

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

# **A biochemical portrait of the nidovirus RNA polymerases and helicase**

**Aartjan te Velthuis**

ISBN: 978-94-6169-230-6

Layout and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

# **A biochemical portrait of the nidovirus RNA polymerases and helicase**

**Proefschrift**

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden  
op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,  
volgens besluit van het College voor Promoties  
te verdedigen op woensdag 16 mei 2012  
klokke 13.45 uur

door

**Arend Jan Wouter te Velthuis**

geboren te Leiden  
in 1983

## Promotiecommissie

### Promotores:

Prof.dr. E.J. Snijder	-	Leids Universitair Medisch Centrum
Prof.dr. N.H. Dekker	-	Technische Universiteit Delft

### Overige leden:

Prof.dr. A.E. Gorbalenya	-	Leids Universitair Medisch Centrum
Prof.dr. R.C. Hoeben	-	Leids Universitair Medisch Centrum
Dr. S.H.E. van den Worm	-	Oregon Health and Science University, Verenigde Staten
Prof.dr. J.K. Ziebuhr	-	Justus Liebig Universität Giessen, Duitsland

Het onderzoek dat in deze dissertatie wordt beschreven is uitgevoerd in de afdeling Medische Microbiologie van het **Leids Universitair Medisch Centrum** (LUMC) en de afdeling Bionanoscience van de **Technische Universiteit Delft** (TUD). Het onderzoek werd hierbij financieel gesteund door de **Nederlandse organisatie voor Wetenschappelijk Onderzoek** (NWO) via Toptalent beurs 021.001.037 en 'Open Access' beurzen 036.001.008, 036.001.250, en 036.001.619 voor de publicatie van hoofdstukken 3, 4 en 5.

The research this thesis expatiates upon was carried out in the Department of Medical Microbiology of the **Leiden University Medical Center** (LUMC) and the Department of Bionanoscience of the **Delft University of Technology** (TUD). For financial support, the **Netherlands Organization for Scientific Research** (NWO) is gratefully acknowledged for Toptalent grant 021.001.037 and Open Access grants 036.001.008, 036.001.250, and 036.001.619 for the publication of chapters 3, 4 and 5.

Voor pap, mam en mijn kleine zusje 'Pien'



<b>TABLE OF CONTENTS</b>		<b>PAGE</b>
	Samenvatting	9
<b>Chapter 1</b>	A more general introduction	13
<b>Chapter 2</b>	Common and unique features of RNA virus polymerases define viral transcription and replication	31
<b>Chapter 3</b>	The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent	49
<b>Chapter 4</b>	Zn <sup>2+</sup> inhibits coronavirus and arterivirus RNA polymerase activity <i>in vitro</i> and zinc ionophores block the replication of these viruses in cell culture	69
<b>Chapter 5</b>	The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both <i>de novo</i> initiation and primer extension	93
<b>Chapter 6</b>	Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data	119
<b>Chapter 7</b>	Nucleotide-dependent switching between active and passive unwinding in a nidovirus superfamily 1 RNA helicase	151
<b>Chapter 8</b>	Dual functionality for the SARS-CoV nsp(7+8) RNA polymerase at the genomic 3' end	179
<b>Chapter 9</b>	Ribavirin triphosphate- and mismatch-stimulated 3'-to-5' exonuclease activity of the SARS coronavirus RNA polymerase nsp12	201
<b>Chapter 10</b>	A discussion of nidovirus RNA synthesis	217
	<b>References</b>	229
	<b>Curriculum Vitae</b>	253
	<b>List of publications</b>	255



## SAMENVATTING

RNA virussen zijn kleine ziekteverwekkers die zich razendsnel kunnen verspreiden en in korte tijd een epidemie of een pandemie kunnen veroorzaken. Ze zijn daardoor dus van grote invloed op het openbare leven, het zorgsysteem en de wereldeconomie. Een bekend voorbeeld hiervan is de uitbraak van SARS, welke in 2003 plaatsvond. Deze ziekte, veroorzaakt door het SARS-coronavirus, kostte binnen enkele maanden bijna duizend mensen het leven en maakte ons op niet mis te verstane wijze duidelijk hoe kwetsbaar we zijn voor dit soort RNA virusinfecties. Bovendien zorgden de neveneffecten van deze pandemie voor een economisch verlies dat volgens de Wereldbank ongeveer 0.5% van de totale wereldeconomie bedroeg.

Gelukkig kon SARS relatief snel en zonder antivirale middelen of vaccins tot staan worden gebracht. Het is echter de vraag of een combinatie van adembescherming en quarantaine van patiënten zoals gebruikt voor SARS in de toekomst zal voldoen om een vergelijkbaar of gevaarlijker virus uit de nidovirus orde onder controle te krijgen. Helaas moeten we constateren dat onze kennis van deze virussen 9 jaar later nog altijd zeer beperkt is en dat goedgekeurde antivirale middelen nog steeds niet voorhanden zijn. Met de resultaten die worden beschreven in dit proefschrift, hoop ik een stukje te hebben bijgedragen aan de karakterisering van deze virusfamilie en - uiteindelijk - tot de ontwikkeling van deze antivirale middelen.

De focus van de in dit proefschrift beschreven resultaten ligt op twee enzymtypen, die van cruciaal belang worden geacht voor de infectiecyclus van nidovirussen: het virale RNA polymerase<sup>1</sup> en het virale helicase<sup>2</sup> enzym. Deze eiwitten zijn het 'kloppend hart' van het grote eiwitcomplex van virale niet-structurele eiwitten (ook wel nsps genaamd), dat zorgt voor de replicatie en transcriptie<sup>3</sup> van het virale genoom<sup>4</sup>. Aangezien er weinig tot niets bekend is over de precieze activiteit van deze enzymen, ligt het voor de hand dat het vergaren van meer kennis over deze enzymen zou kunnen bijdragen aan het beter begrijpen van de replicatiecyclus en evolutie van nidovirussen, en aan de ontwikkeling van antivirale middelen tegen deze pathogenen.

Met het oog op de uitbreiding van deze kennis wordt in **hoofdstuk 3** van dit proefschrift de isolatie en biochemische analyse beschreven van het SARS-coronavirus RNA polymerase enzym, nsp12. Deze analyse laat onder andere zien dat met name het

- 
- 1 **RNA polymerase:** een enzym dat in staat is om genetisch materiaal te vermenigvuldigen.
  - 2 **RNA helicase:** een enzym dat in staat is om de structuur in het genetische materiaal te verminderen en zo o.a. de activiteit van het RNA polymerase enzym te bevorderen.
  - 3 **Transcriptie:** het proces om genetische informatie te 'vertalen' naar een informatie vorm (mRNA) die gebruikt kan worden om meer viraal eiwit te maken en daarmee nieuwe virusdeeltjes.
  - 4 **Genoom:** het genetische materiaal van een organisme of virus.

N-terminale aminozuur<sup>5</sup> van dit eiwit cruciaal is voor de stabiliteit en activiteit van nsp12. Verder wordt er beschreven dat de activiteit van nsp12 sterk afhankelijk is van de sequentie van het template<sup>6</sup> en dat het kopiëren van dit template niet kan plaatsvinden in afwezigheid van een primer<sup>7</sup>. Opmerkelijk genoeg kan, zoals wordt beschreven in **hoofdstuk 4**, deze activiteit volledig geremd worden door eenvoudigweg zink ionen aan de reactie toe te voegen. Biochemische proeven duiden er op dat deze ionen mogelijk werken door het polymerase enzym tot staan te brengen via het verlagen van de affiniteit van nsp12 voor het template.

In **hoofdstuk 5** van dit proefschrift wordt de isolatie beschreven van een tweede SARS-coronavirus enzym dat RNA polymerase activiteit vertoont, het eiwit nsp8, dat opmerkelijk genoeg vijf keer kleiner is dan nsp12. Net als voor nsp12 wordt beschreven dat onnatuurlijke aminozuren in de N-terminus van nsp8 van grote invloed zijn op de activiteit van dit eiwit. Een directe aanwijzing voor een effect op de stabiliteit van het eiwit is er echter niet. Wel zijn er aanwijzingen dat het N-terminale domein van nsp8 van groot belang is voor de complexvorming van nsp8 met nsp7, en dat de aanwezigheid van nsp7 de RNA bindingsaffiniteit en activiteit van nsp8 als RNA polymerase stimuleert.

In **hoofdstuk 6 en 7** wordt vervolgens een zijstap gemaakt naar single-molecule studies<sup>8</sup> met een magnetisch pincet<sup>9</sup>, met als doel een uiterst nauwkeurige meting en beschrijving van de activiteit van het helicase enzym te verkrijgen. Een beperkende factor in de nauwkeurigheid van deze metingen komt voort uit de Browniaanse bewegingen van het paramagnetisch kogeltje dat nodig is om de helicase activiteit meetbaar te maken. Deze Browniaanse bewegingen ontstaan spontaan, doordat snel bewegende watermoleculen tegen het kogeltje botsen, en kunnen met name een rol gaan spelen wanneer grote magnetische krachten nodig zijn om 'handmatig' de structuur van het RNA of DNA in de opstelling te beïnvloeden. Bij dergelijke krachten kunnen de bewegingen van het kogeltje namelijk zo snel worden, dat ze niet meer met de camera te volgen zijn. Om de grens waarbij deze problemen optreden te bepalen en het effect van meetcondities te kunnen voorspellen wordt in **hoofdstuk 6** een simulatie van deze bewegingen beschreven.

Aansluitend wordt deze kennis in **hoofdstuk 7** toegepast om de activiteit van een nidovirus helicase tijdens de ontwinding van een DNA haarspeld structuur te meten. Opmerkelijk genoeg vinden we dat de helicase activiteit sterk afhankelijk is van de

---

5 **N-terminale aminozuur:** het eerste aminozuur van een eiwit.

6 **Template:** de aangeboden genetische informatie, die in een *in vitro* experiment gekopieerd kan worden.

7 **Primer:** een van tevoren geproduceerd "startmolecuul" voor de te maken nucleïnezuur kopie.

8 **Single-molecule analyse:** de analyse van de activiteit van individuele enzymen.

9 **Magnetisch pincet:** een instrument dat er voor is ontwikkeld om enkele RNA of DNA moleculen te 'vangen' tussen een glazen oppervlak en een klein paramagnetisch kogeltje, en de veranderingen in dit RNA of DNA te volgen met een camera.

nucleotidenvolgorde in het template en het nucleoside trifosfaat molecuul<sup>10</sup> dat het helicase enzym gebruikt om energie te generen voor het ontwinden van de haarspeld. Deze bevindingen suggereren dat de activiteit van dit helicase enzym *in vivo* veel dynamischer zou kunnen zijn dan tot op heden werd aangenomen.

In **hoofdstuk 8** wordt terug gegaan naar de activiteit van nsp8 en wordt beschreven dat dit enzym in staat is om twee RNA nucleotiden aan elkaar te koppelen in afwezigheid van een template. Opmerkelijk genoeg lijkt nsp8 daarbij een voorkeur te hebben voor een product dat complementair is aan het uiteinde van het te kopiëren genetische materiaal van het SARS-coronavirus. Daarnaast worden aanwijzingen beschreven waaruit blijkt dat nsp8 het 3'-uiteinde van het virale genoom kan verlengen met adenosine residuen en dat GTP - de nucleotide die complementair is aan het 3'-uiteinde van het genoom - de herkenning en verlenging van het 3'-uiteinde van het genoom kan stimuleren.

Om de hypothese te onderzoeken dat de primer-afhankelijke nsp12 en het *de novo* initiërende nsp8 eventueel met elkaar kunnen samenwerken om een SARS-coronavirus genoom te kopiëren worden in hoofdstuk 8 ook proeven beschreven die deze enzymen samenbrengen in één *in vitro* reactie. Opmerkelijk genoeg worden geen directe aanwijzingen voor een samenwerking van de twee enzymen gevonden. Wel zijn er aanwijzingen dat ze elkaars activiteit juist negatief beïnvloeden. Hieruit zou vervolgens geconcludeerd kunnen worden dat er nog minstens een ander eiwit nodig is om de twee enzymen te reguleren, zodra ze in elkaars aanwezigheid verkeren.

In **hoofdstuk 9** worden tenslotte aanwijzingen beschreven die er op duiden dat nsp12 ook een exonuclease<sup>11</sup> activiteit bevat. Met deze activiteit zou nsp12 bijvoorbeeld fouten uit het zojuist gesynthetiseerde RNA keten kunnen verwijderen om die vervolgens te corrigeren. Inderdaad blijkt uit andere proeven dat de exonuclease activiteit van nsp12 gestimuleerd kan worden door de aanwezigheid van ribavirine, een mutatie-inducerende nucleotide analoog, en een van te voren aangebrachte mutatie in de aangeboden primer. Toekomstige experimenten zullen hopelijk bijdragen aan een verdere ontrafeling van de beschreven eigenschappen en leiden tot het ontdekken van bruikbare remmers voor nidovirus infecties.

---

10 **Nucleoside trifosfaat:** een molecuul dat gebruikt kan worden om nieuw RNA te synthetiseren of energie te leveren voor enzymatische reacties.

11 **Exonuclease:** een enzymactiviteit die er voor zorgt dat het nucleotide dat zich aan het uiteinde van een RNA of DNA molecuul bevindt wordt verwijderd.



# 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<sup>12</sup> 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

---

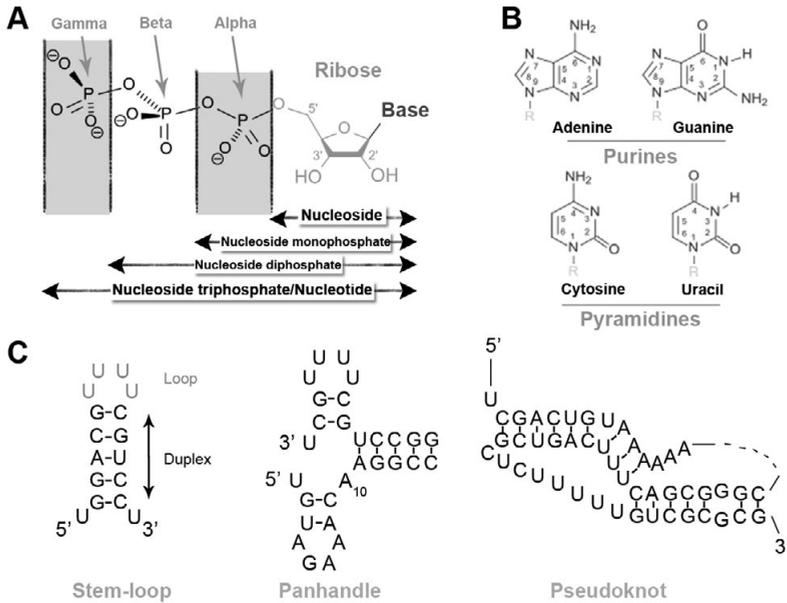
12 **Polymerase:** The term polymerase commonly refers to the enzyme or enzyme complex that copies the genomic content of an organism (*i.e.*, it uses a template). The term can, however, also be used in a more generic context, and refer to enzymes that simply make polymers from monomers in a template 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 *Elysia chlorotica* [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/archaea 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<sup>13</sup> 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<sup>14</sup> (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 cytosine (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<sup>15</sup>. 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].

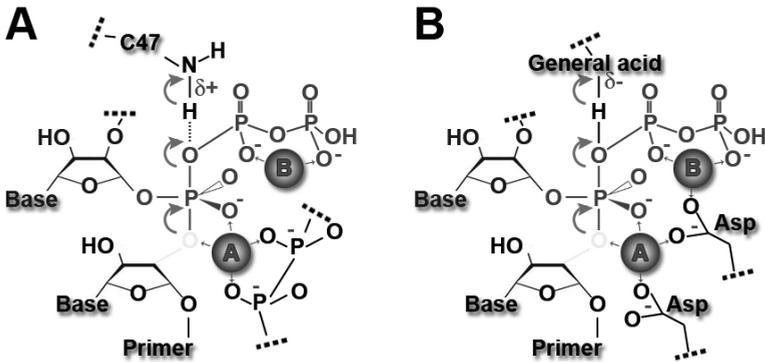
## Polymerising RNA with RNA

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 lncRNAs. 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 ( $\text{min}^{-1}$ ).

So far, no bare, metabolic RNAs<sup>16</sup> 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<sup>17</sup> 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 be 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  $\alpha$ -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.

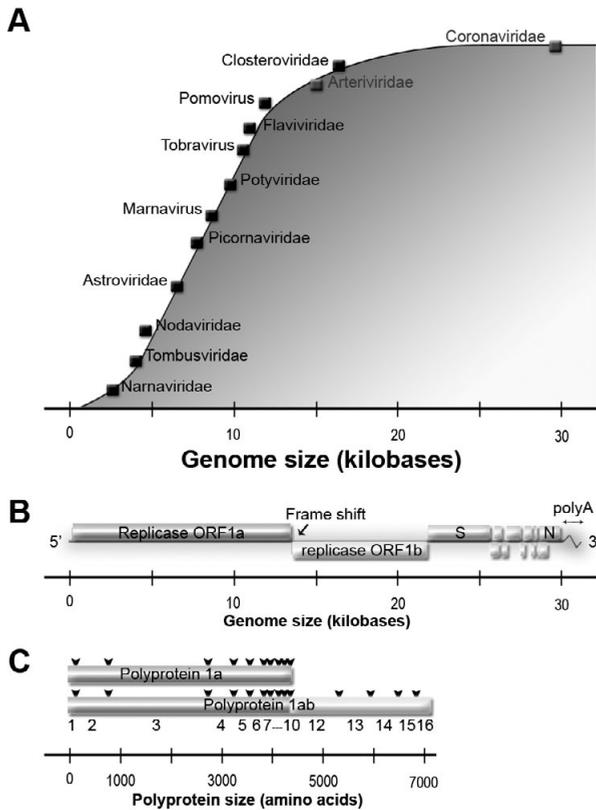
Overall, the ribozyme reaction proceeds in a series of discrete steps: (first) the reactants are aligned to an RNA template by Watson-Crick pairing, (second) the first metal ion activates the 3'-OH of the primer substrate, (third) the 3'-OH of the primer substrate ( $P$ ) attacks the 5'-triphosphate of the ribozyme ( $^{PPP}R$ ), (fourth) the second metal ion stabilises the developing negative charge on the  $\beta$ - and  $\gamma$ -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:



The 'true' RNA-based polymerases that incorporate NTPs onto the 3'-OH of a primer<sup>18</sup> 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 ( $\geq 200$  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 known to date.** (A) The mean genome sizes of various RNA viruses arranged according to genome size reveals that the majority of viruses are constrained 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<sup>19</sup>, reaction rate<sup>20</sup>, processivity<sup>21</sup>, 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<sup>22</sup> 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<sup>23</sup>, 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<sup>24</sup> 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)<sup>25</sup> 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 evolutionary 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]<sup>26</sup>.

---

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.

## 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)<sup>27</sup>, 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*)<sup>28</sup> is unified with the *Torovirinae* in the *Coronaviridae* 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)<sup>29</sup> 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<sup>30</sup> - 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.

# Chapter 2

Common and unique  
features of RNA virus  
polymerases define  
viral transcription  
and replication

Aartjan J.W. te Velthuis<sup>1</sup> and Eric J.  
Snijder<sup>1</sup>

<sup>1</sup> Molecular Virology laboratory,  
Department of Medical  
Microbiology, Center of Infectious  
Diseases, Leiden University  
Medical Center, PO Box 9600,  
2300RC Leiden, The Netherlands.

**ABSTRACT**

All viral RNA polymerases are believed to use the same catalytic mechanism for nucleotide condensation. The protein and RNA elements that are required for these reactions to take place, however, can vary greatly from polymerase to polymerase and can thus significantly affect the efficiency and specificity of viral replication and transcription. Furthermore, the fidelity of nucleotide incorporation, the mechanism employed to initiate RNA synthesis, and the dynamic shuttling between catalysis in the polymerase active site and the (meta- or catabolic) activity of other (enzymatic) domains in the polymerase amino acid chain all contribute to defining key virus properties such as genomic architecture, replication speed, immune-escape, evolution and pathogenicity. Here we will review our current knowledge of what defines RNA polymerase activity and how their properties orchestrate viral RNA synthesis.

## INTRODUCTION

In both the eukaryotic and prokaryotic cell, genetic information flows constantly from the original DNA molecule to the DNA copy and from the DNA 'blueprint' to the RNA messenger or regulatory molecule. However, given the abundance of RNA-based viruses in our ecology, it is likely that some if not most of their exogenous RNA is replicated and transcribed as well. This process can have profound effects on us. For instance, various lines of research have indicated that RNA viruses may have contributed to shaping eukaryotic evolution [112,113,114,115,116,117]. In addition, RNA viruses can benefit us more directly, since they can be employed as research tools or gene therapy agents [118,119,120,121]. Consequently by trying to understand RNA viruses, we may ultimately discover something more about ourselves and the way life evolved on earth, or how we can cure certain diseases. However, the double-edged nature of RNA viruses also makes them cause morbidity and death, and thereby a constant burden on our health systems and economy [65,122,123,124]. Our primary goal for understanding the replication and transcription mechanisms of pathogenic RNA viruses is thus the discovery of new ways to frustrate RNA virus RNA synthesis and to understand how they acquire mutations or even new genes to escape immune systems and current antiviral drugs.

As we will discuss below, RNA virus replication depends strongly on how their genetic material exits and enters the host cell and whether it can be directly translated by cellular ribosomes upon entry. The simplest division that can therefore be made among RNA viruses is a separation between viruses that use single-stranded RNA (ssRNA) as genetic material and viruses that use double-stranded RNA (dsRNA). In addition, we can further divide the ssRNA viruses into viruses that employ positive stranded (+RNA) genomes, *i.e.*, genomes that can be translated into protein immediately upon entry, and negative stranded -RNA genome viruses, *i.e.*, viruses whose genomes require transcription before translation can take place. However, irrespective of the type of viral RNA, the essence of viral replication, and thus the key of our understanding of this process, lies with RNA-dependent RNA polymerase (RdRp).

The RdRp domain is the only essential enzymatic domain that is encoded by all RNA viruses [125,126]. Together with cofactors, which can be virus or host derived (see below), it can form a dedicated type of complex that uses an ssRNA as template for the synthesis of a complementary strand in a fashion that is chemically similar to the replication and transcription of cellular DNA genomes [122,127,128]. However, the fact that RdRps are capable of both transcription and replication makes them significantly different from the task-specific cellular DNA-dependent polymerases. Prokaryotic genomes, for instance, require one DNA-dependent RNA polymerase (DdRp) and five different DNA-dependent DNA polymerases (DdDps) for these two processes. Moreover, eukaryotic cells employ five to six different DdRps, namely Pol I-VI, for roles such as

ribosomal RNA production, or the synthesis of messenger RNAs, small nuclear RNAs, transfer RNAs or siRNAs [129,130,131], and up to fifteen DdDps and one RNA-dependent DNA polymerase (RdDp, or RT) for replication. Consequently, compared to this array of cellular enzymes, RNA virus replication and transcription is very focussed and intricate, thus making it even more remarkable that RNA virus replication and transcription are so perfectly tuned to fit each virus' niche and possibly even more daunting to unravel and understand it. In this review, we will therefore aim to present a comprehensive overview of the RdRp structures, co-factors and RNA regulatory elements that make this happen.

### The general RNA virus polymerase architecture

There is a wealth of literature detailing the diversity of viral RNA polymerases [126,132,133,134,135]. This diversity is particularly evident at the amino acid level, with a number of estimations suggesting a 90% variation across the whole polymerase coding sequence. In contrast, a stronger conservation can be observed among the active site motifs, the catalytic pathway, and the RdRp three-dimensional structure [126,130,136,137,138]. Indeed, structurally, all RdRps resemble the same, cupped right hand organisation (Fig. 1A and 1B) that consists of the three key subdomains that are analogously identified as fingers (including fingertips), palm and thumb (Fig. 1A) [138,139]. This typical polymerase fold is also found among cellular RNA and DNA polymerases [140,141]. Between RNA polymerases, significant differences may be observed in the extremities of thumb and finger subdomains which can in turn give the RdRp a more open structure (Fig. 1B) or a very closed structure (Fig. 1A). Further differences exist between RdRps, since additional N-terminal domains or additional subunits may be present (Fig. 1C-E, see also below).

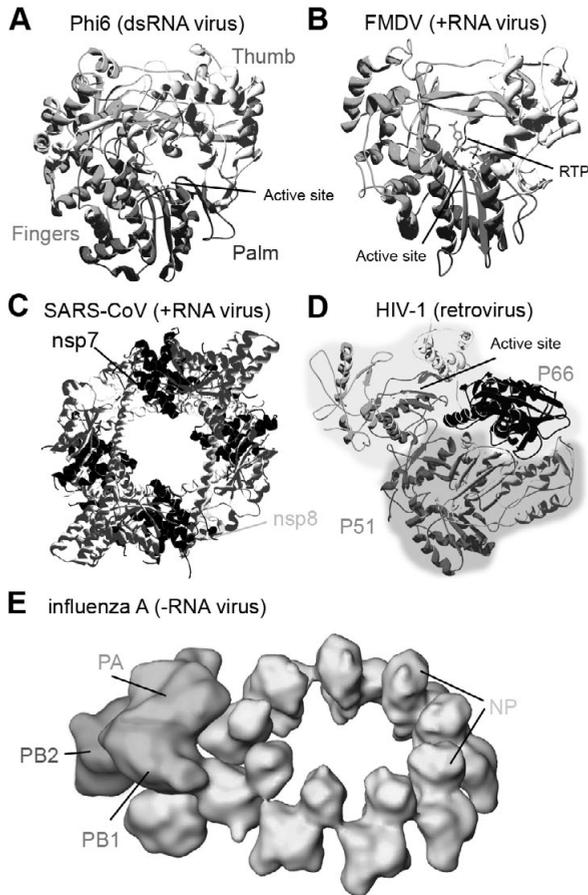
The most key and conserved structure of RdRps (and polymerases in general) is the palm subdomain structure, which is composed of a  $\beta$ -sheet with four anti-parallel strands and two  $\alpha$ -helices (Fig. 1A and B). Inside the palm subdomain, motifs A, B and C are the most prominent features, but only A and C are conserved among all viruses and cells [136,138,139,142]. These two motifs are positioned at the junction of the NTP and template channel (Fig. 2) and constitute together the active site of the RdRp that coordinates the binding of divalent metal ions and the positioning of the incoming nucleotide (Fig. 2B and C). Comparative sequence analysis has shown that motif A typically conforms to a  $Dx_{4-5}D$  consensus sequence, where 'x' represents any non-conserved residue. The C-terminal aspartate is not strictly conserved in this consensus, however, as it is also optional in motif C's  $xDD$  consensus sequence [143]. The upstream aspartates of both motif A and C on the other hand, are strictly required for activity, which easily follows from their spatial juxtaposition in the polymerase active site and their direct involvement in the coordination of the divalent ions that are vital for chemistry [136,138,139,142] (Fig. 2C).

On the palm subdomain, motif B generally contributes to the discrimination between dNTPs and NTPs, and involves at least one key asparagine residue in RdRps [144]. A complementary role in NTP binding has been proposed for motif F, but current evidence suggests that this motif is not completely conserved among all viral RNA virus polymerases, most notably those of the corona- and the retroviruses [89,125]. An additional, particularly prominent motif in the palm domain is motif D. This motif contributes a conserved histidine or lysine residue to the active site that can function as general acid and is thus crucial for activity (Fig. 2C). In addition, this motif has been shown to play a role in controlling the fidelity of nucleotide incorporation and, in particular, the selection for NTPs over dNTPs. Motif D's sequence is thus an important discriminating factor between RNA and DNA polymerases.

### Structural differences and processivity

Regarding the overall RdRp structure, the interaction of the thumb and finger subdomains in RdRps from +RNA and dsRNA viruses creates a more 'closed structure' that surrounds the template-binding channel and effectively allows the RdRp to fully encircle any bound template (Fig. 1) [138,139]. An important contributing part to the relatively closed structure is motif E. This motif is unique among RdRps and it is required to establish an interaction with the nascent strand. In RdRps that can initiate RNA synthesis *de novo*, an additional structure is present to close this site and called the primer-loop or  $\Delta 1$  loop in the hepatitis C virus (HCV) RdRp [145,146]. This structure was shown to be required for the formation and stabilisation of the first dinucleotide product. In addition, it may also contribute to *de novo* RNA synthesis by preventing the template-binding channel from accommodating duplexed or partially duplexed RNAs [145,147]. Alternatively, it was recently proposed that this loop may also function as an interaction interface during oligomerisation of the HCV RdRp in order to assist oligomer-dependent *de novo* initiation [148].

It is believed that an additional benefit of the 'closed structure' is to confer processivity to the RdRp, particularly since the relatively 'open structure' of DdDps requires additional factors to stabilise template binding. The T7 DNA polymerase gp5, for instance, readily dissociates from the template after the incorporation of a mere pair of nucleotides. For T7 to achieve processivity and an ~80 fold increase in affinity for the 3'-OH of the nascent strand, the 12-kDa host protein thioredoxin needs to associate with the polymerase thumb domain and close the polymerase structure [128]. Similar observations have been made for the vaccinia virus DNA polymerase, which needs to associate with a 48-kDa viral protein [149]; the herpes simplex virus polymerase, which recruits a cellular 51-kDa dsDNA binding protein [150]; and the mitochondrial DNA polymerase  $\gamma$ , which associates with a 35-kDa protein [151].

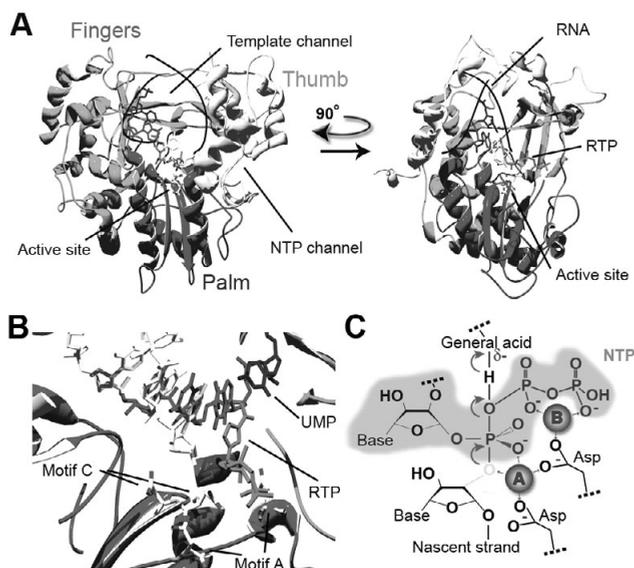


**Figure 1: Structures of viral RdRps and associated proteins.** (A) Structure of the compact Phi6 RdRp P2. The palm, fingers and thumb subdomains are indicated in different shades of grey. Image based on PDB accession 1HI0. (B) Structure of the relatively open FMDV RdRp and the binding of a ribavirin triphosphate (RTP) molecule to the active site on the palm subdomain. Image based on PDB accession 2E9R. Coding as in Fig. 1A. (C) Structure of the SARS-CoV nsp[7+8] hexadecameric protein complex. In this structure, the nsp7 subunits are shaded black, while the nsp8 subunits are grey. Image based on PDB accession 2AHM. (D) Structure of the HIV-1 RT. The P66 and p51 protein subunits are shaded light grey and dark grey. Image based on PDB accession 3V4I. The RNase H domain is shown in black. (E) Cryo electron microscopy-based model of the influenza A ribonucleoprotein complex. This complex consists of the three viral polymerase subunits PA, PB1 and PB2, and an NP-coated viral RNA. Image adapted from Coloma *et al.*

Although similar in structure to the above DNA polymerases, most viral DNA polymerases, all cellular DNA polymerases and the bacteriophage T4 DdDp obtain their processivity from binding to a DNA sliding clamp, a multimeric ring structure that requires ATP for assembly and has a central cavity to accommodate dsDNA [152]. Interestingly, a

similar structure was found in the analysis of two small SARS-coronavirus non-structural proteins (nsps) (Fig. 1C), namely the 22-kDa nsp8 and the 10-kDa nsp7 [153]. Together these subunits were shown to form a hexadecameric ring that is able to bind dsRNA and dsDNA, and may thus play a role in conferring processivity to the SARS-Cov primer-dependent RdRp nsp12 [153,154]. It is presently not yet known whether the nsp12 RdRp lacks the close interlinking of the fingers and thumb domain that is observed in the structures of other +RNA virus RdRps and would thus require a processivity factor. Furthermore, additional biochemical experiments with nsp8 have shown that this protein is capable of RNA both *de novo* and primer extension activity of its own as well [155,156], suggesting that it is not only a very unique RdRp, but that it may function also quite differently than was predicted from the first crystal structure.

The mechanism by which processivity is achieved in segmented minus strand RNA (-RNA) viruses is very different from +RNA virus RdRps and cellular polymerases. Indeed, processivity for these viruses appears to be attributable to the viral nucleocapsid protein (NP), which, although the structure of their heterotrimeric RdRp has yet to be com-



**Figure 2: Catalysis in the RdRp active site.** (A) Structure of the FMDV RdRp with RTP and RNA template bound in the template and NTP channel. On the palm subdomain, the aspartates of the active site are shown as sticks. Image based on PDB accession 2E9R. Coding as in Fig. 1A. (B) Close-up of Fig. 2A, showing the hybridisation of the RTP with the UMP moiety in the template strand. Motif A and C aspartates are shown as sticks to illustrate their involvement in the coordination of the triphosphate group of the RTP molecule. (C) Schematic of the RdRp active site and the role of motif A and C aspartates in binding two divalent metal ions (marked A and B). These metal ions neutralise the charge of the triphosphate of the incoming NTP (shaded grey) and coordinate the formation of the phosphor-diester bond at the 3'-OH of the nascent strand.

pletely solved [157,158], appears to be able to coat the genome segments and minimise the secondary structure of the template RNA to stimulate promoter escape [159]. This notion also appears to be supported by recent cryo-EM models of the viral RdRp (Fig. 1E), which suggest that the enzyme associates with the 25-nt promoter region at the periphery of an N protein-coated, single stranded genome segment [160].

### Additional domains

The functional RdRp domain of most RNA virus polymerases is approximately 600 amino acids in length, which is roughly the size of the prototypic poliovirus 3D<sup>pol</sup> RdRp. However, the number of amino acids that constitute the total RdRp subunit as produced by translation of the viral mRNAs significantly varies among viruses, even among closely related genera. Frequently, these supplementary amino acids are known to present one or more additional domains. In the pestivirus polymerase NS5B subunit, for example, the 720 amino acids also include a small N-terminal domain of still unknown function [146]. Interestingly, up to 90 residues of this N-terminal domain can be deleted from the bovine viral diarrhea virus (BVDV) without any loss of RdRp activity, whereas residues 91-106 are involved in RdRp activity and appear to fold as a separate domain that sits on top of the thumb subdomain [146,161].

Also the related flavivirus NS5 RdRp has an additional N-terminal domain, but this one enlarges the overall subunit to around 900 amino acids (~103 kDa) and harbours a 2'-O-methyltransferase (MTase) activity, which the NS5 subunit can use to generate 5' cap structures on newly synthesised viral genomes [162,163]. Large N-terminal domains are also present in coronaviruses [89], and although no biochemical experiments have been performed so far to test their enzymatic functions, protein expression studies have shown that the N-terminal residues of this domain greatly influence the stability of the nsp12 RdRp [154]. Interestingly, the enzymatic activity of the polymerase N-terminal domain can also be used to distinguish RdRps from DNA polymerases, given that the N-terminal domains of these enzymes typically have 3'→5' exonuclease activity and proofreading abilities [53,55,164,165,166,167,168,169], and that, in line with the relatively high mutation rate of viral genomes, such abilities have not been found among RdRps to date [35,47,170].

RNA virus polymerases may also contain additional domains in their C-terminus or in associated subunits. A prime example of the former is the human immunodeficiency virus (HIV) RT, which has RNase H activity in the p66 polymerase subunit [171] (Fig. 1D). This activity is used to internally cleave the RNA strand in the RNA/DNA hybrid molecules that form during reverse transcription reactions. An example of an RdRp with subunits that contain different enzymatic activity than the core RdRp subunit the influenza A heterotrimeric polymerase. This RdRp assembles via head-to-tail interactions between the viral RdRp subunits PA and PB1 and PB1 and PB2 (Fig. 1E). The RdRp active site is con-

tributed by the PB1 subunit, whereas endonuclease and cap-binding sites are present in PA and PB2, respectively [158,172]. Interestingly, even though the PB1 subunit contains all key RdRp sites, including motifs A-D, it cannot function on its own and requires at least the PB2 subunit to be soluble [158].

### Polymerase dynamics

Apart from combining different enzymatic activities in the same functional protein to improve, *e.g.*, replicase efficiency or to minimise the coding sequence of the genome, the integration of multiple enzymes into one protein unit does impose an important functional consequence on the polymerase dynamics as a whole, particularly when the two (or more) active sites are positioned opposite each other. A striking demonstration of these consequences was visualised in fluorescence resonance energy transfer (FRET) studies of the HIV-1 RT [173,174]. This enzyme needs to catalyse a multi-step process, including i) RNA-templated minus strand DNA synthesis, ii) primer synthesis through the cleavage of polypurine RNA templates, iii) DNA-templated plus strand synthesis, and iv) RNA primer hydrolysis with its RNase H domain to expose the HIV-1 integration sequences, all in order to convert the single stranded viral RNA genome into a dsDNA molecule ready for integration into the host cell genome [175]. In line with the juxtapositioning of the RNase and polymerase active sites, it was demonstrated that the enzyme can orient itself into an RNA hydrolysis position when RNA is bound or, when it encounters a DNA primer, in a position that allows the polymerase active site to be active [174].

These observations mean that a dissociation event is required to move the template strand from one active site to the other. This is in large contrast with the translocation of the nascent strand between the polymerase and exonuclease sites of DNA polymerases, given that this process appears to prefer an intramolecular pathway rather than dissociation into free solution and rebinding [52,176]. Interestingly, the HIV-1 RT can also rapidly swivel between its two binding orientations on the template when polypurine RNA primers are used to prime DNA synthesis. It was hypothesised that only the presence of dNTPs would fix it into a polymerase active state [174].

On a smaller level, the polymerase active site is also structurally highly dynamic and flexible. This was elegantly demonstrated through quantitative analysis of multiple-quantum NMR data of the  $\phi 6$  polymerase p2 and through modelling studies of the poliovirus 3D<sup>pol</sup> [177,178]. For  $\phi 6$  polymerase it was found that the RdRp domain displays essentially two types of motions, namely fast motion ( $k = 1200\text{--}1500\text{ s}^{-1}$ ) and slow motions ( $k = 500\text{--}800\text{ s}^{-1}$ ) [178], an observation that was corroborated in 3D<sup>pol</sup> simulations. Further analysis suggested that the fast motions involve residues that are in close proximity to the template tunnel (Fig. 2A) and the C-terminal part of the RdRp, and it was therefore assumed that particularly these motions contribute to RNA translocation in RdRps [177,178]. Indeed, the rates coincide well with the assumption that translocation

is coupled to pyrophosphate release, which takes place at  $1200 \text{ s}^{-1}$  in T7 RNA polymerase [179]. This also predicts that mutations that affect these fast rates have the ability to greatly impair the processivity of the RdRp.

The slow motions are a completely different story though. Consistent with the general notion that chemistry is the rate-limiting step in the overall polymerase reaction, as was shown through pre-steady state kinetic analysis [180], the slow motions chiefly involve motifs C, D and E of the  $\phi 6$  RdRp [178]. In line with this, the 3D<sup>Pol</sup> simulations displayed little flexibility in the palm domain, whose residues essentially moved on the time scale of catalysis [177]. Interestingly, the motion of the residues lining the NTP channel (Fig. 2A) was synchronised with the palm residues and biased towards the active site, suggesting that they actively pump nucleotides towards the catalytic site. Interestingly, the motions were also exactly anti-correlated with the motions of the nascent RNA-binding channel, which was able to open by  $20 \text{ \AA}$  [177]. These observations do not only suggest that the translocation mode of the RdRp is passive, given that the average minor groove distance of an RNA duplex is  $\sim 8\text{--}11 \text{ \AA}$  [181], but they also imply that during translocation, no solutes can reach the active site through the NTP channel and that the two processes are inherently separated in time.

### Template recognition

Each polymerase is tailored to function within each virus' niche and replication cycle. In essence this principle as well as the flow of replication - and to a certain extent even the structural elements in the RdRp - are effectively determined by the nature of the genome. When a +RNA virus infects a cell, it carries with it a linear, single-stranded genome that has evolved to allow initiation of viral gene expression immediately upon its release from the virion. +RNA genomes may therefore contain a cap-structure at their 5' end, a 3' polyA tail, an internal ribosomal entry site (IRES), internal purine-rich elements that mimic polyA tails, or a combination of these elements [58,182,183,184]. It is not until after translation and the proper processing of the set of viral (poly)proteins that it produces, that a viral RdRp starts viral RNA synthesis [58,134,185,186]. This delay is a small offer, however, because it allowed +RNA viruses to save space, time and resources and they need not package an RdRp together with the genome in the virion such as -RNA viruses. And because no RdRp packaging has to take place, the +RNA virus RdRp was not required to evolve structural elements to facilitate a stable association with the genome.

In contrast, the process of RdRp packaging is a vital step in the replication cycle of viruses with -RNA or dsRNA genomes. This characteristic is dictated by the fact that these molecules are not suited for immediate translation and need to undergo an initial round of transcription before viral proteins can be synthesised by the ribosome. For segmented -RNA viruses, RdRp packaging is achieved through a specific association of

the heterotrimeric polymerase with the vRNA promoter [187], a partially dsRNA structure that spontaneously forms when the 5' and 3' terminal ends of the -RNA genome segments hybridise (Fig. 3A) [188,189]. This promoter is also crucial for transcription initiation, 'cap-snatching', and the initiation of RNA replication or complementary RNA (cRNA) synthesis [188,190,191]. Packaging in dsRNA viruses on the other hand, is achieved by the translocation of single-stranded +RNA segments into a protein shell, called the polymerase complex [192,193]. This translocation activity is specific for the 5' ends of the genome segments and relies on the purine-dependent 5'→3' helicase activity of the packaging/helicase protein [192,194]. Inside the polymerase complex, multiple polymerase subunits (e.g., protein P2 in the *Cystoviridae* [193]) then copy the imported single-stranded genome segments into dsRNAs, recognising and initiating on conserved 3' terminal cysteine residues [195]. In the next round of infection, the complementary strands in these dsRNAs are then used as template for transcription, whereupon the viral mRNAs are translocated from the protein shell into the cytoplasm of the infected cell.

Although required at a different stage in the viral infection cycle, faithful recognition of the genome is also a prerequisite for the RdRps of +RNA viruses. In most +RNA viruses, this process is guided by secondary structures in the genomic RNA, the 3'-end of the viral genome (usually one or more purines) and, in *de novo* initiating RdRps such as the flaviviruses, a specific binding site for GTP. Particularly this latter element plays an important role in assisting the initiation nucleotide at position  $i$  at the 3'-end of the viral genome in attacking the second NTP at position  $i+1$  [40,146,196] (Fig. 3B).

In the replication of the very well studied poliovirus genome, for example, a specific structure at the 5' end of the genome serves as a general promoter for both +RNA and -RNA synthesis [197] (Fig. 3C). This structure is supported by a 5' UTR cloverleaf structure that serves as binding sites for the cellular poly(rC)-binding protein (PCBP) and viral 3CD<sup>pro</sup> polyprotein cleavage intermediate [198,199,200]. Together, these complexes facilitate circularisation of the genome by binding the C-terminal part of the cellular poly(A) binding protein (PABP) [201], a factor with affinity for purine-rich elements and of importance for controlling mRNA stability and translation [202,203]. Although the stoichiometry of the proteins involved has not been studied in great detail yet and the exact role of PCBP has been disputed, the general process of RNA structure-guided circularisation likely assists in triggering self-cleavage of the 3CD<sup>pro</sup> precursor and activation of the RdRp 3D<sup>pol</sup> on the correct template. In addition, 3D<sup>pol</sup> will initiate replication with uridylation of the viral protein VPg using a back-slide mechanism on the *cis*-replication element (CRE) resident inside the 2C coding region as template [197,204,205]. This enzymatic step creates the protein primer VPg(pUpU<sup>OH</sup>) that can hybridise to the viral polyA tail and facilitate a complete duplication of the genomic 3' end by 3D<sup>pol</sup> [206,207]. In addition, the protein primer can bind to the 3' terminal 5'-AA<sup>OH</sup> of the negative strand and thus effectively

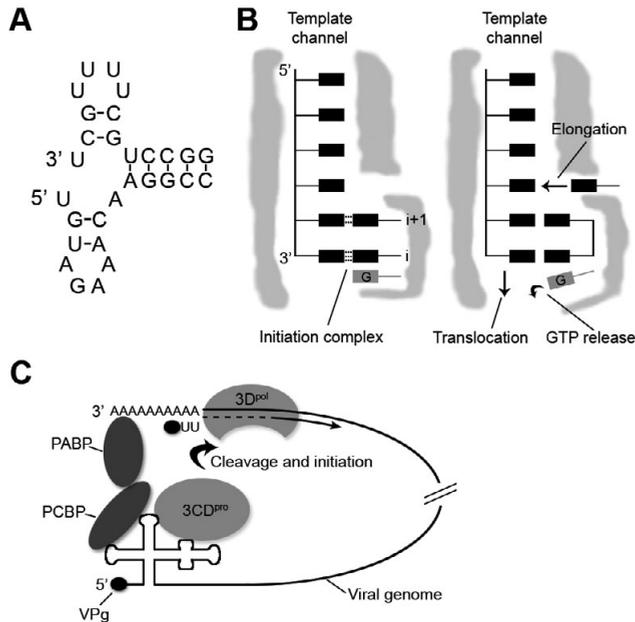
prime +RNA synthesis [197,208,209]. A study of the related foot and mouth disease virus (FMDV) has shown that the terminal uridylate of the VPg(pUpU<sup>OH</sup>) can occupy a similar position as an RNA primer in the active site of the polymerase during elongation [210].

Crucially, to allow positive-strand synthesis after an initial round of -RNA synthesis, the poliovirus 5' cloverleaf structure must refold at the 5'-end of the positive-strand RNA. This may be achieved through spontaneous unfolding of the ds poliovirus replicative intermediate at its 5' or, alternatively, via a 5'-specific helicase-driven unwinding event. Various observations have suggested that the poliovirus 2C protein could carry out this role, given its putative helicase activity based on bioinformatics analyses, its ATPase activity and its interaction with the 3' end of the negative-strand [211,212,213,214]. However, direct evidence for 2C helicase activity remains elusive and other roles for this protein, such as the anchoring of viral RNAs to membrane-bound replication complexes (discussed below) through its membrane domain have been proposed as well [215,216]. A third alternative may be presented by the activity of replicating or transcribing polymerases still present on the (partial) dsRNA replicative intermediate. As has long been recognised for dsDNA replication and transcription, the activity of a polymerase on the ds genome induces the formation of negative supercoils (a more closed helical structure) in front of the polymerase and positive coils (a more open helical structure) behind the polymerase. Interestingly, these changes in local topology have been shown to directly influence the accessibility of the ds template to the transcription machinery [217,218,219], likely due to the fact that a positively coiled template is easier to unwind than its negatively coiled counterpart. This effect may also be present in relatively long dsRNA molecules, particularly if their ends are attached to membranes (see below).

Circularisation of the genome is also an important step in the replication of flavivirus-es. Contrary to picornaviruses, but distantly resembling the interactions in segmented -RNA viruses, this is achieved through long-range interactions of conserved 5' and 3' UTR elements, which effectively position the conserved 5'-CU<sup>OH</sup> 3'-terminal sequence opposite the 5'-terminal end [220,221,222,223]. Biochemical and mutational evidence have shown that particularly the 3'-terminal 5'-CU<sup>OH</sup> of the genome is recognised by the flavivirus RdRp NS5 [224]. However, similar techniques have also demonstrated that a certain fraction of the genome must remain linear in order to support infectivity, possibly to allow translation [225]. Interestingly, RdRp recruitment and initiation of RNA synthesis are believed to take place on a stem-loop/promoter structure in the 5' UTR of the genome, an interaction that likely evolved to support NS5's role in 5' capping of the genome [223,226].

### **The influence of RdRp confinement**

Recognition of the genome is unfortunately not sufficient for effective RNA virus replication. At later stages in the infection cycle, the RdRp also needs to discriminate between



**Figure 3: Template recognition by RdRps.** (A) Structure of the influenza A virus RNA promoter that forms by spontaneous hybridisation of the terminal ends of each viral genome segment. This structure coordinates the binding of the influenza RdRp and its role in replication, transcription and genome segment packaging. (B) Model of *de novo* RNA synthesis by a flavivirus RdRp. The RdRp binds to the 3' end of the viral genome (left panel). Then the template, initiation NTPs, and an additional GTP form an initiation complex that is stabilised by the closed C-terminal loop. After the first polymerase reaction, the GTP is released (right panel) by opening of the C-terminal loop and translocation of the template–nascent strand duplex. A new NTP can enter the active site via the NTP channel to start elongation. (C) A model for the circularisation of the poliovirus genome via PCBP, PABP and the 3CD<sup>pro</sup> cleavage intermediate. This protein bridge ensure that the viral polymerase 3D<sup>pol</sup> is brought into close proximity with the 3' end of the viral genome where it can initiate -RNA synthesis using VPg as protein primer.

strands that are being replicated, strands that are to be translated and strands that will be committed to packaging in new virions. In -RNA replication - with the exception of some rhabdo- and paramyxoviruses - this discrimination is accomplished through the physical separation of the sites of replication and transcription (in the nucleus), and the sites of translation (cytoplasm) [227]. Genomes of +RNA viruses, however, are both replicated and translated in the cytoplasm, which would instigate a direct competition for templates in absence of control mechanisms. Alternatively, they may, in certain scenarios, benefit from the unwinding activity of the ribosomes to stimulate their own replication, similar to the enhancement of bacterial transcription by translating ribosomes [43]. Interestingly though, each +RNA virus replication complex associates with one or more ensphering membranes, which are, in reference to their morphol-

ogy, termed double-membraned vesicles (DMVs), invaginations, spherules, convoluted membranes (CMs) or the reticulovesicular network (RVN) [68,70,228]. Depending on the virus, these alterations are induced in mitochondria [75], endosomal vesicles [72], or the endoplasmic reticulum [67,68,70,229]

The rearrangements of cellular membranes are dramatic, readily identifiable and believed to be the result of a specific membrane modification strategy that helps the virus to confine the localisation of the replicating viral RNAs within the cytoplasm. Indeed, studies investigating the importance of these membranes showed that drugs targeted at specific membrane/transport pathways, such as with Brefeldin A, an inhibitor of the cellular secretory pathway [230,231,232], or mutations in or near membrane-spanning viral proteins can significantly impair the location, shape or activity of the replicase [71,233,234,235]. It has further been shown that viral RdRps can be anchored to or stabilised at virus-induced membranes, possibly via protein factors or via a direct phospholipid interaction [236,237]. RNAs may be anchored to the membranes as well, particularly the ones that are used for replication. In fact, a recent study of the flock house virus showed that the ~50 nm spherules that are formed by this virus contain on average 2-4 replicative intermediates [238].

