in achieving the apparent lower mutation rate of the virus and its larger than average genome as well [89]. However, it is unclear how these features would contribute to these CoV properties or what biochemical properties they actually constitute. More directly relevant seems the prediction that the active site of the SARS-CoV nsp12 protein is endowed with a relatively unobstructed nucleic acid-binding cleft, which implies that, unlike, *e.g.*, the HCV RdRp, the SARS-CoV RdRp can accommodate and extend a primed template [89]. Until this day, however, it has not been demonstrated how SARS-CoV nsp12 would actually obtain this primer, but, interestingly, it was demonstrated that SARS-CoV nsp8 possesses weak, low-fidelity RdRp activity as well [104]. Not unlike cellular primases and *de novo* initiating RNA polymerases, nsp8 was also described to prefer purines over pyrimidines during initiation and to synthesise 6-mer oligonucleotides [23,104], which in turn led to the hypothesis that the two CoV RdRps could cooperate and form something resembling a primase-RdRp complex.

However, given the unprecedented genome size and moderate mutation rate of CoVs, the proposed presence of a low-fidelity polymerase during genome replication is rather counterintuitive. Indeed, a dominant role of such an enzyme could evidently lower the overall fidelity of CoV RNA synthesis, destabilise the genome and consequently impair survival of the virus, all effects that appear to contradict some key CoV properties discussed above. A solution for this disparity may come from the observation that nsp14 harbours 3′-to-5′ exonuclease (ExoN) activity [105] and that CoV mutants lacking this activity show a 15-fold increase in the accumulation of mutations in their genome [106,107].

Together, the negative selection for nucleotide analogues by the RTC and the link between ExoN activity and mutation frequency are indicative of a simple form of an RNA-based proofreading mechanism. Unfortunately, biochemical data for such a proofreading mechanism has so far been elusive, has activity of nsp12 never been convincingly shown, and have no experiments been performed to gain insight into the molecular interplay between the two RdRps. Presently therefore, the scientific model for CoV replication and an explanation for the survival of the largest RNA genome known is still largely obscure.

30 **Ribavirin**: an analogue of both adenosine and guanosine that increases the error frequency of the RTC by elevating the number of transition mutations (U=>C and C=>U) and thus pushes the virus towards error catastrophe and extinction.

Outline of this dissertation

The biochemical and biophysical analysis of viral RNA synthesis is paramount to the discovery and exploration of new avenues to combating viruses. However, it is also vital for advances that may lead to a detailed understanding of the viral replication mechanism or the use of viruses as tools in biochemical studies or vectors for medical purposes. A wealth of such information is already available for HIV, influenza A, HCV and PV replication. A much more limited biochemical knowledgebase, and an even smaller biophysical one, is currently available on the replication of CoVs. Evidently, this hinders the development of antiviral strategies, but it also clouds our ability to model the replication cycle of their extraordinary RNA genome to a greater detail and with more confidence. This thesis will therefore describe a series of experiments that was performed to seek new insights, models and hypotheses regarding the enzymes that form the core of the coronavirus RTC.

This thesis will first describe the features of viral RNA polymerases in general in chapter 2. In it, the similarities and differences between various polymerases will be comprehensively reviewed, and the impact of accessory proteins and effect of dynamics on mutation frequencies discussed. Ultimately, this chapter will illustrate the various strategies that RNA viruses employ to regulate polymerase activity and specificity.

In the next chapter, chapter 3, this thesis will describe the purification of the SARS-CoV nsp that encompasses the classical viral RdRp domain: nsp12. Additionally, this work will outline biochemical experiments to demonstrate and investigate its *in vitro* activity on primed RNA templates of varying length and sequence composition.

With the knowledge that the SARS-CoV (supposedly) main polymerase can be stably and reproducibly purified, this thesis will next discuss the inhibiting effect of zinc ions on the nsp12-RdRp in chapter 4 and offer a comparison with replication complexes that were isolated from infected cells. Furthermore, this chapter will demonstrate the inhibitory effect of a zinc-ionophore, a compound that elevates the zinc ion concentrations ([Zn^{2+}]) in the cell, on the replication of both SARS-CoV and EAV in cell culture, thereby confirming the *in vitro* experiments.

As discussed above, SARS-CoV encodes a second enzyme with polymerase activity: nsp8. In chapter 5, this dissertation will describe the purification of this enzyme and the effect of different fusion tags on the oligomerisation modes of the protein. Furthermore, it will show that different activities, namely *de novo* or primer-dependent polymerase activity, are associated with these forms and that its co-factor nsp7 can stimulate nsp8's primer-dependent activity *in vitro*.

As mentioned above, helicases likely evolved to support the RdRp in the replication of RNA genomes beyond 6 kb in length [35] and, when they are sensitive to stable structures, helicase likely significantly affect polymerase pausing and thereby RTC fidelity and recombination. Presently, the nidovirus helicases have only been studied using

biochemical techniques and through mutational analysis of EAV and human coronavirus 229E [108,109,110,111]. To study the activity of nidovirus helicases in more detail, experiments at the single molecule level will be discussed and performed in chapter 6 and 7, respectively.In chapter 6 this thesis will also expatiate quantitatively upon the problem of reliable force calibration of single molecule experiments that employ magnetic tweezers and the limits of these calibrations.

Having seen the nidovirus helicases at work at the single-molecule level and having established that SARS-CoV encodes two primer-dependent RNA polymerases of which one can putatively act as primase, various questions remain. One concerns the functional part of the genome that has so far been largely uncharted: the polyA tail. Another question concerns the fact that SARS-CoV encodes two polymerases capable of primer-extension: what happens when both proteins are present in the same environment and are provided with a primed template? Will this reveal a form of intrinsic regulation between the two enzymes or will they just directly compete with each other? This is what this thesis will try to answer in chapter 8 by carefully comparing the activities of SARS-Co nsp8 and 12, and assessing the effects of their interactions. In addition, chapter 8 will describe that a distinct polyA polymerase activity is associated with the octameric form of nsp8. This, together with the observed *de novo* initiation activity, inspires an intriguing model in which this enzyme can play a crucial role in +RNA maturation and -RNA initiation.

Finally, chapter 9 describe a 3′-to-5′ exonuclease activity that appears to be associated with the nsp12 N-terminal domain, whereas chapter 10 will discuss the main findings described in this thesis and their implications for present models of nidovirus replication.