Various reasons for the formation of these membrane alterations can be envisioned, including, i) physical support and organisation of the replication complex, ii) tethering of the viral (-)RNA during strand separation, iii) tethering of the viral RNA to induce genome topology, iv) compartmentalisation and elevation of the local viral protein concentration, v) filtering unwanted interactions with the pool of cellular and viral mRNAs committed to translation, and vi) protection of the viral RNA from dsRNA-mediated host defences, such as RNA interference. A putative seventh could be an improved regulation of the RdRp activity. In poliovirus replication, the same 5'-proximal elements that target the poliovirus RdRp to its correct template may also function as molecular switch between replication and translation. The viral protease 3C only associates with the 5' cloverleaf structure as polyprotein cleavage intermediate 3CD<sup>pro</sup> (Fig. 3C) and needs to associate with PABP and PCBP [199]. Interestingly, both 3C and 3CD<sup>pro</sup> can cleave PABP and PCBP, and thereby terminate their interactions with the viral genome [239,240]. However, this step would also prevent circularisation of the genome and thereby its replication. Consequently, the local concentration or the prolonged presence of 3C could function as molecular switch and transfer a viral genome from a replication process to translation or packaging, or *vice versa* [241].

However, proof that the RNA-synthesising activity resides inside the membranous complexes has not been conclusively demonstrated for all +RNA viruses, and the lack thereof together with the absence of a detectable connection between the inside of DMVs with the cytoplasm (and crucially the ribosomes and sites of encapsidation) has questioned the mechanism and location of nidovirus replication [68].

### Fidelity, repair, and non-templated extension

Overall, a vital requirement for every polymerase is the duplication of the genome with high fidelity. Many factors can influence this process, however. For instance, the fidelity of the DNA-dependent DNA synthesis by the HIV-1 RT - this enzymatic phase that takes place after reverse transcription in order to produce a dsDNA ready for integration in the host genome - increased 9-fold when the pH of the *in vitro* reaction was lowered from pH 8.0 to pH 6.5, as determined using a base substitution reversion assay [242]. Similarly, nucleotide analogues such as ribavirin have a dramatic impact on the mutation rate of replication [39,56], as do various divalent metals [180,243], or amino acid mutations that affect the incorporation rate or structural dynamics of the RdRp [177,244]. However, even in absence of such factors, the mutation rate of RNA viruses, and the mutation rate of +RNA viruses in particular, is already several orders of magnitude higher than the mutation rate of DNA genomes [47,170]. This has important medical consequences, since few antivirals can remain effective against +RNA viruses when the high mutation rate facilitates a rapid development of drug resistance.

In spite of their high error frequency, dramatic mutations or deletions can be repaired in the viral genome. It was shown, for instance, that the RdRp of turnip crinkle virus (TCV) is able to repair mutations in the 3' sequence CCUGCCC<sup>OH</sup> sequence of its genome using a 3-step process. Paradoxically, the first step in this process is the non-templated synthesis of an RNA primer, which can then be aligned with the 3' end and extended in a primer-dependent fashion [245]. Different repair mechanisms have been observed for the dengue virus RdRp NS5. This enzyme, which can initiate *de novo* on the 5' end of the genome, was found to correct mutations and deletions in the 3' UTR, provided they still allowed circularisation and were smaller than 6 nt [246]. Although conclusive evidence for the mechanism is still elusive, the authors proposed that the viral RdRp was able to correct these errors through its terminal transferase activity [246]. The use of terminal transferase activity for repair is of course paradoxical, as it generally entails the non-templated and thus mostly random addition of nucleotides to the 3' end of templates [123,124,125] and would thus create a pool of quasispecies. It is nevertheless a possibility that this process is, in part, reminiscent of the strategy used by lesion-repair polymerases, which employ their relaxed fidelity to rescue a stalled replication process, and that a subsequent selection step would enable the virus to filter back the wild-type sequence [246]. Additionally, the RdRp may use homologous recombination or template switching to correct its genome more rapidly.

Terminal transferase activity has been observed for many RdRps and is generally linked to a specific template or nucleotide substrate [247,248,249]. Poliovirus 3D<sup>pol</sup> for instance, is able to catalyse the addition of AMP molecules (*i.e.*, adenylyl transferase, TATase) to the 3' end of the genome [247]. Similar activity has been observed for the alphavirus polymerase [250], suggesting that this activity may be a common feature of RdRp termi-

nation and polyA-tailing of the genome. However, many other RdRps, particularly those of dsRNA viruses, do not portray such strict nucleotide specificities under *in vitro* conditions, such as the pancreatic necrosis virus VP1 polymerase [249]. Furthermore, other reports have implicated internal or terminal genomic polyU sequences as templates for polyA-tail synthesis [251,252,253].

### **Additional polymerases and multimerisation**

In prokaryotic and eukaryotic cells, which evolved more sophisticated genetic codes and a higher number of transcription units than RNA viruses, control over the replication and transcription machinery is enhanced by delegating activities to distinct core enzymes, such as primases [23], leading and lagging strand polymerases [254], and translesion synthesis (TLS) polymerases, which can accommodate damaged DNA in their active site at a loss of specificity to keep the elongation process going [255,256]. But these enzymes also cooperate to complete their function. During eukaryotic replication initiation, for instance, a careful interplay of the proliferating cell nuclear antigen (PCNA) clamp protein, the clamp-loader complex replication factor C (RFC), the minichromosomal maintenance helicase (MCM), and the single-stranded binding protein replication protein A (RPA), launches a total of three polymerases (*i.e.*,  $\alpha$ ,  $\delta$ , and  $\epsilon$ ) and thereby replication. In addition, up to three polymerases may be associated with a single replicative helicase to allow a rapid exchange between polymerases in order to improve the overall processivity of the replication process [257].

Such heavily coordinated processes are often not associated with the seemingly singular viral RdRps, but cooperation and multimerisation have nevertheless been observed between them. For instance, an oligomerisation-driven stimulation of RdRp activity has been reported for the primer-dependent poliovirus 3D<sup>pol</sup> and the primer-independent norovirus RdRp [258,259,260]. In fact, some evidence suggests that the poliovirus 3D<sup>pol</sup> may even form lattices at sufficient local concentration. In addition, the HIV RT is effectively a polymerase pair as well, given that it is a complex of the HIV protein p66 and its cleavage product p51 [261] (Fig. 1D, note the two red palm subdomains). Furthermore, dimerisation of the HCV RdRp has been demonstrated to stimulate *de novo* initiation and the presence of multiple heterotrimeric influenza RdRps may stimulate or even be a prerequisite for transcription [262,263]. Interestingly, oligomerisation among PV RdRps has also been shown to be beneficial for replication even if not all RdRps are catalytically active [258]. Interestingly, such effects could be negated through the introduction of specific mutations in the N-terminal domain of the polymerases, suggesting that this domain is important for polymerase crosstalk.

A more extreme case of the presence of multiple polymerase activities within a viral replication complex and one that may even resemble cellular principles is presented by the coronaviruses (CoVs). These viruses, which are particularly well known for their

uniquely large ~30-kb +RNA genomes, encode at least two distinct polymerase activities, including an RdRp with a canonical, conserved RdRp domain of ~600 amino acids and a non-canonical, multimeric RdRp [154,155,156]. It has been proposed that these two enzymes may cooperate during replication or transcription of the genome to improve the processivity of these reactions. In addition, they may assume roles such as primase and primer-dependent polymerase akin to the cellular replication complex [154,155], putatively in order to achieve better control over initiation and elongation. However, convincing biochemical proof for these concepts has yet to be shown.

### Conclusions and outlook

Although RNA replication varies greatly among RNA viruses and many different strategies exist, common principles clearly exist in the RdRp structure and the organisation of their replication machinery. In addition, it is now also understood that viral RdRps can be targeted by similar approaches, some of which may affect the fidelity of viral replication and thereby virus survival. These factors include nucleotide analogues such as ribavirin, but also amino acids that are located far away from the active site but can nevertheless influence the dynamics of the RdRp structure [177].

However, in spite of these common principles, ultimately the differences in the functionality of the RdRp and the location of the RdRp during the virus lifecycle specify the survival-strategy and pathogenicity of each and every virus. For instance, as mentioned above, some RdRps are able to correct mutations or deletions in the 3' UTR of the viral genome and it has been proposed that this activity evolved to counteract host cell exonucleases. However, other viruses appear to have not evolved this property at all and instead employ the cellular exonucleases for the production of specific subgenomic RNAs, which in turn contribute to the pathogenicity of the viruses. Furthermore, whereas +RNA viruses use membranes to support their replication complexes, -RNA viruses typically coat their viral RNA templates with viral protein, and dsRNA viruses replicate and transcribe inside (sub)viral particles.

Consequently, although a daunting amount of information is already available on viral RdRps and some of that knowledge may be extrapolated to explain and predict functions in other viruses, a specific knowledge of different viral RdRp functions will likely be important if we aim to target and frustrate viral RdRps with (novel) antiviral therapies.



# Chapter 3

The RNA polymerase  
activity of SARS-  
coronavirus nsp12 is  
primer dependent

Aartjan J.W. te Velhuis<sup>1</sup>, Jamie  
J. Arnold<sup>2</sup>, Craig E. Cameron<sup>2</sup>,  
Sjoerd H.E. van den Worm<sup>1</sup> and Eric  
J. Snijder<sup>1</sup>

<sup>1</sup> Molecular Virology laboratory,  
Department of Medical  
Microbiology, Center of Infectious  
Diseases, Leiden University  
Medical Center, PO Box 9600,  
2300RC Leiden, The Netherlands.<sup>2</sup>  
Department of Biochemistry and  
Molecular Biology, Pennsylvania  
State University, University Park,  
Pennsylvania 16802, USA.

**Published in:** Nucleic Acids  
Research, Volume 38, Pages 203-  
214, 2010. Epub October 2009.

## ABSTRACT

An RNA-dependent RNA polymerase (RdRp) is the central catalytic subunit of the RNA-synthesizing machinery of all positive-strand RNA viruses. Usually, RdRp domains are readily identifiable by comparative sequence analysis, but biochemical confirmation and characterisation can be hampered by intrinsic protein properties and technical complications. It is presumed that replication and transcription of the ~30-kb SARS coronavirus (SARS-CoV) RNA genome are catalysed by an RdRp domain in the C-terminal part of nonstructural protein 12 (nsp12), one of sixteen replicase subunits. However, thus far full-length nsp12 has proven refractory to expression in bacterial systems, which has hindered both the biochemical characterization of coronavirus RNA synthesis and RdRp-targeted antiviral drug design. Here, we describe a combined strategy involving bacterial expression of an nsp12 fusion protein and its *in vivo* cleavage to generate and purify stable SARS-CoV nsp12 (106 kDa) with a natural N-terminus and C-terminal hexahistidine tag. This recombinant protein possesses robust *in vitro* RdRp activity, as well as a significant DNA-dependent activity that may facilitate future inhibitor studies. The SARS-CoV nsp12 is primer-dependent on both homo- and heteropolymeric templates, supporting the likeliness of a close enzymatic collaboration with the intriguing RNA primase activity that was recently proposed for coronavirus nsp8.

## INTRODUCTION

The central enzymatic activity in the life cycle of positive-strand RNA (+RNA) viruses is the process of RNA-templated RNA synthesis. In mammalian cells, the replication of +RNA genomes always occurs in the cytoplasm where the viral RNA-synthesising machinery is commonly associated with (modified) cytoplasmic membranes, which presumably provide physical support and may serve to protect the complex from host defense mechanisms [228,235]. The viral enzyme complex contains an RNA-dependent RNA polymerase (RdRp) as its core catalytic subunit for the production of negative strand RNA (-RNA), new genome molecules, and in many virus groups also subgenomic (sg) mRNAs [235,264].

Coronaviruses (CoVs) are a family of +RNA viruses that infect a wide range of vertebrates including humans, as exemplified by Severe Acute Respiratory Syndrome (SARS) and various respiratory infections caused by less virulent family members [65]. The group is also known for its exceptionally large polycistronic genome of approximately 30 kilobases, which is 5' capped and 3' polyadenylated RNA. The 5'-proximal two-thirds of the CoV genome (open reading frames 1a and 1b) are translated into the viral replicase polyproteins pp1a and pp1ab. The 3'-proximal third encodes the viral structural proteins and several so-called accessory proteins, which can only be expressed following the production of a set of four to nine sg mRNAs [51,265]. It is in particular this coronavirus hallmark that has attracted attention over the years, since these sg mRNAs consist of sequences that are noncontiguous in the viral genome: a common 5' leader sequence is attached to body segments representing a variable part of the 3'-proximal genomic region. According to the now widely supported transcription model of Sawicki and Sawicki [266], the arterivirus and coronavirus sg RNAs derive from a discontinuous step during minus-strand RNA synthesis that is guided by specific RNA signals and resembles copy-choice RNA recombination [51,265,267]. This mechanism produces a set of subgenome-length negative-strand templates from which the sg mRNAs are produced. The unusual genome size of coronaviruses and the complexity of their RNA synthesis hint at an RdRp that may have special properties, either by itself or in conjunction with other subunits of the coronavirus replication/transcription complex (RTC).

Coronavirus replication and transcription are driven by the 15 or 16 viral nonstructural proteins (nsps) encoded in the replicase gene, which include many enzyme activities that are rare or lacking in other families of +RNA viruses [35,58]. These nsps are produced during co- and post-translational processing of the pp1a and pp1ab replicase polyproteins by two or three internal viral proteinases [44]. In the years following the 2003 SARS outbreak, bioinformatics, structural biology, (reverse) genetics and biochemical studies have all contributed to the in-depth characterisation of coronavirus nsps in general and those of SARS-coronavirus (SARS-CoV) in particular [35,58,268]. Currently documented

enzyme activities include proteinases [44], a putative RNA primase [155], a superfamily 1 helicase [109], an exo- and an endoribonuclease [105,269], ssRNA binding proteins [270,271] and two methyltransferases [272,273]. Following the sequence analysis of the first coronavirus genomes, more than two decades ago, the putative RdRp domain of the coronavirus RTC had already been identified by comparative sequence analysis [84,274]. The C-terminal two thirds of the 932 amino acid subunit now known as nsp12 were confidently aligned with the conserved motifs of well-known RdRps [84,89]. In view of its unprecedented size among viral RdRp subunits, which commonly consist of 500-600 amino acids, nsp12 may well harbour other functional domains in its as yet uncharacterized N-terminal domain. Biochemical information on the coronavirus RdRp has remained scarce thus far though, in particular since full-length nsp12 was refractory to expression in bacterial systems. Preliminary evidence for the *in vitro* RdRp activity of nsp12 came from a study of the SARS-CoV enzyme by Cheng *et al.* [275]. However, their experiments, employing a glutathione S-transferase (GST)-nsp12 fusion protein, were hampered by protein instability, which resulted in the fragmentation of the protein into three parts. Primer-dependent activity on poly(A) templates was observed in filter binding assays, but the enzyme biochemistry was not further characterized in detail.

Given its pivotal role in viral replication and the efficacy of polymerase inhibitors used to combat other virus infections, the SARS-CoV RdRp is widely regarded as an important and attractive target for the rational design of anti-coronavirus drugs [89,276]. We therefore sought to solve the technical issues described above and develop reliable protocols for stable expression, purification and *in vitro* activity of the full-length coronavirus nsp12. In this paper, we describe an expression strategy utilising *in vivo* cleavage in a bacterial expression system by ubiquitin carboxyl-terminal hydrolase 1 (Ubp1), which releases the small N-terminal fusion partner ubiquitin (Ub) and generates a recombinant nsp12 protein containing its natural N-terminus. This was found to be a critical step on the road towards purification of a stable recombinant SARS-CoV RdRp. We subsequently established that full-length nsp12 can readily synthesise RNA in a primer-dependent fashion on both homo- and heteropolymeric RNA templates and is also efficient at incorporating deoxy-nucleotides on a DNA template. This work therefore provides a solid basis for the further biochemical characterisation of the coronavirus nsp12 and the detailed dissection of its role in viral replication and transcription. Understanding the enzyme's properties will be essential for the development of selective coronavirus RdRp inhibitors and will likely provide novel insights into the intriguing (evolutionary) relation between the RdRp of nidoviruses and other groups of +RNA viruses.

## RESULTS

### Stable expression and purification of full-length SARS-CoV nsp12

A PCR amplicon encoding SARS-CoV nsp12 was first cloned into T7 expression vector pET26-Ub-CHis<sub>6</sub>, which provided the 932-aa viral protein with an N-terminal Ub tag (76 amino acids) and a C-terminal hexahistidine tag (Fig. 1A). To avoid the known risk of T7 RNA polymerase contamination when purifying another RdRp using a T7 promoter-driven expression system (like *E. coli* BL21(DE3)), the Ub-nsp12-CHis<sub>6</sub> fusion gene was subsequently transferred to plasmid pASK3, placing it under control of a tetracycline-inducible promoter. Initially, the protease that was later used to remove the Ub tag was not co-expressed and the expression of the full-length Ub-tagged nsp12-CHis<sub>6</sub> was compared to that of variants containing either maltose binding protein (MBP) or thioredoxin (Thio) as fusion partner. As shown in Fig. 1B and as also documented in the previous work of Cheng *et al.* [275], expression and purification of nsp12 fusion proteins resulted in the abundant and progressive fragmentation of the protein. Based on their sizes in SDS-PAGE, the cleavage products likely were the 64- and 39-kDa nsp12 degradation products as found by Cheng *et al.* when working with their GST-nsp12 fusion construct [275] (Fig. 1B). Therefore, our results confirmed and extended the previously reported observations regarding the instability of SARS-CoV nsp12 when expressed in *E. coli* as part of a fusion protein.

In studies with the poliovirus (PV) RdRp (3D<sup>pol</sup>), it was previously shown that N-terminal foreign sequences, even the addition of a single methionine residue, can have a significant impact on enzyme activity [277,278]. Structural studies subsequently revealed that a native N-terminus is required for the proper folding of the RdRp's "fingers" subdomain and the positioning of the 3D<sup>pol</sup> active site [278]. These findings suggested to us that N-terminal additions might also hamper the proper folding of the SARS-CoV RdRp and could thus be the principal cause of its subsequent degradation. We therefore co-expressed the carboxy-terminal Ub protease Ubp1 [277], which specifically recognises the cleavage site in our fusion protein that is located between the Ub moiety and the N-terminus of nsp12 (LRGG↓SADAS). Proteolytic cleavage of the fusion protein at this site liberated an nsp12-CHis<sub>6</sub> product (Fig. 1A), now containing its natural N-terminus. The expression of the Ubp1 protease was driven by the relatively weak constitutive ADH1 promoter [277]. Hence, to ensure optimal release of nsp12-CHis<sub>6</sub> by the protease, induction of Ub-nsp12-CHis<sub>6</sub> expression was performed only in the middle or late log-phase of bacterial growth. Induction at lower densities or on-column cleavages performed with the cleavable MBP and Thio fusion proteins resulted in significant amounts of degraded nsp12, which was inactive in RdRp assays (data not shown).

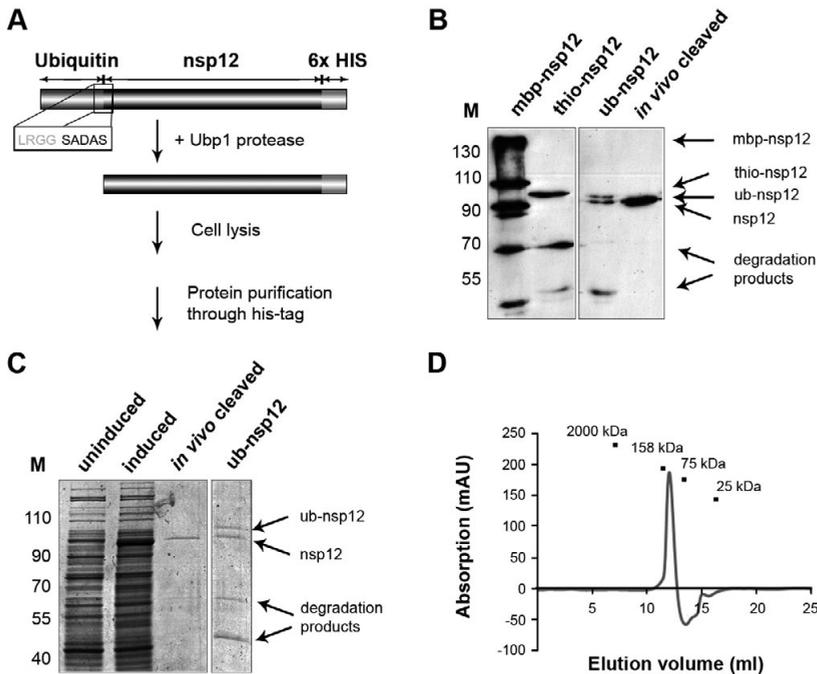
As depicted in Fig. 1B and C, the *in vivo* cleavage product that we were able to routinely purify was of the expected 106-kDa size [58] and was found to be stable, even

after multiple rounds of purification and prolonged storage at  $-20^{\circ}\text{C}$ . Non-denaturing gel filtration yielded a single peak corresponding to a size of about  $140 \pm 10$  kDa as estimated by comparison to a standard curve (Fig. 1E-F). This is different from the calculated mass of 106 kDa, but may be explained by, *e.g.*, a very open or non-globular structure of the protein. Moreover, unexpected behaviour in gel filtration is not without precedent for RdRps, as was, *e.g.*, reported for the monomeric  $\phi 6$  RdRp P2 [195]. Future structural studies involving electron microscopy or crystallography will hopefully allow us to address this question in more detail.

### SARS-CoV nsp12 is a primer-dependent RdRp

+RNA virus replication may proceed through a variety of scenarios, but essentially viral RdRps employ only two fundamentally different mechanisms during the first step of RNA synthesis: initiation can occur either *de novo*, when the 3' hydroxyl group of the first nucleotide essentially is the 'primer' to which a second nucleotide is added, or initiation can depend on an oligonucleotide or protein primer [134]. Imbert *et al.* [155] recently showed that the coronavirus nsp8 subunit, a relatively small protein (22 kDa) that forms a hexadecameric complex with nsp7 [153], can synthesise short oligonucleotides using polyC templates or templates containing a 3' CCG/U motif. Although the exact role of nsp8 remains to be studied in much more detail, one of its proposed functions is that of an 'RNA primase' [155]. According to this hypothesis, and similar to what has been documented for cellular DNA polymerases [23], the oligonucleotide products of nsp8 activity would serve as primers for the viral 'main RdRp', *i.e.*, nsp12. Some other observations were also consistent with the possibility that the SARS-CoV main RdRp belongs to the class of primer-dependent polymerases. Firstly, modelling of the structure of SARS-CoV nsp12 residues 400-900 suggested that the central domain of the protein contains a G motif, a signature present in all primer-dependent RdRps [89]. Secondly, when using the unstable GST-nsp12 fusion protein in combination with a polyA template and a polyU primer, Cheng *et al.* [275] observed a moderate RdRp activity in filter-binding assays. Convincing evidence for the extension of an RNA primer *in vitro*, however, was not documented thus far.

To qualitatively assess the polymerase activity of the nsp12, we monitored the incorporation of nucleotides on a well-defined synthetic RNA template, which was predicted to lack secondary structures that could affect the enzyme's processivity. This template (Fig. 2A) consisted of a 20-nucleotide duplex and a 3' overhang of 10 uridylates, which could serve as template during extension of the 20-mer primer. As shown in Fig. 2, primer extension was revealed by both the incorporation of  $[\alpha\text{-}^{32}\text{P}]\text{ATP}$  (Fig. 2B) and the extension of the  $^{32}\text{P}$ -labelled primer (Fig. 2C). The use of single-stranded or homopolymeric (polyU or polyC) templates did not lead to significant incorporation of nucleotides (data not shown).



**Figure 1: Expression and purification of the SARS-CoV nsp12.** (A) Expression of nsp12 was performed in the presence of the ubiquitin (Ub) protease Ubp1, which removed the N-terminal Ub fusion partner by cleaving after the LRGG signature that formed the junction between the Ub and nsp12 moieties. This *in vivo* cleavage created the native nsp12 amino-terminus with the sequence SADAS. (B) Western blotting using an nsp12-specific polyclonal antiserum identified a single band of the expected mass after purification of the *in vivo* cleaved expression product. Isolation of fusion proteins with different N-terminal tags resulted in common patterns of degradation. Similar results were reported by Cheng *et al.* for a GST-nsp12 fusion protein [260]. (C) Purified nsp12-CHIS<sub>6</sub> was readily visualised by Coomassie staining as a ~106 kDa protein product, in line with the expected mass. Inductions in the absence of Ubp1 expression, however, resulted in a mixture of full-length fusion protein and degradation products, likely as a result of improper folding. No RdRp activity was observed for these preparations. (D) Gel filtration analysis of purified nsp12 in assay buffer showed a single peak, corresponding to a globular molecular weight of  $140 \pm 10$  kDa, suggesting that only the monomeric form of nsp12 is present under the conditions used.

Based on sequence alignments of SARS-CoV nsp12 with RdRps of known structure, it was previously concluded that the canonical aspartic acid residues responsible for the coordination of the divalent cation in the RdRp's active site should be nsp12 residues 618, 760 and 761, the latter two amino acids being part of the nidovirus-specific SDD motif [89,196,279,280]. In other viral RdRps, single or multiple substitutions of the equivalent residues with alanine or asparagine were shown to change or even abolish polymerase activity [281,282], depending on the assay conditions used. For the PV 3D<sup>pol</sup> for example, it was found that mutagenesis of the motif C core sequence, changing it from GDD to GDN, was detrimental in the presence of Mg<sup>2+</sup>, but not when Mn<sup>2+</sup> or Fe<sup>2+</sup> were provided

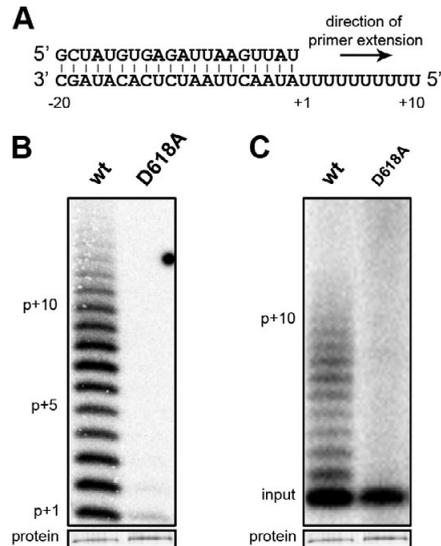
as divalent cation [281]. To obtain an active site mutant that could be used as negative control during further development of our assays, we engineered nsp12 mutant D618A. As portrayed in Fig. 2B and C, the mutant protein displayed only  $8 \pm 3\%$  residual primer extension activity in our assays (Fig. 2B-D), thus confirming the nsp12-specific nature of the RdRp activity described here and supporting the prior hypothesis that this residue is a key player in the enzyme's active site. Further or complete knockdown of activity may be obtained under slightly different assay conditions or in combination with substitution of other active site residues, which remain to be tested in future studies.

### Biochemical determinants of the *in vitro* activity of the SARS-CoV nsp12

It is common practice to employ filter binding assays during optimisation of polymerase activity assays. Although convenient and less time-consuming than gel analysis of reaction products, these assays merely reveal incorporation of label, but do not provide direct information on the length and integrity of the products. Therefore, when exploring the effects of different reaction buffer compositions, we consistently quantified and analysed [ $\alpha$ - $^{32}\text{P}$ ]ATP incorporation using denaturing 20% acrylamide gels. The standard assay buffer consisted of 20 mM Tris-HCl (pH 8.0), 10 mM KCl, 1 mM DTT, 6 mM  $\text{MgCl}_2$ , 5% glycerol, 0.1% Triton-X100 and 50  $\mu\text{M}$  ATP.

First, the influence of reaction temperature on RdRp activity was tested within the range of 20–37°C. Activity at 37°C was 40% higher than at 30°C, which is more commonly used in RdRp assays (Fig. 3A) [196,243,263,283]. However, a more detailed analysis revealed that this increase largely stemmed from enhanced initiation of primer extension rather than elongation, with the latter being decreased by 35% as judged from the amount of full-length product (primer+10, Fig. 3B). Therefore, all subsequent experiments were performed at 30°C. Contrary to expectations however, the major extension product never matched the length of the primer plus 10 nucleotides, suggesting that nsp12 generally terminates before reaching the end of the primed polyU template. The results obtained with other templates will be discussed below.

It is well established that nucleic acid polymerases utilise divalent cations, mostly  $\text{Mg}^{2+}$ , as cofactors to bind and coordinate the incoming nucleotide during the polymerisation reaction [279,280]. As shown in Figs. 3C-D, the activity of nsp12 was positively correlated with the  $\text{Mg}^{2+}$  concentration and reached its optimum at 6 mM. Higher concentrations ( $\geq 20$  mM) ultimately abolished activity (data not shown). The SARS-CoV RdRp does not require  $\text{Mn}^{2+}$  for RNA synthesis, as previously reported for the nsp9-RdRp of the arterivirus equine arteritis virus (EAV), a distant relative in the order of the nidoviruses [196]. In fact, the presence of  $\text{Mn}^{2+}$  in an RdRp assay utilising a polyU template decreased the fidelity of the SARS-CoV nsp12 relative to  $\text{Mg}^{2+}$  and facilitated both transversional and transitional misincorporations (Fig. 4). The specificity for the nucleotide's sugar moiety seemed not affected, however, since dATP was not incorporated when using an RNA



**Figure 2: SARS-CoV nsp12 primer extension assay and active-site mutant.** (A) Schematic showing the structure of the partially double-stranded RNA template with 3' U<sub>10</sub> stretch that served as template for primer extension in the initial nsp12-based nucleotide incorporation assay. (B) Comparison of the RdRp activity of wild-type nsp12 and a D618A active-site mutant, which displayed minimal activity after a 60 min incubation at 30°C. (C) A primer extension assay using a 5' <sup>32</sup>P-labelled primer and ATP confirmed that wild-type nsp12, in contrast to the D618A mutant, was able to elongate the 20-mer primer. The slight reduction of input primer in the D618A lane probably resulted from degradation by RNase activity. The D618A mutant showed 8 ± 3% residual activity (mean of 3 independent experiments). In the loading control (Fig. 2B and C, lower panels) the two nsp12 variants were visualized by silver staining of an SDS-PAGE gel.

template in the presence of Mn<sup>2+</sup>. This discriminates the SARS-CoV RdRp from, *e.g.*, its PV equivalent, for which correct base pairing was more important than the presence of a 2' hydroxyl group [243].

To establish the importance of pH, we tested SARS-CoV RdRp activity in the pH range of 6.0-8.5. We clearly observed an increase of nsp12 overall activity with an optimum at pH 8.0 (Fig. 3E), but differences between pH 7.0 and 8.5 were marginal in terms of both overall activity and amount of full-length product produced (Fig. 3F). Interestingly, pH 7.5 consistently resulted in a significantly larger amount of initiation product, suggesting that higher pH values stimulate the processivity of the enzyme.

In additional assays, we tested the effect of monovalent cations, bovine serum albumin and spermidine (data not shown). No negative influence of potassium or sodium ions was observed below a concentration of 50 mM. Neutral to slightly negative effects were found for both BSA (tested for 0-0.1 mg/ml), which is known to stabilise enzymes and reduce potential side-effects like absorbance to the plastic surface of the reaction tube used. Finally, addition of spermidine, which binds to the phosphate backbone of

nucleic acids and is a known stimulator of various polymerases, did not improve activity in the tested concentration range up to 5 mM.

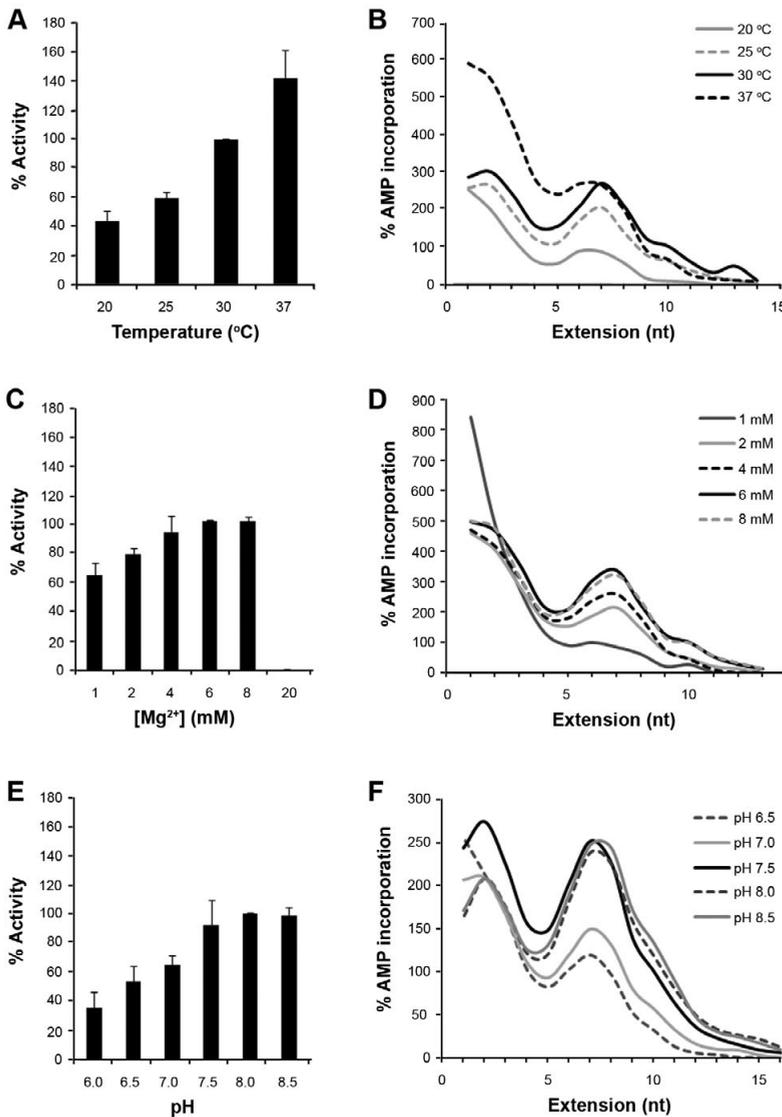
### RNA binding properties of SARS-CoV nsp12

Our combined observations suggested that nsp12 is only active in a primer-dependent fashion and may thus prefer binding to double-stranded RNA (dsRNA). To determine the affinity of nsp12 for dsRNA and single-stranded (ssRNA) we performed electrophoretic mobility shift assays, in which we titrated purified nsp12 in the presence of 0.2 nM 5' <sup>32</sup>P-labelled RNA (Fig. 5). Subsequently, the percentage of bound template and the Hill equation (see Materials and Methods) were used to derive the enzyme's dissociation constants ( $K_d$ ). For dsRNA the  $K_d$  was calculated at  $0.13 \pm 0.03 \mu\text{M}$ , whereas the  $K_d$  for ssRNA was estimated to be  $0.10 \pm 0.02 \mu\text{M}$ .

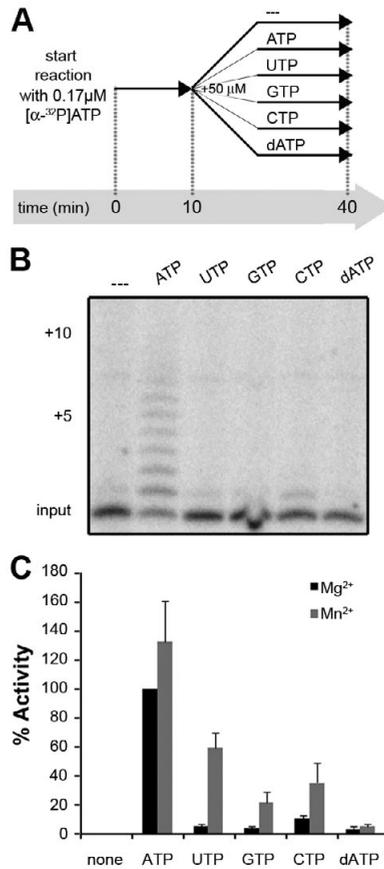
### Nsp12 nucleotide incorporation rates

For various viral RdRps, *in vitro* studies have defined biochemical properties that could not be determined or were largely obscured in *in vivo* assays. Furthermore, following the basic characterisation of individual enzymes, the interplay with templates, co-factors and other subunits of the viral enzyme complex can be addressed. Ultimately, this should facilitate the *in vitro* reconstitution of the holo-enzyme complex and the analysis of its interactions at each step of viral replication and transcription. After the purification of active SARS-CoV nsp12 and exploration of the reaction conditions, our subsequent step towards such a future goal was to evaluate the rate of nucleotide incorporation by the SARS-CoV RdRp. On a primed U<sub>10</sub> template we found relatively poor incorporation rates ( $1.2 \pm 0.4 \text{ nM/min}$ ) and a major product that was 2-3 nucleotides shorter than the expected full-length product (Fig. 6A), with product sizes displaying a similar Gaussian-like distribution as depicted in the graphs in Fig. 3. On the other hand, we also observed minor products that were slightly longer than full-length, suggesting slippage of the polymerase at the end of the template. We did not detect any terminal transferase activity on the templates used (data not shown).

Previously, Cheng *et al.* reported that their unstable GST-nsp12 fusion protein had limited reverse transcriptase activity [275]. In contrast to these prior results, our nsp12 was not able to efficiently use dATP to extend an RNA primer (Fig. 4A and B) and a similar result was obtained with a DNA primer (data not shown). To test whether nsp12 was able to incorporate [ $\alpha$ -<sup>32</sup>P]dAMP when using DNA templates, we employed a primed T<sub>10</sub> template similar to the RNA templates described above. As shown in Fig. 6B, nsp12 was capable of incorporating dAMP when incubated with a DNA template. On average, a minor reduction (~20%) in dAMP incorporation was observed in comparison to the activity on an RNA template. The size distribution of products was more different, and showed significantly less initiation than on an RNA template (compare Fig. 6A and B).



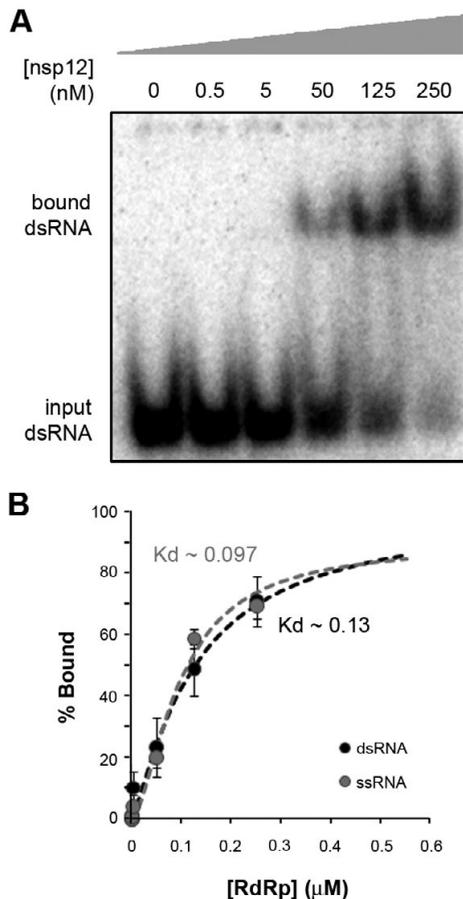
**Figure 3: Biochemical determinants of SARS-CoV nsp12 activity.** (A-B) Tests to determine the influence of temperature indicated that nsp12 incorporates ATP most efficiently at 37°C, although this effect mainly stemmed from increased initiation. Based on these results, 30°C was taken as the standard for subsequent experiments. (C-D) Titration of the Mg<sup>2+</sup> concentration showed that activity reaches its maximum at 6 mM. (E-F) The effect of pH on RdRp activity was evident as well, with lower pHs limiting the activity of the enzyme. All reactions were incubated for 60 min at 30°C, unless otherwise indicated. Error bars in A, C and E represent standard error of the mean (n = 3).



**Figure 4: Nucleotide incorporation fidelity of nsp12.** (A) Experimental set-up of pulse-chase experiments with different nucleotides and a primed polyU template (see Fig. 2A). The reactions were initiated with a limiting concentration of  $[\alpha\text{-}^{32}\text{P}]\text{ATP}$  to allow the incorporation of a first nucleotide and the formation of a stable polymerase-template complex. Subsequently, after 10 min, different unlabelled nucleotides were added to a final concentration of  $50 \mu\text{M}$  to allow elongation for another 30 min. (B) SARS-CoV nsp12 allows only limited transversional and transitional misincorporations. Interestingly, also in the presence of dATP no significant activity was observed, implying that the SARS-CoV RdRp is capable of discriminating between ATP and dATP. (C) Pulse-chase experiments in the presence of  $6 \text{ mM}$   $\text{Mn}^{2+}$  show that manganese ions promote misincorporation of ribonucleotides (both transversions and transitions). The selection against dATP remained unaltered.

The incorporation rate of nsp12 increased dramatically when we doubled the template length ( $V_{\text{U}20} = 27 \pm 3 \text{ nM/min}$ , Fig. 6C and 7), but we again observed a major product that was two nucleotides shorter than the expected full-length product. When the primed poly(U) template was changed to a primed heteromeric  $(\text{CU})_{10}$ , nsp12 activity increased further ( $V_{(\text{CU})10} = 45 \pm 2 \text{ nM/min}$ ) and reached a level comparable to that obtained for other viral RdRps on similar short templates [243]. When the products of the latter reac-

tions were resolved by denaturing PAGE (Fig. 6D) they were found to be longer than template length. Similar products that extending beyond unit length have been often observed when short templates were used for PV 3D<sup>pol</sup> [284] and their generation can be explained by various mechanisms, such as distributive slippage, processive slippage, template switching or terminal transferase activity [284]. Future research will address these mechanisms in more detail.



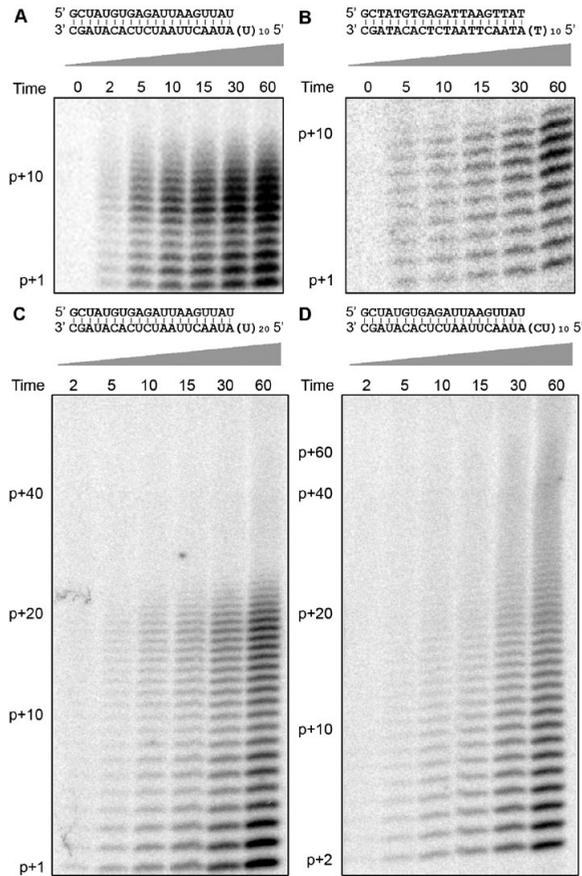
**Figure 5: RNA binding affinity of nsp12.** (A) A fixed concentration (0.2 nM) of radiolabelled dsRNA or ssRNA (data not shown) was titrated with purified nsp12 and the complexes formed were separated from unbound RNA on a native 8% polyacrylamide gel. (B) Free and bound RNA were quantified and fit to the Hill equation (see Materials and Methods), resulting in  $K_d$  values of  $0.13 \pm 0.3 \mu\text{M}$  for dsRNA and  $0.10 \pm 0.2 \mu\text{M}$  for ssRNA binding.  $R^2$  values of these Hill fits were 0.97 and 0.98 for dsRNA and ssRNA, respectively. Error bars represent standard error of the mean ( $n = 3$ ).

## Discussion

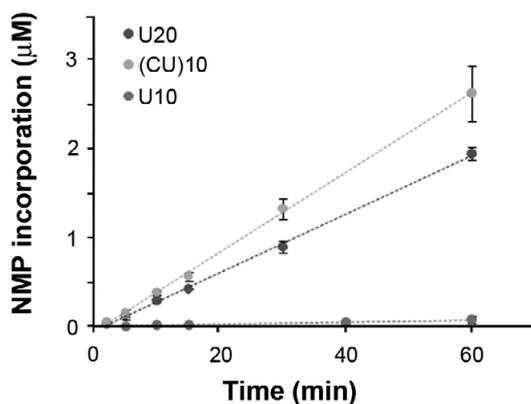
Detailed kinetic and stoichiometric information plays an important role in explaining the molecular basis of RdRp-catalysed ribonucleotide incorporation, the associated mutation rate of viral genomes and, consequently, RNA virus evolution. Furthermore, a comprehensive understanding of the differences between viral RdRps and cellular polymerases is crucial in the design of specific inhibitors of viral RNA synthesis and their subsequent mechanistic characterisation and optimisation. Up till now, all of the above information was particularly limited, if not completely lacking, for the coronavirus RdRp, due to enzyme insolubility, instability and low purification yields.

This study reports the first successful expression and purification of stable, soluble SARS-CoV nsp12, thus paving the way for its further biochemical characterisation. Using a two-step procedure, the protein was purified to near homogeneity, as confirmed by SDS-PAGE and non-denaturing gel filtration. We found that even a single-step purification scheme could produce an enzyme preparation that combined specific activity with reasonable purity (>80-85%). Of particular importance in our expression protocol was the co-expression of the Ubp1 protease that mediated the *in vivo* removal of the N-terminal Ub tag (Fig. 1A) [277]. Apparently, similar to PV 3D<sup>pol</sup> [278], SARS-CoV RdRp does not tolerate additional amino acids at its N-terminus, likely because they induce an unstable and largely inactive fold. Because of the observed sensitivity to N-terminal fusions, it is tempting to speculate that the processing of the pp1ab cleavage generating the nsp12 N-terminus is of great importance for proper folding and activity *in vivo* as well. In line with this notion, it was reported that for the related group 2 coronavirus MHV this cleavage site indeed is indispensable for virus replication [285]. It still remains to be tested whether there is any effect of the carboxy-terminal His-tag on the activity of the enzyme used in this study.

The SARS-CoV nsp12 is a relatively large protein (932 amino acid residues), *e.g.*, much larger than the well described RdRps of HCV (591 amino acids) and PV (461 amino acids), and may have, in addition to the C-terminal RdRp domain, other functional and/or contributing domains in its still uncharacterized N-terminal domain(s). Previously, the PV and HCV RdRps were demonstrated to bind RNA with  $K_d$  values of 2-10  $\mu\text{M}$  [286] and  $\sim 3$   $\mu\text{M}$  [287], respectively. Our data suggest that SARS-CoV nsp12 binds RNA with slightly higher affinity for both dsRNA and ssRNA ( $K_d \sim 0.1$   $\mu\text{M}$ ). The relatively similar affinities for dsRNA and ssRNA suggest that the lack of *de novo* initiation may largely be due to a relatively open structure and the absence of a conserved priming loop in the predicted thumb subdomain [89]. The latter structural element, a  $\beta$ -hairpin in, *e.g.*, hepatitis C virus (HCV) NS5B, is present in many known *de novo*-initiating RNA polymerases. Canonically, it functions by partly obstructing the template binding cleft and is believed to be essential for association with the single-stranded 3' end of the template RNA [89,145,288].



**Figure 6: Analysis of nucleotide incorporation by nsp12.** (A) Reactions with 0.1  $\mu\text{M}$  nsp12, 1  $\mu\text{M}$  template and 50  $\mu\text{M}$  ATP were followed over time and quenched with 50 mM EDTA at the time points indicated. Gel analysis showed that, similar to the curves in Fig. 3, the major accumulating product is the primer (p) extended by seven nucleotides (*i.e.*, p+7) and not the expected full-length product of n+10. (B) To assess whether nsp12 was capable of DNA-templated incorporation of dATP, reactions were performed as in Fig. 6A, but now using a DNA template. The time course illustrates robust [ $\alpha$ - $^{32}\text{P}$ ]dAMP incorporation, implying that the SARS-CoV RdRp is able to utilise both RNA and DNA templates. The incorporation of dAMP was 20% lower than that of AMP on an RNA template. (C) Time courses were additionally performed using an RNA template containing a U<sub>20</sub> template region. Similar to the shorter 10-nt template used in Fig. 6A, the major product here was around two nucleotides shorter than the expected full-length product. (D) Changing the single-stranded template from U<sub>20</sub> to a heteromeric (CU)<sub>10</sub> not only increased the nucleotide incorporation rate by 34%, but also the behaviour of nsp12 on the template. Products of the expected full-length size (p+20) were observed, but also products longer than the template, hinting at, *e.g.*, template switching by nsp12. Future research will address this observation in more detail. In this reaction both 50  $\mu\text{M}$  ATP and GTP were present. The gel image represents the top half of a 20% denaturing PAGE gel.



**Figure 7: Analysis of the nucleotide incorporation rate of nsp12.** Steady-state time courses were performed with 0.1  $\mu\text{M}$  nsp12 as described in Fig. 6. Experimental data was subsequently fit to linear regression to obtain NMP incorporation rates. These clearly illustrated an increase of the incorporation velocity with template length ( $V_{\text{U10}} = 1.2 \pm 0.4$  nM/min and  $V_{\text{U20}} = 27 \pm 3$  nM/min) and from homopolymeric template to copolymeric template ( $V_{(\text{CU})10} = 45 \pm 2$  nM/min).  $R^2$  values were 0.97, 0.99 and 0.99, respectively, and error bars indicate standard error of the mean ( $n = 3$ ).

Contrary to previous GST-pull downs performed for nsp12 [85] and data obtained for other RdRps such as the PV 3D<sup>pol</sup> - which was reported to form multimers and even lattices [286] - we were not able to establish nsp12 multimerisation using non-denaturing gel filtration. Although this suggests that SARS-CoV RdRp exists predominantly as monomer under the conditions used here, it is likely that nsp12 binds to various other viral replicase subunits [289] and cellular co-factors [290] *in vivo*. The coronavirus holoenzyme complex formed in this manner may have more finely tuned characteristics than documented in this study. Additionally, it remains to be tested whether the C-terminal hexahistidine tag interferes with multimerisation. Clearly, these and other questions should be addressed in more detail in future studies.

Upon binding to template and primer, we found nsp12 to have reproducible nucleotide incorporation rates on the templates tested, but incorporation depended significantly on the length of the template. Furthermore, on heteromeric templates the SARS-CoV RdRp also performed better and reached incorporation rates in the same order ( $> \text{nM/min}$ ) as the model polymerases from PV and HCV, when correcting for enzyme and template concentrations [243,291]. Additionally, the SARS-CoV nsp12 is also active on a DNA template, which may greatly facilitate further investigations in view of the superior stability of DNA templates and the fact that various other SARS-CoV enzymes [108,153,270] have documented affinity for DNA as well. It remains to be tested whether SARS-CoV nsp12 is also active on a longer single-stranded template, *e.g.*, RNA templates based on the terminal sequences of the viral genome, and whether it is able to traverse the RNA secondary structures present in such a molecule. Taken together, our

data suggests that our present purification scheme and assay are a solid starting point for further biochemical exploration of the SARS-CoV RdRp and nsp12 targeted inhibitor development.

## Material and methods

### Cloning and expression of recombinant SARS-CoV nsp12

The amino acid sequence of SARS-CoV nsp12, residues S4370 to Q5301 of replicase pp1ab, is fully conserved among all sequenced SARS-CoV isolates. For our study, the nsp12-coding region was amplified by RT-PCR from the genome of SARS-CoV Frankfurt-1 (AY291315). During amplification, the ORF1a/1b -1 ribosomal frameshift near the 5' end of the gene was corrected in a translationally silent manner. Primers used for amplification were SAV424 5'-GCGGGTACCCCGGGTGGATCTGCGGATGCATCAA-3' and SAV425 5'-GCGCGATCGGGATCCCTGCAAGACTGTATGT-3', corresponding to SARS-CoV genome positions 13 327-13 387 and 16 152-16 166, respectively. The PCR amplicon was digested with *Sac*II and *Bam*HI, and ligated into expression vector pET26-Ub-CHis<sub>6</sub> [277]. Following confirmation of the nucleotide sequence of the amplicon and using the flanking *Xba*I and *Xho*I restriction sites, the ubiquitin-nsp12-His<sub>6</sub> fusion gene (Ub-nsp12-CHis<sub>6</sub>) was sub-cloned into vector pASK3 (IBA) and thereby placed under the control of a tetracycline-inducible promoter. This ruled out the known risk of potential contamination with T7 RNA polymerase, as it can occur when using a T7 promoter to drive expression of other RdRps in bacterial expression systems. Nsp12 active-site mutant D618A, carrying an Asp618Ala substitution, was engineered via site-directed mutagenesis according to the QuikChange protocol (Stratagene), using primers SAV526 5'-CCTTATGGGTTGGGCTTATCCAAAATGTG-3' and SAV527 5'-CACATTTTGGATAAGCCCAACCCATAAGGA-3'.

For nsp12 expression, the pASK3-Ub-nsp12-CHis<sub>6</sub> plasmid was transformed to *E. coli* C2523 cells (New England Biolabs) together with the Ubp1 protease expression plasmid pCG1 [277], unless indicated otherwise for particular experiments. Routinely, 250 ml of Luria Broth, containing ampicillin (50 µg/ml) and chloramphenicol (34 µg/ml), was inoculated 1:1000 with 3 ml o/n precultures and cells were grown at 37°C to OD<sub>600</sub> >0.7. Subsequently, the cells were slowly cooled to room temperature, tetracycline was added to a final concentration of 200 ng/ml and the cells were further grown at 20°C for another 16 h. Cells were harvested by centrifugation, washed once with ice-cold PBS and stored at -20°C until protein purification.

SARS-CoV nsp12 expression vectors pKM596-nsp12 (kindly provided by Dr. Isabelle Imbert and Dr. Bruno Canard) and pET102-Thio-nsp12 (a kind gift of Dr. Susanne Pfefferle and Dr. Christian Drosten) were used to transform *E. coli* BL21(DE3) cells and expression was performed using the auto-induction medium ZYM-5052 [292].

### Purification of SARS-CoV nsp12-His

Cell pellets were thawed on ice, resuspended in binding buffer (20 mM Hepes pH 7.4, 500 mM NaCl, 20 mM imidazole, 10% glycerol, 0.05% Tween-20 and 5 mM  $\beta$ -mercaptoethanol) and lysed by sonication. Cell debris was removed by ultra-centrifugation at 20 000  $\times g$  and the cleared supernatant was incubated with Talon beads (Clontech) for 2 h at 4°C. The beads were washed 3 times with 20 volumes of binding buffer, loaded onto spin columns and finally the His-tagged protein was eluted with 300 mM imidazole in binding buffer. The eluate was analysed by SDS-PAGE and typically found to be 80-85% pure. Subsequently, gel filtration on a Superdex 200 column (GE healthcare) in 20 mM Hepes, 500 mM NaCl, 5% glycerol, 5 mM  $\beta$ -mercaptoethanol and 0.1% Tween-20 was used to improve purity to ~95%. To estimate the molecular mass of nsp12, a standard curve for hydrodynamic elution was made using chymotrypsinogen A, conalbumin, aldolase and dextran blue (all purchased from GE Healthcare) as size markers and fitted to a linear equation ( $R^2 = 0.99$ ).

After the gel filtration step, the nsp12-containing fractions were pooled, concentrated with Ultrafree 10kDa filter columns (Millipore) and dialysed overnight against 1000 volumes of dialysis buffer (20 mM Hepes pH 7.4, 100 mM NaCl, 0.1% Triton-X100, 1 mM DTT and 50% glycerol) in cellulose tubing (SnakeSkin, Pierce). Protein concentrations were determined using a Bradford assay and a bovine serum albumin (BSA) standard curve, and the purified protein was stored at -20°C. When using expression vectors pKM596-nsp12 and pET102-Thio-nsp12, protein purification procedures were essentially similar to those described above, except that amylose resin (New England Biolabs) was used to bind the pKM596-nsp12 product, which was eluted using a buffer containing 10 mM amylose instead of imidazole.

### SDS-PAGE and Western blotting

Cells were lysed in Laemmli's sample buffer [293], separated on 8% SDS-PAGE-gels and either electroblotted to PVDF membrane (GE Healthcare) or stained directly with Coomassie G-250 according to standard protocols. Prior to incubation with a SARS-CoV nsp12-specific rabbit antiserum (a gift from Dr. Mark Denison) [294], blot membranes were blocked with 10% skimmed milk and 0.02% Tween-20 in PBS. Blots were washed with 0.02% Tween-20 in PBS, incubated with a swine-anti-rabbit horseradish peroxidase conjugate diluted in 5% skimmed milk, 0.02% Tween-20 and PBS, washed and subjected to ECL Plus chemoluminescence detection (GE Healthcare).

### Preparation of templates

RNA oligonucleotides and their DNA counterparts SAV557R 5' - GCUAUGUGAGAU-UAAGUUUAU -3' and SAV481R 5'- UUUUUUUUUUAUAACUUAUCUCACAUAGC -3', SAV555R 5'- UUUUUUUUUUUUUUUUUUUUAUAACUUAUCUCACAUAGC -3' and

SAV556R 5'- UCUCUCUCUCUCUCUCUCUCAUAACUUAUCUCACAUAGC -3' were purchased from Eurogentec, excised as a single band from 7 M urea/15% polyacrylamide gels, eluted overnight in deionised water and desalted using NAP-10 columns (GE healthcare). To anneal RNA duplexes, oligonucleotide mixtures in annealing buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl and 5 mM EDTA) were heat denatured, slowly cooled to room temperature and subsequently purified from 15% non-denaturing polyacrylamide gels.

### RdRp activity assays

To test the *in vitro* RdRp activity of SARS-CoV nsp12-CHis<sub>67</sub>, 0.1 μM of the protein (unless otherwise indicated) was incubated at 30°C in the presence of 1 μM template, 0.17 μM [ $\alpha$ -<sup>32</sup>P]ATP (0.5 μCi/μl; Perkin-Elmer), 50 μM ATP (Roche), 20 mM Tris-HCl (pH 8.0), 10 mM NaCl, 10 mM KCl, 6 mM MgCl<sub>2</sub>, 5% glycerol, 0.01% Triton-X100, 1 mM DTT and 0.5 units of RNaseOUT (Invitrogen). Glycerol and NaCl were introduced with the addition of enzyme due to their presence in the nsp12 storage buffer. RdRp reactions were terminated by adding an equal volume of 90% formamide, 50 mM EDTA pH 8.0, 0.01% bromophenol blue and xylene cyanol. Reaction products were analysed on 7 M Urea, 20% polyacrylamide gels buffered with 1x TBE (90 mM Tris, 90 mM H<sub>3</sub>BO<sub>3</sub> and 2 mM EDTA). Gels were run for 5 h at 2000 V and analysed using phosphorimaging. Plates were scanned on a Typhoon variable mode scanner (GE Healthcare) and products were quantified with ImageQuant TL 7.0 (GE Healthcare). To calculate the amount of [ $\alpha$ -<sup>32</sup>P]AMP incorporated from the phosphorimager signal, a 10<sup>-2</sup>-10<sup>-5</sup> dilution series of the [ $\alpha$ -<sup>32</sup>P]ATP stock was spotted in triplicate on Whatman filter and exposed alongside the PAGE gel of interest. Using the standard curve, phosphorimager counts were converted into the amount of label incorporation and kinetic traces were fit to a linear function with the Matlab 2008b Curve Fitting Toolbox (Mathworks).

### Electrophoretic mobility shift assays

A dilution series of SARS-CoV RdRp in storage buffer was incubated for 10 min at room temperature with 0.2 nM of <sup>32</sup>P-labelled oligonucleotide SAV557R (see above), in the presence or absence of the complementary oligonucleotide SAV481R. Samples were then directly loaded on to 8% acrylamide gels containing 5% glycerol and 0.5x TBE and were run at 100 V for 2 h at 4°C. Gels were dried on Whatman filter paper and bands were quantified by phosphorimaging. Using the Matlab 2008b Curve Fitting Toolbox, the percentage of RNA bound was fit to the Hill equation  $RNA_{\text{bound}} = b * [nsp12]^n / (K_d^n + [nsp12]^n)$ , where b is the upper binding limit, [nsp12] is the nsp12 concentration, n is the Hill coefficient and K<sub>d</sub> is the enzyme's dissociation constant.

## Acknowledgements

The authors acknowledge Dr. Alexander Gorbalenya, Dr. Mark Denison, Dr. Nynke Dekker, Steve Cramer, Dr. Susanne Pfefferle, Dr. Christian Drosten, Dr. Isabelle Imbert and Dr. Bruno Canard for helpful discussions or providing reagents. This work was supported by the Netherlands Organization for Scientific Research (NWO) through Toptalent grant 021.001.037 and ECHO grant 700.55.002 from the Council for Chemical Sciences (NWO-CW). JA and CC were supported, in part, by grant AI45818 from NIAID/NIH.

# Chapter 4

Zn<sup>2+</sup> inhibits coronavirus  
and arterivirus RNA  
polymerase activity  
*in vitro* and zinc  
ionophores block the  
replication of these  
viruses in cell culture

Aartjan J.W. te Velthuis<sup>1</sup>, Sjoerd  
H.E. van den Worm<sup>1</sup>, Amy C. Sims<sup>2</sup>,  
Ralph S. Baric<sup>2</sup>, Eric J. Snijder<sup>1</sup> and  
Martijn J. van Hemert<sup>1</sup>

<sup>1</sup> Molecular Virology Laboratory,  
Department of Medical  
Microbiology, Center of Infectious  
Diseases, Leiden University  
Medical Center, PO Box 9600,  
2300RC Leiden, The Netherlands.

<sup>2</sup> Departments of Epidemiology  
and Microbiology and  
Immunology, University of North  
Carolina, Chapel Hill, North  
Carolina, USA.

**Published in:** PLoS Pathogens,  
Volume 6, e1001176, 2010

## ABSTRACT

Increasing the intracellular  $Zn^{2+}$  concentration with zinc-ionophores like pyrithione (PT) can efficiently impair the replication of a variety of RNA viruses, including poliovirus and influenza virus. For some viruses this effect has been attributed to interference with viral polyprotein processing. In this study we demonstrate that the combination of  $Zn^{2+}$  and PT at low concentrations (2  $\mu M$   $Zn^{2+}$  and 2  $\mu M$  PT) inhibits the replication of SARS-coronavirus (SARS-CoV) and equine arteritis virus (EAV) in cell culture. The RNA synthesis of these two distantly related nidoviruses is catalysed by an RNA-dependent RNA polymerase (RdRp), which is the core enzyme of their multiprotein replication and transcription complex (RTC). Using an activity assay for RTCs isolated from cells infected with SARS-CoV or EAV - thus eliminating the need for PT to transport  $Zn^{2+}$  across the plasma membrane - we show that  $Zn^{2+}$  efficiently inhibits the RNA-synthesising activity of the RTCs of both viruses. Enzymatic studies using recombinant RdRps (SARS-CoV nsp12 and EAV nsp9) purified from *E. coli* subsequently revealed that  $Zn^{2+}$  directly inhibited the *in vitro* activity of both nidovirus polymerases. More specifically,  $Zn^{2+}$  was found to block the initiation step of EAV RNA synthesis, whereas in the case of the SARS-CoV RdRp elongation was inhibited and template binding reduced. By chelating  $Zn^{2+}$  with MgEDTA, the inhibitory effect of the divalent cation could be reversed, which provides a novel experimental tool for *in vitro* studies of the molecular details of nidovirus replication and transcription.

## INTRODUCTION

Zinc ions are involved in many different cellular processes and have proven crucial for the proper folding and activity of various cellular enzymes and transcription factors. Zn<sup>2+</sup> is probably an important cofactor for numerous viral proteins as well. Nevertheless, the intracellular concentration of free Zn<sup>2+</sup> is maintained at a relatively low level by metallothioneins, likely due to the fact that Zn<sup>2+</sup> can serve as intracellular second messenger and may trigger apoptosis or a decrease in protein synthesis at elevated concentrations [295,296,297]. Interestingly, in cell culture studies, high Zn<sup>2+</sup> concentrations and the addition of compounds that stimulate cellular import of Zn<sup>2+</sup>, such as hinokitol (HK), pyrrolidine dithiocarbamate (PDTC) and pyrithione (PT), were found to inhibit the replication of various RNA viruses, including influenza virus [298], respiratory syncytial virus [299] and several picornaviruses [300,301,302,303,304,305]. Although these previous studies provided limited mechanistic information, this suggests that intracellular Zn<sup>2+</sup> levels affect a common step in the replicative cycle of these viruses.

In cell culture, PT stimulates Zn<sup>2+</sup> uptake within minutes and inhibits RNA virus replication through a mechanism that has only been studied in reasonable detail for picornaviruses [305,306]. *In vitro* studies with purified rhinovirus and poliovirus 3C proteases revealed that protease activity was inhibited by Zn<sup>2+</sup> [93,307], which is in line with the inhibition of polyprotein processing by zinc ions that was observed in cells infected with human rhinovirus and coxsackievirus B3 [305]. The replication of segmented negative-strand RNA viruses such as influenza virus, however, does not depend on polyprotein processing and the effect of PDTC-mediated Zn<sup>2+</sup> import was therefore hypothesised to result from inhibition of the viral RNA-dependent RNA polymerase (RdRp) and cellular cofactors [298]. Moreover, an inhibitory effect of Zn<sup>2+</sup> on the activity of purified RdRps from rhinoviruses and hepatitis C virus was noted, but not investigated in any detail [308,309].

Details on the effect of zinc ions are currently largely unknown for nidoviruses. This large group of positive-strand RNA (+RNA) viruses includes major pathogens of humans and livestock, such as severe acute respiratory syndrome coronavirus (SARS-CoV), other human coronaviruses, the arteriviruses equine arteritis virus (EAV), and porcine reproductive and respiratory syndrome virus (PRRSV) [35,65]. The common ancestry of nidoviruses is reflected in their similar genome organisation and expression strategy, and in the conservation of a number of key enzymatic functions in their large replicase polyproteins [58]. A hallmark of the corona- and arterivirus replicative cycle is the transcription of a 5'- and 3'-coterminal nested set of subgenomic (sg) mRNAs from which the viral structural and accessory protein genes are expressed [51,265].

Analogous to picornaviruses [307,310], zinc ions were demonstrated to inhibit certain proteolytic cleavages in the processing of the coronavirus replicase polyproteins in

infected cells and cell-free systems [311,312]. In this study we report that the zinc-ionophore pyrithione (PT) in combination with  $Zn^{2+}$  is a potent inhibitor of the replication of SARS-coronavirus (SARS-CoV) and equine arteritis virus (EAV) in cell culture. To assess whether - besides a possible effect on proteolytic processing - nidovirus RTC subunits and RNA synthesis are directly affected by  $Zn^{2+}$ , we employed *in vitro* systems for SARS-CoV and EAV RNA synthesis that are based on membrane-associated RTCs isolated from infected cells (from here on referred to as RTC assays) [290,313]. In addition, we used *in vitro* recombinant RdRp assays to directly study the effect of zinc ions on the RdRps of SARS-CoV and EAV [154,196].

Using these independent *in vitro* approaches, we were able to demonstrate that  $Zn^{2+}$  directly impairs nidovirus RNA synthesis, since it had a strong inhibitory effect in both RTC and RdRp assays. Interestingly, the  $Zn^{2+}$ -mediated inhibition could be reversed through the addition of a  $Zn^{2+}$  chelator (MgEDTA). We therefore applied this compound to stop and restart the *in vitro* RNA-synthesising activity at will. This convenient tool allowed us to study various mechanistic aspects of arteri- and coronavirus RNA synthesis in more detail. Additionally, the zinc-mediated inhibition of nidovirus RNA synthesis described here may provide an interesting basis to further explore the use of zinc-ionophores in antiviral therapy.

## RESULTS

### Zinc and pyrithione inhibit nidovirus replication *in vivo*

Zinc ions are involved in many different cellular processes, but the concentration of free  $Zn^{2+}$  is maintained at a relatively low level by metallothioneins [295].  $Zn^{2+}$  and compounds that stimulate import of  $Zn^{2+}$  into cells, such as PT, were previously found to inhibit replication of several picornaviruses, including rhinoviruses, foot-and-mouth disease virus, coxsackievirus, and mengovirus in cell culture [300,301,302,303,304,305]. To determine whether  $Zn^{2+}$  has a similar effect on nidoviruses, we investigated the effect of PT and  $Zn^{2+}$  on the replication of EAV and SARS-CoV in Vero-E6 cells, using reporter viruses that express green fluorescent proteins (GFP), *i.e.*, EAV-GFP [314] and SARS-CoV-GFP [315]. EAV-GFP encodes an N-terminal fusion of GFP to the viral nonstructural protein 2 (nsp2), one of the cleavage products of the replicase polyproteins, and thus provides a direct readout for translation of the replicase gene. In SARS-CoV-GFP, reporter expression occurs from sg mRNA 7, following the replacement of two accessory protein-coding genes (ORFs 7a and 7b) that are dispensable for replication in cell culture.

We first assessed the cytotoxicity of a range of PT concentrations (0-32  $\mu$ M) in the presence of 0 to 8  $\mu$ M  $ZnOAc_2$ . Treatment with PT of concentrations up to 32  $\mu$ M in combination with <4  $\mu$ M  $ZnOAc_2$  did not reduce the viability of mock-infected cells

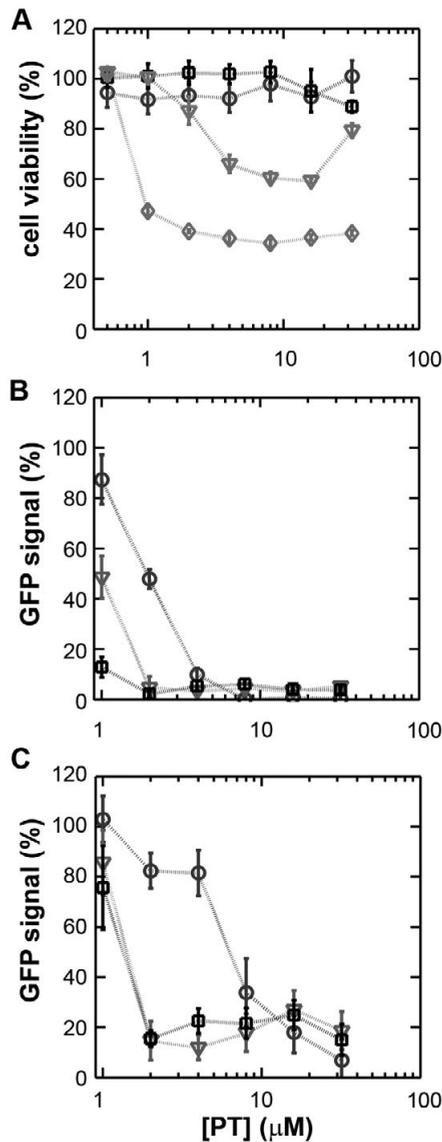
after 18 h (Fig. 1A), as measured by the colorimetric MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) viability assay. As elevated Zn<sup>2+</sup> concentrations are known to inhibit cellular translation, we also used metabolic labelling with <sup>35</sup>S-methionine to assess the effect of PT and Zn<sup>2+</sup> on cellular protein synthesis. Incubation of Vero-E6 cells for 18 h with the combinations of PT and Zn<sup>2+</sup> mentioned above, followed by a 2-h metabolic labelling, revealed no change in overall cellular protein synthesis when the concentration of ZnOAc<sub>2</sub> was <4 μM (data not shown).

Using these non-cytotoxic conditions we subsequently tested the effect of PT and ZnOAc<sub>2</sub> on EAV-GFP and SARS-CoV-GFP replication. To this end, Vero-E6 cells in 96-well plates were infected with a multiplicity of infection (m.o.i.) of 4. One hour post infection (h p.i.), between 0 and 32 μM of PT and 0, 1, or 2 μM ZnOAc<sub>2</sub> were added to the culture medium. At 17 h p.i., a time point at which GFP expression in untreated infected cells reaches its maximum for both viruses, cells were fixed, and GFP fluorescence was quantified.

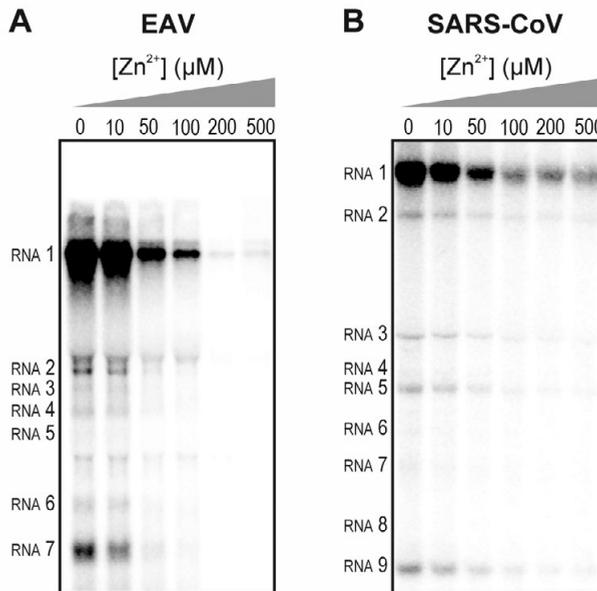
The reporter gene expression of both SARS-CoV-GFP and EAV-GFP was already significantly inhibited in a dose-dependent manner by the addition of PT alone (Fig. 1B and C). This effect was significantly enhanced when 2 μM of Zn<sup>2+</sup> was added to the medium. We found that addition of ZnOAc<sub>2</sub> alone also reduced virus replication, but only at levels that were close to the 50% cytotoxicity concentration (CC<sub>50</sub>) of ZnOAc<sub>2</sub> in Vero-E6 cells (~70 μM, data not shown). This is likely due to the poor solubility of Zn<sup>2+</sup> in phosphate-containing medium and the inefficient uptake of Zn<sup>2+</sup> by cells in the absence of zinc-ionophores. The combination of 2 μM PT and 2 μM ZnOAc<sub>2</sub> resulted in a 98 ± 1% and 85 ± 3% reduction of the GFP signal for EAV-GFP and SARS-CoV-GFP, respectively. No cytotoxicity was observed for this combination of PT and ZnOAc<sub>2</sub> concentrations. From the dose-response curves in Fig. 1, a CC<sub>50</sub> value of 82 μM was calculated for PT in the presence of 2 μM zinc. Half maximal inhibitory concentrations (IC<sub>50</sub>) of 1.4 μM and 0.5 μM and selectivity indices of 59 and 164 were calculated for SARS-CoV and EAV, respectively.

### Zn<sup>2+</sup> reversibly inhibits the RNA-synthesising activity of isolated nidovirus RTCs

We previously developed assays to study the *in vitro* RNA-synthesising activity of RTCs isolated from cells infected with SARS-CoV or EAV [290,313]. In these RTC assays [ $\alpha$ -<sup>32</sup>P] CMP is incorporated into both genomic (replication) and sg mRNA (transcription) (Fig. 2). This allowed us to monitor the synthesis of the same viral RNA molecules that can be detected by hybridisation of RNA from nidovirus-infected cells. A benefit of these assays is that the activity does not depend on continued protein synthesis and that it allows us to study viral RNA synthesis independent of other aspects of the viral replicative cycle [290]. To investigate whether the inhibitory effect of PT and zinc ions on nidovirus replication in cell culture is reflected in a direct effect of Zn<sup>2+</sup> on viral RNA



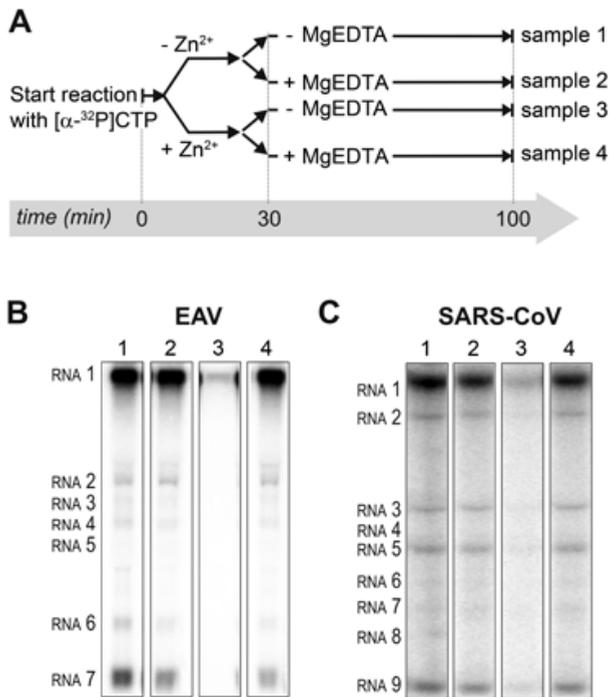
**Figure 1: The zinc ionophore pyrrithione inhibits nidovirus replication in cell culture.** (A) Cytotoxicity of PT in Vero-E6 cells in the absence (circles) or presence of 2 (squares), 4 (triangles), or 8  $\mu\text{M}$   $\text{ZnOAc}_2$  as determined by the MTS assay after 18 hours of exposure. (B) Dose-response curves showing the effect of PT and  $\text{Zn}^{2+}$  on the GFP fluorescence in Vero-E6 cells infected with a GFP-expressing EAV reporter strain at 17 h p.i. Data were normalised to GFP expression in infected, untreated control cultures (100%). The different  $\text{Zn}^{2+}$  concentrations added to the medium were 0 (circles), 1 (triangles), or 2  $\mu\text{M}$   $\text{ZnOAc}_2$  (squares). (C) Effect of PT and  $\text{Zn}^{2+}$  on the GFP fluorescence in Vero-E6 cells infected with a GFP-expressing SARS-CoV reporter strain at 17 h p.i. Data were normalised to GFP expression in infected untreated control cells (100%). Coding for different  $\text{Zn}^{2+}$  concentrations as in Fig. 1B. Error bars indicate the standard deviation ( $n = 4$ ).



**Figure 2: Inhibition of the *in vitro* RNA-synthesising activity of isolated RTCs by Zn<sup>2+</sup>.** Incorporation of [ $\alpha$ -<sup>32</sup>P]CMP into viral RNA by EAV (A) and SARS-CoV (B) in RTC assays in the presence of various Zn<sup>2+</sup> concentrations, as indicated above each lane.

synthesis, we tested the effect of Zn<sup>2+</sup> addition on RTC activity. For both EAV (Fig. 2A) and SARS-CoV (Fig. 2B), a dose-dependent decrease in the amount of RNA synthesised was observed when ZnOAc<sub>2</sub> was present. For both viruses, a more than 50% reduction of overall RNA-synthesis was observed at a Zn<sup>2+</sup> concentration of 50 μM, while less than 5% activity remained at a Zn<sup>2+</sup> concentration of 500 μM. Both genome synthesis and sg mRNA production were equally affected.

To test whether the inhibition of RTC activity by Zn<sup>2+</sup> was reversible, RTC reactions were started in the presence or absence of 500 μM Zn<sup>2+</sup>. After 30 min, these reactions were split into two aliquots and magnesium-saturated EDTA (MgEDTA) was added to one of the tubes to a final concentration of 1 mM (Fig. 3A). We used MgEDTA as Zn<sup>2+</sup> chelator in these *in vitro* assays, because it specifically chelates Zn<sup>2+</sup> while releasing Mg<sup>2+</sup>, due to the higher stability constant of the ZnEDTA complex. Uncomplexed EDTA inhibited RTC activity in all reactions (data not shown), most likely by chelating the Mg<sup>2+</sup> that is crucial for RdRp activity [154,196], whereas MgEDTA had no effects on control reactions without Zn<sup>2+</sup> (Fig. 3B, compare lane 1 and 2). As shown in Fig. 2, the EAV RTC activity that was inhibited by Zn<sup>2+</sup> (Fig. 3B and 3C, lane 3) could be restored by the addition of MgEDTA (Fig. 3B, lane 4) to a level observed for control reactions without Zn<sup>2+</sup> (Fig. 3B, lane 1). Compared to the untreated control, the EAV RTC assay produced approximately 30% less RNA, which was consistent with the 30% shorter reaction time after the addition of

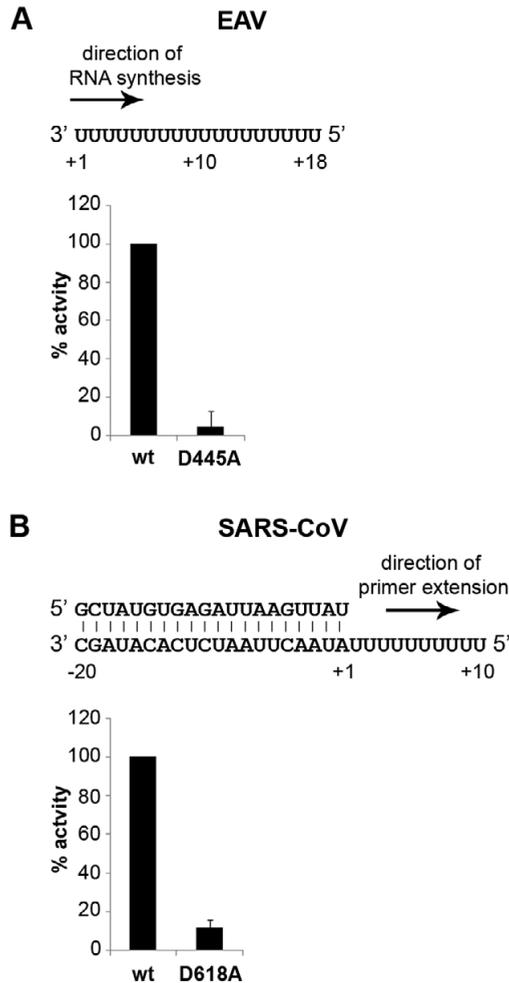


**Figure 3: Inhibition of nidovirus RTC activity by Zn $^{2+}$  can be reversed by chelation.** (A) Schematic representation of the *in vitro* assays with isolated RTCs, which were initiated with [ $\alpha$ - $^{32}$ P] CTP, either in the absence (sample 1 and 2) or presence of 500  $\mu$ M Zn $^{2+}$ . After a 30-min incubation at 30°C, both the untreated and Zn $^{2+}$ -treated samples were split into two aliquots, and 1 mM of the Zn $^{2+}$  chelator MgEDTA was added to samples 2 and 4. All reactions were subsequently incubated for another 70 min before termination. (B) Analysis of RNA products synthesised in assays with EAV RTCs. Numbers above the lanes refer to the sample numbers described under 3A. (C) *In vitro* activity assay with SARS-CoV RTCs.

the MgEDTA (100 versus 70 min for lanes 1 and 4, respectively). Surprisingly, SARS-CoV RTC assays that were consecutively supplemented with Zn $^{2+}$  and MgEDTA incorporated slightly more [ $\alpha$ - $^{32}$ P]CMP compared to untreated control reactions (Fig. 3C; compare lane 1 and 4). This effect was not due to chelation of the Zn $^{2+}$  already present in the post-nuclear supernatant (PNS) of SARS-CoV-infected cells, as this increase was not observed when MgEDTA was added to a control reaction without additional Zn $^{2+}$  (Fig. 3C, lane 2).

### Zinc ions affect the *in vitro* activity of recombinant nidovirus RdRps

To establish whether inhibition of RTC activity might be due to a direct effect of Zn $^{2+}$  on nidovirus RdRp activity, we tested the effect of Zn $^{2+}$  on the activity of the purified recombinant RdRps of SARS-CoV (nsp12) and EAV (nsp9) using previously developed RdRp assays [154,196]. Using an 18-mer polyU template, the EAV RdRp incorporated [ $\alpha$ - $^{32}$ P] AMP into RNA products of up to 18 nt in length (Fig. 4A). Initiation was *de novo*, which



**Figure 4: EAV and SARS-CoV RdRp assays with wild-type enzyme and active-site mutants.** (A) The EAV polymerase was incapable of primer extension and required a free 3' end and poly(U) residues to initiate. Nucleotide incorporating activity of the wild-type enzyme and D445A mutant of nsp9 on an 18-mer poly(U) template confirmed the specificity of our assay. (B) SARS-CoV nsp12 RdRp assays were performed with an RNA duplex with a 5' U<sub>10</sub> overhang as template. The bar graph shows the nucleotide incorporating activities of wild-type and D618A nsp12. Error bars represent standard error of the mean (n = 3).

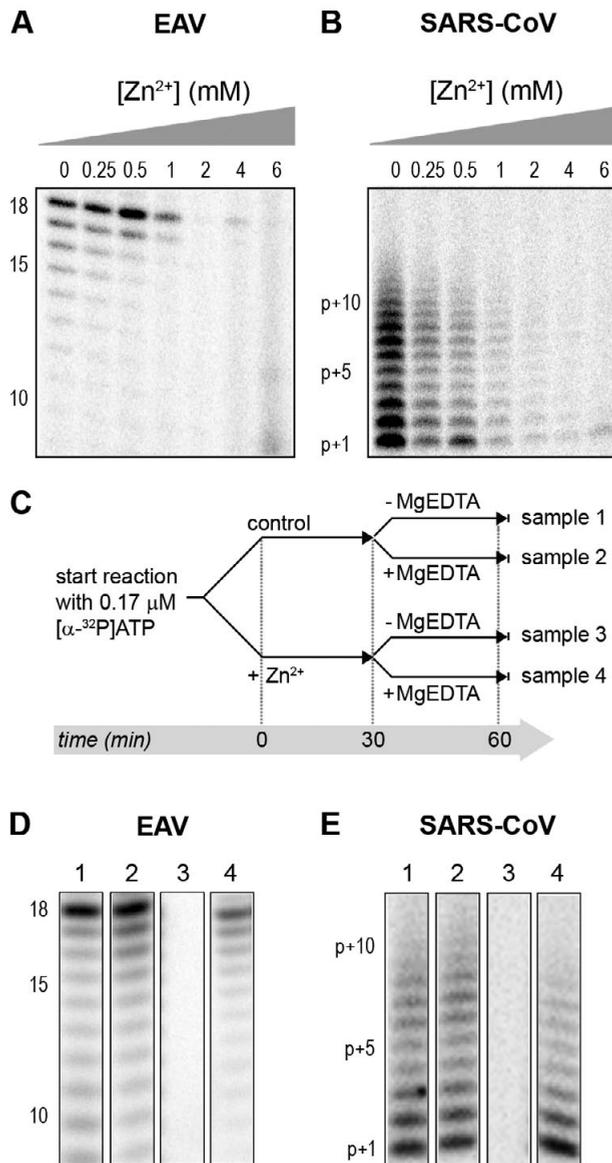
is in line with previous observations and the presence of a conserved priming loop in the nsp9 sequence [196]. Unlike the EAV RdRp nsp9, the *in vitro* activity of the SARS-CoV RdRp nsp12 - which lacks a priming loop- was shown to be strictly primer-dependent [154]. Thus, to study the RdRp activity of SARS-CoV nsp12, a primed polyU template was used (Fig. 4B), thereby allowing us to sample [ $\alpha$ -<sup>32</sup>P]AMP incorporation as described previously [154]. As specificity controls, we used the previously described SARS-CoV

nsp12 mutant D618A [154], which contains an aspartate to alanine substitution in motif A of the RdRp active site, and EAV nsp9-D445A, in which we engineered an aspartate to alanine substitution at the corresponding site of EAV nsp9 [89,196]. Both mutant RdRps showed greatly reduced [ $\alpha$ - $^{32}$ P]AMP incorporation in our assays (Fig. 4), confirming once again that the radiolabelled RNA products derived from nidovirus RdRp activity.

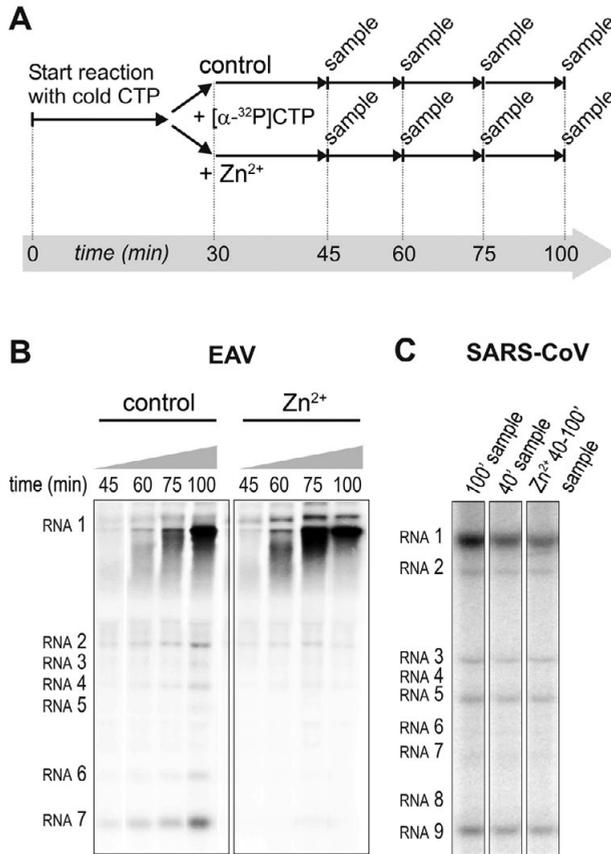
Addition of ZnOAc<sub>2</sub> to RdRp assays resulted in a strong, dose-dependent inhibition of enzymatic activity for both the EAV and SARS-CoV enzyme (Fig. 5A and B, respectively), similar to what was observed in RTC assays. In fact, compared to other divalent metal ions such Co<sup>2+</sup> and Ca<sup>2+</sup>, which typically bind to amino acid side chains containing oxygen atoms rather than sulfur groups, Zn<sup>2+</sup> was the most efficient inhibitor of SARS-CoV nsp12 RdRp activity (Supplemental Fig. S1). To test whether, as in the RTC assay, the RdRp inhibition by zinc ions was reversible, RdRp assays were pre-incubated with 6 mM Zn<sup>2+</sup>, a concentration that consistently gave >95% inhibition. After 30 min, 8 mM MgEDTA was added to both a control reaction and the reaction inhibited with ZnOAc<sub>2</sub>, and samples were incubated for another 30 min (Fig. 5C). As shown in Fig. 5D, the inhibition of EAV RdRp activity by Zn<sup>2+</sup> could be reversed by chelation of Zn<sup>2+</sup> (Fig. 5D; compare lanes 3 and 4). The amount of product synthesised was consistently 60 ± 5% of that synthesised in a 60-min control reaction (Fig. 5D; compare lanes 1 and 4), which was within the expected range given the shorter reaction time. The inhibition of the SARS-CoV RdRp was reversible as well. During the 30-min incubation after the addition of MgEDTA, SARS-CoV nsp12 incorporated 40 ± 5% of the label incorporated during a standard 60-min reaction (Fig. 5E). This was slightly lower than the expected yield and may be caused by the elevated Mg<sup>2+</sup> concentration, which was shown to be suboptimal for nsp12 activity [154] and results from the release of Mg<sup>2+</sup> from MgEDTA upon chelation of Zn<sup>2+</sup>.

### Differential effect of Zn<sup>2+</sup> on the initiation and elongation phase of nidovirus RNA synthesis

For EAV, close inspection of the RdRp assays revealed a less pronounced effect of Zn<sup>2+</sup> on the generation of full-length 18-nt products than on the synthesis of smaller reaction intermediates (Fig. 5A). This suggested that Zn<sup>2+</sup> specifically inhibited the initiation step of EAV RNA synthesis. To test this hypothesis, an RTC assay was incubated for 30 min with unlabelled CTP (initiation), after which the reaction was split in two. Then, [ $\alpha$ - $^{32}$ P]CTP was added to both tubes (pulse), 500  $\mu$ M Zn<sup>2+</sup> was added to one of the tubes, and samples were taken at different time points during the reaction (Fig. 6A). Fig. 6B shows that in the presence of Zn<sup>2+</sup> [ $\alpha$ - $^{32}$ P]CMP was predominantly incorporated into nascent RNA molecules that were already past the initiation phase at the moment that Zn<sup>2+</sup> was added to the reaction. No new initiation occurred, as was indicated by the smear of short radiolabelled products that progressively shifted up towards the position of full-length genomic RNA. This suggested that Zn<sup>2+</sup> does not affect the elongation phase of EAV



**Figure 5: The activity of the RdRps of EAV and SARS-CoV is reversibly inhibited by Zn<sup>2+</sup>.** RdRp activity of purified EAV nsp9 (**A**) and SARS-CoV nsp12 (**B**) in the presence of various Zn<sup>2+</sup> concentrations, as indicated above the lanes. (**C**) Schematic representation of the experiment to test whether Zn<sup>2+</sup>-mediated inhibition of RdRp activity could be reversed with MgEDTA. RdRp reactions, either untreated controls (sample 1 and 2) or reactions containing 6 mM Zn<sup>2+</sup> (samples 3 and 4) were incubated for 30 min. Both Zn<sup>2+</sup>-containing and control samples were split into two aliquots and 6 mM MgEDTA was added to sample 2 and 4. All reactions were incubated for an additional 30 min and then terminated. Reaction products of the RdRp assays with EAV nsp9 and SARS-CoV nsp12 are shown in (**D**) and (**E**), respectively. Numbers above the lanes refer to the sample numbers described under 5C.



**Figure 6: Effect of  $Zn^{2+}$  on initiation and elongation in *in vitro* assays with isolated EAV and SARS-CoV RTCs. (A)** An *in vitro* RTC assay with isolated EAV RTCs was allowed to initiate with unlabelled NTPs (initiation). After 30 min,  $[\alpha\text{-}^{32}\text{P}]\text{CTP}$  was added (pulse), the reaction was split into two equal volumes, and  $Zn^{2+}$  was added to a final concentration of 0.5 mM to one of the tubes. At the time points indicated, samples were taken and incorporation of  $[\alpha\text{-}^{32}\text{P}]\text{CMP}$  into viral RNA was analysed. **(B)** Radiolabelled EAV RNA synthesised at the time points indicated above the lanes in the presence and absence of  $Zn^{2+}$ . **(C)** Radiolabelled RNA synthesised by isolated SARS-CoV RTCs in reactions terminated after 100 (lane 1) and 40 (lane 2) min. Reaction products of a reaction to which 500  $\mu\text{M}$   $Zn^{2+}$  was added after 40 min, and that was terminated at  $t=100$  are shown in lane 3.

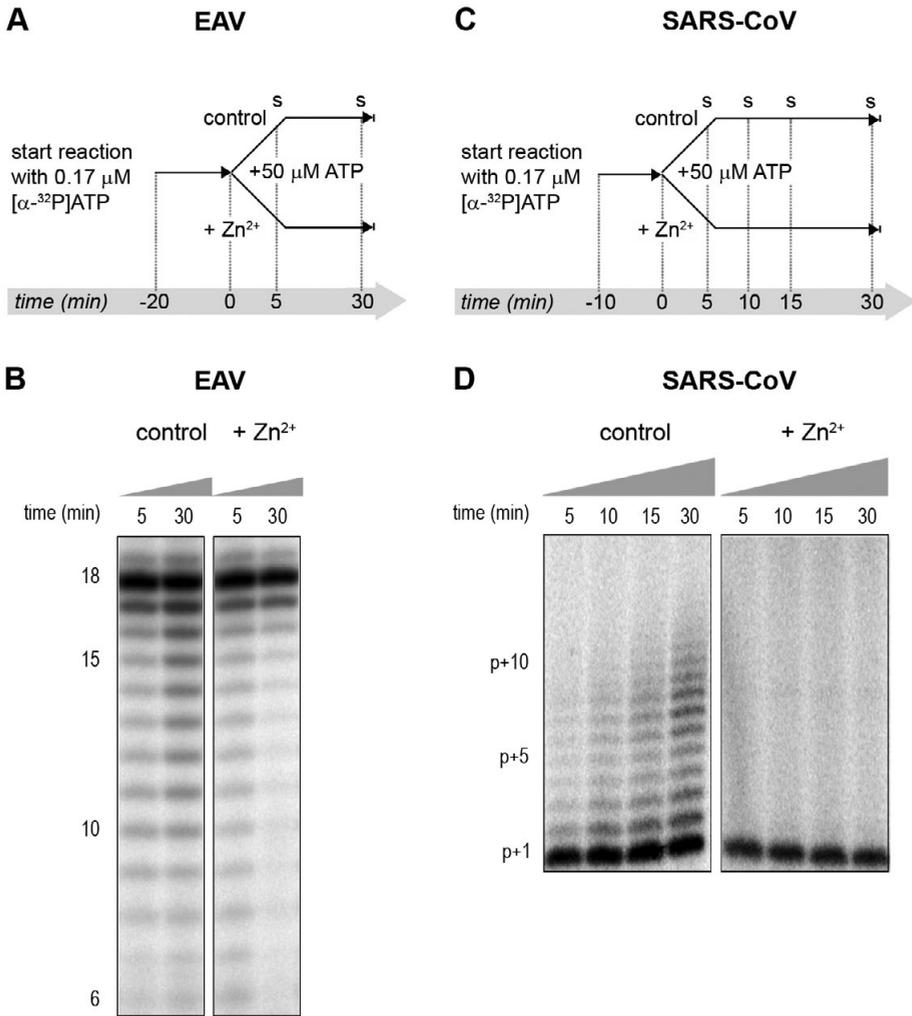
RNA synthesis and that it specifically inhibits initiation. This also explains the relatively weak signal intensity of the smaller sg mRNA bands (e.g., compare the relative change in signal of RNA2 to RNA7) produced in the presence of  $Zn^{2+}$ , since multiple initiation events are required on these short molecules to obtain signal intensities similar to those resulting from a single initiation event on the long genomic RNA, e.g., 16 times more in the case of RNA7. In contrast to EAV, the effect of  $Zn^{2+}$  on RNA synthesis by SARS-CoV RTCs was not limited to initiation, but appeared to impair the elongation phase as well,

given that the addition of Zn<sup>2+</sup> completely blocked further incorporation of [ $\alpha$ -<sup>32</sup>P]CMP when added 40 min after the start of the reaction (Fig. 6C).

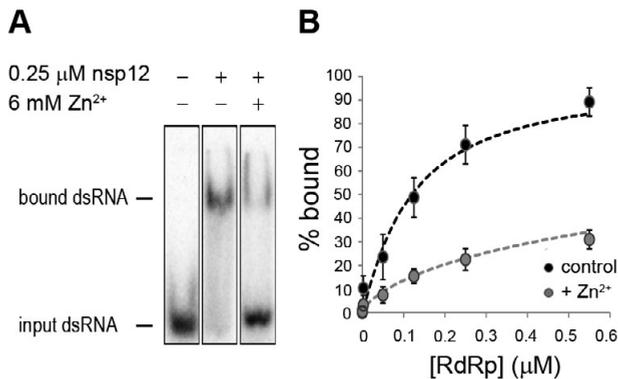
In the RdRp assays, the short templates used made it technically impossible to do experiments similar to those performed with complete RTCs. However, we previously noticed that at low concentrations of [ $\alpha$ -<sup>32</sup>P]ATP (~0.17  $\mu$ M) SARS-CoV nsp12 RdRp activity was restricted to the addition of only a single nucleotide to the primer [154]. EAV nsp9 mainly produced very short (2-3 nt long) abortive RNA products and only a fraction of full length products, as is common for *de novo* initiating RdRps [196]. This allowed us to separately study the effect of Zn<sup>2+</sup> on initiation and elongation by performing an experiment in which a pulse with a low concentration of [ $\alpha$ -<sup>32</sup>P]ATP was followed by a chase in the presence of 50  $\mu$ M of unlabelled ATP, which increased processivity and allowed us to study elongation (Fig. 7A and C) as described previously [154]. The results of these experiments were in agreement with those obtained with isolated RTCs. For EAV, with initiation and dinucleotide synthesis completely inhibited by the presence of 6 mM Zn<sup>2+</sup> (Supplemental Fig. S2A), the amount of reaction intermediates shorter than 18 nt diminished with time, while products from templates on which the RdRp had already initiated were elongated to full-length 18-nt molecules (Fig. 7B, right panel). This was consistent with the observation that the EAV RdRp remained capable of extending the synthetic dinucleotide ApA to trinucleotides in the presence of Zn<sup>2+</sup> (Supplemental Fig. S2B). Likely due to the absence of reinitiation in the reactions shown in Fig. 7B, the low processivity of the EAV RdRp, and the substrate competition between the remaining [ $\alpha$ -<sup>32</sup>P]ATP and the >200 fold excess of unlabelled ATP, the differences between the 5- and 30-min time points were small. In the absence of Zn<sup>2+</sup>, the RdRp continued to initiate as indicated by the ladder of smaller-sized RNA molecules below the full-length product (Fig. 7B, left panel) and the time course shown in Supplemental Fig. S2A. In contrast, the addition of Zn<sup>2+</sup> to a SARS-CoV RdRp reaction also blocked elongation, since extension of the radiolabelled primer as observed in the absence of Zn<sup>2+</sup> (Fig. 7D, left panel) no longer occurred (Fig. 7D, right panel).

### Zinc affects SARS-CoV RdRp template binding

To assess whether Zn<sup>2+</sup> affects the interaction of recombinant SARS-CoV nsp12 with the template used in our assays, we performed electromobility shift assays (EMSA) in the presence and absence of Zn<sup>2+</sup> (Fig. 8A). To measure the binding affinity of the RdRp for the template, we determined the fraction of bound template at various protein concentrations and observed a 3-4 fold reduction in RNA binding when Zn<sup>2+</sup> was present in the assay (Fig. 8B). We also assessed whether pre-incubation of the RdRp or RNA with Zn<sup>2+</sup> was a requirement for this drop in binding affinity, but found no significant difference with experiments not involving such a preincubation (data not shown).



**Figure 7: The effect of  $\text{Zn}^{2+}$  on initiation and elongation activity of purified EAV and SARS-CoV RdRps.** (A) An EAV RdRp reaction was initiated in the presence of [ $\alpha$ - $^{32}\text{P}$ ]ATP under conditions that do not allow elongation, *i.e.*, low ATP concentration. After 20 min, the reaction was split into two equal volumes, and  $\text{Zn}^{2+}$  was added to one of the tubes. A chase with 50  $\mu\text{M}$  unlabelled ATP, which allows elongation, was performed on both reactions and samples were taken after 5 and 30 min. (B) EAV RdRp reaction products that accumulated in the presence and absence of  $\text{Zn}^{2+}$  (indicated above the lanes) after a 5- and 30-min chase with unlabelled ATP. The length of the reaction products in nt is indicated next to the gel. (C) A SARS-CoV RdRp reaction was initiated in the presence of 0.17  $\mu\text{M}$  [ $\alpha$ - $^{32}\text{P}$ ]ATP, which limits elongation. After 10 min, the reaction was split into two equal volumes, and  $\text{Zn}^{2+}$  was added to one of the tubes. A chase with 50  $\mu\text{M}$  unlabelled ATP was performed on both reactions and samples were taken after 5, 10, 15, and 30 min. (D) SARS-CoV RdRp reaction products formed at the chase times indicated above the lanes in the presence and absence of  $\text{Zn}^{2+}$ . The length of the reaction products in nt is indicated next to the gel (p is the primer length).



**Figure 8: The effect of Zn<sup>2+</sup> on SARS-CoV nsp12 template binding.** (A) Electrophoretic mobility shift assay with radiolabelled dsRNA and nsp12 in the presence and absence of Zn<sup>2+</sup> (indicated above the lanes). The position of unbound and nsp12-bound RNA in the gel is marked on the left of the panel. (B) Determination of the nsp12 affinity for RNA in the presence and absence of Zn<sup>2+</sup>. A fixed amount of RNA was incubated with an increasing amount of nsp12. This revealed a 3-4 fold reduction in the percentage of bound RNA in the presence of zinc ions (grey) relative to the percentage of bound RNA in the absence of zinc ions (black). Error bars represent standard error of the mean (n = 3).

No binding was observed when a similar RNA binding assay was performed with purified EAV RdRp. Likewise, nsp9 did not bind RNA in pull-down experiments with Talon-beads, His<sub>6</sub>-tagged nsp9, and radiolabelled EAV genomic RNA or various short RNA templates including polyU, whereas we were able to detect binding of a control protein (SARS-CoV nsp8, which has demonstrated RNA and DNA binding activity [153]) using this assay. It presently remains unclear why we are not able to detect the binding of recombinant EAV nsp9 to an RNA template.

## Discussion

Although a variety of compounds have been studied, registered antivirals are currently still lacking for the effective treatment of SARS and other nidovirus-related diseases [316]. RdRps are suitable targets for antiviral drug development as their activity is strictly virus-specific and may be blocked without severely affecting key cellular functions. Several inhibitors developed against the polymerases of, *e.g.*, human immunodeficiency virus (HIV) and hepatitis C virus are currently being used in antiviral therapy or clinical trials [276,317,318]. Therefore, advancing our molecular knowledge of nidovirus RdRps and the larger enzyme complexes that they are part of, and utilising the potential of recently developed *in vitro* RdRp assays [154,196,290,313] could ultimately aid in the development of effective antiviral strategies.

Zinc ions and zinc-ionophores, such as PT and PDTC, have previously been described as potent inhibitors of various RNA viruses. We therefore investigated whether PT-

stimulated import of zinc ions into cells also inhibited the replication of nidoviruses in cell culture. Using GFP-expressing EAV and SARS-CoV [314,315], we found that the combination of 2  $\mu\text{M}$  PT and 2  $\mu\text{M}$   $\text{Zn}^{2+}$  efficiently inhibited their replication, while not causing detectable cytotoxicity (Fig. 1). Inhibition of replication by PT and  $\text{Zn}^{2+}$  at similar concentrations (2-10  $\mu\text{M}$ ) was previously observed for several picornaviruses such as rhinoviruses, foot-and-mouth disease virus, coxsackievirus, and mengovirus [300,301,302,303,304,305].

The inhibitory effect of  $\text{Zn}^{2+}$  on the replication of picornaviruses appeared to be due to interference with viral polyprotein processing. In infections with the coronavirus mouse hepatitis virus (MHV),  $\text{Zn}^{2+}$  also interfered with some of the replicase polyproteins cleavages [312], albeit at a much higher concentration (100  $\mu\text{M}$   $\text{Zn}^{2+}$ ) than used in our studies. Since impaired replicase processing will indirectly affect viral RNA synthesis in the infected cell, we used two recently developed *in vitro* assays to investigate whether  $\text{Zn}^{2+}$  also affects nidovirus RNA synthesis directly. Our *in vitro* studies revealed a strong inhibitory effect of zinc ions on the RNA-synthesising activity of isolated EAV and SARS-CoV RTCs. Assays with recombinant enzymes subsequently demonstrated that this was likely due to direct inhibition of RdRp function. The inhibitory effect could be reversed by chelating the zinc ions, which provides an interesting experimental (on/off) approach to study nidovirus RNA synthesis. Addition of  $\text{Zn}^{2+}$  following initiation of EAV RNA synthesis had little or no effect on NTP incorporation in molecules whose synthesis had already been initiated in the absence of  $\text{Zn}^{2+}$  (Fig. 6 and 7), indicating that  $\text{Zn}^{2+}$  does not affect elongation and does not increase the termination frequency, as was previously found for  $\text{Mn}^{2+}$  [313]. Therefore,  $\text{Zn}^{2+}$  appears to be a specific inhibitor of the initiation phase of EAV RNA synthesis. In contrast,  $\text{Zn}^{2+}$  inhibited SARS-CoV RdRp activity also during the elongation phase of RNA synthesis, probably by directly affecting template binding (Fig. 8). In coronaviruses, zinc ions thus appear to inhibit both the proper proteolytic processing of replicase polyproteins [311,312] and RdRp activity (this study). Contrary to the RTC assays, millimolar instead of micromolar concentrations of  $\text{ZnOAc}_2$  were required for a nearly complete inhibition of nucleotide incorporation in RdRp assays.

It has been well established that DNA and RNA polymerases use conserved aspartate residues in motifs A and C to bind divalent metal ions like  $\text{Mg}^{2+}$ , which subsequently coordinate incoming nucleotides during the polymerisation reaction [279,280].  $\text{Mg}^{2+}$  is also the divalent metal ion that is required for the *in vitro* activity of isolated SARS-CoV and EAV RTCs and recombinant RdRps [154,196,290,313], although *de novo* initiation of EAV nsp9 is primarily  $\text{Mn}^{2+}$ -dependent.  $\text{Zn}^{2+}$  could not substitute for  $\text{Mg}^{2+}$  or  $\text{Mn}^{2+}$  as cofactor as it was incapable of supporting the polymerase activity of nidovirus RTCs and RdRps in the absence of  $\text{Mg}^{2+}$  (data not shown), as was also reported for the poliovirus RdRp [243]. Moreover, inhibition of nidovirus RdRp activity by  $\text{Zn}^{2+}$  was even observed at low concentrations and in the presence of a more than 25-fold excess of  $\text{Mg}^{2+}$ , sug-

gesting that either the affinity of the active site for Zn<sup>2+</sup> is much higher or that Zn<sup>2+</sup> does not compete for Mg<sup>2+</sup>-binding and binds to another zinc(-specific) binding site in the protein.

Specific protein domains or pockets that contain zinc ions may be involved in protein-protein interactions, protein-RNA/DNA interactions, or conformational changes in enzyme structures. Zinc-binding domains commonly consist of at least three conserved cysteine and/or histidine residues within a stretch of ~10-30 amino acids, such as in zinc-finger motifs and metalloproteases [92,296,319]. However, in RdRps there are only few precedents for the presence of zinc-binding pockets, such as those identified in the crystal structure of the Dengue RdRp [145]. Sequence analysis of the EAV nsp9 amino acid sequence revealed that it lacks patches rich in conserved cysteines and/or histidines. In contrast, inspection of the SARS-CoV nsp12 amino acid sequence revealed two such patches, namely H295-C301-C306-H309-C310 and C799-H810-C813-H816. A crystal structure for nsp12 is presently unavailable, but a predicted structure that represents the C-terminal two-thirds of the enzyme has been published [89]. Interestingly, in this model, C799, H810, C813 and H816 are in a spatial arrangement resembling that of the Zn<sup>2+</sup> coordinating residues in the Zn<sub>2</sub> zinc-binding pocket found in motif E of the Dengue virus RdRp (see Supplemental Fig. S3). Clearly, an in-depth analysis of nidovirus RdRps, *e.g.*, through structural analysis and subsequent mutational studies targeting aforementioned cysteines and histidines, is required to provide further insight into and a structural basis for the Zn<sup>2+</sup>-induced inhibitory effects on RdRp activity documented in this study. Such studies may, however, be complicated when Zn<sup>2+</sup> binding proves to be very transient in nature and not detectable with currently available methods.

In summary, the combination of zinc ions and the zinc-ionophore PT efficiently inhibits nidovirus replication in cell culture. This provides an interesting basis for further studies into the use of zinc-ionophores as antiviral compounds, although systemic effects have to be considered [320,321] and a water-soluble zinc-ionophore may be better suited, given the apparent lack of systemic toxicity of such a compound at concentrations that were effective against tumors in a mouse xenograft model [322]. *In vitro*, the reversible inhibition of the RdRp by Zn<sup>2+</sup> has also provided us with a convenient research tool to gain more insight into the molecular details of (nido)viral RNA synthesis, and revealed novel mechanistic differences between the RdRps of SARS-CoV and EAV.

## Materials and methods

### Cells and viruses

Vero-E6 cells were cultured and infected with SARS-CoV (strain Frankfurt-1; accession nr. AY291315) or SARS-CoV-GFP as described previously [67]. All procedures involving live SARS-CoV were performed in the biosafety level 3 facility at Leiden University Medical

Center. BHK-21 or Vero-E6 cells were cultured and infected with EAV (Bucyrus strain; accession nr. NC\_002532) or EAV-GFP [314] as described elsewhere [313].

### Effect of zinc ions on nidovirus replication in cell culture

One day prior to infection, Vero-E6 cells were seeded in transparent or black (low fluorescence) 96-well clusters at 10,000 cells per well. The next day, cells were infected with SARS-CoV-GFP or EAV-GFP with an m.o.i. of 4, and 1 h p.i. the inoculum was removed and 100  $\mu$ l of medium containing 2% fetal calf serum (FCS) was added to each well. In some experiments 0-32  $\mu$ M of pyriothione (Sigma) was added in addition to 0-2  $\mu$ M ZnOAc<sub>2</sub>. Infected cells were fixed at 17 h p.i. by aspirating the medium and adding 3% paraformaldehyde in PBS. After washing with PBS, GFP expression was quantified by measuring fluorescence with a LB940 Mithras plate reader (Berthold) at 485 nm. To determine toxicity of ZnOAc<sub>2</sub> and PT, cells were exposed to 0-32  $\mu$ M PT and 0-8  $\mu$ M ZnOAc<sub>2</sub>. After 18 h incubation, cell viability was determined with the Cell Titer 96 AQ MTS assay (Promega). EC<sub>50</sub> and CC<sub>50</sub> values were calculated with Graphpad Prism 5 using the nonlinear regression model.

### RNA templates and oligonucleotides

RNA oligonucleotides SAV557R (5'-GCUAUGUGAGAUUAAGUUU-3'), SAV481R (5'-UUUUUUUUUUUAACUUAUCUCACAUAGC-3') and poly(U)<sub>18</sub> (5'-UUUUUUUUUUUUUUUUUU-3') were purchased from Eurogentec, purified from 7 M Urea/15% PAGE gels and desalted through NAP-10 columns (GE healthcare). To anneal the RNA duplex SAV557R/SAV481R, oligonucleotides were mixed at equimolar ratios in annealing buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl and 5 mM EDTA), denatured by heating to 90°C and allowed to slowly cool to room temperature after which they were purified from 15% non-denaturing PAGE gels.

### *In vitro* viral RNA synthesis assay with isolated RTCs

SARS-CoV and EAV RTCs were isolated from infected cells and assayed for activity *in vitro* as described previously [290,313]. To assess the effect of Zn<sup>2+</sup>, 1  $\mu$ l of a ZnOAc<sub>2</sub> stock solution was added to standard 28- $\mu$ l reactions, resulting in final Zn<sup>2+</sup> concentrations of 10-500  $\mu$ M. When Zn<sup>2+</sup> had to be chelated in the course of the reaction, magnesium-saturated EDTA (MgEDTA) was added to a final concentration of 1 mM. After RNA isolation, the <sup>32</sup>P-labelled reaction products were separated on denaturing 1% (SARS-CoV) or 1.5% (EAV) agarose formaldehyde gels. The incorporation of [ $\alpha$ -<sup>32</sup>P]CMP into viral RNA was quantified by phosphorimaging of the dried gels using a Typhoon scanner (GE Healthcare) and the ImageQuant TL 7 software (GE Healthcare).

### Expression and purification of nidovirus RdRps

SARS-CoV nsp12 and EAV nsp9 were purified essentially as described elsewhere [154,196], but with modifications for nsp9. In short, *E. coli* BL21(DE3) with plasmid pDEST14-nsp9-CH was grown in auto-induction medium ZYM-5052 [292] for 6 hours at 37°C and a further 16 hours at 20°C. After lysis in buffer A (20 mM HEPES pH 7.4, 200 mM NaCl, 20 mM imidazole, and 0.05% Tween-20) the supernatant was applied to a HisTrap column (GE Healthcare). Elution was performed with a gradient of 20-250 mM imidazole in buffer A. The nsp9-containing fraction was further purified by gel filtration in 20 mM HEPES, 300 mM NaCl and 0.1% Tween-20 on a Superdex 200 column (GE Healthcare). The fractions containing nsp9-CH were pooled, dialysed against 1000 volumes of buffer B (20 mM HEPES, 100 mM NaCl, 1 mM DTT and 50% glycerol) and stored at -20°C. RdRps with a D618A (SARS-CoV) or D445A (EAV) mutation were obtained by site-directed mutagenesis of the wild-type (wt) plasmid pDEST14-nsp9-CH [196] with oligonucleotides 5'-TACTGCCTTGAAACAGCCCTGGAGAGTTGTGAT-3' and 5'-ATCACAACCTCTCCAGGGCTGTTTCAAGGCAGTA-3', and plasmid pASK3-Ub-nsp12-CHis<sub>6</sub> with oligonucleotides 5'-CCTTATGGGTTGGGCTTATCCAAAATGTG-3' and 5'-CACATTTTGGATAAGCCCAACCCATAAGGA-3', as described elsewhere [154]. Mutant proteins were purified parallel to the wt enzymes.

### RdRp assays with purified enzymes

Standard reaction conditions for the RdRp assay with 0.1 μM of purified SARS-CoV nsp12 are described elsewhere [154]. To study the effect of Zn<sup>2+</sup> in this assay, 0.5 μl of a dilution series of 0-80 mM ZnOAc<sub>2</sub> was added to the 5 μl reaction mixture, yielding final Zn<sup>2+</sup> concentrations of 0-8 mM. The EAV RdRp assay contained 1 μM nsp9, 1 μM RNA template poly(U)<sub>18</sub>, 0.17 μM [α-<sup>32</sup>P]ATP (0.5 μCi/μl; Perkin-Elmer), 50 μM ATP, 20 mM Tris-HCl (pH 8.0), 10 mM NaCl, 10 mM KCl, 1 mM MnCl<sub>2</sub>, 4 mM MgOAc<sub>2</sub>, 5% glycerol, 0.1% Triton-X100, 1 mM DTT and 0.5 units RNaseOUT. ZnOAc<sub>2</sub> was added to the reaction to give a final concentration of 0-6 mM. To chelate Zn<sup>2+</sup> during reactions, MgEDTA was added to a final concentration of 8 mM. Reactions were terminated after 1 hour and analysed as described [154].

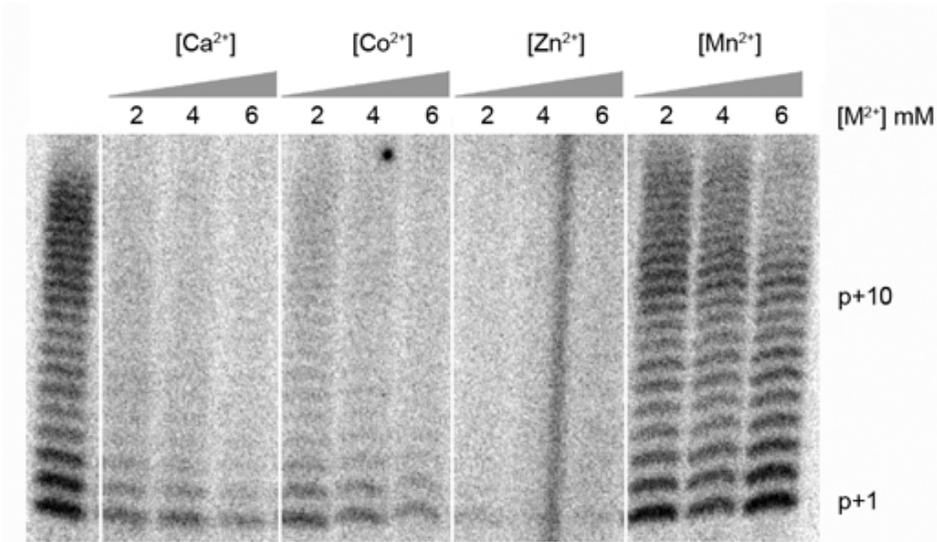
### SARS-CoV nsp12 electrophoretic mobility shift assay

SARS-CoV RdRp was incubated with 0.2 nM 5' <sup>32</sup>P-labelled SAV557R/SAV481R RNA duplex, for 10 minutes at 30°C either in presence or absence of 6 mM ZnOAc<sub>2</sub>. Reactions were analysed as described previously [154].

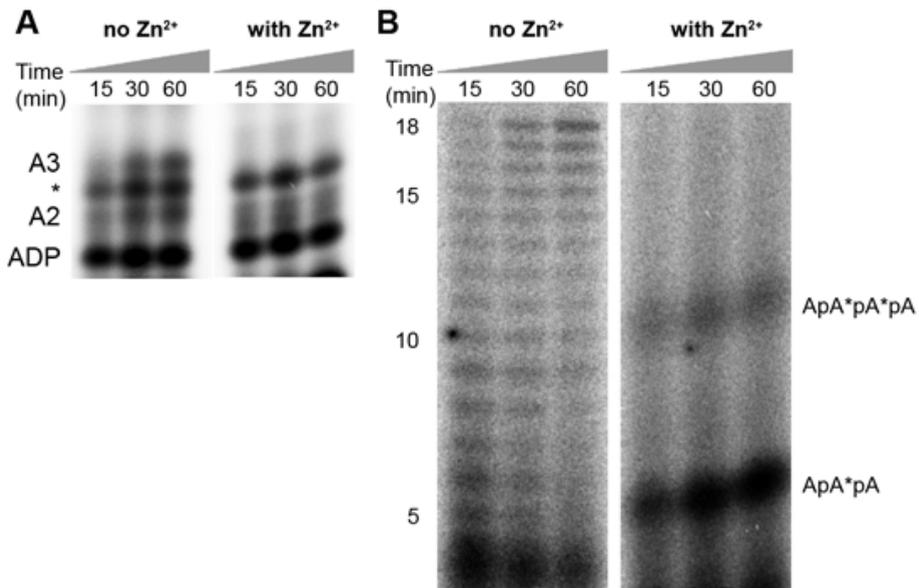
## **Acknowledgements**

This work was supported by the Netherlands Organization for Scientific Research (NWO) with grants from the Council for Chemical Sciences (NWO-CW grant 700.55.002) and an NWO Toptalent grant (021.001.037).

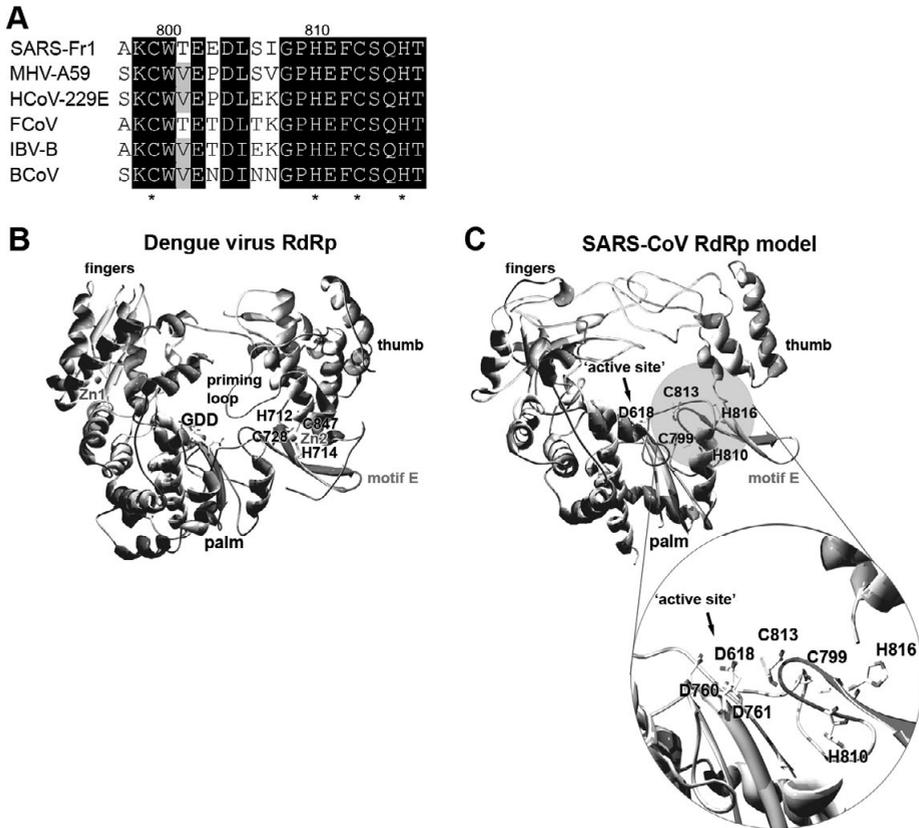
## CHAPTER 4 - SUPPLEMENTAL INFORMATION



**Figure S1: Effect of various divalent cations on the RdRp activity of SARS-CoV nsp12.** Purified recombinant SARS-CoV nsp12 was incubated with a primed template, ATP, and [ $\alpha$ -<sup>32</sup>P]ATP in the presence of either 6 mM Mg<sup>2+</sup> only (lane 1), and with increasing concentrations of a second divalent metal (M<sup>2+</sup>), specifically: 2–6 mM Ca<sup>2+</sup> (lane 2–4), 2–6 mM Co<sup>2+</sup> (lane 5–7), 2–6 mM Zn<sup>2+</sup> (lane 8–10), or 2–6 mM Mn<sup>2+</sup> (lane 11–13). The strongest inhibition was observed for Zn<sup>2+</sup>. For more details on the SARS-CoV nsp12 RdRp assay, see the main text.



**Figure S2: Effect of Zn<sup>2+</sup> on the dinucleotide extension activity of EAV nsp9.** Purified recombinant EAV nsp9 was incubated with a U<sup>18</sup> template in the presence of [ $\alpha$ -<sup>32</sup>P]ATP, ATP, 4 mM Mg<sup>2+</sup>, 1 mM Mn<sup>2+</sup>, and 1  $\mu$ M ApA. **(A)** Reaction mixtures were split into two aliquots, one of which was supplemented with 6 mM Zn<sup>2+</sup>, and samples were taken at the time points (minutes) indicated above the lanes. In the absence of Zn<sup>2+</sup>, EAV nsp9 initiates *de novo* and produces di- and trinucleotides, indicated with A2 and A3, respectively. A non-specific band, unrelated to RdRp activity, between A2 and A3 is indicated with an asterisk. In the presence of 6 mM Zn<sup>2+</sup>, the synthesis of dinucleotides and trinucleotides was blocked. **(B)** When performing the assay described under (A) in the absence of Zn<sup>2+</sup>, a full-length product of 18 nucleotides is formed. This product is not observed when the assay is performed in the presence of 6 mM Zn<sup>2+</sup>, but nsp9 was capable of elongating the provided dinucleotide primer ApA into tri- (ApA\*pA) and tetranucleotide ((ApA\*pA\*pA) products (the asterisk indicates radiolabelled phosphates). Due to the absence of a 5' triphosphate group, these reaction products migrate much slower in the 20% acrylamide and 7 M urea gel used for this analysis. See the main text for additional experimental details on the EAV nsp9 RdRp assay.



**Figure S3: Putative zinc-binding residues in the predicted structure of SARS-CoV nsp12 and comparison with the structure of the zinc-containing Dengue virus RdRp domain. (A)** Sequence alignment of coronavirus RdRps showing conservation of four potential zinc-binding residues amino acids (C799-H810-C813-H816 in SARS-CoV; indicated with asterisks) in the C-terminal region of coronavirus nsp12. Black shading indicates complete conservation among coronaviruses. The coronavirus RdRp sequences were aligned with Muscle 3.6. The aligned sequences and NCBI accession numbers are the following: mouse hepatitis virus strain A59 (MHV\_A59; NP\_068668), human CoV 229E (HCoV\_229E; NP\_068668), infectious bronchitis virus strain Beaudette (IBV\_B; P0C6Y1), bovine coronavirus (BCoV; NP\_742138.1), feline coronavirus (FeCoV; YP\_239353.1), and SARS-CoV strain Frankfurt-1 (SARS\_Fr1; AAP33696). **(B)** Crystal structure of the Dengue virus RdRp domain showing the position of four cysteine and histidine residues that form Zn<sup>2+</sup>-binding pocket Zn2, located close to motif E (depicted in light grey). A second Zn<sup>2+</sup>-binding pocket (Zn1) and the two zinc ions identified in the crystal structure are indicated in light-grey. **(C)** Predicted three-dimensional structure model of SARS-CoV nsp12 (Xu *et al.*, Nucl. Acids Res. 31: 7117–7130), based on PDB code 1O5S, rendered with Swiss-PdbViewer 4.01 and POV-Ray 3.6. The positions of the conserved cysteine and histidine residues indicated in panel A (C799-H810-C813-H816) close to motif E (depicted in light grey) and RdRp active-site residues (D618, D760 and D761) are indicated. The spatial arrangement of these cysteines and histidines in this model strikingly resembles the positioning of the metal ion-coordinating residues of Zn-binding pocket Zn2 in the Dengue virus RdRp domain (see panel B).



# Chapter 5

The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both *de novo* initiation and primer extension

Aartjan J.W. te Velhuis<sup>1</sup>, Sjoerd H.E. van den Worm<sup>1#</sup> and Eric J. Snijder<sup>1</sup>

Molecular Virology laboratory,  
Department of Medical  
Microbiology, Center of Infectious  
Diseases, Leiden University  
Medical Center, PO Box 9600,  
2300RC Leiden, The Netherlands.

# Present address: Vaccine and  
Gene Therapy Institute, Oregon  
Health and Science University,  
Beaverton, Oregon, USA

**Published in:** Nucleic Acids  
Research, Volume 40, Pages 1037-  
1047, 2012. Epub October 2011.

## ABSTRACT

Uniquely among RNA viruses, replication of the ~30-kilobase SARS-coronavirus genome is believed to involve two RNA-dependent RNA polymerase (RdRp) activities. The first is primer-dependent and associated with the 106-kDa nonstructural protein 12 (nsp12), whereas the second is catalysed by the 22-kDa nsp8. This latter enzyme is capable of *de novo* initiation and has been proposed to operate as a primase. Interestingly, this protein has only been crystallised together with the 10-kDa nsp7, forming a hexadecameric, dsRNA-encircling ring structure (*i.e.*, nsp(7+8), consisting of 8 copies of both nsps). To better understand the implications of these structural characteristics for nsp8-driven RNA synthesis, we studied the prerequisites for the formation of the nsp(7+8) complex and its polymerase activity. We found that in particular the exposure of nsp8's natural N-terminal residue was paramount for both the protein's ability to associate with nsp7 and for boosting its RdRp activity. Moreover, this "improved" recombinant nsp8 was capable of extending primed RNA templates, a property that had gone unnoticed thus far. The latter activity is, however, ~20-fold weaker than that of the primer-dependent nsp12-RdRp at equal monomer concentrations. Finally, site-directed mutagenesis of conserved D/ExD/E motifs was employed to identify residues crucial for nsp(7+8) RdRp activity.

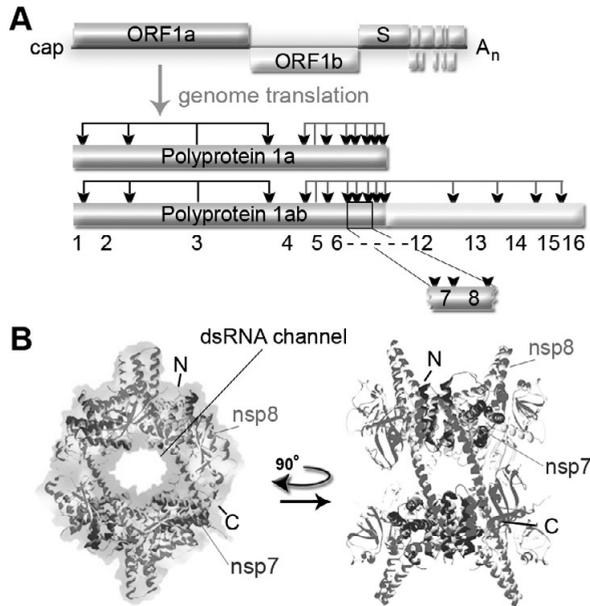
## INTRODUCTION

In the replicative cycle of RNA viruses, the crucially important process of RNA-templated RNA synthesis is generally performed by an RNA-synthesizing complex of viral enzymes [235,264]. Commonly, its core subunit is a single RNA-dependent RNA polymerase (RdRp) that drives the production of template strands for replication, new genome molecules, and - in many RNA virus groups - also subgenomic (sg) mRNAs. This canonical RdRp is structurally conserved among RNA viruses and widely accepted to drive catalysis of phosphodiester bond formation via a well-established reaction mechanism involving two metal ions that are coordinated by aspartate residues in its motifs A and C [89,279,281].

Uniquely among RNA viruses however, current evidence suggests that at least two RdRp activities are encoded by the genomes of members of the coronavirus (CoV) family, the +RNA virus group that infects a wide range of vertebrates and is renowned for its exceptionally large polycistronic genome of approximately 30 kilobases [65]. Both CoV RdRps belong to the set of 16 nonstructural proteins (nsps) that are produced through proteolytic processing of the pp1a and pp1ab replicase precursor polyproteins, which both derive from translation of the genomic RNA [58,323]. For the Severe Acute Respiratory Syndrome-associated coronavirus (SARS-CoV), which emerged in 2003 and caused worldwide concern due to the ~10% mortality rate associated with infection of humans [65,324], the two replicase subunits with RdRp activity have been studied in some detail. The first is the 106-kDa nsp12, which contains the canonical viral RdRp motifs in its C-terminal part and employs a primer-dependent initiation mechanism [154,325]. The second polymerase, the 22-kDa nsp8, is unique for CoVs and was reported to be only capable of *de novo* RNA synthesis on ssRNA templates, albeit with low fidelity [155]. Together, these observations inspired a hypothesis in which nsp8 would serve as an RNA primase, *i.e.*, would synthesise short oligonucleotide primers for subsequent extension by the nsp12 “main RdRp” [155].

In spite of this attractive model, however, many questions regarding CoV RNA synthesis remain unanswered thus far. For instance, it is unclear whether the homomeric form of nsp8, for which *in vitro* RdRp activity was previously documented [155], actually occurs *in vivo*, as nsp8 was also shown to co-crystallise and form a unique hexadecameric ring-structure with the 10-kDa nsp7 subunit, which resides immediately upstream in the replicase polyprotein precursors (Fig. 1) [153]. In a similar fashion, it is presently unknown whether the postulated double-stranded RNA (dsRNA) binding channel of this complex plays a role in the RdRp activity of nsp8 and whether this activity is influenced by nsp7, particularly given the observed low fidelity and low processivity of nsp8 [155].

To investigate the properties of the nsp7+nsp8 (nsp(7+8)) hexadecamer in more detail, and seek answers to the above questions, we here generated and purified re-



**Figure 1: SARS-coronavirus genome organization and structure of the nsp7+nsp8 hexadecamer.** (A) The coronavirus genome contains two large 5'-proximal ORFs (ORF1a and 1b) that encode the two replicase polyproteins, whose mature products bring about the formation of the viral replication and transcription complex. Both polyproteins are cleaved (cleavage sites indicated with arrow heads) by the proteinase activities of nsp3 (left hand side) and nsp5 (right hand side), which releases the mature nsps. Also indicated are the 5' cap structure and the 3' polyA tail (A<sub>n</sub>). (B) The SARS-CoV nsp8 crystal structure (pdb 2AHM) resembles a "golf club-like" shape. This nsp8 conformation connects to a much larger, hexadecameric structure that is composed of seven additional nsp8 subunits (light grey) and eight nsp7 subunits (dark grey). The hollow hexadecameric ring structure has a positively charged channel (darker grey background shading) that was proposed to mediate RNA binding. The outside of the structure is predominantly negatively charged (light grey background shading).

combinant forms of SARS-CoV nsp8 and nsp(7+8) that have natural N-terminal residues. This technical refinement was found to greatly improve nsp8's ability to associate with nsp7. Moreover, and in contrast to previous observations [155], exposure of the natural N-terminus proved crucial for the enzymatic activity of the complex on partially double-stranded RNA templates, demonstrating that nsp(7+8) is capable of primer-dependent RdRp activity as well. Site-directed mutagenesis of nsp8 in the context of the nsp(7+8) complex identified a conserved D/ExD/E motif that is important for catalysis *in vitro*, possibly providing a first indication of the location of the presently unknown nsp8 active site. Overall, these results define the SARS-CoV nsp(7+8) complex as an intriguing multimeric RNA polymerase that is capable of primer extension.

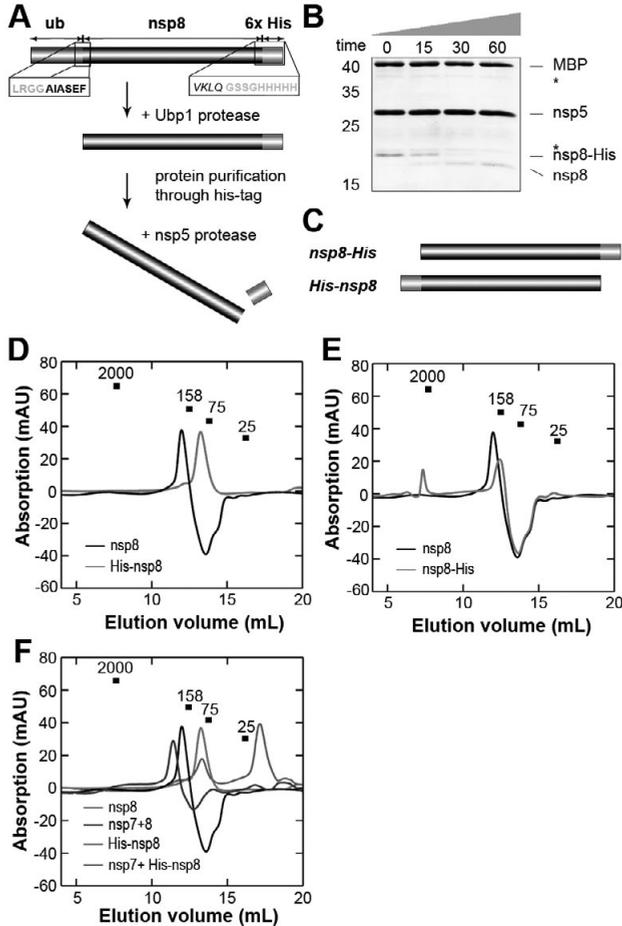
## RESULTS

### N-terminal processing defines nsp8 multimerisation and nsp(7+8) complex formation

SARS-CoV nsp7 and nsp8 were previously reported to interact and form a hollow ring structure that is composed of an intricate nsp8 octamer supported by eight copies of nsp7 [153,289] (Fig. 1B). Based on the large diameter, positive charge of the hexadecamer's channel and *in silico* docking, it was proposed to be able to encircle dsRNA (Fig. 1B). However, the functional significance of the compound interactions between nsp7 and nsp8 is poorly understood, as are the polymerase activities associated with monomeric nsp8 or nsp8-containing multimers. So far, strategies for the purification of recombinant nsp8 have involved the use of affinity tags (*e.g.*, His<sub>6</sub> or glutathione-S-transferase (GST) [153,155]) that were fused to one terminus to facilitate protein recovery. Inadvertently though, such tags or other exogenous sequences may significantly impede the correct folding of enzymes and thus alter their stability or activity, as exemplified by studies of the poliovirus (3D<sup>pol</sup>) and SARS-CoV (nsp12) RdRp subunits [154,277,278]. To circumvent this issue, we developed a protocol in which SARS-CoV nsp8 was expressed as a ubiquitin (ub) fusion protein carrying a C-terminal His<sub>6</sub>-tag (ub-nsp8-His), which was subsequently processed at both termini in two steps. The first step was co-translational and involved the release of the N-terminal ub fusion partner by the co-expressed ubiquitin carboxyl-terminal hydrolase 1 (Ubp1, Fig. 2A) [154,277]. The second proteolytic step, catalysed by a recombinant form of the SARS-CoV nsp5 main protease [326], removed the C-terminal His<sub>6</sub>-tag and was performed either in solution (Fig. 2A and 2B) or when nsp8-His was immobilised to Talon beads. This procedure yielded SARS-CoV nsp8 with its exact natural N- and C-terminus (replicase residues Ala-3920 and Gln-4117, respectively; Fig. 2A), the product that is normally liberated by the nsp5-driven autoprocessing of the SARS-CoV replicase polyproteins [44].

In accordance with the octameric state observed in cross-linking experiments using glutaraldehyde (Fig. S1) or ethylene glycolbis [153], the hydrodynamic profile of the untagged nsp8 corresponded to a mass of ~160 kDa (Fig. 2D). To identify and explain differences with previously published observations, we also produced and characterised N- and C-terminally tagged forms of nsp8 (Fig. 2C). Importantly, under the same assay conditions, the N-terminally His<sub>6</sub>-tagged nsp8 (His-nsp8) that was used in the original nsp8 RdRp activity study [155] showed a marked difference in multimerisation behaviour (Fig. 2D and Fig. S1). On the other hand, little difference was observed between untagged nsp8 and a C-terminally His<sub>6</sub>-tagged version of the protein (nsp8-His; Fig. 2E).

To investigate whether nsp7 could influence the change in multimerisation behaviour, we next added separately purified and C-terminally processed nsp7 to the different nsp8 preparations. Interestingly, we found that nsp8 and nsp8-His could both associate with



**Figure 2: Purification and multimerisation of recombinant SARS-CoV nsp7-8 precursor and different nsp8 variants.** (A) Expression of nsp8 in the presence of the ubiquitin protease Ubp1 to liberate the natural N-terminal sequence (AIASEF), followed by purification and cleavage by recombinant SARS-CoV nsp5 main protease to remove the C-terminal His<sub>6</sub>-tag and its upstream GSSG linker. (B) Eighteen per cent SDS-PAGE analysis of nsp5-treated, purified nsp8-His demonstrates near-complete release of the C-terminal His<sub>6</sub>-tag within 60 min. The maltose binding protein (MBP) was added to the reaction to serve as an independent loading control. Asterisks indicate nonspecific bands. (C) In addition to the tag-less nsp8 and nsp8-His, we also produced the N-terminally His<sub>6</sub>-tagged nsp8 (His-nsp8) used by Imbert *et al.* [155]. (D) Comparative gel filtration analysis of nsp8 (22 kDa as a monomer) versus His-nsp8 and (E) nsp8 versus nsp8-His. In all three cases, nsp8 formed multimers in solution, but the apparent molecular mass of complexes formed by both nsp8 and nsp8-His was ~2 fold higher than for complexes formed by His-nsp8. (F) Comparative analysis of nsp8, nsp(7+8), His-nsp8 and nsp7+nsp8-His. Only nsp(7+8) showed a molecular weight shift to the ~225-kDa size range with a standard deviation of 15-kDa (n = 3). This size is indicative of hexadecamer formation, whereas the analysis of nsp7+nsp8-His showed dominant peaks of nsp8-His and nsp7 (which is ~10 kDa as a monomer).

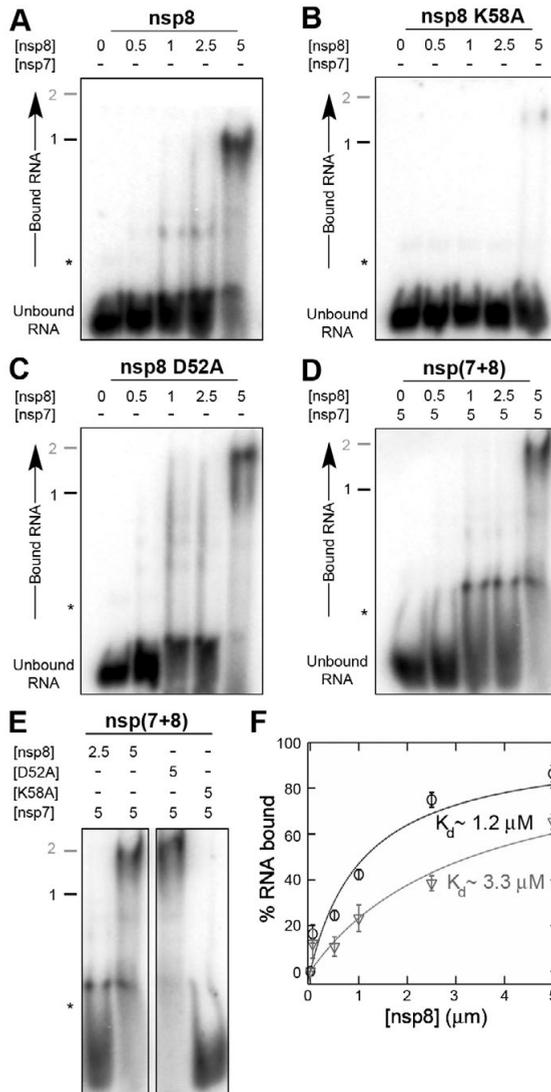
this protein, in accordance with published data [153], but that His-nsp8 was unable to do so within the frame of our experimental conditions (Fig. 2F). Consequently, although various lines of evidence support the observation that nsp7 and nsp8 can form a hexadecamer, it now appears that the correct N-terminal processing of nsp8 is a significant factor in determining the final oligomeric state of the protein.

### SARS-CoV nsp7 enhances RNA binding by nsp8

A unique feature of the hexadecameric SARS-CoV nsp(7+8) structure is the fact that it does not derive from stacking of its protein subunits, but rather from stable interconnections of the “golf club-like” nsp8 molecules (Fig. 1B) [153]. The structural support of the nsp8 octamer by eight copies of nsp7 thus appears to be redundant, in line with the critical role for the nsp8 N-terminal domain described above. We surmised therefore that the additional complexity must have evolved to improve nsp8’s function and set out to compare the RNA binding capabilities of the purified nsp8 octamer and nsp(7+8) hexadecamer.

By analysing the steady-state ribonucleotide-protein (RNP) complexes formed through binding of nsp8 to 5′ <sup>32</sup>P-labelled dsRNA (Fig. 3A), we estimated the nsp8 dissociation constant ( $K_d$ ) for dsRNA to be ~3.3 μM (Fig. 3F), which is about ~25 fold higher than the apparent  $K_d$  of nsp12 under comparable conditions [154]. A comprehensive analysis of the influence of nsp7 on nsp8-dependent RNA binding required an nsp8 mutant that was incapable of RNA binding. To this end, we engineered an alanine substitution of the conserved residue K58, which resides in nsp8’s proposed dsRNA-binding channel (residues 55-78 [153]). As is evident from the electromobility shift assay in Fig. 3B, this mutation was sufficient to significantly disrupt RNA binding. As a control, we also performed an aspartate-to-alanine substitution at position 52, which is partially conserved, yet not expected to participate in RNA backbone binding due to its negative charge and position just outside the proposed RNA binding channel. Indeed, the D52A mutation only induced a migratory shift of the dominant RNP signal towards the anode, likely as a result of the lost negative charge (Fig. 3C).

With the results obtained with these control proteins in mind, we next explored the contribution of nsp7 to RNA binding by the nsp(7+8) complex. We used a fixed concentration of nsp7 and added either wild-type or mutant nsp8 up to the point where the nsp7:nsp8 ratio reached equimolarity. No RNA binding was observed in the absence of nsp8, but upon nsp(7+8) complex formation the amount of bound dsRNA rapidly increased (Fig. 3D). Indicative of successful complex formation, we also observed a shift in the molecular weight of the major RNP complex formed (Fig. 3D). Western blot analysis confirmed that both nsp7 and nsp8 were present at this position in the gel (not shown), but due to the generally unpredictable migration behaviour of proteins and RNPs in native PAGE, it was not possible to assess whether this band indeed corresponded to the



**Figure 3: SARS-CoV nsp7 stimulates nsp8-dependent RNA binding.** (A) Five prime  $^{32}\text{P}$ -labelled dsRNA was incubated with increasing concentrations (0–5  $\mu\text{M}$ ) of wild-type nsp8, (B) nsp8 K58A, (C) or nsp8 D52A. Clearly, mutation of K58 to alanine significantly reduced the binding affinity of nsp8, whereas mutation of D52 to alanine did not. We also noted that the change in charge due to the mutation (up to eightfold in the octamer) resulted in an upward shift of the dominant RNP band, relative to the dominant RNP in panel 3A (labelled with black 1). (D) Five prime  $^{32}\text{P}$ -labelled duplex RNA was incubated with a fixed concentration of nsp7 (5  $\mu\text{M}$ ) and increasing concentrations of wild-type nsp8 (0–5  $\mu\text{M}$ ). Note the migration shift of the dominant ribonucleotide-protein (RNP) complex in the presence of nsp7 (compare RNPs labelled with black 1 and grey 2). (E) Addition of an equimolar amount of nsp7 to the nsp8 mutants D52A and K58A stimulated binding of dsRNA. For reference the 2:1 and 1:1 ratios of wild-type nsp8 and nsp7 are shown in the left panel. Asterisks indicate nonspecific bands. (F) RNA-binding curves for nsp8 in the absence (grey triangles) or presence of a fixed (5  $\mu\text{M}$ ) concentration of nsp7 (black circles). Lines represent fits to the Hill equation, while error bars represent standard deviations ( $n = 3$ ).

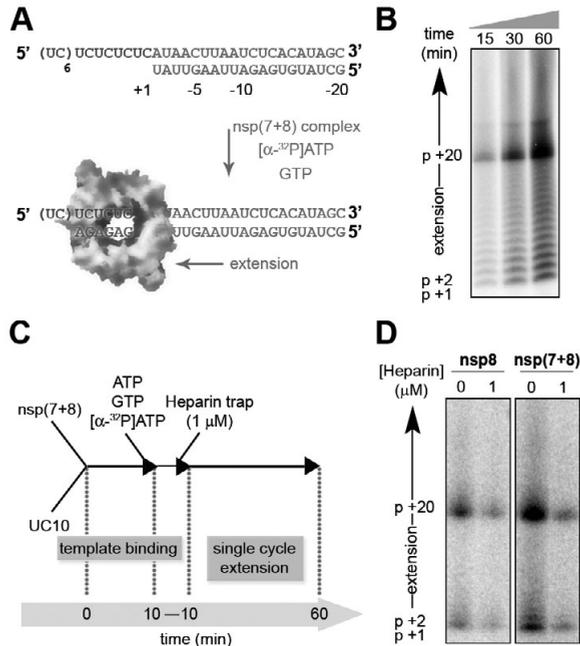
nsp(7+8) hexadecamer. The  $K_d$  of the nsp(7+8) complex was estimated at  $\sim 1.2 \mu\text{M}$ , about 3-fold lower than that of nsp8 alone (Fig. 3F).

When we next added an equimolar amount of nsp7 to the nsp8 RNA-binding mutant K58A, we observed a minor increase in the binding affinity for RNA (compare Fig. 3B with Fig. 3E). Mutant D52A, on the other hand, behaved similar to the wild-type protein (Fig. 3E). Together, these results complement the observation that various positively charged nsp7 residues line the inside of the nsp8-scaffolded RNA binding channel [153], and they provide the first direct evidence for a functional role of nsp7 in the SARS-CoV nsp(7+8) structure.

### The nsp(7+8) complex has primer extension activity

Given nsp(7+8)'s ability to bind dsRNA, we wondered whether this protein complex would also be catalytically active on this type of template and able to incorporate nucleoside monophosphates (NMPs) into partially double-stranded RNA molecules, *i.e.*, primed templates. We therefore examined the ability of nsp8 to extend a 20-nt primer that was pre-annealed to a heteromeric template with relatively low secondary structure, to rule out potential adverse effects of hairpins (Fig. 4A). Interestingly and in contrast to previous observations [155], the nsp(7+8) complex readily extended the primer up to template length, resulting in the formation of a 40-base pair RNA duplex (Fig. 4B).

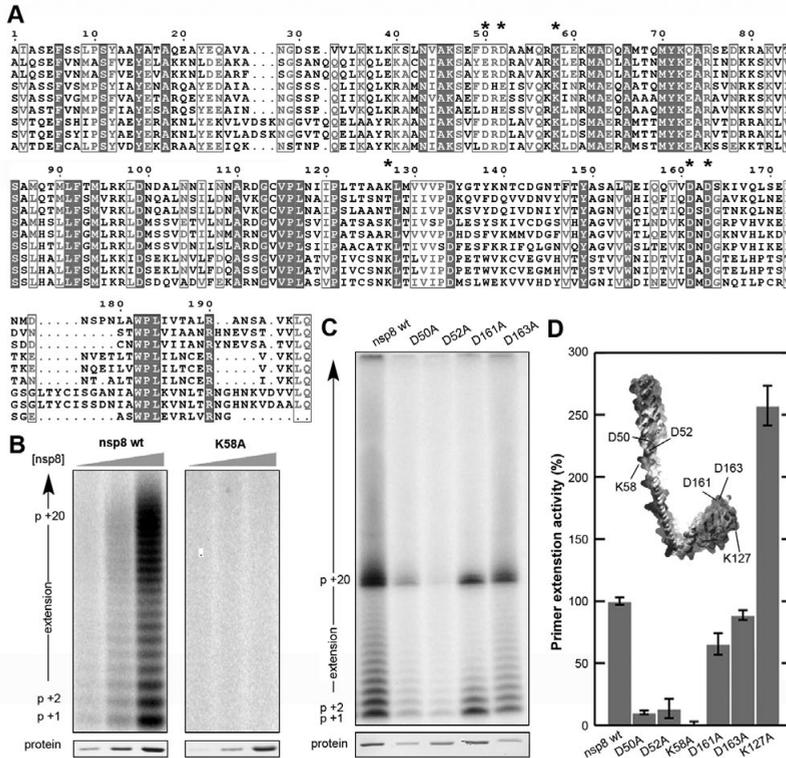
The negatively charged and helical polymer heparin is able to occupy the binding sites of RNA and DNA polymerases, and can thus directly compete with RNA and DNA templates. To verify that the full-length and longer RNA products were derived from single nsp(7+8) complexes bound to the template (*i.e.*, from a processive activity), and not from multiple binding and extension events (*i.e.*, a distributive activity), we performed the primer extension reaction in the presence of heparin to trap any unbound nsp(7+8). We first tested the concentration required to saturate all nsp(7+8) complexes in the reaction by titrating 0-100  $\mu\text{M}$  into the reaction (Fig. S2A) and observed that the incorporation levels were stable above 1  $\mu\text{M}$  (Fig. S2B), suggesting that these reactions represent single initiation-extension events. We next assessed whether the activity of nsp8 or nsp(7+8) was distributive or processive by quantifying the incorporated signal in full-length or longer products in the presence of 1  $\mu\text{M}$  heparin (Fig. 4C). As shown in Fig. 4D,  $66 \pm 4\%$  (mean  $\pm$  standard deviation) of the nsp8 products were full length compared to  $61 \pm 2\%$  of the nsp(7+8) products, suggesting that both enzymes complexes are mostly processive and that nsp7 does not confer additional processivity to nsp8. Interestingly, both nsp8 and nsp(7+8) are able to extend the RNA primers beyond template length in the presence of heparin (Fig. 4D and Fig. S2B), suggesting that these extensions result from terminal transferase activity and not from template switching, as was previously observed for poliovirus 3D<sup>pol</sup> [284].



**Figure 4: The nsp(7+8) complex has primer extension activity.** (A) Schematic presentation of the nsp8 primer extension assay, in which [α-<sup>32</sup>P]AMP and GMP are incorporated into a primed RNA template. (B) Incorporation of [α-<sup>32</sup>P]AMP by the nsp(7+8) complex. Samples were taken at the indicated time points and resolved on a 20% PAGE/7M urea gel. (C) Schematic presentation of the single-cycle reaction. Template and nsp(7+8) complex were pre-incubated for 10 min before nucleotides were added. The mixture was then rapidly split into equal aliquots that were immediately mixed with heparin to trap unbound or released enzyme. (D) Samples were taken after 60 min and resolved on 20% PAGE/7M Urea.

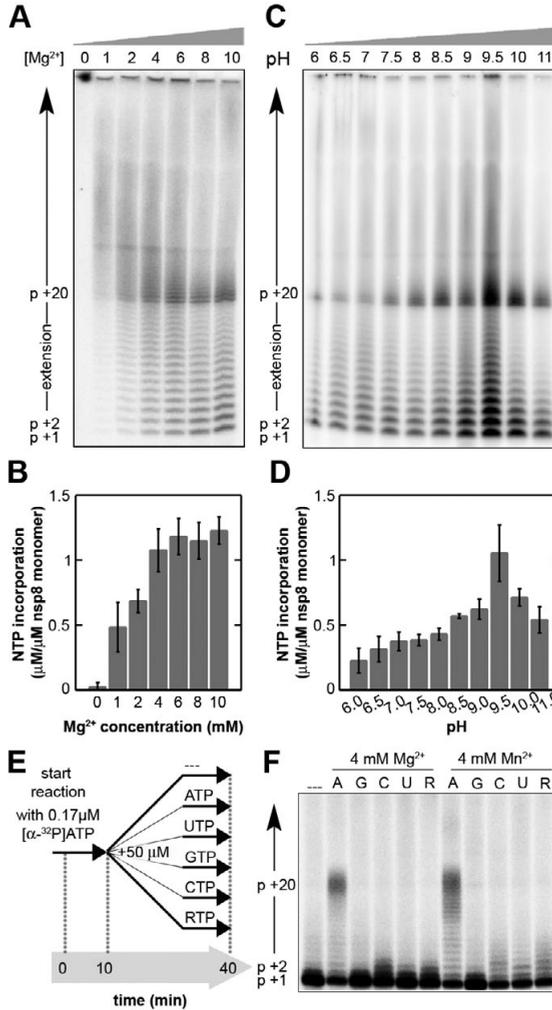
### The nsp(7+8) complex requires a D/ExD/E motif for catalysis

Intrigued by the primer extension activity of the SARS-CoV nsp(7+8) complex described above, we next designed a set of mutations to verify that the activity indeed was nsp(7+8) derived and to identify the most critical residues for activity in the complex. We first tested RNA-binding mutant K58A (Fig. 2) at varying concentrations and observed a ~95% loss of nucleotide incorporation activity compared to the wild-type protein (Fig. 5). Other likely candidates for a direct role in RdRp catalysis generally are Mg<sup>2+</sup>-coordinating aspartate residues and lysine or histidine residues that can function as general acid [279]. In canonical RNA polymerases, the aspartates commonly reside in motifs A and C [84,279], while in DNA-dependent RNA primases they are usually found in a central D/ExD/E motif [328]. Given the absence of classical RdRp A and C motifs in the nsp8 sequence [155], we screened an alignment of CoV nsp8 sequences for conserved D/ExD/E motifs. Interestingly, we found such a motif in both the N-terminal and the C-terminal domain (Fig. 5A). Subsequent alanine substitution of the N-terminal D/



**Figure 5: Mutagenesis of SARS-CoV nsp8.** (A) Alignment of nsp8 sequences from representative alpha-, beta- and gammacoronaviruses. Fully conserved residues are shaded dark grey, while partially conserved residues are boxed. The residues targeted by mutagenesis are indicated with asterisks. Please see Material and Methods for the Genbank accession numbers associated with the presented sequences. Sequences are presented in the order SARS-CoV, MHV, HCoV OC43, HCoV NL63, HCoV 229E, Bat-CoV HKU8, IBV, Turkey CoV and BW-CoV. (B) To verify that the observed extension activity was nsp8-dependent, we tested the incorporation of AMP into the primed  $U_{20}$  template by 1, 5 or 10  $\mu\text{M}$  of wild-type nsp8 or template-binding mutant K58A. Mutation of K58 resulted in a  $\sim 95\%$  reduction of AMP incorporation. (C) To assess the importance of the two D/ExD/E motifs in nsp8, we engineered alanine substitution mutants of these residues and tested their primer extension activity on the primed  $U_{10}$  template (see Fig. 4). Reactions were stopped after 60 min and compared to the activity of the wild-type nsp(7+8) complex on a 20% PAGE/7M urea gel. The bottom panel shows the nsp8 protein concentration present in each of the reactions. (D) Quantification of the primer extension activities on the  $U_{10}$  template of the D/ExD/E alanine substitution mutants and control substitutions K58A and K127A. Values are normalised to the protein concentration. Error bars represent standard deviations ( $n=3$ ).

ExD/E motif, composed of D50 and D52 in SARS-CoV, greatly affected primer extension activity on the  $U_{10}$  template as shown in Fig. 5C. Mutation of the downstream domain (residues D161 and D163 in SARS-CoV), however, had a much smaller effect on polymerase activity, suggesting that this C-terminal D/ExD/E motif is not critical for catalysis. Controls included mutant K58A and a mutant carrying a lysine-to-alanine substitution



**Figure 6: Influence of divalent ions and pH on nsp(7+8) activity.** (A) To test the influence of magnesium ions on nsp(7+8) activity, we performed nsp(7+8) primer extension reactions at 0-10 mM Mg<sup>2+</sup>. (B) Quantification of the results presented in Fig. 6A, presented as the amount of NTP incorporated per μM nsp8 monomer. Error bars represent standard deviations (n=3). (C) The influence of the pH on nsp(7+8) activity was tested for a pH range of 6-11. A clear optimum was observed around 9.5. (D) Quantification of the results in Fig. 6C, presented as the amount of NTP incorporated per μM nsp8 monomer. Error bars indicate standard deviations (n=3). (E) Schematic presentation of the pulse-chase experiment that was used to test the nsp(7+8) nucleotide incorporation specificity on a primed poly(U) template (see Table 1). The reactions were initiated with a limiting concentration of [α-<sup>32</sup>P]ATP to allow the formation of a stable polymerase-template complex. Unlabelled nucleotides were used at a final concentration of 50 μM. (F) SARS-CoV nsp(7+8) allowed only limited transversional and transitional mutations. Use of manganese ions as cofactor for polymerase activity resulted in a minor, though noticeable loss of fidelity. Lane 1 represents the input signal to which no unlabelled nucleotides were added. Nucleoside triphosphates are abbreviated to single letters (i.e., A for ATP, G for GTP, U for UTP, C for CTP, and R for RTP).

of the non-conserved residue 127. In line with the observation of the U<sub>20</sub> template and its conservation in CoVs, the loss of a lysine at position 58 resulted in a near complete loss of RdRp activity, whereas mutation of K127 positively influenced RNA synthesis (Fig. 5).

### **Influence of divalent ions and protons on nsp(7+8) activity**

As outlined above, magnesium ions are well-known cofactors of nucleic acid polymerases and assist in the coordination and activation of incoming nucleoside triphosphates. Also the activity of SARS-CoV nsp(7+8) was found to be positively correlated with the Mg<sup>2+</sup> concentration, albeit with a broad optimum running from 4-10 mM (Fig. 6A). At this optimum, nsp(7+8) incorporates ~1 μM NMP into the primed template per μM of monomeric nsp7 and nsp8 present in the reaction.

Similar to the presence of divalent cations, the pH greatly affects the activity of RdRps and has been shown to play a role in both catalysis and fidelity [242,279]. To investigate the influence of the pH on nsp(7+8), we tested the activity of the complex in a pH range of 6-11. As shown in Fig. 6C, we observed a sharp optimum at pH 9.5, which is considerably higher than the optimum that was previously observed for the SARS-CoV nsp12-RdRp and the His-nsp8 homomer (pH optimum 7.5 and 8.0, respectively) [154,155].

Interestingly, the primer extension activity of nsp(7+8) did not require manganese ions as was previously reported for the His-nsp8 homomer [155]. In fact, similar to the SARS-CoV nsp12-RdRp [154], the addition of Mn<sup>2+</sup> was found to reduce the fidelity of nsp(7+8) and induce both transversional and transitional misincorporations in a pulse-chase experiment (Fig. 6E and 6F). Interestingly, the assay also revealed a discrimination against the widely used ATP and GTP analogue ribavirin triphosphate (RTP) [56,94]. Whether this may offer an explanation for SARS-CoV's relative resistance to this antiviral drug [329,330] remains an open question for future research.

### **N-terminal extensions other than nsp7 frustrate the primer extension activity of nsp8**

The primer-extension and terminal transferase activity documented in Fig. 4 for the complex containing the untagged nsp8 was not observed by Imbert *et al.* when they first purified and analysed His-nsp8 [155]. To investigate whether this difference could be attributed to complex formation with nsp7 or the removal of the affinity tag, we performed the primer extension assay with three different recombinant nsp8 versions of which the gel filtration analysis is documented in Fig. 2. Interestingly, for all three variants primer-extension activity was observed (Fig. 7A), but the activity was most pronounced for nsp8-His and the untagged nsp8 (Fig. 7A). To estimate the effect of nsp7 on the nsp8-driven primer extension activity, we performed a direct comparison of the two enzyme complexes and found that the activity of nsp8 alone was >2 fold

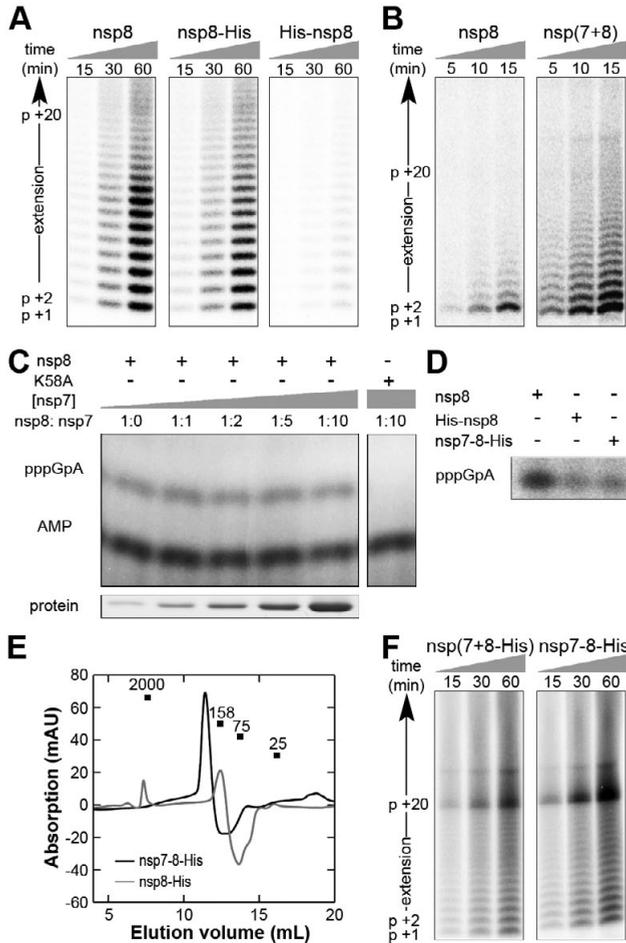
lower than when nsp7 and nsp8 were present at equal molarity in the reaction (Fig. 4D and Fig. 7B). A similar comparison was performed for the *de novo* activity of nsp8, using the assay published by Imbert *et al.* [155] and taking the first dinucleotide (pppGpA) product as readout. Interestingly, both nsp8 and nsp(7+8) synthesised equal amounts of the pppGpA dinucleotide (Fig. 7C), suggesting that the effect of nsp8 is limited to the primer-extension activity of nsp8. In addition, we observed that the *de novo* initiation activity of nsp8 was ~2 fold higher than that of His-nsp8 (Fig. 7D).

Our comparative study revealed that the N-terminal His<sub>6</sub>-tag of His-nsp8 greatly influences the primer-extension activity of nsp8 (Fig. 7A), its multimerisation profile and its association with nsp7 (Fig. 2). To test if this inhibitory effect was His<sub>6</sub>-tag specific, we assessed the activity of a ub-nsp8-His fusion protein. At the same time, control reactions were performed in which we i) followed the activity of this protein as it was being processed by a recombinant form of the ubiquitin-cleaving nsp2 protease of equine arteritis virus [88] or ii) monitored the activity of nsp8-His. As shown in Fig. S3, the presence of the ub-tag decreased nsp8 activity to a level that was comparable to that of N-terminally His<sub>6</sub>-tagged nsp8. Upon cleavage by EAV nsp2, however, a partial recovery of the primer extension activity was observed (Fig. S3). Unfortunately, we were not able to perform the same experiment with purified ub-nsp8, since our recombinant nsp5 removed the N-terminal ub-tag with similar efficiency as the C-terminal His<sub>6</sub>-tag (Fig. S4).

Extrapolating to the situation in the viral pp1a and pp1ab precursor polyproteins, in which the nsp8 N-terminus is initially fused to nsp7 (Fig. 1A), our observations suggested that nsp8 may thus be inactive in the polyprotein context. This would constitute a form of regulation of viral enzyme activity that is not without precedent, since also the poliovirus 3Dpol is inactive as long as it is fused to the 3C protease in the 3CD precursor [331]. To verify this hypothesis, we expressed nsp7-8-His and tested this protein for RdRp activity. Interestingly, this fusion protein, a potential intermediate of CoV replicase polyprotein processing and a multimer in solution (Fig. 7E), showed primer extension activities that were comparable to or higher than the activity of nsp(7+8-His) (Fig. 7F). The *de novo* initiation activity of nsp7-8-His was, however, ~2 fold lower than the activity of nsp8 and nsp(7+8) (Fig. 7D). In conclusion, this result clearly underlines that the two N-terminal fusion partners other than nsp7 are specifically detrimental to SARS-CoV nsp8 primer-dependent RdRp activity *in vitro*. It also demonstrates that nsp8 alone may be sufficient to act as a primase.

## Discussion

The complex replication and transcription process that coronaviruses initiate upon infection involves up to 16 viral nsps and at least one host factor [51,79,290]. Both individually as well as in complex with each other, these subunits engage in numerous protein-protein interactions [85,289] and embody various enzymatic activities, includ-



**Figure 7: Influence of His<sub>6</sub>-tags and nsp7 on the RdRp activity of SARS-CoV nsp8.** (A) The UC<sub>10</sub> template (see Fig. 4) was incubated with 1 μM of wild-type nsp8 or either of three nsp8 variants to investigate the influence of the His<sub>6</sub>-tag. Samples were taken at the indicated time points and checked for [<sup>α-32</sup>P]AMP incorporation by 20% PAGE/7M urea analysis. (B) Side-by-side comparison of the primer-extension activities of nsp8 and nsp(7+8). Shorter incubations are shown to better demonstrate the difference in activity. (C) *De novo* activity of nsp8 and nsp(7+8) on template AFMB131 (see Table 1), using the synthesis of the first dinucleotide pppGpA, as previously described by Imbertet *al.* [155], as readout. Nsp8 template binding mutant K58A was used as negative control. The AMP contaminant present in the used [<sup>α-32</sup>P]ATP label is marked as loading control and size reference. (D) Side-by-side comparison of the *de novo* initiation activities of nps8, His-nsp8 and nsp7-8-His. (E) Elution profile of the nsp7-8-His fusion protein relative to nsp8-His. (F) Primer-extension activities of putative cleavage intermediate nsp7-8 on the U<sub>20</sub> template (see Fig. 4 and Table 1).

ing proteolytic [44,332], ATPase [108], and 5' cap modifying reactions [333]. Remarkably though, the mechanism and enzymes required to catalyse RNA synthesis in the CoV RTC

remain very poorly understood. Moreover, uniquely among RNA viruses which generally employ a single RNA polymerase to drive their RNA synthesis [235,264], the polymerase activity assays and nsp8 mutagenesis documented in this and other studies suggest that, in addition to the presumed nsp12 “main RdRp”, other polymerase activities could play a critical role in the synthesis of SARS-CoV RNAs [154,155,285,334].

Following up on the description of an nsp8- and nsp7-containing hexadecameric ring structure [153] and the nsp8-associated polymerase activity [155], we here demonstrate that the nsp(7+8) hexadecamer is the most probable conformation of the second SARS-CoV polymerase, given the near-complete association of nsp7 and nsp8 when mixed 1:1 in solution (Fig. 2F). Significant for our understanding of CoV RNA synthesis, we find that this complex is capable of binding dsRNA molecules and extending partially double-stranded RNA templates. This activity is therefore essentially comparable to the activity reported for the nsp12-RdRp [154].

A direct comparison with the nsp12 activity is difficult, however. In the course of a one-hour reaction, 0.1  $\mu\text{M}$  monomeric nsp12-RdRp incorporates  $\sim 2 \mu\text{M}$  NMP into a primed  $(\text{CU})_{10}$  template [154]. The nsp(7+8) complex, at a 1  $\mu\text{M}$  concentration of nsp7 and nsp8 monomers, incorporates  $\sim 1 \mu\text{M}$  NMP. Per monomer, the activity difference is therefore 20-fold, but if we assume that most nsp7 and nsp8 monomers assemble into hexadecamers and that each hexadecamer contributes only one functional active site per incorporation event, the difference would be much smaller and only  $\sim 2.5$  fold. Presently, however, we do not yet have an estimate for the efficiency and stability of the nsp(7+8) complex, nor do we know the number of active sites in the complex that determine its overall activity.

Mutagenesis of nsp8 was performed to identify residues that may contribute to the catalytic centre of the nsp(7+8) polymerase, while differently tagged nsp8 recombinant proteins were constructed to explain some striking differences with previous observations. These efforts resulted in two intriguing observations. First, mutation of the conserved N-terminal D/ExD/E motif, comprising D50 and D52 in SARS-CoV, abolished RdRp activity, whereas mutation of the C-terminal motif, including SARS-CoV residues D161 and D163, did not affect polymerase activity (Fig. 5). Given the general importance of acidic residues for metal-ion coordination in polymerase active sites [89,279,281,328] and the D/ExD/E consensus sequence in coronaviruses at positions 50-52, we now postulate that these residues are part of the  $\text{Mg}^{2+}$ -binding active site in spite of the more conserved nature of D161 and D163 (Fig. 5), and their position in the nsp(7+8) structure (see below for further discussion).

Second, the presence of N-terminal extensions other than nsp7, such as ubiquitin and  $\text{His}_6$ , severely affected the primer extension activity of nsp8 (Fig. 7), potentially by changing its oligomeric state (Fig. 2). However, the relatively strong activity of nsp7-8 (Fig. 7), a potential naturally occurring replicase processing intermediate, implies that

nsp8's activity is unlikely to be directly controlled by an N-terminal cleavage event, as was observed for, *e.g.*, the poliovirus polymerase [331]. In addition, these observations suggest that a more diverse array of nsp8-containing RdRps may be involved in CoV replication and transcription.

Interestingly, our study and the published nsp(7+8) structural data [153] deviate at four main points. First, we observe that in the published nsp(7+8) crystal structure four of the eight N-terminal D/ExD/E motifs in the complex reside at the border of partially unresolved N-terminal nsp8 domains, where the coordinates of up to 49 nsp8 residues and 5 exogenous amino acids derived from the removed GST fusion partner were not determined. In light of our own finding that unnatural N-terminal extensions severely impair nsp8's RdRp activity (Fig. 5), this suggests that the published crystal structure may not represent an active conformation of the nsp(7+8) polymerase. Second, we observe that residues D50 and D52, which are both crucial for nsp(7+8) activity, are residing in an  $\alpha$ -helix in the nsp(7+8) structure (Fig. S5), whereas in canonical primases and polymerases, the catalytic centre is preferentially located on  $\beta$ -strands or turns [155,328]. Third, we note that  $Mg^{2+}$  was lacking from the published nsp(7+8) crystal structure [153], even though it is required for nsp(7+8) activity. Fourth and last, we observe that a 1:1 ratio of nsp7:nsp8 is sufficient to capture all nsp8 in a higher molecular weight complex (Fig. 2F) whereas previously a 2:1 ratio was required [153], potentially due to the additional N-terminal residues that altered the dynamics of complex formation.

The (functional) implications of these observations are not clear at present, but additional structural studies will likely be required to address these issues in detail, and gain insights that may aid in explaining the *in vitro* results presented here. Likely, such experiments will also offer further information regarding the residues that are involved in nucleotide positioning,  $Mg^{2+}$  coordination and RdRp chemistry.

In summary, our results provide important novel insights into the functionality of the SARS-CoV hexadecameric nsp(7+8) complex and demonstrate its activity as an RNA polymerase. In addition, our experiments and controls revealed and address a number of disparities between previous claims and hypotheses [155], and our own observations. The "primase hypothesis" previously formulated by Imbert and co-workers [155] remains an intriguing model to explain the initiation of SARS-CoV RNA synthesis and is a topic that will be addressed in detail elsewhere. Nevertheless, based on the primer extension activity of nps7+8 on non-structured RNA templates, we can no longer exclude the possibility that nsp(7+8) may synthesise substantially longer products than mere oligonucleotide primers *in vivo*, possibly stimulated by the presence of additional viral protein factors that could, *e.g.*, provide RNA-unwinding activity. Consequently, it is now a distinct possibility that CoV RNA synthesis involves structurally different and functionally separable RNA synthesising complexes (*e.g.*, containing nsp12 or nsp(7+8)), each possessing their own dedicated RdRp characteristics and function in viral plus or

minus strand RNA synthesis. It will therefore be crucial to study whether these different polymerase activities are part of the same enzyme complex and, if so, whether they can influence each other's activity or are subject to additional control mechanisms.

## Material and methods

### Cloning, mutagenesis and expression

For SARS-CoV nsp7-nsp8 expression, the sequence encoding amino acids 3837-4117 of the SARS-CoV replicase pp1a was amplified by reverse transcription-polymerase chain reaction (RT-PCR) from the genome of SARS-CoV isolate Frankfurt-1 (Genbank accession number AY291315). The primers used were SAV704 and SAV429 (Table S1). For nsp8 expression, the sequence encoding pp1a residues 3920 to 4117 was amplified by RT-PCR using SAV428 and SAV429 as primers (Table S1). Both PCR products were digested with *Sac*II and *Bam*HI, and ligated into expression vector pASK3-Ub-CHis<sub>6</sub> [154]. This vector was originally derived from the pET26-Ub-CHis<sub>6</sub> vector [277], but drives expression of N-terminally ubiquitin-tagged and C-terminally His<sub>6</sub>-tagged fusion proteins via a tetracyclin-inducible promoter, to rule out the potential T7 polymerase contaminations that are known to cause false positive results when using T7 promoter-driven systems for recombinant RdRp expression. All described nsp8 mutants were engineered via site-directed mutagenesis according to the QuikChange protocol (Stratagene) using the primers listed in Table S2.

For nsp7-8 or nsp8 expression, *Escherichia coli* C2523 cells (New England Biolabs) were transformed with the plasmids pASK3-Ub-nsp7-8-CHis<sub>6</sub> or pASK3-Ub-nsp8-CHis<sub>6</sub> together with the Ubp1 protease expression plasmid pCG1 [277]. Routinely, 50 ml of Luria Broth, containing ampicillin (50 µg/ml) and chloramphenicol (34 µg/ml), was inoculated 1:1000 with o/n precultures, and cells were grown to OD<sub>600</sub> >0.8 at 37°C. Subsequently, the cells were slowly cooled to 20°C, followed by induction with anhydrotetracycline (Fluka) at a final concentration of 200 ng/ml for 16 h. Expression at 20°C was, however, only crucial for the preparation of certain nsp8 mutants and similar yields of active wild-type protein could be obtained by expression at 37°C for 3-4 h. Cells were harvested by centrifugation and stored at 20°C until protein purification was started.

The expression of SARS-CoV nsp7 with a C-terminal His<sub>6</sub>-tag (nsp7-His) was achieved from plasmid pDEST14-nsp7-His<sub>6</sub> according to the protocol previously described for EAV nsp9 [325]. SARS-CoV nsp5-His<sub>6</sub> (nsp5-His) was expressed as a self-cleaving maltose binding protein (MBP)-fusion protein and was purified via its C-terminal His<sub>6</sub>-tag [326]. The pASK3-His-nsp8 plasmid for expression of the N-terminally His<sub>6</sub>-tagged nsp8 was kindly provided by Dr. Imbert and Dr. Canard (University of Marseille, France).

### Purification of SARS-CoV nsp8, nsp7-8 and nsp7

Bacterial pellets were thawed on ice, resuspended in buffer A (20 mM Hepes pH 7.4, 10 mM imidazole, 0.05% Tween-20, 5 mM  $\beta$ -mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche)) containing 500 mM NaCl, and lysed by sonication. The supernatant was cleared by ultracentrifugation at 20,000 *g* for 30 min and subsequently incubated with Talon beads (Clontech) for 2 h at 4°C. The beads were washed four times 15 min with 20 volumes of binding buffer. Ultimately, the C-terminally His<sub>6</sub>-tagged proteins were eluted with 150 mM imidazole in buffer A containing 150 mM NaCl, or cleaved off of the column during a 3-h digestion with SARS-CoV nsp5 in the presence of 4 mM MgCl<sub>2</sub>.

The eluates were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and typically found to be >90% pure. Elution fractions containing nsp8-, nsp7-8, or nsp7 were subsequently pooled, dialysed, stored and analysed as described previously for SARS-CoV nsp12 [154].

### Chemical cross-linking

To study SARS-CoV nsp(7+8) complex formation, different nsp8:nsp7 ratios were mixed in binding buffer (20 mM Hepes pH 7.5, 50 mM NaCl, 5% glycerol, 0.1% Triton X-100, and 1 mM DTT) to give a final reaction volume of 10  $\mu$ l. The proteins were pre-incubated for 10 min at 20°C, after which cross-linking was initiated through the addition of 0.5  $\mu$ l of a freshly prepared 2.5% glutaraldehyde solution. The reactions were incubated for a further 5 min at 30°C and then terminated with 1  $\mu$ l 1 M Tris pH 8.0. Analysis of complex formation was performed on SDS-PAGE gels, which were stained with Coomassie G-250 dye.

### Template binding assays

A dilution series of 0-5  $\mu$ M SARS-CoV nsp8 in storage buffer was incubated for 10 min at 20°C with 0.2 nM of <sup>32</sup>P-labelled duplex RNA. Subsequently, samples were directly loaded onto 8% polyacrylamide gels containing 5% glycerol and 0.5x TGE (25 mM Tris, 190 mM glycine and 10 mM EDTA) buffer and run at 150 V for 1 h at 4°C. Gels were dried on Whatman filter paper and bands were quantified by phosphorimaging using a Typhoon variable mode scanner (GE Healthcare) and ImageQuant TL 7.0 software (GE Healthcare) as described elsewhere [154]. Using the Matlab 2009a Curve Fitting Toolbox, the percentage of bound RNA was fit to the Hill equation, which is defined as:

$$RNA_{bound} = b \times [nsp8]^n / (K_d^n + [nsp8]^n).$$

Here *b* is the upper binding limit, [*nsp8*] the nsp8 concentration, *n* the Hill coefficient and *K<sub>d</sub>* the dissociation constant.

### Polymerase activity assays

The oligoribonucleotide substrates used for polymerase assays are listed in Table 1 and were prepared as described previously [154]. Primer-extension assays for nsp8, the nsp7-8 polyprotein, and the nsp(7+8) complex were essentially performed as described previously for SARS-CoV nsp12 [154,325]. In each primer-extension reaction, typically 1  $\mu\text{M}$  wild-type or mutant nsp8 was incubated with 4 mM  $\text{MgCl}_2$ , 50  $\mu\text{M}$  GTP, 50  $\mu\text{M}$  ATP, 0.17  $\mu\text{M}$  [ $\alpha$ - $^{32}\text{P}$ ]ATP, 1mM DTT, 0.1% Triton X-100, 10 mM KCl and 20 mM Tris (pH 9.5). At most 10 mM NaCl and 5% glycerol were introduced with the nsp8 storage buffer. Gels were run and analysed as described previously [154]. To convert the phosphorimager signal into the amount of [ $\alpha$ - $^{32}\text{P}$ ]AMP incorporated, a  $10^{-2}$  to  $10^{-5}$  dilution series of the [ $\alpha$ - $^{32}\text{P}$ ]ATP stock was spotted in triplicate on Whatman filter paper and exposed alongside the PAGE gel. The amount of incorporated label was ultimately corrected for the concentration of competing, unlabelled nucleotides present in the reaction mixture.

*De novo* initiation assays were essentially performed as described by Imbert *et al.* [155], with small modifications for optimisation. Briefly, 1  $\mu\text{M}$  wild-type or mutant nsp8 was incubated with 4 mM  $\text{MgCl}_2$ , 1 mM  $\text{MnCl}_2$ , 1 mM GTP, 5  $\mu\text{M}$  ATP, 0.17  $\mu\text{M}$  [ $\alpha$ - $^{32}\text{P}$ ]ATP and 1  $\mu\text{M}$  of oligo AFMB131.

### Sequence alignment

Alignments of nsp8 sequences were made using Muscle [327]. Sequences used included the alphacoronaviruses human CoV 229E (NC\_002645), human CoV NL63 (NC\_005831), and bat CoV HKU8 (NC\_010438); the betacoronaviruses SARS-CoV Frankfurt-1 (AY291315), mouse hepatitis virus A59 (MHV, NC\_001849), and human CoV OC43 (NC\_005147); and the gammacoronaviruses beluga whale CoV SW1 (NC010646), turkey CoV (NC\_010800), and avian infectious bronchitis virus (IBV, AJ311317).

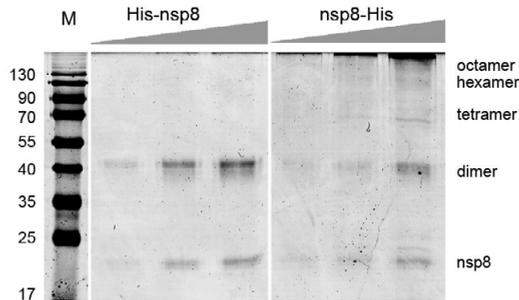
**Table 1:** Oligoribonucleotides used for activity assays

RNA oligo	Purpose	Sequence
SAV555	(UC) <sub>10</sub> template	5'-UUUUUUUUUUUUUUUUUUUUUAACUAAUCUCACAUAGC-3'
SAV556	(U) <sub>20</sub> template	5'-UCUCUCUCUCUCUCUCUCUCAUAACUAAUCUCACAUAGC-3'
SAV557	primer	5'-GCUAUGUGAGAUUAAGUUAU-3'
AFMB131	<i>de novo</i> assay template	5'-UAUAAUCCAAA-3'

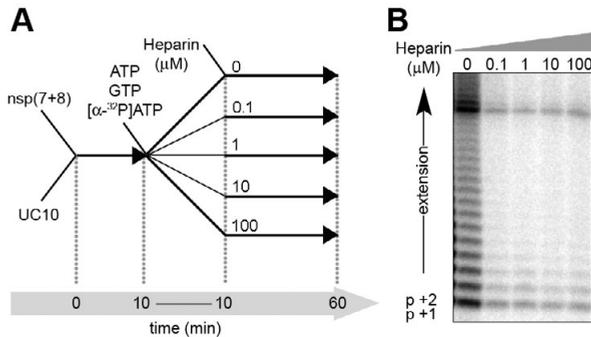
## Acknowledgements

The authors thank Dr. Danny Nedialkova, Lorenzo Subissi, Dr. Isabelle Imbert, Dr. Bruno Canard, and Dr. Alexander Gorbalenya for stimulating discussions; Linda Boomaars-van der Zanden and Dr. Clara Posthuma for assistance with nsp5 purification; Puck van Kasteren and Dr. Marjolein Kikkert for providing the EAV nsp2 protease; and Jos van Vugt for his initial work on nsp8 in our lab. This work was supported by the Netherlands Organization for Scientific Research (NWO) through Toptalent grant 021.001.037 and ECHO grant 700.55.002 from the Council for Chemical Sciences (NWO-CW).

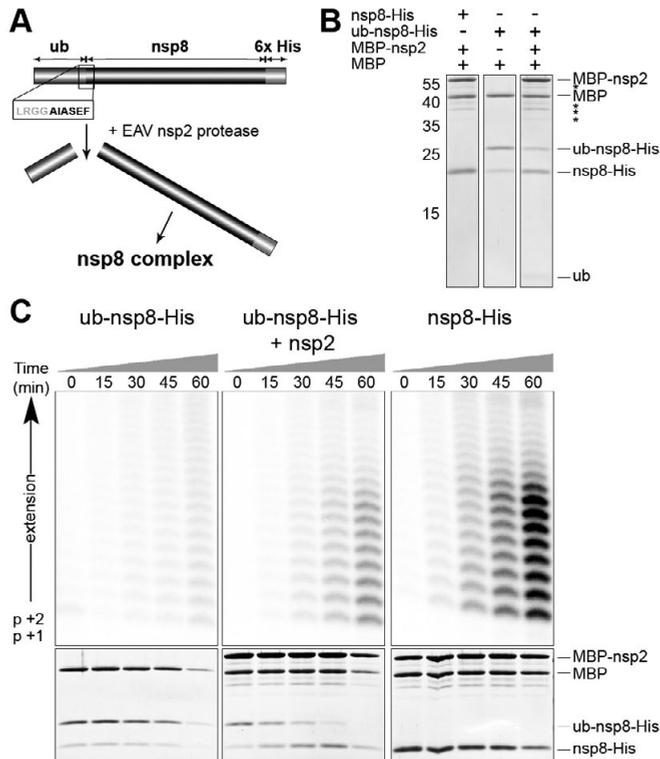
## CHAPTER 5 - SUPPLEMENTAL INFORMATION



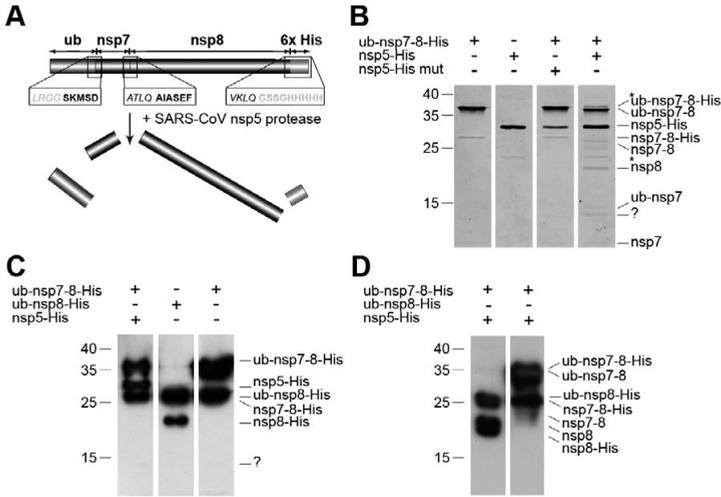
**Figure S1: Analysis of nsp8 multimerisation via chemical cross-linking.** Cross-linking analysis of N-terminally His<sub>6</sub>-tagged nsp8 and C-terminally His<sub>6</sub>-tagged nsp8. Increasing concentrations of nsp8 were incubated in HEPES buffer (pH 7.5) in the presence of 0.12% glutaraldehyde for 5 min. Subsequent SDS-PAGE analysis and staining with Coomassie G-250 dye shows that only the C-terminally tagged protein with native N-terminus forms higher order multimers, whereas the N-terminally tagged protein reveals solely mono- and dimers.



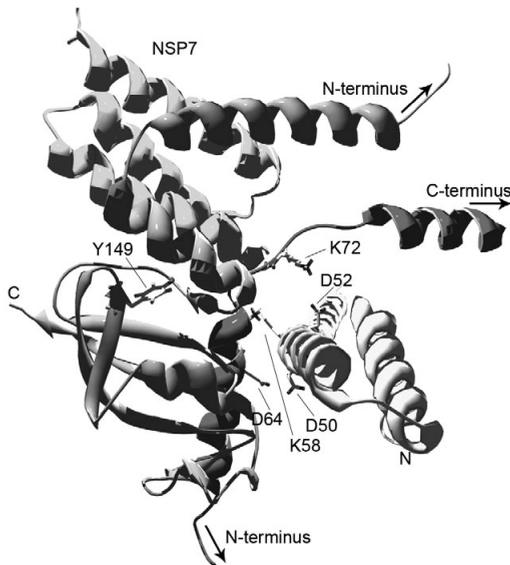
**Figure S2: Inhibition of nsp(7+8) activity with heparin.** (A) Schematic presentation of the single-cycle reaction. Template and nsp(7+8) complex were pre-incubated for 10 min before nucleotides were added. The mixture was then rapidly split into equal aliquots that were immediately mixed with different concentrations of heparin. (B) Samples were taken after 60 min and resolved on 20% PAGE/7M Urea.



**Figure S3: Activation of nsp8 RdRp activity by removal of an N-terminal fusion partner. (A)** To study whether the interference by non-nsp7 N-terminal extensions was reversible, we purified nsp8 with an N-terminal ubiquitin extension. Addition of either purified EAV nsp2 or Ubp1 would subsequently result in hydrolysis of the fusion protein C-terminal of the LRGG site. **(B)** Analysis of ub-nsp8-His cleavage products by SDS-PAGE and Coomassie G-250 staining. Asterisks indicate unspecific bands **(C)** Time-course of  $[\alpha\text{-}^{32}\text{P}]$ AMP incorporation of ub-nsp8-His, ub-nsp8-His in the presence of EAV nsp2, or nsp8-His that was *in vivo* cleaved by Ubp1. Lower panels demonstrate the stability or cleavage of ub-nsp7-8-his over time through SDS-PAGE analysis. In all reactions, MBP was added an independent loading control.



**Figure S4: Processing of ub-nsp7-8-His by the SARS-CoV nsp5 main protease.** (A) Due to the presence of the natural nsp7-8 cleavage sites in the ub-nsp7-nsp8-His construct, the polyprotein could be processed into mature subunits by the SARS-CoV nsp5 main proteinase. The question mark indicates an unidentified cleavage product. (B) SDS-PAGE analysis followed by Coomassie G-250 dye staining demonstrates that addition of nsp5 to purified ub-nsp7-nsp8-His results in cleavage at the sites indicated in Fig. 2C. Asterisk indicates non-specific band. The question mark indicates an unidentified cleavage product. (C) Western blot analysis of the protein samples used in Fig. 2D using an anti-His<sub>6</sub> monoclonal antibody and (D) an anti-nsp8 monoclonal antibody.



**Figure S5: Position of (putative) critical residues in the published nsp(7+8) structure.** Position of the critical residues, including D50, D52 in the nsp(7+8) hexadecamer. For reference, the conserved residues K58A, D62, K72A, and Y149, and the location of the termini are also indicated. The residues of the different termini are hidden to simplify the image. The nsp7 subunit that is in direct contact with the three nsp8 subunits is shown in the top left part of the structure.

**Table S1:** DNA primers that were used for cloning of nsp8 and nsp7-8

Primer	Sequence
SAV704	5'-GCGGGTACCCCGCGGTGGATCTAAAATGTCTGACGTAAAGTGCACA-3'
SAV429	5'-GCGCGATCGGGATCCCTGTAGTTTAAACAGCT-3'
SAV428	5'-GCGGGTACCCCGCGGTGGAGCTATTGCTTCAGAAT-3'

**Table S2: DNA primers** that were used for mutagenesis of nsp8

nsp8 mutation	PCR primers	Sequence
<b>D50A</b>	SAV574	5'-GCTAAATCTGAGTTTGCCCGTGATGCTGCCATG-3'
	SAV575	5'-CATGGCAGCATCACGGGCAAACCTCAGATTTAGC-3'
<b>D52A</b>	SAV590	5'-TCTGAGTTTGACCGTGCTGCTGCCATGCAACGC-3'
	SAV591	5'-GCGTTGCATGGCAGCAGCACGGTCAAACCTCAGT-3'
<b>K58A</b>	SAV402	5'-GCCATGCAACGCGCTTTGGAAAAGATGG-3'
	SAV403	5'-CCATCTTTTCAAAGCGCGTTGCATGGC-3'
<b>K127A</b>	SAV501	5'-GACTACAGCAGCCGCACTCATGGTTGTTG-3'
	SAV502	5'-CAACAACCATGAGTGCGGCTGCTGTATGC-3'
<b>D161A</b>	SAV503	5'-CCAGCAAGTTGTTGCTGCGGATAGCAAGA-3'
	SAV504	5'-TCTTGCTATCCGAGCAACAACCTGCTGG-3'
<b>D163A</b>	SAV505	5'-AGTTGTTGATGCGGCTAGCAAGATTGTTTC-3'
	SAV506	5'-GAACAATCTTGCTAGCCGCATCAACAACCT-3'



# Chapter 6

Quantitative guidelines  
for force calibration  
through spectral  
analysis of magnetic  
tweezers data

Aartjan J.W. te Velthuis<sup>1,2+</sup>, Jacob  
W.J. Kerssemakers<sup>1+</sup>, Jan Lipfert<sup>1</sup>,  
Nynke H. Dekker<sup>1</sup>

<sup>1</sup>Department of Bionanoscience,  
Kavli Institute of Nanoscience,  
Delft University of Technology,  
Lorentzweg 1, 2628 CJ Delft, The  
Netherlands

<sup>2</sup>Department of Medical  
Microbiology, Molecular Virology  
Laboratory, Leiden University  
Medical Center, Albinusdreef 2,  
2333 ZA Leiden, The Netherlands

+ These authors contributed  
equally to this work

**Published in:** Biophysical Journal,  
Volume 99, Pages 1292-1302,  
2010.

**Abstract**

Single-molecule techniques are powerful tools that can be used to study the kinetics and mechanics of a variety of enzymes and their complexes. Force spectroscopy, for example, can be used to control the force applied to a single molecule and thereby facilitate the investigation of real-time nucleic acid-protein interactions. In magnetic tweezers, which offer straightforward control and compatibility with fluorescence measurements or parallel tracking modes, force measurement typically relies on the analysis of positional fluctuations through video microscopy. Significant errors in force estimates, however, may arise from incorrect spectral analysis of the Brownian motion in the magnetic tweezers. Here we investigated physical and analytical optimisation procedures that can be used to improve the range over which forces can be reliably measured. To systematically probe the limitations of magnetic tweezers spectral analysis, we have developed a magnetic tweezers simulator, whose outcome was validated with experimental data. Using this simulator, we evaluate methods to correctly perform force experiments and provide guidelines for correct force calibration under configurations that can be encountered in typical magnetic tweezers experiments.

## Introduction

Recently, single-molecule techniques such as atomic force microscopy (AFM), tethered-particle microscopy (TPM), optical tweezers (OT), and magnetic tweezers (MT) have become more and more frequently applied to further our understanding of biological processes. For example, they have become invaluable tools for studies of nucleic acid replication and repair, and have provided information on details of enzyme kinetics such as translocation times, step sizes and the coupling in the mechano-chemical reaction cycle. In addition, they have revealed mechanical properties of both nucleic acid polymers, and provided insight into processes such as recombination [335,336,337,338,339].

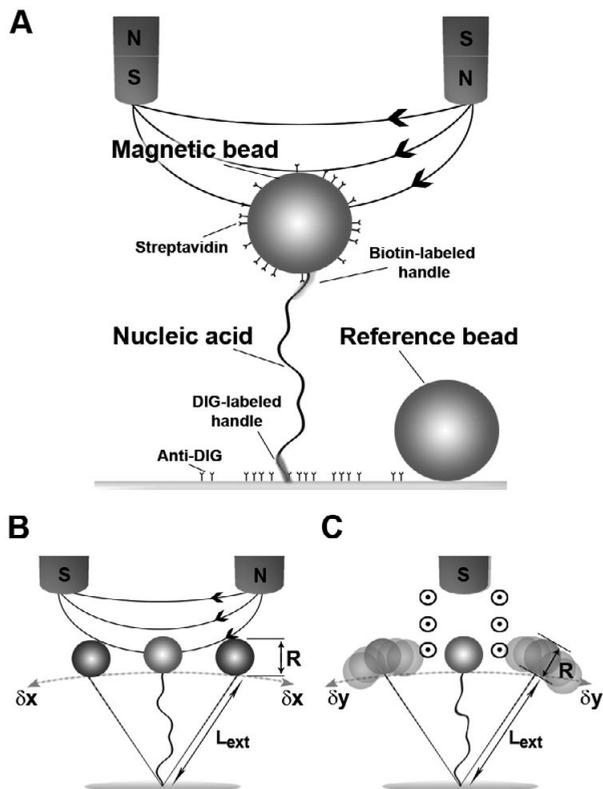
Compared to other single-molecule techniques that allow manipulation of externally applied forces, MT are particularly suited for experiments that cover a force range from femto- up to tens of piconewtons [340,341], need operation at a constant force, or require parallel measurements when enzyme processivity is limited [342]. Moreover, MT set-ups are relatively straightforward to construct, and consist primarily of an inverted microscope, a set of magnets, and a flow cell. In this flow cell, superparamagnetic beads are physically tethered to a surface via a nucleic acid or protein tether (Fig. 1A) [337,341,343,344], while the magnets positioned above the flow cell create an external, controllable magnetic field. The gradient of this field supplies a force in the Z-direction (Fig. 1), whereas rotation of the magnetic field applies torque to the tether and introduces supercoils [337,345,346]. The ability to apply both force and torque facilitates experiments aimed at studying the different structural forms of nucleic acids [337,347,348]. Additionally, this dual mode of manipulation has provided insight into the force- and torque-dependent properties of, *e.g.*, topoisomerases and chromatin [339,349,350,351,352].

In the MT, the motions of the bead take place round an equilibrium position due to Brownian motion, and are constrained by both the applied force and the flexibility of the nucleic acid tether (Fig. 1B) [337]. The applied force can be computed from the field configuration and the magnetisation of the beads according to:

$$\vec{F} = \frac{1}{2} \vec{\nabla} (\vec{m} \cdot \vec{B}) \quad [1]$$

where  $\vec{m}$  is the magnetisation of the bead in the external magnetic field  $\vec{B}$  [341]. However, practical use of this relationship may be limited due to complexity in the field configuration or variability in bead size and magnetisation density.

An alternate approach to force calibration is to employ measurements of the Brownian motion of the bead itself: the variance of the bead motion in *x* and *y* is inversely related to the applied force [337,341]. Depending on the type of instrumentation employed in the magnetic tweezers, however, this approach has limitations. In particular, the movements of the bead in *x* and *y* are typically recorded with a charge-coupled device (CCD)



**Figure 1. Motion of a tethered particle.** (A) In a magnetic tweezers set-up a nucleic acid polymer is tethered by digoxigenin (DIG) and biotin-labelled handles between an antibody-coated surface and a streptavidin-coated paramagnetic bead. A non-magnetic bead attached to the surface functions as reference bead during particle tracking in  $x$ ,  $y$  and  $z$  to subtract low-frequency mechanical drifts from the microscope. (B) An illustration of the motions of a tethered bead in  $x$  with the magnetic field lines (black lines) pointing predominantly in the  $x$ -direction. The magnetisation consequently forces the bead to maintain alignment with the field. Such an alignment with the magnetic field only takes place at large enough forces as discussed elsewhere [364]. (C) A schematic of the motions of the tethered bead in  $y$  with the field lines (black circles) oriented predominantly perpendicular to the  $y$ -direction. In this direction, the beads are able to freely rotate.

camera. However, finite camera acquisition frequencies may introduce artifacts in variance measurements due to camera blurring and aliasing, as has been noted elsewhere [353,354,355]. This may lead to systematic biases, particularly for short constructs such as nucleic acid hairpins [356] in which the tethered bead-nucleic acid system's response time (inversely proportional to the system's natural frequency) approaches or is shorter than the time the camera shutter is open. Corrections for these deleterious effects of camera blurring and aliasing can be implemented by examining the bead's fluctuations in the spectral domain [353].

In theory the approach to accurately measure the bead variance by correction for systematic acquisition biases can be expected to work successfully up until the point where the natural frequency of the system approaches the Nyquist frequency of the camera ( $f_{Nyq}$  equal to half the sampling frequency) [353,357]. However, the limitations of this approach have not been systematically explored for MT experiments, and it is presently not documented within which force regime this approach can be reliably applied. A tool to quantitatively explore the correction limits, chart the effects of MT setup parameters, and define where trade-offs appear between the magnitude of the calibration error and the height of the applied force is thus desirable. This would allow users to make the biases of the system explicit and then optimise conditions for a given application without the need to extensively test configurations. Particularly experiments in which the use of small beads is required to minimise low frequency noise or high forces need to be applied to unwind double-stranded regions in the template (*i.e.*, > 14 pN) would highly benefit from such carefully investigated guidelines [356,358,359].

Here, we have systematically evaluated the limitations of MT spectral analysis and the efficacy of camera correction. To do so, we have developed an MT simulator that simulates the Brownian dynamics [360] of a tethered bead molecule in three dimensions, generating  $x(t)$ ,  $y(t)$ , and  $z(t)$  traces as a function of user-defined parameters. An emulated camera then simulates the sampling effects that are commonly encountered in real-time measurements, such as blurring and aliasing. Analysis of the simulated camera output demonstrates that it is in quantitative agreement with experimental data taken for tethered constructs under known magnetic forces. When deconvolution of the modified spectrum is used to recover the correct signal at a defined pulling force, our analysis reveals that accurate measurements (< 10% error) can be conducted in the absence of camera aliasing and blurring correction until the natural frequency of the tethered bead system reaches ~50% of  $f_{Nyq}$ . This limit is raised to ~80% when spectral corrections are applied, although higher forces can be measured when the error constraint is relaxed. As the natural frequencies in MT are directly related to bead size, contour length and sample viscosity, higher forces can alternatively be measured by optimising these parameters for a given MT setup. By systematically varying these parameters, we explored the upper limits of force measurements with <10% error. Overall, our analysis gives insight into the systematic biases in MT experiments and provides general guidelines that can be used to correctly design MT experiments.

## Theoretical background

In an MT, a magnetic force ( $f_{mag}$ ) is applied on a nucleic acid or protein tether in the Z-direction. This force is effectively constant over the range of motion explored by the tether.  $f_{mag}$  is counteracted by the restoring force originating from the tether, which gives the total potential energy of the system:

$$E_p = E_{tether} + E_{magnet} = A(L_{ext}) - F_{mag} \cdot z \quad [2]$$

where  $A(L_{ext})$  is the energy stored in the nucleic acid as function of the extension of the tether  $L_{ext}$ . In equilibrium, all partial derivatives must be equal to zero and we can thus write  $(x, y, z) = (0, 0, L_{ext}) = \vec{r}_0$ . From this first order condition it follows that:

$$F_{mag} = \frac{\partial A}{\partial L_{ext}} \quad [3]$$

The tethered bead system is constantly forced out of this equilibrium by Brownian motion, which creates movements both in the direction of the field (which we take to be the  $x$ -direction) and perpendicular to the field (which we define as the  $y$ -direction). Note that in the  $x$ -direction, the orientation of the magnetic bead is constrained by its alignment with the magnetic field, thus preventing rotation about its axis (Fig 1B). In the  $y$ -direction, conversely, the orientation of the bead is unconstrained, and the bead will thus align itself with the nucleic acid tether, making  $L_{ext}$  appear longer by an amount equal to the radius of the bead (Fig. 1C). When we evaluate the second partial derivatives, the total potential energy around the equilibrium position is given to second order by [361]:

$$E_p(\vec{r}) \approx E_p(\vec{r}_0) + \frac{1}{2} k_x \delta x^2 + \frac{1}{2} k_y \delta y^2 + \frac{1}{2} k_z \delta z^2$$

$$E_p(\vec{r}) \approx E_p(\vec{r}_0) + \frac{1}{2} \left( \frac{F}{L_{ext}} \right) \delta x^2 + \frac{1}{2} \left( \frac{F}{L_{ext} + R} \right) \delta y^2 + \frac{1}{2} \left( \frac{\partial F}{\partial L_{ext}} \right) \delta z^2 \quad [4]$$

where  $R$  is the bead radius, and  $k_x, k_y, k_z$  are the trap stiffness in  $x, y$ , and  $z$ , respectively. In the  $z$ -direction, the motions of the bead are dependent on the local variation of the restoring force, which can be calculated from models of polymer elasticity. For a double stranded nucleic acid tether we here use the  $z$ -derivative of the inextensible worm-like chain (WLC) model as an approximation for the spring constant  $k_z$ . The three stiffnesses in the MT are:

$$K_x = \frac{F}{L_{ext}} \quad [5]$$

$$K_y = \frac{F}{L_{ext} + R} \quad [6]$$

$$k_z = \frac{\partial F(L_{ext})}{\partial L_{ext}} = \frac{K_B T}{2L_p L_0} \cdot \left( 2 + \left( 1 - \frac{L_{ext}}{L_0} \right)^{-3} \right) \quad [7]$$

where  $L_p$  is the persistence length,  $L_0$  the contour length of the nucleic acid tether,  $K_B$  the Boltzmann constant, and  $T$  the absolute temperature. Note that for simulations of different tethers, *e.g.*, consisting of ssDNA or protein, other spring constants for the  $z$ -direction may need to be used. With  $\langle \delta x^2 \rangle$ , the variance of the bead excursions in the  $x$ -direction, the equipartition theorem directly provides an estimation of the force, as described previously [337]:

$$F = \frac{k_B T L_{ext}}{\langle \delta x^2 \rangle} \quad [8]$$

## Experimental parameters

### Experimental configuration of magnetic tweezers

We use a MT setup similar to that developed and described by Strick *et al.* [337] (Fig. 1) as published elsewhere [362]. Briefly, a CCD camera with a sampling frequency of 120 Hz (TM-6710CL, PULNix America Inc.) was used to track superparamagnetic beads of 2.8  $\mu\text{m}$  in diameter (M-280 beads, Life Sciences) tethered to the surface. As magnets, we used gold-plated (Ni-Cu-Ni-Au), 5 x 5 x 5 mm neodymium-iron-boron (NdFeB) permanent magnets (SuperMagnet, Germany). As tether, we used a 7.9 or 3.6 kb dsDNA construct with multiple biotin and digoxigenin labels at the ends [363]. Flow cells were made from microscope coverslips with parafilm spacers. Surfaces were passivated with bovine serum albumin (10 mg/ml), while measurements were performed in phosphate buffered saline at room temperature. Reference beads were attached to the surface and imaged simultaneously with measurement beads to correct for mechanical drift. Bead fluctuations were tracked with an accuracy of  $\sim 5$  nm.

### Simulation of magnetic tweezers data

The simulation of the tethered beads and spectral analysis was performed in LabView 8.6. Typical parameter values are listed in table S1, unless indicated otherwise for specific simulations.

## Results

In an MT experiment, a DNA or RNA construct tethers a bead to a flow cell surface (Fig. 1A). Due to Brownian motion, this bead is displaced from its equilibrium position, but constrained by both the applied force (governed by the bead's internal magnetisation and the magnetic field) and the flexibility of the nucleic acid tether (Fig. 1B and C). The Brownian motion of the bead in  $x$  is inversely related to the applied force and may thus be used to determine the force imposed onto the system (Eq. 8). MT experimental data, however, is typically obtained via image acquisition through a CCD camera, which may result in erroneous measurement of the force when tracking high frequency Brownian movements.

To study the adverse effects of image acquisition on force measurements and determine how well they can be corrected for, we have developed an MT simulator to which the effects of the camera can be applied. This simulated data was then compared to experimental traces of uncorrected MT data (Section I). We subsequently examined how well camera effects due to finite camera acquisition frequencies can be deconvoluted, and the extent to which this procedure limits the force calibration of MT experiments (Section II). Furthermore, we investigated the effects of experimental conditions and how their optimisation could improve the force calibration, and present these findings as general guidelines for the spectral analysis of MT data (Section III).

### I: Magnetic tweezers simulations

#### Equation of motion of the tethered bead

To simulate a tethered bead system, we start with the Einstein-Ornstein-Uhlenbeck theory of Brownian motion [360], which defines a bead's movements in a harmonic trap with the associated Langevin equation:

$$m \cdot \ddot{x}(t) + \gamma \cdot \dot{x}(t) + k \cdot F_{therm} \quad [9]$$

where  $x(t)$  is the position of the particle as function of time,  $m$  the inertial mass,  $k$  the trap stiffness,  $F_{therm}$  the thermal force on the particle by random collisions with water molecules, and  $\gamma$  the friction coefficient, which is itself defined as  $\gamma = 6 \pi \eta R = k_b T / D$ , with  $D$  representing the bead's diffusion constant, and  $\eta$  the dynamic viscosity. Since the loss of kinetic energy through friction takes place over a very short time interval,  $t_{inert} \equiv m / \gamma$  ( $\sim 10^{-6}$  sec), the inertial term is negligible. Defining  $\omega$  as the frequency in radians per second, the theoretical power spectrum  $P(\omega)$  of the Brownian fluctuations is given by [353]:

$$P(\omega) \propto \left| x^2(\omega) \right| = \frac{2\gamma k_B T}{\gamma^2 \omega^2 + k^2} \quad [10]$$

whose integral can be fit with an arctangent. The radial cut-off frequency of the system  $\omega_c$  is defined as the frequency at which  $P(\omega)$  is half its maximal value:

$$\omega_c = \frac{k}{\gamma} = \frac{kD}{k_B T} \quad [11]$$

We will also frequently refer to the cut-off frequency in Hz, defined as  $f_c = \omega_c / 2\pi$ , since this better facilitates comparison with experimental and simulated data. We note that  $\gamma$  is influenced by  $R$ , as discussed in the supplemental text, and that an increase of the effective  $R$  near surfaces lowers the cut-off frequency of the system. Additionally, we note that at low forces the motion of the bead may change to coupled translation-rotations due to interactions with the surface [364]. Since we focus our study on high forces, these effects have been omitted for simplicity.

For our simulation of motion in the MT, we assume that the motion of a tethered bead is induced by Langevin-force impulses of duration  $\Delta t$ , whose effective displacement ( $\delta x_{\text{Langevin}}$ ) is drawn from a Gaussian distribution with  $\langle \delta x \rangle = 0$  and  $\langle (\delta x)^2 \rangle = 2D\Delta t$  as detailed elsewhere for tethered particle motion [365,366]. We can thus write for  $\Delta x$ :

$$\Delta x = \frac{-k_x x D}{k_B T} \Delta t + \text{Langevin}(\Delta t) \quad [12]$$

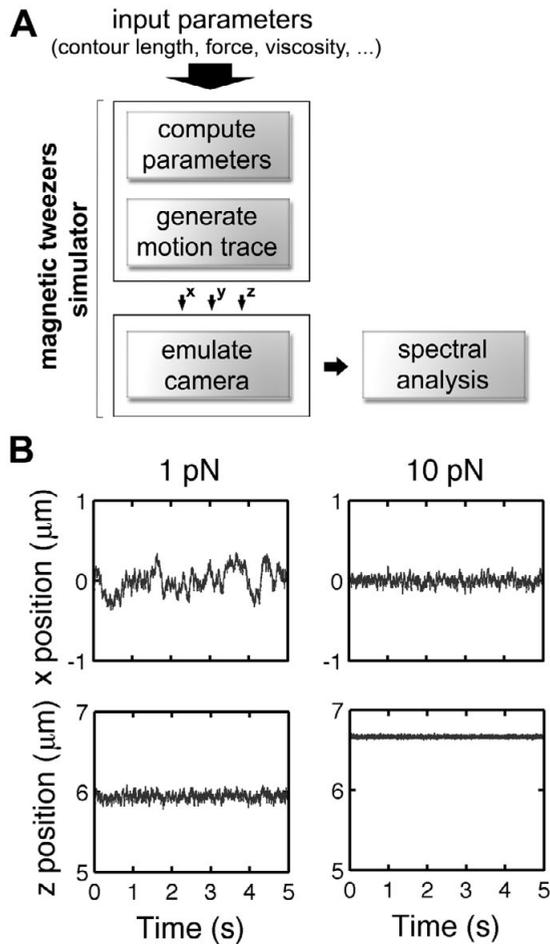
Using Eq. 5-7 we can write three equations of motion in MT:

$$\Delta x(t) = \frac{F}{L_{\text{ext}}} \cdot \frac{-xD}{k_B T} \Delta t \delta x_{\text{Langevin}}(\Delta t) \quad [13]$$

$$\Delta y(t) = \frac{F}{L_{\text{ext}} + R} \cdot \frac{-yD}{k_B T} \Delta t \delta x_{\text{Langevin}}(\Delta t) \quad [14]$$

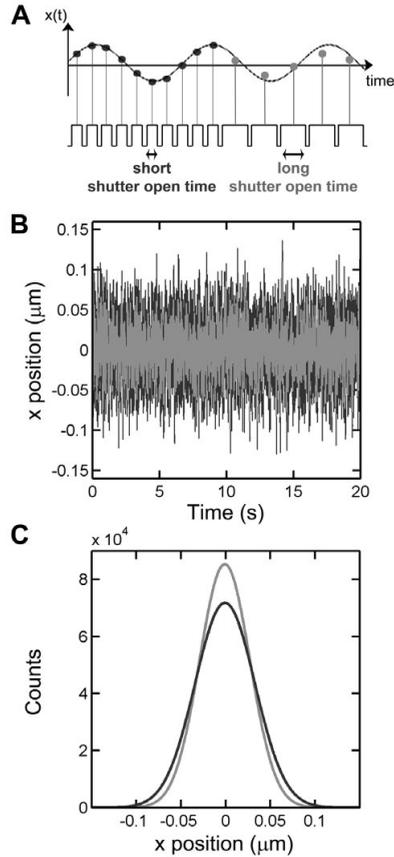
$$\Delta z(t) = \frac{k_B T}{2L_p L_0} \cdot \left( 2 + \left( 1 - \frac{L_{\text{ext}}}{L_0} \right)^{-3} \right) \cdot \frac{-zD}{k_B T} \Delta t \delta x_{\text{Langevin}}(\Delta t) \quad [15]$$

These equations allowed us to simulate a series of tethered bead excursions in  $x$ ,  $y$ , and  $z$ -directions, at user-defined forces and tether parameters (Fig. 2). We typically simulated traces with a time step  $\Delta t$  of 0.005 ms, which is small enough to correctly simulate the



**Figure 2. Simulation of a tethered particle.** (A) Schematic presentation of simulation and analysis procedure, showing that the magnetic tweezers simulator consists of three modules: 1) parameter computation, 2) motion simulation and 3) camera emulation. (B) Simulated traces of Brownian motion of a tethered bead in three dimensions. The input parameters for these simulations were  $1.4 \mu\text{m}$  for the bead radius ( $R$ ),  $7 \mu\text{m}$  for the contour length ( $L_0$ ),  $50 \text{ nm}$  for the persistence length ( $L_p$ ) and either a 1 or 10 pN force applied in  $z$ . Changing the applied force readily results in different amplitudes in all three dimensions and a shift of the bead position in  $z$ .

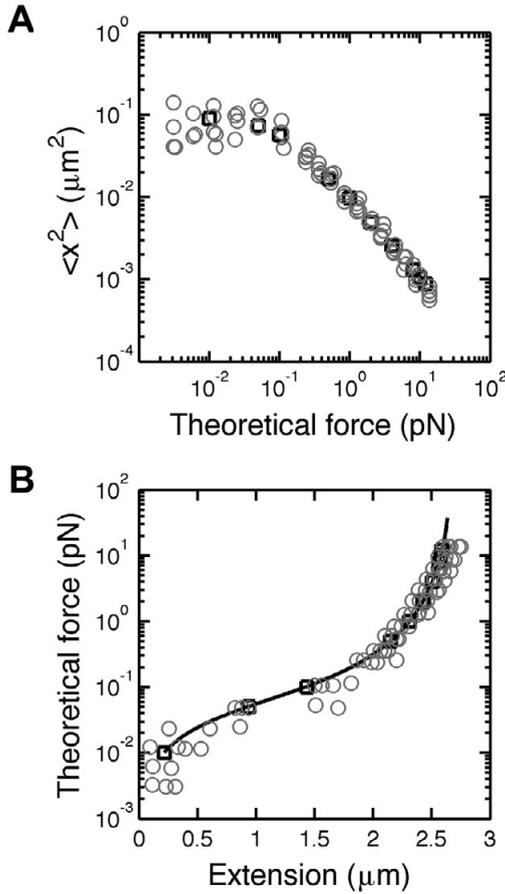
Brownian motion under high forces (Fig. S1A, S1C). Shortening  $\Delta t$  did not significantly affect the outcomes of the spectral analysis (Fig. S1C).



**Figure 3. Effect of camera integration time on measurement of tethered bead fluctuations.** (A) Illustration of the effect of camera sampling. Lower sampling frequencies combined with long shutter times result in averaging effects, thereby reducing the observed bead movements (light grey) relative to the true movements. (B) Simulated trace of a bead's position in  $x$  as function of time at 10 pN stretching force (dark grey trace). Sampling this simulated trace with an emulated camera with  $f_s$  of 120 Hz and shutter time of 8.3 ms clearly reduces the variance in  $x$  (light grey trace). (C) This effect is also evident when the variance in  $x$  is analysed with a Gaussian fit to the histograms of the bead traces (same coding as in Fig. 3B). The simulations shown in this figure were performed as defined in table S1, using a constant force of 10 pN.

### The simulated variances are in agreement with experimental data

To obtain simulated data that was comparable to the experimental data derived from video microscopy, we emulated the effects of image acquisition by a CCD camera. In general, a camera samples at a finite camera or acquisition frequency  $f_s$  and integrates each sample point over a finite integration time or shutter time  $W$ , leading to motion blur and aliasing (see Section II for details). To emulate the effects introduced by the CCD camera, we sampled the  $x(t)$ ,  $y(t)$  and  $z(t)$  data at a fixed  $f_s$  and time-averaged each



**Figure 4. Verification of simulator output with experimental data.** (A) The output of the simulations was tested by i) plotting the simulated mean variance in  $x$  as function of  $F_{sim}$  (squares) and ii) comparing it to the experimentally obtained variances for DNA-tethered beads against the applied  $F_{mag}$  (circles). The latter force was computed from beads confirmed to have uniform magnetisation as described elsewhere [341]. Note, to clearly distinguish the forces computed here from those forces deduced via analysis of the bead fluctuations (see equation 8), we have labelled them here as ‘theoretical force’. (B) The output of the simulation was also tested by plotting the simulated extensions (squares) and experimental extension (circles) against the theoretical force (see note in A). Both agree well within the entropic regime of the inextensible WLC (<10 pN), as expected. The fit parameters for the inextensible worm-like chain model are  $L_0 = 2.7 \mu\text{m}$  and  $L_p = 51 \text{ nm}$ .

sample over a rectangular window of width  $W$ . For simplicity, we assume that  $f_s$  and the shutter frequency  $f_e$  ( $f_e = 1/W$ ) are equal, *i.e.*  $f_e = f_s$  (see table S1). To ensure that the observed differences between the original and camera-averaged variances did not derive from a statistical error  $\epsilon$  of force measurement at low forces, we used sufficiently long traces for our analysis (see Fig. S1B and supplemental text of chapter 6). As shown in

Fig. 3, camera averaging resulted in a reduction of the measured variance  $\text{var}(x_m)$  of the bead's position compared to the true variance  $\text{var}(x)$ :  $\text{var}(x_m) \leq \text{var}(x)$ .

To verify that the time-averaging was implemented correctly in the simulations, we compared time-averaged data from simulations and experiments by plotting i) the simulated, time averaged variances against the simulated force ( $F_{sim}$ ) and ii) the experimentally measured variances against the magnetic force ( $F_{mag}$ ). It should be noted that the latter is only possible when the calculated  $F_{mag}$  (computed using Eq. 1) is known to be an accurate measure of the force that was applied experimentally, as described previously for uniform beads in combination with the field configuration specified in Fig. 1 [341]. As shown in Fig. 4, we find excellent agreement between the simulated bead variances (Fig. 4A, squares) and the experimental data (Fig. 4A, circles). The force-extension behaviour of the simulated tethers (Fig. 4B, squares) also agrees well with the experimentally obtained data (Fig. 4B, circles) within the entropic regime of the inextensible WLC model ( $<10$  pN), as expected.

## II: Approach to correct for camera blurring and aliasing

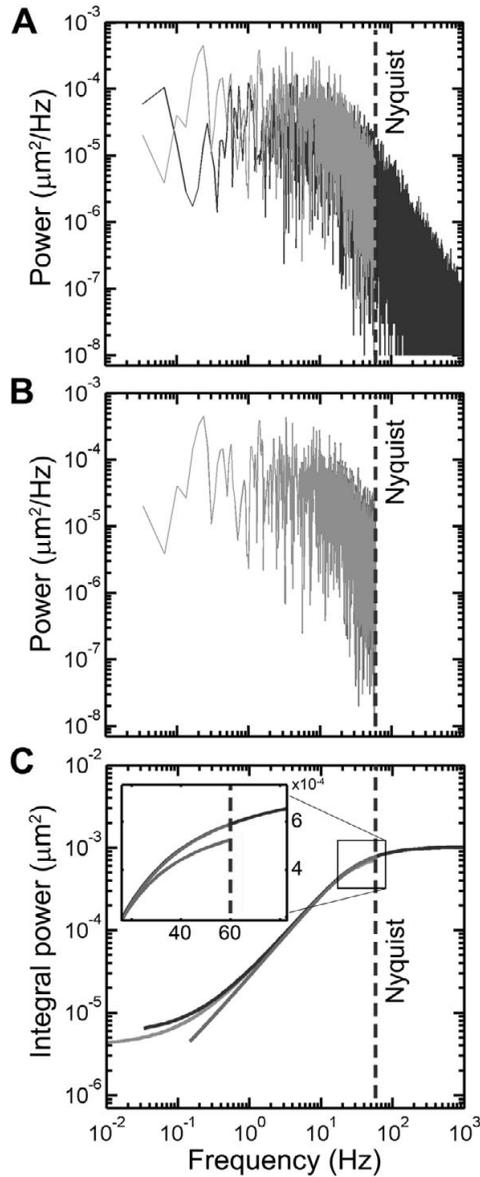
### Method to correct for effects of camera on measured spectrum

According to the Nyquist theorem, camera sampling with a finite  $f_s$  can bring about significant errors when tracking frequencies higher than half  $f_s$ . In general, sampling Brownian motion with a camera results in two effects. The first is time averaging or motion blur, and leads to  $\text{var}(x_m)$  being lower than  $\text{var}(x)$  as mentioned above (Fig. 3B). The second is aliasing, which results in a contribution to the frequencies of the measured spectrum by signal components that are higher than  $f_s$ . This principle is also known as back-folding, and although it does not affect the power of the spectrum it does change its shape, and consequently the cut-off frequency of the system.

If we assume that the integration window of the camera is rectangular (Fig. 3A), we define  $P_{window}(\omega)$ , the power spectrum of the moving average window, by:

$$P_{window}(\omega) = \left( \frac{\sin(\omega W / 2)}{\omega W / 2} \right)^2 \quad [16]$$

To retrieve the underlying, correct variance from the incorrect, measured  $\text{var}(x_m)$ , we start with the measured power spectrum  $P_m$  iteratively correct it (see Eq. 17-19 below), and fit the integral of the spectrum with an arctangent until the fitting error reaches a value below  $10^{-4} \mu\text{m}^2/\text{Hz}$ . The correction steps we include are the following: first, we account for the finite exposure effect of the camera, which increases the power in the



**Figure 5. Means of correcting the camera-sampled fluctuations.** (A) Simulation of the power spectral density of the bead fluctuations in  $x(t)$  (dark grey points: without camera emulation; light grey points: with camera emulation) results in a loss of signal beyond  $F_{Nyq}$  (here  $60\text{ Hz}$ ) and averaging effects over the measured frequencies. (B) The integrated spectra can be fit to an arctangent to yield the variance in  $x$ . As is clear from the inset, the figure shows that the integral to the corrected power spectrum (dark grey) produces the same result as the fit to the integral of the original power spectrum (black), whereas the fit to the integral of the camera-sampled spectrum (light grey) clearly differs from the fit to the integral of the original power spectrum.

higher frequencies that was lost by camera-based sampling (Fig. 5, compare original black spectrum to sampled light grey spectrum, and the sampled spectrum to the corrected dark grey spectrum) by dividing the measured spectrum  $P_m$  through the blur correction term  $C_{blur}$  [353]:

$$C_{blur}(\omega) = \left( \frac{\sin(\omega W / 2)}{\omega W / 2} \right)^2 \quad [17]$$

Second, we account for potentially back-folded frequency components from the measured spectrum with the alias correction term  $C_{alias}$ , which is subtracted from the measured spectrum (see Eq. 19 below). Although in principle this correction has to be performed for an infinite number of terms, the first two anti-aliasing terms (*i.e.*,  $n = 1$  and  $n = -1$ ) provide sufficient estimation of the alias effects of the camera given the rapid decay in contribution of higher order terms. Indeed we find that based on the data presented in Fig. 6, the effect of accounting for the first two terms is  $\sim 30\%$  and  $\sim 1\%$ , respectively. We write for  $C_{alias}$  [353]:

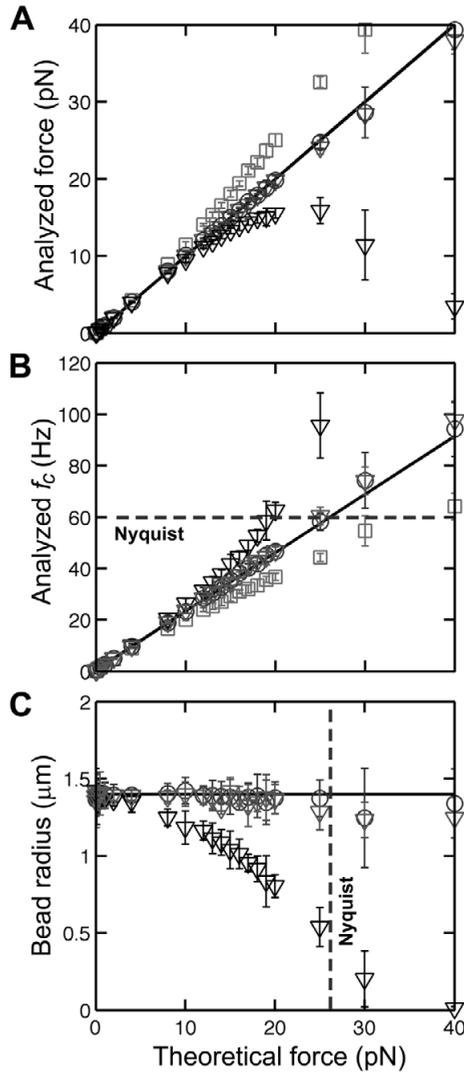
$$C_{alias}(\omega) = \sum_{n=-1,+1} \left( \frac{2\gamma k_B T}{\gamma^2 (\omega + n \cdot \omega_s) k^2} \right) \cdot \left( \frac{\sin(\omega + n \cdot \omega_s) W / 2}{(\omega + n \cdot \omega_s) W / 2} \right)^2 \quad [18]$$

where  $\omega_s$  is the radial sampling frequency ( $\omega_s = 2\pi f_s$ ). Additionally, each anti-aliasing term is itself corrected for blurring through multiplication by a  $C_{blur}$  term. Using Eq. 17 and 18, we can now correct the measured power spectrum and obtain a correctly fitted var ( $x$ ) after integration of the spectrum (Fig. 5C), using:

$$P_{corrected}(\omega) = \frac{P_m - C_{alias}}{C_{blur}} \quad [19]$$

### Performance of camera correction terms in MT simulations

Having validated the MT simulator, we now have a tool to explore the effect of the various steps in the correction procedure on computing var( $x$ ) from var( $x_m$ ). In particular, we will systematically apply either no correction, account for blur effects only by iteratively dividing  $P_m$  through Eq. 17, or account for both blur and anti-aliasing using Eq. 19, and plot the forces and cut-off frequencies obtained through spectral analysis against the theoretical  $F$  and  $f_c$  that were used as inputs in the simulation. Ideally, the theoretical



**Figure 6. Effectiveness of spectral corrections of the fluctuations.** (A) Simulated force, and the force deduced from the transverse fluctuations through spectral analysis. (B) Simulated force and the cut-off frequency ( $f_c$ ) derived from spectral analysis. (C) Deduction of  $R$  via spectral analysis of the fluctuations, plotted as function of the force. The deviation between the measured  $R$  as obtained after spectral analysis and the expected value for  $R$  can be used as a measure for the extent to which the spectral analysis yields correct values for the forces. In A-C, the following coding is employed: input parameters (black line), analysis in the absence of any correction for camera sampling (squares), analysis with blur correction (black triangles), analysis with blur and the  $n = -1$  alias correction (green triangles), and analysis including the complete correction including blur, alias  $n = 1$ , and alias  $n = -1$  corrections (circles). We find that the alias  $n = 1$ , and alias  $n = -1$  corrections contribute 30% and 1% to the correction, respectively. Error bars represent SD of 5 complete repeats of data simulation and analysis.

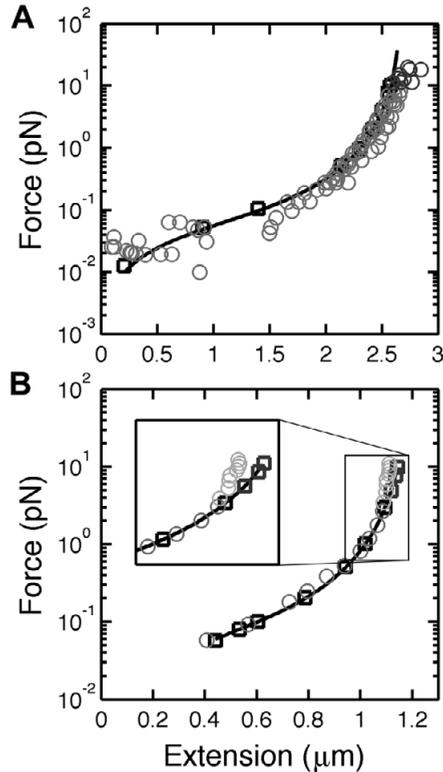
force and cut-off frequencies should be equal to the values obtained from the analysis. As an additional test for the correction efficiency, we will also derive  $R$  from the spectral analysis, which follows directly from  $f_c$  via  $f_c = k / (12\pi^2\eta R)$  (see Eq. 11). This  $R$  should be constant and equal to the bead radius that we define in our simulation. In experimental situations, the measured  $R$  also presents a simple confidence measure, provided that the true bead radius is known (e.g., provided by the manufacturer).

When we apply no correction to the recorded spectrum, we observe an overestimation of the force due to averaging effects of the camera. Specifically, an error of 10% is reached at forces of 14 pN for a 2.7  $\mu\text{m}$  dsDNA construct, and this error becomes larger when more force is applied (Fig. 6A, squares). At these forces, the natural frequency of the system exceeds 32 Hz, or approximately one-fourth of  $f_s$  (Fig. 6B). If we apply the blur correction only, we observe a systematic underestimation of the force when  $F > 14$  pN (Fig. 6A, B, triangles). We thus find that applying solely a blur correction to the recorded spectrum does not provide a more reliable calibration of the force than spectral analysis without blur correction. However, if we apply the aliasing correction only (i.e., no blur correction), we find already a  $>10\%$  overestimation of the force at 1 pN (data not shown).

Fortunately, when we subsequently account for both blur and the  $n = -1$  term of the anti-alias (Eq. 18), a significant reduction in the error is observed, which thus facilitates calibration of higher forces with the same confidence (Fig. 6A, triangles). This effect is apparent in both the correction of the force (Fig. 6A) and the estimation of  $R$ , our additional control (Fig. 6C). In fact, we can observe in Fig. 6B and 6C that if both blur and aliasing corrections are taken into account, reliable measurements of the mean applied force can be performed for cut-off frequencies much closer to  $f_{Nyq}$  or even beyond  $f_{Nyq}$ . When we specifically look at the mean force obtained within  $<10\%$  error and small standard deviations, we see that we can correctly measure forces up to  $\sim 19$  pN, or  $f_c \sim 46$  Hz. Finally, when we account for the  $n = 1$  term in the anti alias as well, we observe a small additional increase ( $\sim 1\%$ ) in force-measurement confidence (Fig. 6A, circles), which is in accordance with the rapid decay in contribution of the alias terms. Overall, we demonstrate that when all camera correction terms are applied, force measurements can be accurately performed within 10% error for tethered bead systems with cut-off frequencies less than  $47 \pm 2$  Hz (e.g.,  $20 \pm 1$  pN for a 2.7  $\mu\text{m}$  DNA tether). This is close to 80% of  $f_{Nyq}$ .

### Correction of experimental data

Having explored the spectral correction procedure on simulated data, we next apply this procedure to experimental MT traces. To this end, we obtained experimental data for two relatively short DNA construct lengths namely, 7.9 kb ( $\sim 2.7$   $\mu\text{m}$ ) and 3.6 kb ( $\sim 1.2$   $\mu\text{m}$ ). We subsequently analysed  $x(t)$  data acquired with these molecules and corrected  $P_m$  for both blur and aliasing effects. Since we already confirmed that our MT simulator



**Figure 7. Application of spectral corrections to experimental data.** (A) Force extension curve obtained after spectral analysis of tethered bead simulations (squares) or experimental data (circles) for a dsDNA molecule with a contour length of  $2.7 \mu\text{m}$  ( $\sim 7.9$  kbp). Also shown is the fit of the inextensible worm-like chain model (fit parameters are  $L_0 = 2.7 \mu\text{m}$  and  $L_p = 49$  nm). Data points beyond the applicable range of the inextensible WLC model are indicated in dark grey circles. The natural frequency did not exceed 47 Hz (see Fig. S2). (B) Force extension curve obtained after spectral analysis of tethered bead simulations (squares) or experimental data (circles) of a dsDNA molecule with an  $L_0$  of  $1.2 \mu\text{m}$  ( $\sim 3.6$  kbp) show good agreement below cut-off frequencies below 47 Hz ( $\sim 3$  pN). At higher forces, the points between the simulated data, the fit to inextensible WLC model (black line; parameters are:  $L_0 = 1.2 \mu\text{m}$  and  $L_p = 45$  nm) and experimental data start to diverge (indicated with different shades of grey). Both simulated and experimental data were corrected for blurring and aliasing effects.

was able to correctly reproduce the relation between the force, extension and Brownian motion of a tethered magnetic bead within the entropic regime of the WLC (Fig. 4), we simulated data sets specific for the experimental configurations used and were thus able to assess the efficacy of the spectral analysis of experimental data through direct comparison.

For the 7.9 kb dsDNA construct, we observe good agreement between the simulated and experimental data sets below forces of 10 pN (Fig. 7A, compare circles and squares). We also observe that  $R$  obtained from the analysis of the experimental data was within

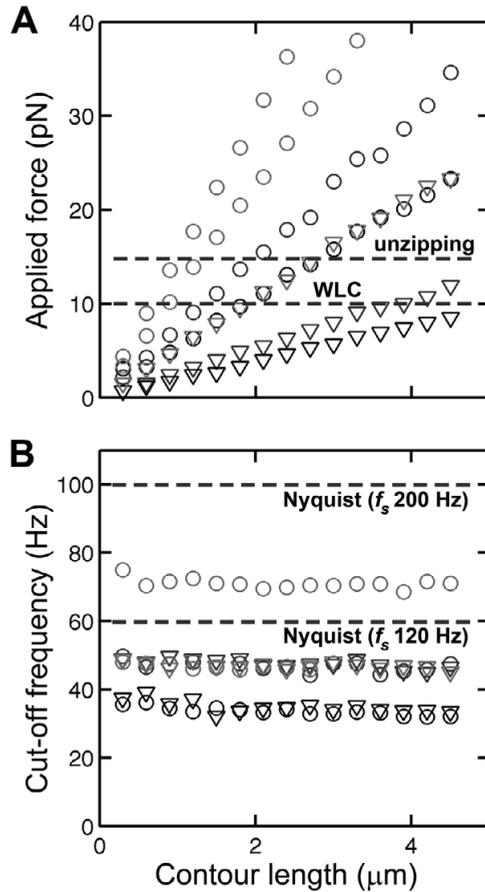
the expected error (Fig. S4). At forces higher than 10 pN, deviations can be observed between the experimental and simulated force-extension data (Fig. 7A, compare circles and squares). This is, however, in accordance with the inextensible WLC fit model used for the simulation as demonstrated by the fit to the experimental data (Fig. 7A, black line). When we analyse the data for the 3.6 kb construct, we find that deviations between the simulated and experimental data already start to occur at  $\sim 3$  pN (Fig. 7B, compare squares with circles). At this force,  $f_c$  exceeds 47 Hz. As established using the MT simulator, this  $f_c$  corresponds to the 10% error threshold that is reached by the spectral analysis, thus explaining the deviation. Additionally, we also observe that at higher forces (*i.e.*,  $> 3$  pN) significant deviations become apparent between the bead radius obtained via spectral analysis and the original  $R$  (not shown). The 7.9 kb tether remained well below this threshold over the whole force range used (Fig. S2). This thus demonstrates that we can correctly measure forces in experimental MT setups, but that we have to be aware of the limitations of the spectral analysis, as identified using the MT simulator. Additionally, these results reveal that one should be able to measure higher forces with higher confidence when the  $L_o$  of the tether increases, and, consequently,  $k_x$  and  $f_c$  decrease.

### III: Guidelines to spectral analysis

Validation of the correction procedure with experimental data puts us in a position to explore the limits of our approach and to provide guidelines for optimising tethered bead calibrations. This is particularly vital when relatively high forces have to be measured with good accuracy (*e.g.*, errors  $< 10\%$ ). When short tethers are not essential for the experiment, a long  $L_o$  will generally provide good results, as shown in Fig. 8A. In this plot, the 10% error limit is presented as a function of  $L_o$ . We readily see that when  $L_o > 2 \mu\text{m}$  and  $R = 1.4 \mu\text{m}$ , forces of 10 pN can be reliably measured with 10% accuracy without camera correction procedures (Fig. 8, circles). We can also observe in Fig. 8 that the 10% limit rises linearly with  $L_o$ .

When high forces have to be achieved for relatively short  $L_o$ , either the application of i) spectral corrections, ii) the implementation of faster camera frequencies, iii) high viscosity buffers, and iv) larger beads are options to be considered. As shown in Fig. 8A, when we first apply both the blur and alias correction we can reliably measure forces (error  $< 10\%$ ) that are 1.4-fold higher compared to the situation in which camera corrections are not implemented (Fig. 8A, compare dark and light circles). More specifically, we can now measure forces close to 8 pN with  $1 \mu\text{m}$  tethers ( $\sim 3$  kb dsDNA,  $R = 1.4 \mu\text{m}$ ). We find similar differences between uncorrected and corrected forces when beads are used with a  $0.5 \mu\text{m}$  radius (Fig. 8A, triangles).

The accurate measurement of even higher forces can be achieved by increasing the sampling frequency of the camera (Fig. 8A, grey circles). As an example, we emulated a sampling frequency of 200 Hz and observed that the 10% error limit was raised to  $\sim 12$



**Figure 8. Force measurements at a 10% error limit using various experimental parameters.**

(A) DNA constructs tethered to beads with a radius of 1.4  $\mu\text{m}$  (circles) or a radius of 0.5  $\mu\text{m}$  (triangles) were used to create 10% error plots under varying conditions. Curves are shown for uncorrected spectral analysis (black), camera-corrected analysis (dark grey), camera-corrected analysis for traces simulated at a two-fold higher viscosity (light grey), and camera-corrected analysis for traces sampled at 200 Hz instead of 120 Hz (second light grey curve from left). Indicated by dotted lines are the range of validity of the inextensible WLC model (<10 pN) and the unzipping force required to open a typical DNA hairpin. (B) The cut-off frequencies for the tethered bead systems simulated in A. This illustrates that applying spectral analysis corrections for blurring and aliasing allows one to measure within 10% error for cut-off frequencies close to 80% of  $F_{\text{Nyq}}$ .

pN for 1  $\mu\text{m}$  tethers. This is an improvement of  $\sim 1.5$  fold relative to the blur and alias corrected force measurements. The third alternative to improve correct measurement of higher forces is increasing the viscosity of the calibration buffer (Fig. 8A, circles). In general, the force that can be accurately measured increases linearly with increasing viscosity, *e.g.*, doubling the viscosity raises the force limit by a factor of 2, thereby raising

the 10% error limit to  $\sim 15$  pN for  $1 \mu\text{m}$  tethers. Such higher viscosities are relatively easy to achieve, since *e.g.*, 25% glycerol raises the viscosity by a factor of  $\sim 2$  at room temperature. However, higher viscosities may not be suitable to all experiments as they may change enzyme behaviour and complicate cross-comparisons with the literature. This may be solved by first performing a magnet and construct calibration at high viscosity, before the actual experiment is done in a low viscosity buffer following buffer exchange [367].

The fourth alternative is an increase of bead size. As an example, we used here beads with a radius of  $1.4$  or  $0.5 \mu\text{m}$  for our force measurements and can readily see that increasing the bead radius gives a linear increase of the force limit for the same error magnitude (Fig. 8A, compare triangles and circles). The range of bead sizes is of course limited by what manufacturers supply. Furthermore, the use of larger beads may be less desirable when short tethers are used, since this may change the geometry of the bead's behaviour in the magnetic field (*e.g.*, rolling along its attachment) and introduce interactions of the bead with the surface of the flow cell [364]. Moreover, larger bead sizes are accompanied by poorer time resolution and increased Brownian motion, since  $\gamma$  is directly related to  $R$  (see section I). The choice of  $R$  will thus be a trade-off between the force range and correct experimental design.

## Discussion and conclusions

The forces that can be applied to tethered beads depend both on the magnetisation of the beads and the magnets of the tweezers setup [341]. However, reliable measurement of the applied forces depends i) on the correcting of experimental frequency specific artifacts, such as acoustical noise peaks and low frequency drift, and ii) how well camera corrections can be applied to allow measurements near  $f_{Nyq}$ . As noted elsewhere,  $f_{Nyq}$  does not always presents a solid upper limit to spectral analysis since correction procedures employing varying shutter speeds and oscillating light-sources have been demonstrated to reconstruct variances beyond  $f_{Nyq}$  [353,368]. However, to date these procedures have not been investigated for MT experiments and we chose to limit our analysis here to the direct corrections of a single power spectrum as light modulation i) reduces the overall light intensity and this the signal-to-noise ratio, ii) requires additional hard- and software, and iii) necessitates several measurements at different modulation frequencies to build up a spectrum.

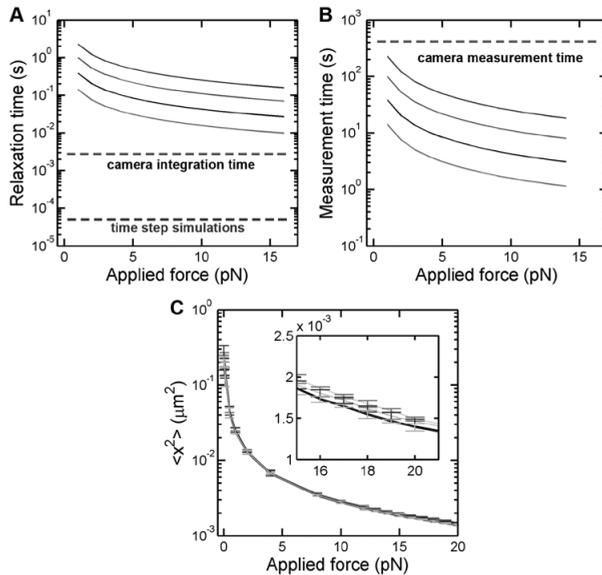
We here explored the limits of MT spectral analysis for a wide range of configurations, varying either the applied force, contour length of the tether, bead size, camera sampling frequency and the buffer viscosity. We observe that it is crucial (error  $> 10\%$ ) to correct the sampled Brownian spectrum when the natural frequency of the system is about a quarter of the sampling frequency of the camera. We also find that the spectral correction should take both blurring and the first order alias into account. When it is required

for the particular experiment to accurately measure higher forces - or when the natural frequency exceeds 80% of  $f_{Nyq}$ , which can simply be calculated from Eq. 11 - we find that increasing the sampling frequency of the camera gives a good optimisation of the force range (table S2). Alternatively, bead sizes and buffer viscosities can be increased since both give linear improvements of the force range within the same error magnitude (Fig. 8B), provided that the tracking error does not exceed the sampling error (see Fig. S3 and supplemental text). Ultimately, bead sizes obtained through spectral analysis at lower forces can be used to better calibrate very high forces when the bead size is then treated as a fixed fitting parameter. Such additional though more intensive procedures can be likely be used to extent the spectral analysis even closer or beyond  $f_{Nyq}$  [353,368].

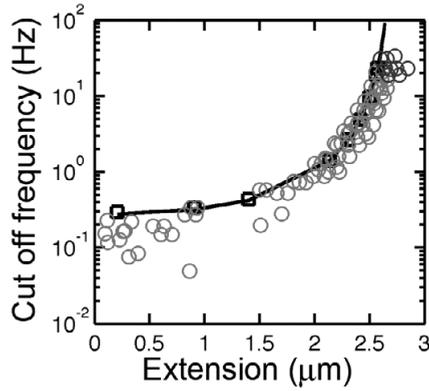
### Acknowledgments

We thank Susanne Hage, Iwijn de Vlaminck and Gary Skinner for assistance with DNA construction, helpful discussions and providing experimental data. This work was supported by the Netherlands Organization for Scientific Research (NWO) via a Veni (to JL), Vidi (to ND), and Toptalent grant (to AV), and by the European Science Foundation via a EURYI grant (to ND).

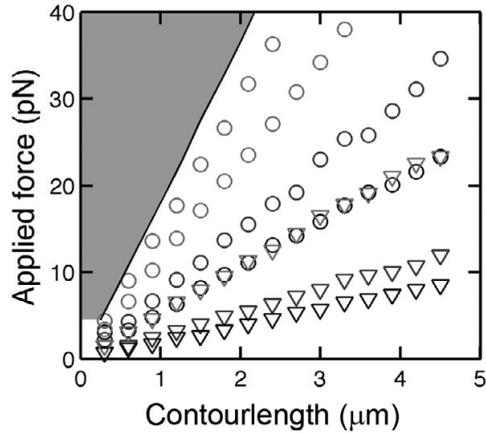
## CHAPTER 6 - SUPPLEMENTAL INFORMATION



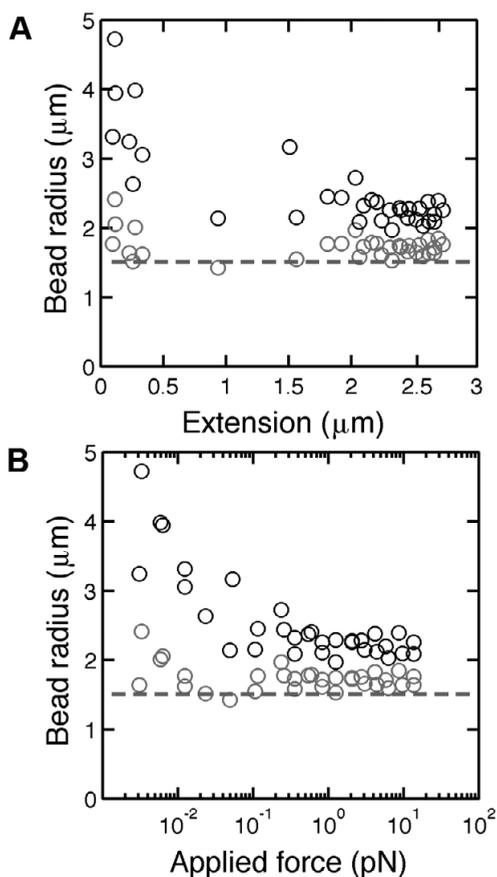
**Figure S1. Time scales in the simulation.** (A) Simulations showing the relaxation times in the magnetic tweezers for tethered systems of different contour lengths ( $L_0 = 16 \mu\text{m}$ , top curve;  $L_0 = 7 \mu\text{m}$ , second curve from top;  $L_0 = 2.7 \mu\text{m}$ , black curve;  $L_0 = 1 \mu\text{m}$ , bottom curve) as a function of the input applied force. The dashed dark grey line indicates the simulation time step employed in the simulations of magnetic tweezers data, whereas the dashed light grey line indicates the integration time of the camera. Traces were generated with parameters  $R = 1.4 \mu\text{m}$  for the bead and  $L_p = 50$  nm for the persistence length of the tether. (B) Simulations showing the required time durations for accurate measurement ( $<10\%$  statistical error; see also supplemental text) of the force in the magnetic tweezers for tethered systems of different contour lengths as in Fig. S1A. The dashed grey line indicates the overall measurement time by the camera (*i.e.*, 5 min). (C) The simulated bead variances obtained with different simulation time steps ( $\Delta t = 0.05$  ms, dark grey;  $\Delta t = 0.01$  ms, light grey;  $\Delta t = 0.005$  ms, ultralight grey; theoretical deviations, black line) plotted against the applied simulated force. Overall, we observe no significant dependence of the simulated variances on the size of the time step in the range investigated. Only for the highest forces, do we observe that long simulation time steps lead to a small deviation from the theoretical relation between the force and the bead displacement (equation 2), see inset. In practice, we employ a time step  $\Delta t = 0.005$ , which provides a good approximation to the theoretical curve.



**Figure S2. Cut-off frequencies for experimental 7.9 kb DNA.** Cut off frequencies obtained after spectral analysis as function of the tether extension. Simulated data is indicated in squares, whereas experimental data is represented as circles. Indicated in dark grey are the data points that lie outside the inextensible WLC fit (see main text, Fig. 7A).



**Figure S3. Limits in the estimation of the force and a 5 nm tracking error.** DNA constructs tethered to beads with a radius of 1.4  $\mu\text{m}$  (circles) or a radius of 1  $\mu\text{m}$  (triangles) were used to create 10% error plots under varying conditions. Curves are shown for uncorrected spectral analysis, camera-corrected analysis, camera-corrected analysis for traces simulated at a two-fold higher viscosity, and camera-corrected analysis for traces sampled at 200 Hz instead of 120 Hz. Coding as in main Fig. 8. Indicated by the black line is the force that can maximally be measured with 10% accuracy when the tracking error of the camera is 5 nm. The grey-shaded area represents force measurements with tracking errors  $< 5$  nm.



**Figure S4. Bead radius estimation in the presence or absence of viscosity corrections. (A)** Bead radii extracted from experimental data following spectral analysis, plotted as function of the extension of the tether. As shown, when the spectral analysis is performed in the absence (black circles), the estimated bead radius is significantly larger than the bead radius used for the experiment. When the spectral analysis is performed in the presence (light grey circles) of viscosity corrections, a much better estimation of the bead radius is obtained. According to the manufacturer, the beads employed had a radius of  $\sim 1.4 \mu\text{m}$  (indicated by the grey dotted line). The dsDNA tether used had a contour length of  $2.7 \mu\text{m}$ . **(B)** Bead radii extracted from experimental data following spectral analysis, plotted as function of the applied force. As the results of the spectral analysis demonstrate, the viscosity correction (black circles) allows us to recover the correct bead radius at all forces except forces below  $0.001 \text{ pN}$ . When the spectral analysis is performed in the absence (light grey circles) of viscosity corrections, however, the estimation of the bead radius is significantly larger at all forces. This procedure not only illustrates the necessity of applying viscosity corrections at low and relatively high forces, it also shows how the bead radius can be used to check the performance of the spectral analysis. As in Fig. S4A, the beads employed had a radius of  $\sim 1.4 \mu\text{m}$  (indicated by the grey dotted line).

### Sampling time vs. measurement error

Both the time for which we simulate our magnetic tweezers experiments and how long we sample real data with a CCD camera influence the statistical uncertainty in any force measurement [369]. How long should we thus measure to get a proper force measurement, assuming we take into account an expected or target force  $F$ , a DNA extension  $L_{\text{ext}}$  and  $a$ , for example, 10% acceptable statistical error? We assume we measure  $N$  points during time  $T$ . Since these points should be independent, they need to be taken at least one relaxation time  $\tau$  apart, *i.e.*  $N = T/\tau$ . For the  $x$ -direction, this relaxation time is

$$\tau = \frac{\gamma}{k_x} = \frac{12\pi^2\eta R_0 L_{\text{ext}}}{F} \quad [1]$$

Starting from equation 8 of the main manuscript, the error can be deduced from:

$$F = \frac{k_B T L_{\text{ext}}}{\langle \delta x^2 \rangle} \quad [2]$$

The associated error propagation can then be described as:

$$\frac{\delta F}{F} = \sqrt{\left[ \frac{\delta L_{\text{ext}}}{L_{\text{ext}}} \right]^2 + \left[ \frac{\delta \langle \delta x^2 \rangle}{\langle \delta x^2 \rangle} \right]^2} \quad [3]$$

In the magnetic tweezers, the DNA extension  $L_{\text{ext}}$  depends on the force applied in the  $z$ -direction. Since we measure both  $L_{\text{ext}}$  and the fluctuations of the signal over  $N$  sampling points with a certain standard error  $\sigma$ , we obtain:

$$\left[ \frac{\delta L_{\text{ext}}}{L_{\text{ext}}} \right]^2 = \frac{\sigma_z^2}{N L_{\text{ext}}^2} \quad [4]$$

and

$$\left[ \frac{\delta \langle \delta x^2 \rangle}{\langle \delta x^2 \rangle} \right]^2 = \left[ \frac{\sigma_x^2 / N}{\sigma_x^2} \right]^2 = 1 / N^2 \quad [5]$$

where

$$\sigma_x^2 = k_B T / k_x \text{ and } \sigma_z^2 = k_B T / k_z$$

with  $k_B T$  and the stiffnesses as defined in the main text with equations 5 and 7. When we combine equations 3-5 this subsequently gives:

$$\frac{\delta F}{F} = \sqrt{\frac{\sigma_z^2}{NL_{ext}^2} + \frac{1}{N^2}} = \frac{1}{N} \sqrt{N \cdot \left[ \frac{\delta_z}{L_{ext}} \right]^2 + 1} \quad [6]$$

Using

$$k_z(F) = \left. \frac{\delta F_{WLC}}{\delta_z} \right|_F \quad [7]$$

We obtain:

$$\frac{\delta F}{F} = \frac{1}{\sqrt{N}} \sqrt{\frac{k_B T}{L_{ext}^2} \left[ \frac{\delta F_{WLC}}{\delta_z} \right]^{-1} + 1} \quad [8]$$

This relationship allows us to determine the number of independent sample points

$N$  as function of  $F$ , given a desired accuracy

$$\frac{\delta F}{F} = \varepsilon$$

We deduce that:

$$N = \frac{1}{\varepsilon^2} \left( \frac{k_B T}{L_{ext}^2} \left[ \frac{\delta F_{WLC}}{\delta_z} \right]^{-1} + 1 \right) \quad [9]$$

which can be reexpressed in terms of the sampling time:

$$T = \tau \cdot \frac{1}{\epsilon^2} \cdot \left( \frac{k_B T}{L_{\text{ext}}^2} \left[ \frac{\delta F_{\text{WLC}}}{\delta z} \right]^{-1} + 1 \right)$$

Substituting  $\tau$  equation 1, we obtain our final expression:

$$T = \frac{1}{\epsilon^2} \frac{12\pi^2 \eta R_0 L_{\text{ext}}}{F} \left( \frac{k_B T}{L_{\text{ext}}^2} \left[ \frac{\delta F_{\text{WLC}}}{\delta z} \right]^{-1} + 1 \right) \quad [10]$$

As a rule of thumb, the first term between brackets expresses the contribution from the uncertainty in  $L_{\text{ext}}$ . This term approaches unity for short (<1 micron) tethers at low forces (<0.1 pN) and  $N$  has to be >100 for  $\epsilon = 0.1$ .

### Tracking error

If we measure the bead positions with a certain tracking error  $\sigma^2$  and variance  $\sigma_T^2$ . Sampling  $N$  points will therefore result in an additive error in force measurements, giving:

$$F' = \frac{k_B T L_{\text{ext}}}{\sigma^2 + \sigma_T^2} \quad [11]$$

Thus,

$$\frac{F' - F}{F} + \frac{F'}{F} - 1 = \frac{\sigma^2}{\sigma^2 + \sigma_T^2} - 1 = \frac{1}{\sigma^2 / \sigma_T^2} \quad [12]$$

This thus amounts to a bias  $\epsilon_T$  in the force:

$$\epsilon_T = 1 \cdot \left( \frac{k_B T L_{\text{ext}}(F)}{F \cdot \sigma_T^2} + 1 \right)^{-1} \quad [13]$$

### Corrections to the bulk viscosity in proximity of the surface

In the proximity of the flow cell surface, the Stokes drag on the tethered bead increases [370]. This increase becomes significant once the distance  $z$  to the surface reaches the same order of magnitude as the bead's radius, as described by Happel and Brenner [371].

If one continues to use the bulk viscosity ( $\eta$ ) contribution to the drag mathematically, the measurement bead radius ( $R_m$ ) will appear larger than the expected radius  $R_0$  by the viscosity correction factor  $C_\eta$ :

$$R_m = R_0 \cdot C_\eta \quad [14]$$

where

$$c_\eta(z, R) = \left[ 1 - \frac{9}{16} \left( \frac{R}{L_{\text{ext}}} \right) + \frac{1}{8} \left( \frac{R}{L_{\text{ext}}} \right)^3 - \frac{45}{256} \left( \frac{R}{L_{\text{ext}}} \right)^4 + \frac{1}{16} \left( \frac{R}{L_{\text{ext}}} \right)^5 \right]^{-1} \quad [15]$$

Use of  $R_m$  in the computation of the system's cut-off frequency (see equation 11 of the main text) would lead to an erroneous value. To recover the correct  $R_0$ , we compute:

$$R_0 = \frac{R_m}{C_\eta} \approx R_m \cdot \left[ 1 - \frac{9}{16} \frac{R_0}{z} \right] \Leftrightarrow R_0 = R_0 \cdot \left[ 1 + \frac{9}{16} \frac{R_m}{L_{\text{ext}}} \right]^{-1} \quad [16]$$

where we have ignored quadratic and higher-order terms. However, in our bead tracking procedure we do not directly observe  $z$  (the distance to the surface), but the height difference (extension) between the tethered bead and a reference bead (Fig. 1A). If we assume that these have the same radius, we can transform Eq. 16 accordingly and solve it to yield  $R_0$  as a function of  $R_m$  and  $L_{\text{ext}}$ :

$$R_0 = \frac{1}{2} \left[ \left( \frac{7}{16} R_m - L_{\text{ext}} + \sqrt{\left( \frac{7}{16} R_m - L_{\text{ext}} \right)^2 + 4L_{\text{ext}} R_m} \right) \right]^2 + 1 \quad [17]$$

### Camera integration window

Sampling Brownian motion with a camera results in two effects. The first is time averaging or motion blur, which leads to suppression of the signal. The experimentally measured variance  $\text{var}(x_m)$  is therefore lower than the true variance that is required for precise force calibration (Fig. 3B main text). The second is aliasing, which results in a contribution to the frequencies of the measured spectrum by signals that are higher than the sampling frequency. This principle is also known as back-folding, and although it does not affect the power of the spectrum it does change its shape, and consequently

the cut-off frequency of the system. In the following section, we will discuss how we can correct for these effects in the spectral domain.

If we assume that the integration window of the camera is rectangular (Fig. 3A main text), we can define the relation between the measured bead position  $x_m(t)$  and the true bead position  $x(t)$  as:

$$x_m(t) = \int_{t-W}^t x(t') H(t-t') dt' \quad [18]$$

where  $H(t)$  is defined as:

$$H(t) = \begin{cases} \frac{1}{W} & 0 < t \leq W \\ 0 & \text{Elsewhere} \end{cases} \quad [19]$$

The power spectrum of the moving average window  $P_{window}(\omega)$  is given by:

$$P_{window}(\omega) = \left( \frac{\sin(\omega W / 2)}{\omega W / 2} \right)^2 \quad [20]$$

Convolution of this spectrum with the spectrum of the tethered bead (see main text, equation 10) yields an expression for the measured variance:

$$\text{var}(x_m) = \frac{1}{2\pi} \int P(\omega) P_{window}(\omega) d\omega \quad [21]$$

**Table S1**

Parameter settings	Parameter value
Viscosity	0.001 kg/m*s
Force	0.01-10 pN
Bead radius	1.4 $\mu\text{m}$
Contour length	2.7 $\mu\text{m}$
Persistence length	50 nm
Simulation time step	0.005 ms
Sampling frequency	120 Hz
Shutter time	8.33 ms

**Table S2**

Camera correction	Bead radius ( $\mu\text{m}$ )	Viscosity (kg/m*s)	Sampling frequency (Hz)	Increase in F (error 10%)	Relative increase in F (error 10%)
No	0.5	0.001	120	-	-
Yes	0.5	0.001	120	$1.4 \pm 0.05^a$	-
Yes	0.5	0.001	200	$1.5 \pm 0.03^b$	$0.9 \pm 0.03^b$
Yes	1.4	0.001	120	$2.8 \pm 0.04^b$	$1.0 \pm 0.04^b$
Yes	0.5	0.002	120	$2.0 \pm 0.05^b$	$1.0 \pm 0.05^b$

a: relative to spectral analysis without camera correction

b: relative to spectral analysis with camera correction



# Chapter 7

Nucleoside-dependent  
switching between  
active and passive  
unwinding in a  
nidovirus superfamily  
1 RNA helicase

Aartjan J.W. te Velthuis<sup>1,2</sup>, David  
Dulin<sup>1</sup>, Danny D. Nedialkova<sup>2#</sup>,  
Suzanne S. Hage<sup>1</sup>, Eric J. Snijder<sup>2</sup>,  
and Nynke H. Dekker<sup>1</sup>

<sup>1</sup> Department of Bionanoscience,  
Kavli Institute of Nanoscience,  
Delft University of Technology,  
Lorentzweg 1, 2628 CJ Delft, The  
Netherlands

<sup>2</sup> Department of Medical  
Microbiology, Molecular Virology  
Laboratory, Leiden University  
Medical Center, Albinusdreef 2,  
2333 ZA Leiden, The Netherlands

# Present address: Max  
Planck Institute for Molecular  
Biomedicine, Röntgenstraße 20,  
48149 Münster, Germany

## Abstract

RNA helicases are enzymes that are critically involved in the replication and transcription of RNA viruses. Although most RNA viruses encode so-called superfamily (SF) 1-type helicases, real-time kinetics data is currently only available on the SF2 type helicase NS3 of the hepatitis C virus. It is therefore unknown whether other viral RNA helicases use a comparable mechanism to couple ATP hydrolysis to the unwinding double-stranded helices or employ a different strategy to separate the strands. Here we present a single-molecule approach to study the unwinding activity of nsp10, the SFI type helicase of the nidovirus prototype equine arteritis virus (EAV). We find that nsp10 is able to use all four NTPs for the unwinding of double-stranded templates, but that the use of ATP, GTP and UTP results in high pause frequencies and a strong sequence-dependency, reminiscent of a passive unwinding mechanism. Strikingly, the hydrolysis of CTP makes the enzyme relatively insensitive to the local sequence and less prone to pausing. These results suggest that the mode of chemo-mechanical coupling during duplex unwinding is not a fixed parameter and dependent on the nucleotide present in the active site.

## Introduction

RNA helicases are important agents in the metabolism of nucleic acid and play key roles in transcription, translation, and replication in eukaryotic cells, bacteria and both DNA and RNA viruses [372,373]. In addition, RNA helicases have been implicated in the expansion of viral positive-stranded RNA (+RNA) genomes from a few hundred bases to up to ~31 kilo bases [35,37] and are often strictly required during the RNA virus life cycle, which makes them valuable targets for antiviral therapy [276,372]. Phylogenetic analysis and biochemical studies have shown that particularly the superfamily (SF) 1-type helicase motifs are well-conserved among RNA viruses. Although this type of helicase domain is believed to be structurally and functionally different from other SF types [373,374,375], detailed knowledge of the unwinding dynamics of its RNA viral representatives is currently very limited and mostly based on extrapolation from the well-studied SF2 type helicase NS3 of the hepatitis C virus (HCV) [376,377,378,379].

SF1 type helicases are encoded by a variety of +RNA viruses [84,374], including lethal and economically important pathogens such as the severe acute respiratory syndrome coronavirus (SARS-CoV) and the arterivirus porcine reproductive and respiratory syndrome virus (PRRSV) [65,380]. Both viruses belong to the nidovirus order, which constitutes a group of viruses that is unique in that they employ a discontinuous RNA synthesis process to generate a set of subgenomic mRNAs that is used to express their 3' encoded open reading frames (ORFs). Moreover, they encode a complex set of replicative proteins or non-structural proteins (nsps) to catalyse RNA synthesis [35,58]. Not all enzymatic functions are crucial for nidoviral RNA replication, but those that are, are all conserved in equine arteritis virus (EAV). Aided by its more compact genome, inability to infect humans and robust reverse genetics system, this prototypic arterivirus has become one of the best-studied nidovirus models [58,381].

Of the conserved enzymatic functions in nidoviruses, the RNA helicase domain is the most conserved. Like most helicases, this helicase is assumed to couple ATP hydrolysis to the translocation along one strand of a double helix in order to displace the other strand [108,110]. Biochemical evidence suggests that the nidovirus helicases have a strict 5'-to-3' directionality during unwinding, but it is presently not known how their translocation is coupled to unwinding. Based on extensive theoretical, structural, biochemical and single-molecule studies, it is assumed that helicases generally use either a passive mechanism, in which case the transient fraying of the fork is trapped by the translocating helicase, or an active mechanism that relies on ATP hydrolysis to directly destabilise the hydrogen bonds at the fork [382,383,384,385,386,387,388]. As a consequence, a passive helicase is strongly controlled by the free energy of fork melting ( $\Delta G^\circ_{\text{melt}}$  which depends on the sum of the free energy of hydrogen bond formation and the stacking energy [15,389]) and therefore the GC-content of the template. An active

helicase on the other hand will be relatively insensitive to the nucleic acid sequence that it has to unwind [382].

To gain insight into the unwinding dynamics of the nidovirus RNA helicase and the abilities of SF1 RNA helicases in general, we here show a magnetic tweezers (MT)-based single molecule assay that allowed us to follow the 5'-to-3' strand displacement dynamics of EAV nsp10. Specifically, we used nsp10's ability to unwind RNA and DNA with equal efficiency and tethered a ~1 kb long DNA hairpin between a paramagnetic bead and glass-surface, and monitored the nsp10-driven change in the tether's end-to-end distance. We observed that nsp10 was able to couple hydrolysis of all four RNA nucleotides to the unwinding of this hairpin, but that the unwinding rate and pause frequency was strongly dependent on the nucleotide used. Furthermore, when we related the unwinding rate and pauses to the template sequence, we found that nsp10 was sensitive to the GC-content when hydrolysing ATP, GTP and UTP, but insensitive in the presence of CTP. These results suggest that this type of viral helicase is able to switch between a passive and active mode and that the coupling between NTP hydrolysis, translocation and active fork stabilisation is more flexible than was previously assumed.

## Results and discussion

### The EAV RNA polymerase nsp9 does not have strand-displacement activity

The EAV replication and transcription complex (RTC) is assumed to be anchored to cellular membranes, and to consist of a host factor and at least 13 viral non-structural proteins [313,380]. Similar to many other +RNA viruses, these enzymes include an RNA-dependent RNA polymerase (RdRp, nsp9) and the RNA helicase (nsp10) [390]. However, unlike various other well-studied viral RTCs, such as those of poliovirus or HCV, it is not known whether the RdRp has strand-displacement activity of its own, with the helicase playing merely a supporting role. To investigate this, we developed an EAV nsp9 assay in which we could test whether the processivity of this *de novo*-initiating RdRp [196] could be frustrated by placing a duplex region downstream of the initiation site in its template.

As was previously shown, EAV nsp9 prefers to initiate RNA synthesis on polyU repeats in the presence of [ $\alpha$ - $^{32}$ P]ATP [196,325]. However, when all nucleotides are present in the reaction, only low-molecular weight RNA products are formed on polyU<sub>18</sub> (Fig. S1A), suggesting that polymerisation is inherently limited on this template. To test whether processivity could be stimulated on heteromeric templates, we designed a template (denoted Hel11R) with a 3' 10-nt long polyU initiation site and a downstream 20-nt long heteromeric region with relatively limited secondary structure. As shown in Fig. S1A and S1B, nsp9 was able to synthesise significantly longer products on Hel11R, using either [ $\alpha$ - $^{32}$ P]ATP or [ $\alpha$ - $^{32}$ P]CTP incorporation as readout. Interestingly, we also observed that none of the products was of template length (*i.e.*, a 30-mer) and the [ $\alpha$ - $^{32}$ P]CMP

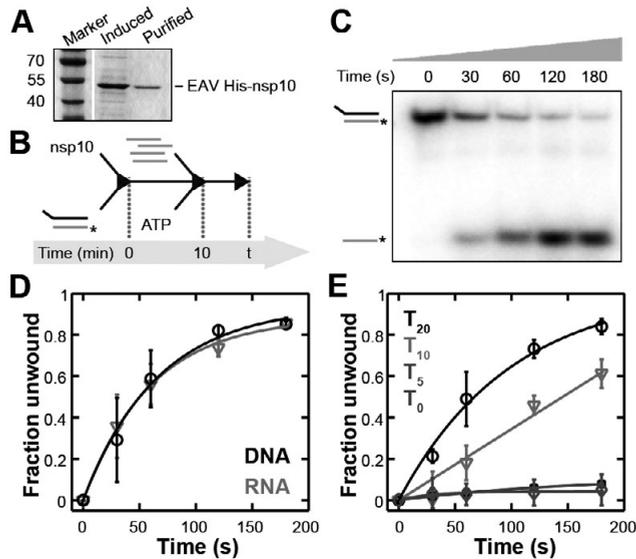
incorporation signal only appeared in >4-mer products, suggesting that EAV nsp9 was initiating internally on the 3' uridine repeat and specifically using the heteromeric region for extension.

To test the influence of dsRNA structures on the EAV nsp9 activity, we hybridised an oligonucleotide that was fully complementary to the heteromeric part of the template (denoted Hel1R). A 5' polyU tail was added to this template to create a forked substrate and rule out that high molecular weight products were formed through primer-dependent extension of the hybridised oligonucleotide. Strikingly, all extension activity was lost in this duplex configuration, whereas initiation products were still present (Fig. S1B and S1C). Similar results were obtained with a non-tailed, but 5' 3'dA-blocked complementary oligo. These results thus lead us to suggest that unlike many other viral RdRps, EAV nsp9 does not have strand displacement activity of its own and would thus have to fully rely on the helicase activity of nsp10 or other (cellular) enzymes with helicase activity during replication.

### Biochemical characteristics of EAV nsp10

Previous studies of the EAV RNA helicase nsp10 were performed with maltose binding protein (MBP)-tagged nsp10 preparations and demonstrated that, like the coronavirus helicase nsp13 [109], EAV nsp10 was able to unwind both RNA and DNA with a 5'-to-3' polarity in the presence of ATP [110]. However, these experiments focused only qualitative rather than quantitative kinetic results, which could have potentially masked differences in the unwinding efficiency. In addition, MBP is a relatively large tag (42 kDa) compared to the 51-kDa nsp10 and could have modulated the dynamics of EAV nsp10. We therefore readdressed the substrate specificity and biochemical characteristics of EAV nsp10 using purified N-terminally his<sub>6</sub>-tagged EAV nsp10 (Fig. 1A).

Using single cycle assays (Fig. 1B), we tested nsp10's ability to displace a 5' <sup>32</sup>P-labelled RNA or DNA oligonucleotide that was part of a dsRNA or dsDNA substrate in the presence of optimal 4 mM Mg<sup>2+</sup> and near-saturating 2 mM ATP concentrations (Fig. S2). Aliquots were taken at various time points and analysed by native PAGE and autoradiography as exemplified in Fig. 1C. In line with previous observations, the unwinding efficiencies on DNA and RNA were essentially comparable and burst rates on DNA and RNA templates ( $k_{DNA}$  and  $k_{RNA}$ ) obtained from single exponential fits to the single-cycle data revealed near-identical values:  $k_{DNA} = 0.015 \pm 0.007 \text{ s}^{-1}$  and  $k_{RNA} = 0.016 \pm 0.004 \text{ s}^{-1}$  (fitted value  $\pm$  95% confidence interval). We next explored the influence of the single-stranded 5' loading region on the initiation of dsDNA unwinding by varying the length of the 5' tail from 0 to 20 nt. In line with observations previously made for SARS-CoV nsp13 and other helicases [391,392], the unwinding rates decreased with a reduction of the tail length (Fig. 1E) and we found that nsp10 required at least a 10-nt single stranded loading platform to display significant unwinding activity.

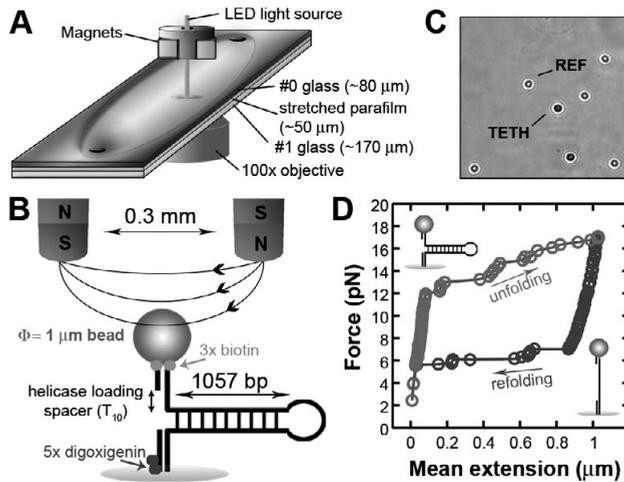


**Figure 1: Template requirements of EAV nsp10.** (A) SDS-PAGE analysis of purified N-terminally His<sub>6</sub>-tagged EAV nsp10. The migration speed of his<sub>6</sub>-nsp10 was in line with its expected molecular weight. (B) Schematic of the steps in the single-cycle assay. First, helicase (nsp10) and template consisting of a 5' <sup>32</sup>P-labelled release strand (asterisk-labelled grey line) and an unlabelled loading strand (black line) were allowed to form a complex, before ATP and the unlabelled trap strand (grey lines) were added to the reaction. Next, samples were taken at time points t and these were resolved by native PAGE. (C) Example time course of nsp10 helicase activity in the presence of 2 mM ATP on dsRNA with a 20-mer overhang. Samples were resolved on 15% native PAGE and analysed by autoradiography. (D) Comparison of nsp10 unwinding activity on dsRNA (light grey triangles) and dsDNA (black circles) in the presence of 2 mM ATP. Both templates had a 20-mer overhang. The dsRNA curve was fit with using Eq. 1 with  $K_{RNA} = 0.015 \pm 0.007$  and  $A = 0.94 \pm 0.06$ . For the dsDNA curve variables were  $k_{DNA} = 0.016 \pm 0.004$  and  $A = 0.88 \pm 0.08$ . (E) To test the dependency of the EAV nsp10 unwinding activity on its ability to load, the 5' single-stranded overhang of the loading strand was varied from a blunt-end (squares), a 5-nt T-tail (lozenges), a 10-nt T-tail (triangles), to a 20-nt T-tail (circles).  $K_{T20}$  was  $0.009 \pm 0.003$  and  $A_{T20}$  0.99. All experiments were performed in the presence of 2 mM ATP.

### Design of magnetic tweezers construct and instrument calibration

To extend our study of EAV nsp10 in more detail and observe its unwinding dynamics in real-time, we employed our knowledge of its template specificity to design a MT experiment. Given the observation that EAV nsp10 unwinds dsDNA and dsRNA with similar efficiency, we opted to use a long DNA hairpin tether instead of a short RNA hairpin. Furthermore, to get insight in the behaviour of this enzyme on a natural sequence, we partly based this hairpin on the EAV genomic 3' end.

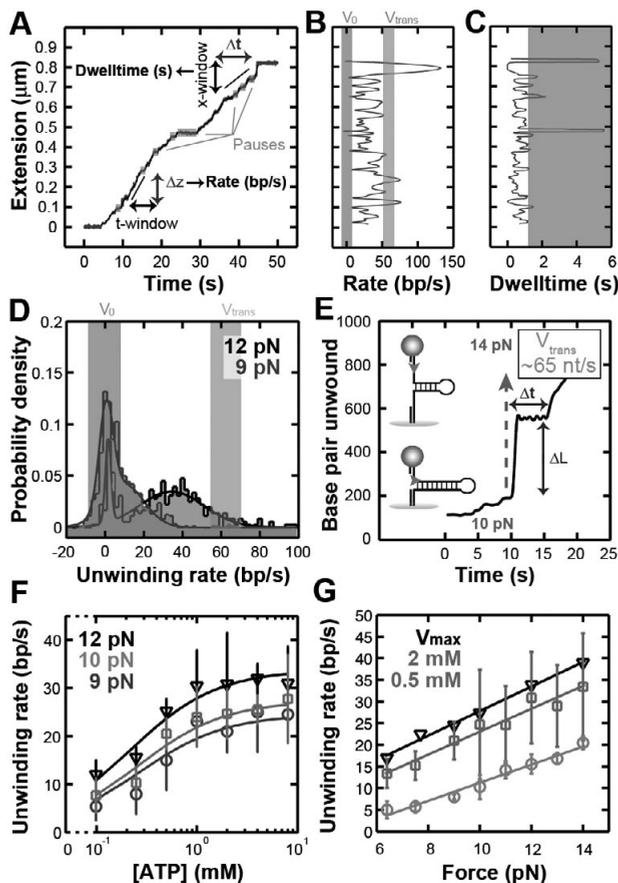
Our MT setup consisted of a glass flow cell (Fig. 2A) in which the DNA hairpin was fixed between a paramagnetic bead and the glass surface through a pair of short DNA handles labelled with biotin and digoxigenin (Fig. 2B). On the 5'-side of the hairpin, a



**Figure 2: Magnetic tweezers setup.** (A) Experimental setup of magnetic tweezers instrument. Two magnets are used to exert a force on paramagnetic beads that are tethered to the bottom surface of a thin flow cell consisting of stretched parafilm sandwiched between #0 glass (top) and #1 glass (bottom). Figure is not to scale. (B) The streptavidin-coated paramagnetic beads of 1  $\mu\text{m}$  in diameter are tethered to the glass surface through a single DNA hairpin and multiple digoxigenin-anti-dig interactions. Permanent magnets are placed in a vertical alignment to achieve a sufficiently strong magnetic field to unwind the hairpin. For details about the hairpin construction, see Fig. S10. (C) Example microscopic field of view with reference beads (REF) and tethered beads (TETH). (D) Mechanical unfolding of the 1057-nt long DNA hairpin is achieved by bringing the magnets closer to the top surface of the top glass slide of the flow cell. The hairpin is fully unfolded at 16-17 pN.

short,  $T_{10}$  ssDNA region was created to provide a minimal platform for EAV nsp10 loading, thus minimising the chance of loading multiple active nsp10 molecules onto the template. Bound tethers were manipulated through a magnetic field created between two horizontally aligned permanent magnets and the bead (Fig. 2A and 2B).

As discussed elsewhere, the camera acquisition frequency ( $f_s$ , here 100 Hz) limits the calibration of the applied force [393], while the Brownian motion of the tethered bead and the intrinsic resolution of the MT instrument essentially limit the resolution at which the tether length can be measured to  $\sim 1$  nm [394]. However, this limit can only be achieved through the use of smaller beads and shorter nucleic acid tethers, which in turn constrain the force range that can be applied to the tether and the confidence with which the magnetic field can be calibrated [393]. In an attempt to optimise all conditions for our experiments, we employed monodisperse 1  $\mu\text{m}$  beads so we could use a calibration of the magnetic field to estimate the applied force within 10% error [393]. In addition, we used a DNA hairpin with relatively short handles to minimise the initial tether length, and a stronger magnetic by reducing the magnet gap-size to 0.3 mm and decreasing the non-accessible range of our flow cell (*i.e.*, the distance between



**Figure 3. dsDNA unwinding by EAV nsp10 in the presence of ATP. (A)** Analysis of the extension traces is performed by a dwell time analysis using a fixed 10-bp sliding window in  $z$  and a rate analysis by applying a fixed 0.9-s (90 points) sliding window in  $t$ . Pauses identified using either of these methods are indicated in dark grey. Experiment was performed at 10 pN force and 2 mM ATP. **(B)** Velocity and **(C)** dwell time analysis of the nsp10 activity along the activity trace. Pauses are indicated in dark grey and the translocation velocity ( $V_{trans}$ ) in light grey. **(D)** Analysis of the instantaneous unwinding rates of EAV nsp10 at 2 mM ATP shows a pronounced peak around  $0 \pm 3.4$  bp/s (mean  $\pm$  sd) for 9 pN (indicated in dark grey) in addition to the unwinding peak centred around 21 bp/s at 9 pN. The unwinding peak at 12 pN is centred around 31 bp/s. The light grey bar indicates  $V_{trans}$ . **(E)** To measure the translocation rate of EAV nsp10, the helicase was first allowed to initiate at a constant force of 10 pN. Then the dsDNA template was quickly converted into a partial ssDNA template by increasing the applied force to 14 pN. At this force, the hairpin forms a stably folded intermediate. The distance covered by the helicase ( $\Delta L$ ) and the time required to do so ( $\Delta t$ ) provide us with a measure of  $V_{trans}$ , which we estimated at  $65 \pm 13$  nt/s. **(F)** Mean unwinding rates as function of the ATP concentration for EAV nsp10 at 9 (circles), 10 (squares), and 12 pN force (triangles). At all forces investigated, the unwinding rate obeys Michaelis-Menten kinetics with  $K_m = 0.24\text{--}0.26$  mM. **(G)** Unwinding rates at 0.5 (circles) and 2 mM ATP (squares) and maximum unwinding velocities ( $V_{max}$ , triangles) obtained from Michaelis-Menten kinetics plotted as function of the applied magnetic force.

the top of the tethered bead and bottom of the magnets, a value that would be zero if we they could physically touch) to 130  $\mu\text{m}$  (Fig. 2A and 2B). As shown in Fig. S3 and S4, these modifications allowed us to apply forces up to  $20 \pm 2$  pN and track reference beads through microscope-based image analysis (Fig. 2C) with a vertical standard deviation of  $\sim 1.4$  nm and tethered molecules with a resolution of  $\sim 3.0$  nm in  $z$  (Fig. S4).

### Single-molecule assay and ATP dependency

Unwinding or mechanical unfolding of the hairpin results in an increase in the distance between the flow cell surface and the magnetic bead. Additionally, a force that is applied perpendicular to the fork junction lowers the free energy that the helicase has to overcome to unwind the duplex with  $\Delta G_f$ . Consequently, at forces where  $\Delta G_f > \Delta G_{\text{melt}}^\circ$  the hairpin will spontaneously unfold. To identify the force at which our 1057-bp hairpin was stable, we first mechanically unfolded the hairpin in absence of helicase (Fig. 2D). Analysis of the mechanical unfolding demonstrated that the hairpin remained stably folded up to forces of  $12 \pm 1$  pN (Fig. 2D). In addition, we observed that over  $16 \pm 2$  pN was required to fully unwind the DNA duplex and that refolding of the hairpin occurred at  $6.4 \pm 1$  pN (Fig. 2D).

When we next added ATP and EAV nsp10 to the flow cell, we observed changes in bead height. The most common event at 9-14 pN and 2 mM ATP was a steady increase in extension up to the entire single-stranded length of the tether (Fig. 3A). No reziping or refolding was observed when the helicase reached this position. Typically, the instantaneous unwinding rate - computed from the slope of a running window along the time axis (Fig. 3A and 3B) - and the dwell time of the helicase fluctuated along the template, and we occasionally observed interruptions or pauses in the extension (Fig. 3). The latter we defined as those positions in the template where the unwinding rate was within  $0 \pm 7$  bp/s (mean  $\pm 2$  standard deviations; Fig. 3D, orange area). At lower forces and reduced ATP concentrations, we also observed a rapid refolding of the hairpin (Fig. S5), likely because the helicase was more prone to dissociation under these conditions. To ensure that we were measuring the kinetics of a single nsp10 helicase, we tested dilutions of the helicase preparations and found consistent behaviour up to the lowest concentration tested, *i.e.*, 2 nM of helicase monomer (Fig. S5E and S5F).

To investigate whether the nsp10 helicase unwinding rate would ultimately approach the translocation rate ( $V_{\text{trans}}$ ), we performed an experiment in which the helicase was first allowed to initiate unwinding at a constant force of  $\sim 10$  pN. Once helicase activity was detected, we rapidly increased the applied force to  $\sim 14$  pN (Fig. 3E), which resulted in a stably folded intermediate that had opened by  $\sim 550$  bp relative to the fork junction of the fully folded hairpin (Fig. 2D). We next recorded the time  $\Delta t$  that the helicase required to travel along the created single stranded template with length  $\Delta L$  and initiate a second unwinding event at the new fork (Fig. 3E). The helicase translocation rate was

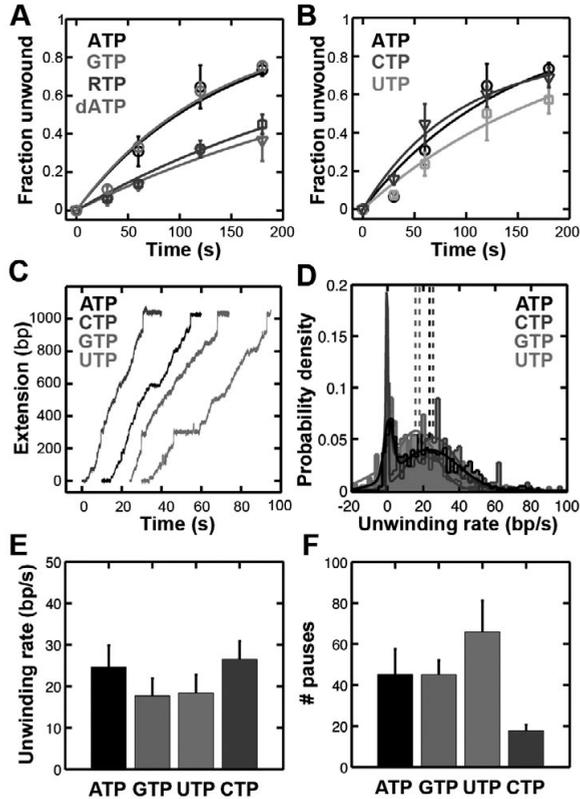
obtained from  $\Delta L / \Delta t$  and found to be  $\sim 65 \pm 13$  nt/s. When we now apply this value to local unwinding rate observed in Fig. 3B, we find that occasionally, the helicase did indeed approach this rate. In addition, we also observe extension locations where the unwinding rate appeared to have exceeded the translocation rate. Closer inspection shows however, that these fast events coincide with abrupt changes in the extension, which suggest that they are derived from short spontaneous unfolding events of the hairpin whose rates (typically  $\sim 1000$  bp/s [395]) got averaged to lower rates by our 0.9 s sliding window.

The direct measurement of the helicase activity on the double-stranded template gives insight into the coordination between ATP hydrolysis and unwinding. Specifically, by varying the ATP concentration ( $[ATP]$ ) we were able to obtain the Michaelis-Menten constant,  $K_m$  of ATP. This value was  $\sim 0.26$  mM and independent on the applied force (Fig. 3F). Furthermore and consistent with the reduction of the energy barrier for duplex unwinding, an increase in the applied force gave a strong linear increase in the unwinding rate over the force range that could be measured (Fig. 3G). This increase was consistent among various ATP concentrations and the  $V_{max}$  obtained for the Michaelis-Menten fits. Interestingly, previous single-molecule studies observed fundamentally different behaviours for the relation between the unwinding rate and the applied force, including no force-dependence at all for the SF2 helicase NS3 of HCV [376] and a non-linear dependence for the hexameric helicases of T7 and T4 [383,395]. Comparable dependencies were found in our dwell time analysis, which used a sliding window of 10 nt over the extension axis (Fig. S6).

### **EAV nsp10 is a promiscuous helicase**

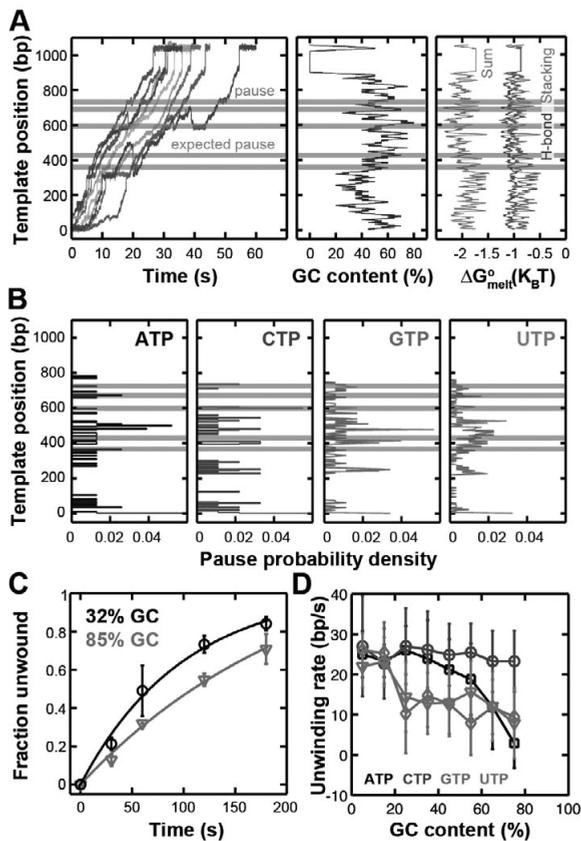
It is known that some helicases are not strictly specific for ATP hydrolysis when they unwind a nucleic acid template [373]. For instance, the hexameric T7 helicase can utilise all dNTPs and NTPs except CTP and GTP in both bulk and single-molecule experiments [396,397]. Interestingly, we found in our bulk experiments that EAV nsp10 was highly promiscuous as well, and able to utilise ATP, dATP and CTP with similar efficiency for unwinding activity (Fig. 4A and 4B). Lower efficiencies were observed when we performed the reaction in the presence of UTP, GTP or ribavirin triphosphate, a ATP and GTP analogue commonly used to treat viral infections [56].

We next focussed on nsp10's ability to use the ribonucleotides in our single-molecule assay. When experiments were conducted in the presence of 2 mM CTP, we essentially found similar mean unwinding velocities as in the presence of 2 mM ATP (Fig. 4C). In line with our bulk experiments, lower unwinding velocities were observed in the presence of UTP and GTP (Fig. 4D). However, processive unwinding was strikingly more often interrupted with pauses (Fig. 4E) and refolding events in the presence of ATP, UTP and GTP compared to CTP. Although we believe that the latter may be attributed to the helicase



**Figure 4. Use of different NTPs modulates the nsp10 dsDNA unwinding rate.** (A) Bulk assays were performed to test nsp10's ability to use different nucleotides for dsRNA unwinding. Clearly, both ATP and dATP could be used with similar efficiency, suggesting that the enzyme does not discriminate between the sugar moiety on the NTPs. Differences were observed for the other purines GTP and RTP at an equal concentration of 2 mM. (B) Nsp10 was also able to efficiently unwind dsRNA in the presence of 2 mM CTP, but showed slightly lower kinetics in the presence of 2 mM UTP. All 4 traces were taken at 10 pN. (C) Single-molecule unwinding traces of EAV nsp10 in the presence of 2 mM of CTP (dark grey), ATP (black), GTP (light grey) and UTP (ultralightgrey). Experiments were performed at 10 pN assisting force. (D) Probability densities of the unwinding velocities of the EAV nsp10 in the presence of 2 mM ATP, CTP, GTP or UTP. Experiments were performed at 10 pN. Dashed lines indicate the centres of the Gaussian fits. (E) On average, we found that CTP and ATP gave similar mean instantaneous unwinding rates at 10 pN, whereas unwinding in the presence of GTP and UTP was lower on average, but not significantly. Error bars represent standard deviations. (F) Analysis of the average number of pauses per unwinding event at 10 pN demonstrated significant differences between unwinding in the presence of CTP and the other three nucleotides.

dissociating from the template, the former may also be related to sequence-dependent pausing and the unwinding mechanism that the helicase uses to separate the strands (see below).



**Figure 5. EAV nsp10 can switch between passive and active nucleic acid unwinding. (A)** Example of aligned traces taken for unwinding at 10 pN and 2 mM CTP (left panel). Identified pauses that corresponded to a high GC content in the template (middle panel) and a correspondingly low  $\Delta G^{\circ}_{\text{melt}}$  are indicated in grey (right panel). The GC content and  $\Delta G^{\circ}_{\text{melt}}$  were calculated for a running window of 10 bp.  $\Delta G^{\circ}_{\text{melt}}$  was calculated as the sum of the hydrogen bonding energy (dark grey) and stacking energy (light grey) [15,389]. **(B)** At certain positions, no pauses were observed in CTP traces (2<sup>nd</sup> panel), but clearly evident in helicase unwinding events obtained in the presence of ATP (1<sup>st</sup> panel), GTP (3<sup>rd</sup> panel) and UTP (4<sup>th</sup> panel). We have marked several of these different events with bars in panels A and B. **(C)** Unwinding efficiency of nsp10 on dsRNA templates containing a low (black circles) or high GC (grey triangles) content. Both experiments were performed in the presence 2 mM ATP. **(D)** Unwinding velocities were associated with their template position and corresponding GC-content over a 10-bp window. Unwinding rates were subsequently binned by their GC-content and fitted with the sum of three Gaussians to extract the mean unwinding rate and standard deviation (see Fig. S8). All experiments were performed under 10 pN assisting force.

### Active and passive unwinding mechanisms

The above results of the fluctuating local unwinding rate and the difference between the translocation velocity on single-stranded nucleic acid should allow us to define whether nsp10 was using a passive or active unwinding mechanism in the presence of ATP. Such an estimate is important for the interpretation of the helicase function, because passive

helicases are typically involved in replication and allow for a tighter coupling between the helicase and the replicative polymerase than active helicases [383,388,395,398,399]. In eukaryotic cells, this strategy may facilitate co-ordination between leading and lagging strand synthesis, while active mechanisms appear to be mostly used for processes that are not strictly dependent on this co-ordination, such as DNA repair [382,400].

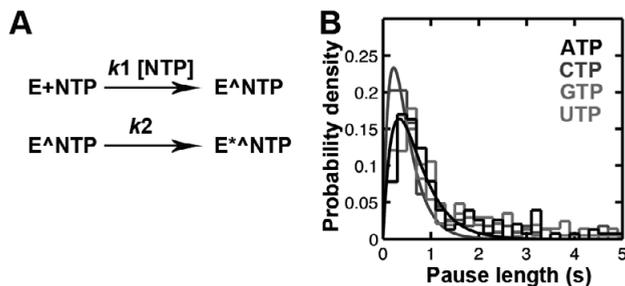
Previous single-molecule studies of helicases have attempted to define the helicase unwinding mechanism by fitting the observed unwinding velocity as a function of  $\Delta G_{\text{melt}}^{\circ}$ , the helicase forward and backward translocation rates  $k^+$  and  $k^-$ , the  $\Delta G_p$ , the helicase step size  $n$  per ATP hydrolysis event, and an helicase-fork interaction potential  $\Delta G_{\text{int}}$  in order to describe the helicase's ability to actively destabilise the fork [386]. Crucially, however, the fitting approach may lead to fundamentally different interpretations of  $\Delta G_{\text{int}}$  depending on the assumed step size and the backward translocation rate [382], and did indeed result in different outcomes for relatively similar enzymes [383,395].

A more robust method is to interpret the helicase unwinding rate as a function of the sequence of the template, *i.e.*, the ratio between the velocity at high GC-content ( $V_{\text{GC}}$ ) and 0 GC-content ( $V_{\text{AT}}$ ) [382]. Specifically, if we follow Manosas *et al.* and set the threshold that separates an active from a passive mechanism at  $\Delta G_{\text{int}} = 1 K_B T$  or about half the averaged  $\Delta G_{\text{melt}}^{\circ}$  (Fig. 5A), we can predict that an active helicase advancing with 1-bp steps and a negligible backward translocation rate will loose at most 75% of its unwinding rate when going from a AT-rich to a GC-rich sequence [382]. Consequently,  $V_{\text{GC}} / V_{\text{AT}}$  will therefore exceed 25% for active helicases, while passive helicases should have a sequence-dependent unwinding ratio that is much smaller than 25%.

### The unwinding mechanism of EAV nsp10 depends on the nucleoside triphosphate used for hydrolysis

To investigate the sequence dependence of EAV nsp10, we first aligned the helicase traces to improve the accuracy in relating the unwinding rate to the local GC content. Typically, we found that we could align the traces by their end point with 1 nm standard deviation (Fig. S8). We next used the pause probability density as function of the template position to get an estimate of the average drift and the accuracy of the alignment along the rest of the template. As shown in Fig. 4B and S8, we were able to identify pauses at GC-rich positions in the middle of the template with a standard deviation of 8 nm, suggesting that the error in sequence identification was  $\sim 5$  nm per 250 nm (approximately 6 bp per 300 bp). This was comparable to the standard deviation of a tethered bead position in absence of helicase over the course of the longest experiments (Fig. S8).

Strikingly, the side-by-side comparison of the pause probability densities of the helicase in the presence of the four NTPs shows that they are not equal Fig. 4B). In general, we observe that the helicase has a probability to pause upstream of GC-rich sequences. However, this behaviour is only consistent for helicase unwinding in the presence of



**Figure 6: Pausing by nsp10 depends on two steps and nsp10 can switch between passive and active unwinding.** (A) Similar to HCV NS3 [376], pauses that occur during template unwinding by EAV nsp10 can be fit with two irreversible kinetics events, one of which may present a NTP-dependent binding step, while the other may present a NTP-independent structural change in the enzyme, *e.g.*, in one of the two RNA binding domain. The first is a pseudo first order reaction involving nucleotide binding ( $k_1$  [NTP]), while the second is NTP independent ( $k_2$ ). (B) Pause duration histograms of nsp10 pause steps (0.2 s bins) in the presence of 2 mM ATP, CTP, GTP or UTP can be fitted with the an equation describing the sum of the two kinetic schemes of Fig. 6A as described previously [376]. Overall, the histograms of ATP, GTP and UTP can be well fit with  $k_1 = 1.5\text{-}1.7 \cdot 10^3 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_2 = 0.6\text{-}0.7 \text{ s}^{-1}$ . The CTP histogram is best fit with  $k_1 = 2.9\text{-}3.0 \cdot 10^3 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_2 = 0.6\text{-}0.7 \text{ s}^{-1}$ , suggesting that the nucleotide-dependent step is discriminating between the nucleotides.

ATP, UTP and GTP, and not in the presence of CTP (Fig. 4B). Also in our bulk experiments do we observe that helicase unwinding is more efficient on templates with a lower GC-content (Fig. 4C). When we next plot the unwinding rate of the helicase as function of the GC-content, we find that although the helicase loses 90% of its velocity when hydrolysing ATP, UTP or GTP on a GC-rich sequence, it is relatively insensitive to the template sequence when using CTP (Fig. 4D, Fig. S8). Consequently, when we apply the 25% threshold proposed by Manosas *et al.*, we find that the helicase is effectively unwinding the template in an active mode in the presence of CTP, while it appears to unwind the template with a passive mechanism in the presence of the other nucleotides.

This observation appears to be supported by the analysis of the helicase pause durations in the presence of the different nucleotides (Fig. 6). Interestingly, we found that not only the unwinding velocity, but also the pause frequency during unwinding events is dependent on the ATP concentration (Fig. S9). This suggests that similar to HCV NS3, the exit from an EAV nsp10 pause is also nucleotide dependent. The shapes of the pause duration histograms are, however, not single exponentials and can be fit if we assume a kinetic mechanism that involves two steps of which only the first is nucleotide dependent (Fig. 6A). If we fit our data with such a two-step model, we find that that all nucleotides have the same nucleotide-independent rate, but that although the ATP-, GTP- and UTP-dependent pauses have the same rate for nucleotide binding, the nucleotide-dependent rate CTP rate is  $\sim 2$  fold higher (Fig. 6B).

## Concluding remarks

The work presented here provides a study of the nucleotide hydrolysis and nucleotide specificity of the SF1 helicase of EAV nsp10 and the coupling of these events to the mechanism that this helicase uses for nucleic acid unwinding. Our study indicates that in the presence of ATP, the most abundant nucleotide in the cell at 1-5 mM [401], nsp10 is dependent on the applied force and using a passive unwinding mechanism to displace the two template strands. In the presence of CTP on the other hand, whose intracellular concentration is ~0.2 mM, nsp10 appeared to unwind the double stranded template using an active mechanism [402].

Given the significantly higher intracellular ATP at homeostasis it appears likely that the default unwinding mechanism of EAV nsp10 is passive. Nevertheless, analysis of the unwinding mechanism of various helicases has so far always assumed that the coupling between hydrolysis, translocation and unwinding is fixed and either passive or active [382,383,386]. Here we demonstrate that this need not be the case and that their pairing may switch depending on the bound nucleotide. Interestingly, it was recently shown that the hexameric T7 helicase processively unwinds dsDNA in the presence of dTTP, whereas ATP-based unwinding induces backsliding as a result of the additional 2'-OH group on the ribose sugar [397]. Structural analysis revealed that this 2'-OH may displace the side-chain of nucleotide-binding residue Y535 and thereby alter the structural dynamics of the helicase and thus its affinity for the template [397]. Such changes are certainly not without precedent, as structural dynamics analyses of RNA polymerases have shown that local perturbations in the structure may induce significant changes in the overall behaviour of the enzyme and, in polymerases, the fidelity of nucleotide incorporation [177].

Structural data is presently unavailable for the nidovirus helicases, which makes it impossible to verify whether amino acid side-chains in the nucleotide-binding pocket could be repositioned upon CTP binding or whether this may affect the structural dynamics of the enzyme. We speculate, however, that such effects could well explain the switch that we observe here and hypothesise that a similar nucleotide-dependent coupling may be present in the active site of other helicases. Further, if local perturbations can influence the unwinding mode of the helicase, this would open up the existence of (co)factors that induce allosteric changes that favour one unwinding mechanism over the other and are thus able to regulate or frustrate the helicase's co-operation with enzymes in, *e.g.*, the replisome or the viral RTC.

## Material and methods

### Cloning, expression and purification of EAV nsp10

The coding region for EAV nsp10 in the Bucyrus genome (Genbank accession number NC\_002532) was amplified by PCR from the EAV cDNA clone [403] using primers E779 and E789 (see Table S1). The PCR product was ligated into the Gateway cloning vector pDONR (Invitrogen) and subsequently subcloned into the T7 driven expression vector pDEST14 (Invitrogen) according to the manufacturer's protocol, thereby creating pDEST14-EAV-His-nsp10.

For expression, BL21(DE3) cells were transformed with pDEST14-EAV-His-nsp10 and grown on Luria Broth (LB) agar plates containing ampicillin (50 µg/ml). Next, 2000 ml LB-5052 [292] containing ampicillin (50 µg/ml) was inoculated with 50% of the colonies on the agar plate and grown to  $OD_{600} > 1.0$  at 37°C. Subsequently, the cells harvested and resuspended in 2000 ml LB containing ampicillin (50 µg/ml) and IPTG (1 mM) and grown at 37°C for a further 3–4 h. Cells were finally harvested by centrifugation and stored at 20°C until protein purification was started.

For purification, bacterial pellets were thawed on ice, resuspended in buffer A (20 mM Tris pH 8.5, 20 mM imidazole, 0.05% Tween-20, 5% glycerol, 5 mM β-mercaptoethanol and EDTA-free protease inhibitor cocktail (Roche)) containing 500 mM NaCl, and lysed by sonication. The supernatant was cleared by ultracentrifugation at 20,000 *g* for 30 min and subsequently incubated with Talon beads (Clontech) for 2 h at 4°C. The beads were washed four times 15 min with 20 volumes of binding buffer. Ultimately, the N-terminally His<sub>6</sub>-tagged helicase was eluted with 150 mM imidazole in buffer A containing 500 mM NaCl and immediately gelfiltered with a supradex 200 column (GE Healthcare) in buffer A containing 300 mM NaCl. Nsp10 containing fractions were subsequently pooled, dialysed, stored and analysed as described previously for SARS-CoV nsp12 [154].

### Polymerase and unwinding assays

EAV nsp9 was purified and analysed as described previously [325]. For EAV nsp10 unwinding assays, a total of 100 nM of purified EAV nsp10 was incubated with 1 nM 5' <sup>32</sup>P-labelled RNA or DNA duplex substrate at 22 °C for 10 minutes in reaction buffer (Tris-HCl pH 7.5, 20 mM NaCl, 0.01 mg/ml BSA, 5% glycerol, 0.01% Triton X100). Prior to the addition of 2 MgCl<sub>2</sub> and 2 ATP, unless stated otherwise, the reactions were cooled down to 15 °C and 1 µM unlabelled release strand added to trap unbound helicase (Fig. 1B). Samples were taken at the indicated time points and quenched with an equal volume of quenching buffer (20 mM Tris-HCl pH 8.0, 50 mM EDTA, 60% glycerol, 0.5% SDS and 0.1% bromophenol blue)

After gel electrophoresis on 15% native PAGE gels, the fraction unwound was analysed via autoradiography. Next, the fraction of unwound nucleic acid was plotted against the

reaction time and fitted to a single-exponential equation as described previously using Eq. 1 [391,404]:

$$F(t) = A(1 - e^{-kt}) + C \quad [\text{Eq. 1}]$$

Here,  $F(t)$  is the fraction of template unwound at time  $t$  in seconds,  $A$  the amount of unwound product,  $C$  the correction for the amount of unwound product present in the beginning of the reaction, and  $k$  the pseudo first order rate constant of nucleic acid unwinding.

### DNA hairpin construction

The hairpin representing the 3' 990 nt of the EAV genome was amplified by PCR from the EAV cDNA clone [403] using primers 513 and 514. This product was digested with KpnI and XhoI, and hybridised to hairpin oligos 515-518 (see table S1 and Fig. S10) in annealing buffer (10mM Tris pH 7.9, 50 mM NaCl, 1 mM DTT, and 10 mM MgCl<sub>2</sub>). After ligation with T4 DNA ligase (New England Biolabs) in annealing buffer with 1 mM ATP, the products were gel purified and labelled with digoxigenin-UMP at their 3' end using Klenow polymerase. Final products were cleaned using YM100 columns (Millipore) and stored at -20 °C in TE (20 mM Tris-HCl pH 8.0, 10 mM EDTA).

### Magnetic tweezers and flow cell configuration

We used a custom built MT setup similar to that developed and described by Strick *et al.* [337], albeit it with technical modifications. Briefly, a CCD camera with a acquisition frequency of 100 Hz (Dalsa) was used to track monodisperse superparamagnetic beads of  $1.0 \pm 0.1 \mu\text{m}$  in diameter (MyOne beads, Life Sciences) tethered to the surface via the hairpin construct ( $\sim 1.0 \mu\text{m}$  fully unzipped, see Fig. 2D). Focusing of the 100x oil-immersion objective (Olympus) was achieved with a pifoc (P-726.1CD, Physik Instrumente) that was fixed onto a 25-mm thick aluminium base plate and encased in a stage (Merzhäuser).

As magnets, we used two gold-plated (Ni-Cu-Ni-Au), 5 x 5 x 5 mm neodymium-iron-boron (NdFeB) permanent magnets (SuperMagne, Germany) that were placed 0.3 mm apart in an aluminium holder. Flow cells were constructed using a #1 nitrocellulose-covered microscope coverslip as bottom slide and a #0 coverslip as top slide (VWR International). As spacer between the two slides we used manually stretched parafilm, which we found to have an average thickness of  $50 \pm 5 \mu\text{m}$ . Overall, this layout resulted in a flow cell thickness of  $\sim 300 \mu\text{m}$  and a minimal magnet-to-surface distance of  $\sim 130 \mu\text{m}$  (Fig. 2B).

### Magnetic tweezers calibration

To be able to correct for mechanical drift during the MT experiments, 1.0  $\mu\text{m}$  polystyrene beads (Polysciences) were baked onto the surface by first adsorbing them to the bottom slide of the flow cell in 10 mM Tris pH 8.0 and 100 mM  $\text{MgCl}_2$  and next fixing them through baking at 100  $^\circ\text{C}$  for 3 min. During measurements, mechanical drift was minimised by correcting the tethered bead height relative to the 1- $\mu\text{m}$  reference bead every 20 frames. Before use, the inner surfaces of the flow cell were passivated with bovine serum albumin (BSA, 10 mg/mL). The extension-to-nucleotide conversion was calibrated as described previously (Fig. S3D) [383,395]. The reference bead fluctuations could be tracked with a standard deviation of  $\sim 1.4$  nm in z, and  $\sim 0.6$  nm in x- and y-directions (Fig. S4). Tethered MyOne beads could be tracked with a standard deviation of  $\sim 3$  nm within the force range of our experiments (Fig. S4). To calibrate the decay of the magnetic field and the applied magnetic force, a 20.7 kbp dsDNA ( $\sim 7.0$   $\mu\text{m}$ ) tethered to MyOne beads was used (Fig. S3). Spectral corrections were performed as described previously [393].

### Single-molecule helicase experiments

Hairpin DNA tethers were prepared by binding 1 ng of hairpin DNA to 2  $\mu\text{l}$  MyOne beads in 10  $\mu\text{l}$  binding buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 100 mM NaCl, 0.01 mg/ml BSA and 0.01% Tween-20) at 20  $^\circ\text{C}$  for 15 min. Tether-bead complexes were taken up in a final volume of 100  $\mu\text{l}$  and introduced into the flow cell at 10 ml/min and incubated for 60 min in absence of the magnetic field. Flow cells were next washed with 0.5 ml helicase buffer (20 mM Tris-HCl pH 7.5, 20 mM NaCl, 4 mM  $\text{MgCl}_2$ , 0.01% Triton X-100 and 0.01 mg/ml BSA) before the introduction of 2mM ATP and 20 nM helicase in helicase buffer, unless noted otherwise for specific experiments.

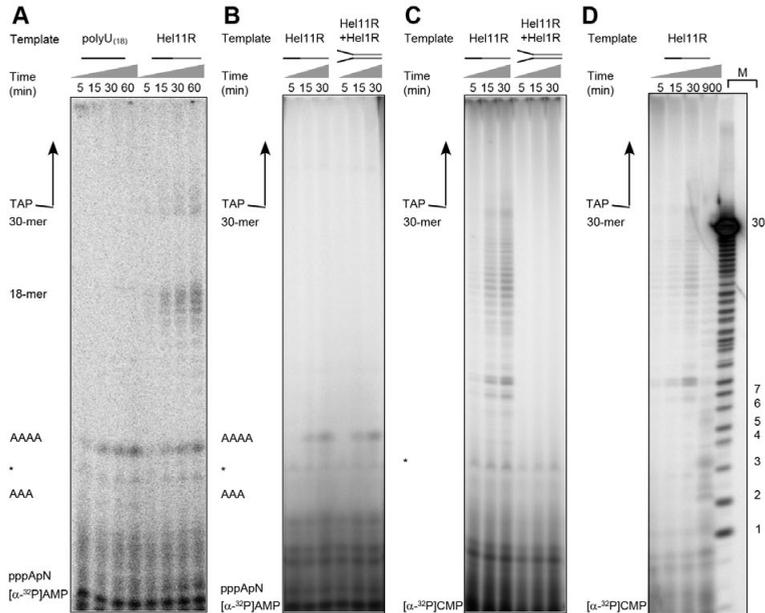
### Data analysis

Helicase unwinding events were identified by thresholding the end-to-end change in z to 2 standard deviations beyond the mean z-position at the beginning of each measurement. To estimate instantaneous helicase unwinding velocities from the end-to-end measurements, traces were first median filtered at 1 Hz and aligned by their fully unwound extension where possible. Slopes were then fitted to the data points in fixed 90-point windows, which were moved by 10 points along the time axis. Pauses and non-zero velocities were identified by fitting unwinding rate histograms of 2-bp/s-bins with the sum of three Gaussians (one for negative rates, one for pauses, and one for positive rates). Pause positions were subsequently extracted by thresholding the velocities at consecutive time axis windows to 2 standard deviations of the zero velocity peak. Dwell times were extracted from the traces by moving a 10-bp window along the extension axis with standard deviation intervals of the fluctuations in the z-position.

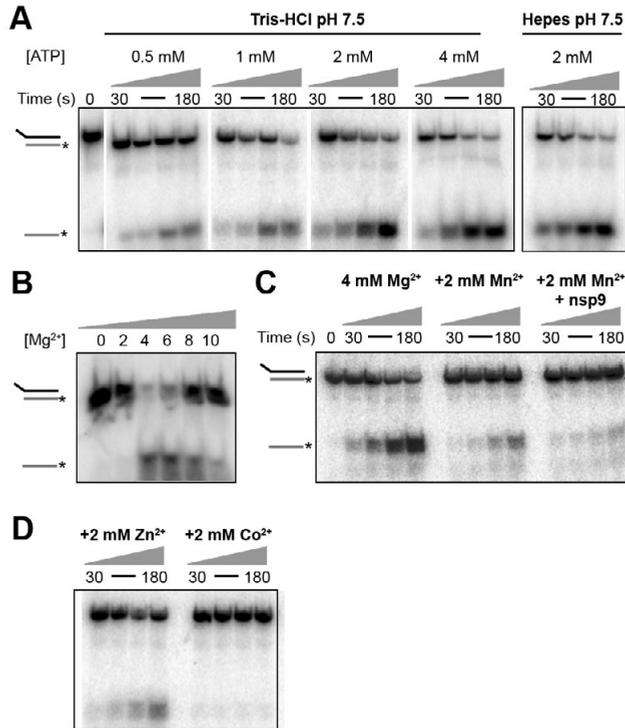
## Acknowledgements

The authors thank Dr. Jan Lipfert, and Dr. Iwijn de Vlamincx for stimulating discussions and assistance with programming, Dr. Jacob Kerssemakers for discussions and his assistance with the development of the flow cell, and Dr. Michelle Spiering and Dr. Stephen Benkovic for sending hairpin DNA for testing. This work was supported by the Netherlands Organization for Scientific Research (NWO) through Toptalent grant 021.001.037 to AV and TOP-GO grant 700.10.352 to ES and ND.

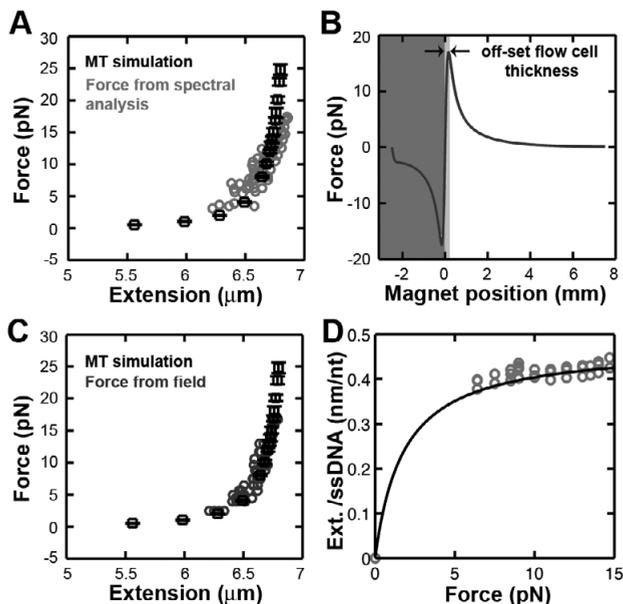
## CHAPTER 7 - SUPPLEMENTAL INFORMATION



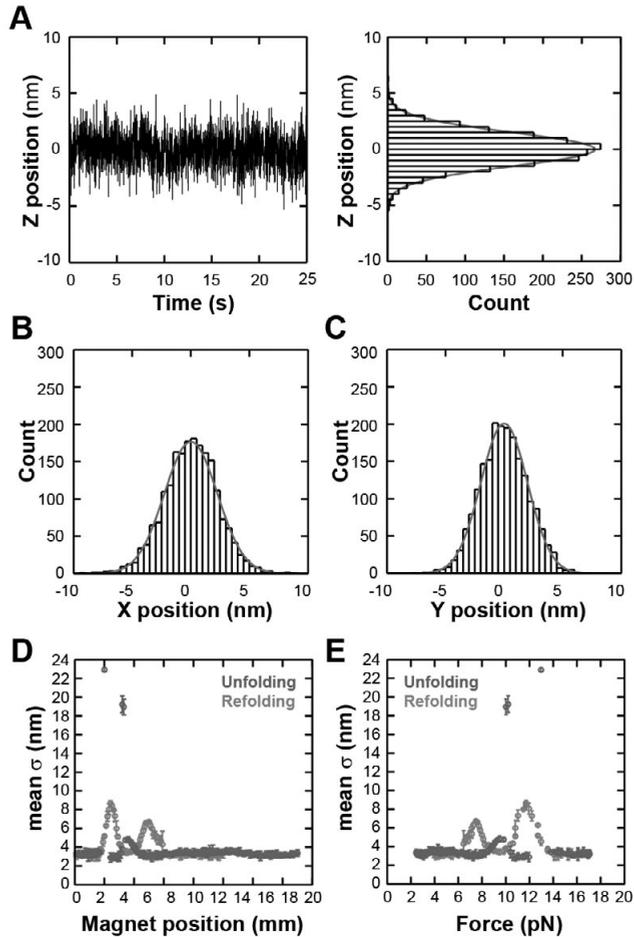
**Figure S1. Activity of EAV nsp9 is influenced by dsRNA.** (A) Purified EAV nsp9 [325] can synthesise oligomeric RNA products in the presence of uridine repeats (black line in template schematic) in the polyU<sub>18</sub> or the Hel11R template, and 0.17  $\mu\text{M}$  [ $\alpha$ -<sup>32</sup>P]ATP, 50  $\mu\text{M}$  ATP and 100  $\mu\text{M}$  CTP, GTP and UTP. (B) Similar oligomeric products are formed in the presence of only [ $\alpha$ -<sup>32</sup>P]ATP, 50  $\mu\text{M}$  ATP and Hel11R (left panel), or [ $\alpha$ -<sup>32</sup>P]ATP, 50  $\mu\text{M}$  ATP and the partially duplex template consisting of oligos Hel11R and Hel1R (right panel). (C) When [ $\alpha$ -<sup>32</sup>P]ATP is replaced with [ $\alpha$ -<sup>32</sup>P]CTP RNA products >5-mers are observed only when nsp9 is incubated with the single stranded Hel11R, but not with the oligo duplex. Together with the results of Fig. S1B, these observations suggest that the duplex specifically inhibits extension. (D) nsp9 polymerase activity on the Hel11R template relative to the NaOH degradation of a 5' <sup>32</sup>p-labelled Hel11R to indicate the expected RNA product length if nsp9 was initiating RNA synthesis on the terminal 3' U of the templates. We note that the size distribution of the smaller 5' <sup>32</sup>p-labelled products should not be compared to the migration pattern of synthesised RNA products (which contain 5' triphosphates) due to their different negative charge and migration velocity in the electric field of the 20% 7M Urea gel.



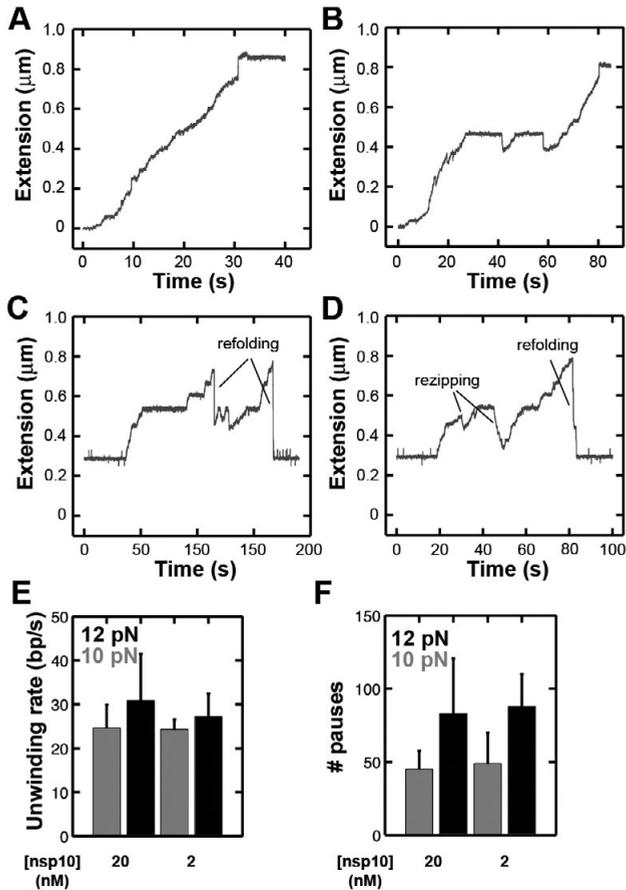
**Figure S2. Biochemical characteristics of EAV nsp10.** (A) Purified nsp10 was incubated with 5' tailed RNA duplexes and assessed for its ability to displace the 5' <sup>32</sup>P-labelled bottom strand (grey and labelled with asterisk) in the presence of various concentrations of ATP, 4 mM MgCl<sub>2</sub> and either a 20mM Tris-HCl or 20 mM Hepes buffer. (B) The ability of nsp10 to unwind dsRNA in the presence of various concentrations of Mg<sup>2+</sup>, 2 mM ATP and 20mM Tris-HCl. (C) When Mn<sup>2+</sup> - required for *de novo* RNA synthesis of EAV nsp9 *in vitro* - was added to helicase reactions containing 2 mM ATP, the ability of nsp10 to unwind the provided dsRNA template was significantly reduced both in the presence and absence of nsp9 polymerase. (D) Similar to the effect of Mn<sup>2+</sup>, the addition of Zn<sup>2+</sup> or Co<sup>2+</sup> could interfere with nsp10's unwinding efficiency in the presence of 2 mM ATP.



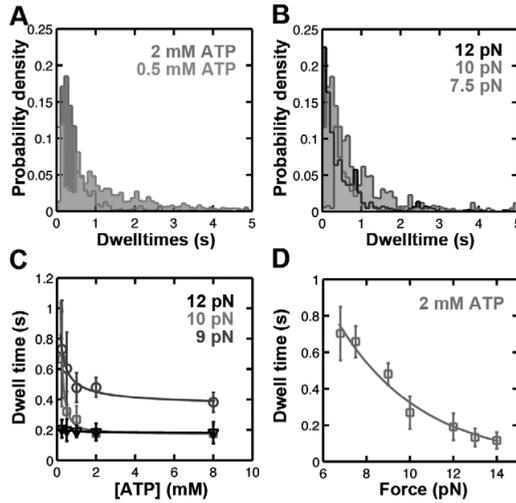
**Figure S3. Force calibration and nm to bp conversion.** (A) It is generally feasible to measure the force exerted by the magnetic field on the DNA molecule by recording the fluctuations of the bead's position and analysing these positions used previously discussed spectral analysis procedures [393]. However, given the short contour length of our 1057-kb hairpin in both the folded and unfolded configuration, the high cut-off frequency associated with this tethered-bead system, and the limited acquisition frequency of our camera (100 Hz) we were not able to perform a reliable direct measurement (*i.e.*, with an error <10%) of the applied magnetic force on these tethered beads. We therefore calibrated the magnetic field and used the low polydispersity (<10%) of the MyOne beads to estimate the applied magnetic force. To calibrate the magnetic field in our flow cell, we tethered 20.7-kb (~7  $\mu\text{m}$ ) dsDNA molecules to MyOne beads and measured their extension as function of the magnet position. The applied magnetic force was then computed using spectral analysis and plotted against the extension for three tethers (light grey circles). For reference, a simulated force-extension curve (including spectral analysis of the simulated data) of a ~7  $\mu\text{m}$  dsDNA tether is shown (black squares). Error bars depict standard deviation of 5 simulations. (B) Magnetic field gradient computed for various magnet positions as described previously [405], using vertical 5x5 mm magnets that were placed 0.3 mm apart and 1  $\mu\text{m}$  MyOne paramagnetic beads. The flow cell thickness is indicated in light grey (C) dsDNA molecules of 21-kb (~7  $\mu\text{m}$ ) were tethered to MyOne beads and we measured their extension as function of the magnet position. Using the field gradient computed in Fig. S3B, we then plotted the magnetic force against the measured extension (dark grey circles). For reference, a simulated force-extension curve of a ~7  $\mu\text{m}$  dsDNA tether is shown (black squares). (D) When the maximum measured change in extension is known for a particular applied force, the tether's extension in nm can be converted into the number unwound base pairs since the maximum end-to-end change is proportional to the full length of the hairpin, *i.e.* 1057 bp. Since every unwound bp releases two nucleotides, this gives us directly the extension per nucleotide change. The relation between the extension-to-nucleotide conversion and the force is best approximated with a polynomial as described previously as secondary structures in the ssDNA may lead to deviations from the freely-jointed chain (FJC) model [383].



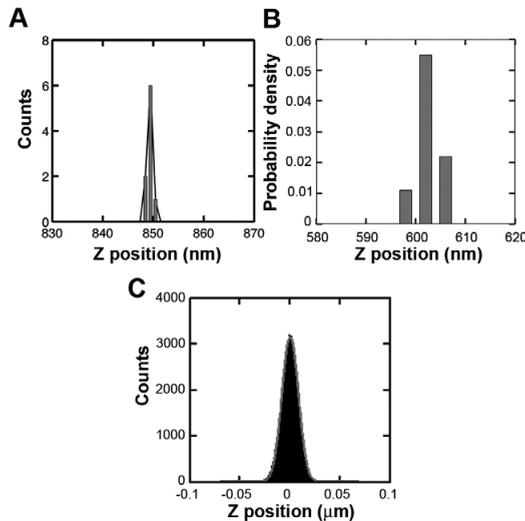
**Figure S4. Tracking resolution of reference and tethered beads.** (A) The left panel shows an example trace of the displacements of a reference bead in Z. The right panel illustrates the distribution of these displacements around the tether point '0'. A single Gaussian was fit to this distribution with a standard deviation of 1.4 nm. (B) Distributions of reference bead displacements in X or (C) Y. Single Gaussians were fit with standard deviations of 0.65 and 0.55 nm, respectively. (D) The tracking resolution for tethered beads plotted against the magnet position. (E) The tracking resolution for tethered beads plotted against the applied force.



**Figure S5. Variations in hairpin unwinding behaviour and refolding.** (A and B) Although the helicase activity of EAV nsp10 predominantly resulted in the complete unwinding of the 1057-nt long hairpin, we also observed (C and D) events in which unwinding was abruptly terminated and the hairpin refolded (annealing of the hairpin after dissociation of the helicase or helicase back-sliding) or rezipped (hairpin closing in the wake of the helicase after template switching). All traces were observed at 2 mM ATP and 10 pN force. (E and F) Although we found that the unwinding velocity was dependent on the force and ATP concentration (Fig. 3), it was not dependent on the nsp10 concentration, suggesting that our measurements were performed under single-molecule conditions.

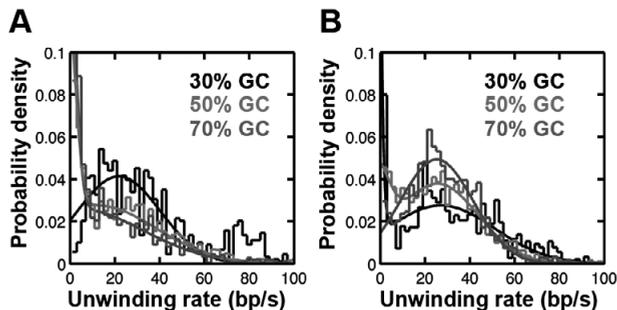


**Figure S6. Dwell time analysis of EAV nsp10 at different ATP concentrations and forces.** (A) EAV nsp10 dwell time distributions for traces obtained at 10 pN force and 2 mM ATP or 0.5 mM ATP, show that the dwell times increase with lower ATP concentrations. Bin sizes were set at 0.1 seconds. (B) Dwell time distributions taken at a constant ATP concentration of 2 mM show that the dwell times are also force dependent. Bin sizes were set at 0.1 second. (C) EAV nsp10 mean dwell times are plotted as a function of the ATP concentration and fitted with a power-law ( $y = ax^b + c$ ). Fit values used were: 12 pN  $a = 0.014$ ,  $b = -0.54$  and  $c = 0.17$ ; 10 pN  $a = 0.056$ ,  $b = -1.56$ , and  $c = 0.17$ ; 9 pN  $a = 0.162$ ,  $b = -0.33$ , and  $c = 0.19$ . (D) Mean dwell times are inversely correlated with the applied magnetic force. Data points from traces taken at 2 mM ATP. The dependency of the dwell times on the force can be well-fitted with a single exponential decay as shown here, or a single polynomial.

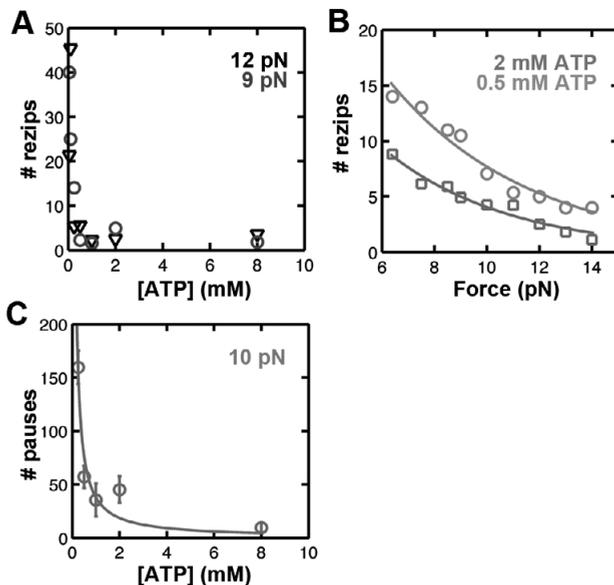


**Figure S7. Alignment of helicase traces.** (A) Alignment of the unwinding events observed on 9 different molecules shows that the standard deviation at the fully unwound position is approximately 1 nm. (B) Due to drift, the resolution falls along the sequence and is reduced by ~5 nm every 250 nm. (C) Histogram of the z-positions of a tethered bead in absence of helicase. This

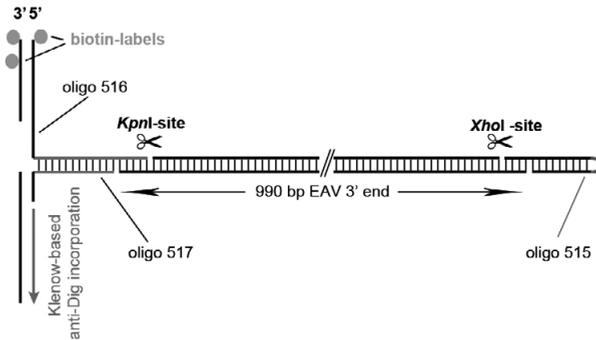
trace was taken for  $\sim 11$  minutes and the z-position distribution can be fit with a single Gaussian distribution with a standard deviation of 11 nm.



**Figure S8. EAV nsp10 unwinding velocity distributions at varying GC-contents.** (A) The unwinding velocities at 2 mM ATP or CTP (B) were binned according to their association with the template position and that position's GC-content. To extract the mean unwinding rate, the unwinding rates were fitted with the sum of three Gaussians. All data was taken at 10 pN force.



**Figure S9. Rezip and pausing frequency of EAV nsp10 in the presence of ATP.** (A) The reipping frequency per unwinding event is dependent on the ATP concentration and (B) the applied force. The latter dependency can be fit with a single exponential decay. (C) If we keep the applied force constant and vary the ATP concentration, we also observe that the pausing frequency is dependent on [ATP]. This dependency can be fit with a single exponential decay.



**Figure S10. Construction of the 1057-kb DNA hairpin.** To assemble the hairpin, the primers listed in table S1 were used in accordance to the method presented in the Material and Methods

**Table S1.** Oligos for hairpin construction.

Oligo	Purpose	Sequence (5' to 3')
E779	Cloning nsp10 fw	CATGCCATGGGCCATCACCATCACCATCACAGTGCCGTGTGCACAGTTGTGG
E789	Cloning nsp10 rv	GGATCTCGAGTTATTGCTTTTCCCAGCCACAGG
514	Fw PCR primer	TGCT <b>GGTACC</b> TACGGCAGCAAAGTCAACC
513	Rv PCR primer	AAACGACGGCCAGTGCCAAGCTCCTC
515	Hairpin loop	P- <b>TCGAGA</b> AAGCGAGCGAAAGCTCGCTTC
516	5'-handle	BIO- AACCAAGTCATTCTGAGAATAGTGTATGCCG-CGACCGAGTTGCTCTTGCCTTTT TTTTTATGCTCTT-TACAACCGGTTGACTGCTTCAGGGGTCGATCCCGCTTT- <b>GTAC</b>
	5'-handle complementary part	ACACTATTCACAGAATGACTTGTT-BIO
517	3'-handle	P- AAAGCGGGATCGACCCCTGAAGCAGTCAACCGTTGTA-AAGAGCATGCCAGATG GTAAGCCCTCCCGTATCGTAGTTATC-TACACGACGGGAGTCAGGCAACTATGGATG AACGA
518	3'-handle complementary part	TGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATC-CATAGTTGCCTGACTCCCGT CGTGTAGATAACTACGATA-CGGGAGGGCTTACCATCTGGC



# Chapter 8

Dual functionality for  
the SARS-CoV nsp(7+8)  
RNA polymerase at  
the genomic 3' end

Aartjan J.W. te Velhuis<sup>1</sup>, Isabelle  
Imbert<sup>2</sup>, Bruno Canard<sup>2</sup>, Alexander  
E. Gorbalenya<sup>1</sup>, Sjoerd H.E. van den  
Worm<sup>1</sup> and Eric J. Snijder<sup>1</sup>

<sup>1</sup> Molecular Virology laboratory,  
Department of Medical  
Microbiology, Center of Infectious  
Diseases, Leiden University  
Medical Center, PO Box 9600,  
2300RC Leiden, The Netherlands.

<sup>2</sup> Architecture et Fonction des  
Macromolécules Biologiques,  
UMR 6098 Centre National de  
la Recherche Scientifique and  
Universités d'Aix-Marseille I et II,  
Case 925, 13288 Marseille Cedex  
9, France.

# Present address: Vaccine and  
Gene Therapy Institute, Oregon  
Health and Science University,  
Beaverton, Oregon, USA

## Abstract

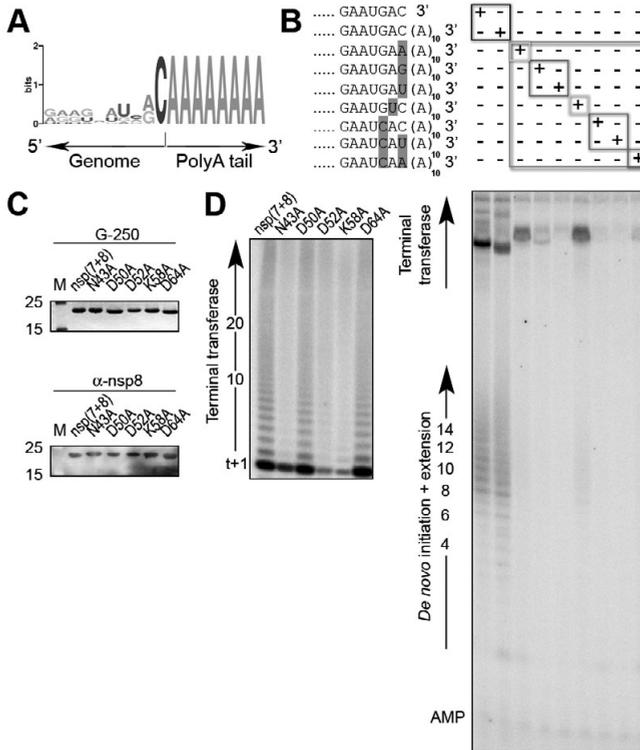
Despite identification of two virus-encoded RNA-dependent RNA polymerases (RdRps) genome, many critical steps in coronavirus RNA synthesis remain to be elucidated at the molecular level. Here, using the SARS-coronavirus model, we describe intriguing biochemical observations suggesting that the initiation of minus-strand RNA synthesis and the polyadenylation of both genome-length and subgenomic mRNAs may depend on a unique platform that involves an interplay of the nsp(7+8) RNA polymerase with a conserved 3'-terminal signature of the viral genome. Specifically, our *in vitro* studies implicate the 3'-terminal GAC<sup>OH</sup>, which can be polyadenylated or used as template for *de novo* initiation of minus strand synthesis. We hypothesise that this specificity is partly achieved through the template-independent formation of a 5'-pppGpU<sup>OH</sup> product that is complementary to this sequence. Lastly we explore here the intrinsic regulation between the two SARS-CoV encoded RdRps and the implications thereof for regulation of their activities.

## Introduction

The initiation of RNA synthesis, *i.e.* the formation of a dinucleotide product through condensation of the first nucleotide bond, is a crucially important function of the viral RNA-dependent RNA polymerases (RdRp) during genome replication. Accordingly, initiation is typically coupled to the recognition of specific cis-acting “promoter elements” in the 3' untranslated region (UTR) of the viral genome in order to secure (i) specificity for the viral promoter in the infected cell, and (ii) negative self-regulation when these RNA sequences were not faithfully copied in the previous cycle [188,197,406,407,408]. So far, a number of initiation mechanisms have been documented that guide RNA viral RdRps to accurate duplication of these important RNA sequences, including i) internal prime-realign initiation [190], ii) *de novo* initiation directed by the ‘priming loop’ of the RdRp and RNA structures [139], and iii) protein-primed initiation [206,207,208]. In addition, non-templated initiation of RNA synthesis has been proposed as a fourth mechanism and is thought to be used by the RdRps of non-segmented negative strand RNA (-RNA) viruses and the dsRNA rotaviruses. Conclusive biochemical support for the latter mechanism, however, is presently still lacking [409,410].

In addition to various replication elements, the genomic 3' UTR contains a polyadenyl (polyA) tail in many positive stranded (+RNA) viruses. Although it is believed that this structure is chiefly involved in regulating the translational efficacy and stability of RNA molecules in the cytoplasm of the cell, it may also contribute to viral replication [183,251,252,411]. In RNA viruses, mRNA polyadenylation is typically based on mechanisms such as i) repeated ‘stuttering’ on a short polyU tract in the viral minus strand [253,412,413], ii) direct copying of a 20 nt long polyU sequence encoded at the 5' end of the minus strand RNA [251,408], or iii) non-templated, terminal extension of the viral mRNA [247,248,250]. Additionally, whereas cellular polyA tails, are mainly synthesised in a template-independent manner by nuclear polyA polymerases (PAPs) following the 3' processing of pre-mRNAs by the nuclear cleavage and polyadenylation specificity factor (CPSF) [414,415] most viral +RNA genomes appear to be polyadenylated at the site of viral replication in the cytoplasm.

In contrast to the above examples, it is completely unknown which initiation or polyadenylation mechanisms are used during the complex RNA synthesis of coronaviruses (CoVs). This +RNA virus group infects a wide range of vertebrates and is renowned for its exceptionally large polycistronic genome of approximately 30 kilobases (Fig. 1A) [65]. Moreover, CoV genome replication must operate in balance with the synthesis of four to nine sg mRNAs [51,265] - crucial molecules that are 5'-capped and 3'-polyadenylated like the genome. They are thought to be produced from complementary subgenome-length -RNA templates that derive from discontinuous negative strand synthesis [51,265,416,417,418,419,420] and it is likely that the synthesis of sg RNAs utilises the same initiation and extension signals as those used for anti-genome synthesis and



**Figure 1: SARS-CoVnsp(7+8) has two distinct activities on the SARS-CoV genomic 3' end.** (A) Conservation of the genomic 3' residues of the CoV genome. For sequence alignment see Fig. S1. (B) To test whether nsp(7+8) could use the SARS-CoV genomic 3' end as template, purified nsp(7+8) was incubated with [ $\alpha$ - $^{32}$ P]ATP as readout, 10  $\mu$ M ATP, 500  $\mu$ M GTP, 100  $\mu$ M of CTP and UTP, and a 111 nt long RNA template that was based on the SARS-CoV genomic 3' end. As controls, side-directed mutagenesis was used to introduce mutations in the 3' end template. Mutations are clustered based on their effect on nsp(7+8) activity. The AMP contaminant present in the radioactive label is indicated as internal loading control. (C) To test whether the terminal transferase activity on the 3' genomic end was specific, mutations were introduced in conserved nsp8 residues. Mutant proteins were purified using the same protocol used for the wild type enzyme and analysed by SDS-PAGE. Upper panel shows Coomassie G-250 staining. Lower panel shows Western-blot analysis with a monoclonal nsp8 antibody. (D) Terminal transferase activity of the nsp8 mutants in the presence of nsp7 relative to the nsp(7+8) RdRp. Mutation of residues D52A and K58A resulted in a  $\sim$ 80% loss of terminal transferase activity. The activity of mutant N43A was reduced by  $\sim$ 30%, while mutants D50A and D64A showed wild type-like activities.

genome polyadenylation. Coronavirus genomes are further characterised by the strict conservation of a 3' terminal cytosine just upstream of the polyA tail (Fig. 1A, Fig. S1). The mechanism and purpose of its conservation are unclear, however, just like it remains to be elucidated how the previously defined RNA sequences and structures in the 3' end of the CoV genome minister control over the initiation of viral minus strand synthesis [334,421,422].

Complicating matters further, current evidence suggests that, uniquely among RNA viruses, CoV RNA synthesis involves not one, but two viral proteins with RdRp activity [154,155,156]. Both these proteins belong to the set of 16 CoV nonstructural proteins (nsps) that are produced through autoproteolytic cleavage of the pp1a and pp1ab precursor polyproteins [44,58] and co-localise with the membrane-associated viral replication machinery in infected cell [68,294]. Presently, these enzymes have only been studied in some detail for the Severe Acute Respiratory Syndrome-associated coronavirus (SARS-CoV) [154,155,156], which emerged in 2003 and caused worldwide concern due to the ~10% mortality rate associated with its infection of humans [65,324]. Of the two CoV RdRps, the first was found to employ a primer-dependent initiation mechanism and to be residing in the conserved canonical RdRp domain of the 106-kDa nsp12 [154,325]. The second on the other hand, the 22-kDa nsp8, was reported to be capable of *de novo* RNA synthesis and primer extension, both alone and in complex with the 10-kDa nsp7, *i.e.* nsp(7+8) [155,156].

In spite of identification of the two CoV RdRps, many fundamental questions regarding CoV RNA synthesis remain unanswered. For instance, it is unclear what viral template is used by nsp8 for *de novo* initiation and whether this includes specific RNA signals in the genomic 3' UTR. In addition, it is unknown whether nsp8's *de novo* activity may be able to support the primer-dependent activity of nsp12. To investigate the RdRp properties of nsp8 in more detail and seek answers to these questions, we here studied the activity of the SARS-CoV nsp(7+8) polymerase and found it capable of both non-templated dinucleotide synthesis and subsequent extension of these products into ~10-nt long RNAs using a template representing the 3' end of the viral genome. Interestingly, 3'-terminal extension of this template with adenylyl moieties was also observed, while mutation of the 5'-GAC<sup>OH</sup> signature specific for the 3' end of the SARS-CoV genome or amino acid substitutions in SARS-CoV nsp8 could abolish or separate the initiation of -RNA strand synthesis and +RNA polyadenylation *in vitro*. Overall, we propose that these results define the CoV nsp(7+8) complex as a unique RNA polymerase that recognises genomic 3'-terminal sequences and plays a critical role in both initiation of minus strand synthesis, the first catalytic step of SARS-CoV replication, as well as the termination of plus-strand synthesis.

## Results

### SARS-CoV nsp(7+8) has two activities on genome 3' end-derived templates

SARS-CoV nsp7 and nsp8 were reported to interact and form the previously crystallised hollow ring structure that is composed of an intricate nsp8 octamer supported by eight copies of nsp7 [153,289]. In addition, the nsp(7+8) complex was shown to have RNA polymerase activity that is catalysed by residues in the nsp8 subunit [155,156]. Previous

biochemical analysis of recombinant SARS-CoV His-nsp8 revealed the enzyme's preference for *de novo* initiation on RNA templates containing a 5'-(G/U)CC<sup>OH</sup> signature [155]. Interestingly, the coronavirus genome contains a fully conserved terminal cytosine as well (Fig. 1A and S1). Together with the compensating mutations identified in nsp8 after mutation of the 3' end of the mouse hepatitis virus genome, this suggests that nsp8 may interact with the genomic 3' end during -RNA synthesis [334].

To explore this hypothesis we incubated purified SARS-CoV nsp(7+8) with 111-nt long transcripts representing the SARS-CoV genomic 3' end. In addition, point mutations were introduced into this RNA sequence to investigate template-dependence and the role of the terminal cytosine in template recognition. Compellingly, nsp(7+8) was readily able to synthesise RNA products of >8 nt in size on templates containing the wild-type genomic sequence, both in the presence or absence of a polyA<sub>10</sub> tail (Fig. 1B). Mutation of the 3'-terminal cytosine to any other nucleoside, however, abolished [ $\alpha$ -<sup>32</sup>P]AMP incorporation into short RNA products (Fig. 1B). Strikingly though, some residual [ $\alpha$ -<sup>32</sup>P]AMP incorporation was still observable for products beyond template length when the 3' cytosine was mutated to adenosine (Fig. 1B). We assumed that this activity was identical to the terminal transferase activity of nsp(7+8) on primed templates [156] and further gel analysis of the products from assays using the wild-type, polyA-tailed template did indeed reveal a classic tailing pattern (Fig. 1D). In addition, the observed products appeared to be longer on mutant templates than on the wild type 3' end, suggesting that the formation of the short RNAs competes with the terminal transferase activity on the 3' genomic end. Similar results were obtained when we mutated the 3' terminal 5'-GAC<sup>OH</sup> to 5'-GUC<sup>OH</sup> (Fig. 1B), whereas a loss of both activities was observed for all other substitutions in the 3' terminal 5'-GAC<sup>OH</sup> sequence (Fig. 1B).

### **Mutational analysis of SARS-CoV nsp(7+8) identifies positions involved in terminal transferase activity**

To verify that the observed terminal transferase-like activity was nsp(7+8) derived, mutations were engineered in nsp8 via site-directed mutagenesis. Mutant proteins were purified in parallel with the wild-type protein and analysed by SDS-PAGE, Western blot and gel filtration (Fig. 1C). When we subsequently investigated the terminal transferase activity of these enzymes, we found that substitution of aspartate-52 to alanine (D52A) and lysine-58 to alanine (K58A) significantly impaired the observed activity (Fig. 1D). Mutation of aspartate-50 and the non-conserved aspartate-64 to alanine (D50A and D64A) on the other hand did not affect the activity at all, whereas alanine substitution of the conserved asparagine-43 resulted in a ~30% loss of activity. In part, these observations are in line with the importance of D52 for nsp(7+8) primer-extension activity and the role of K58 in RNA binding [156].

### SARS-CoV nsp(7+8) terminal transferase activity prefers ATP, but requires GTP for template specificity

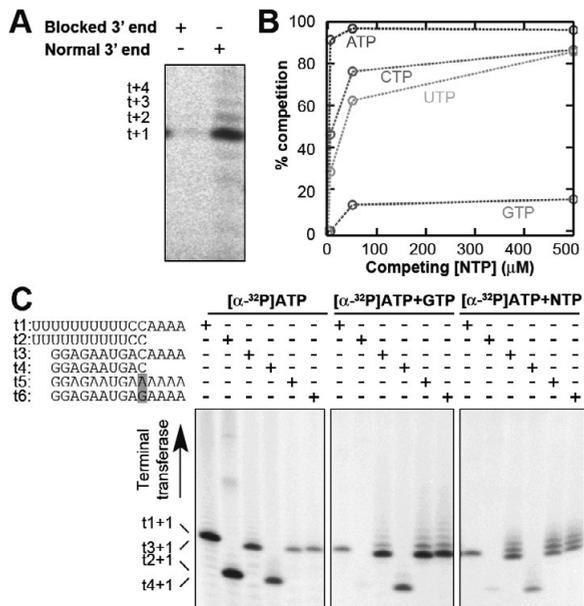
As observed in Fig. 1B and described elsewhere [156], activity reminiscent of terminal transferase activity was observed for SARS-CoV nsp(7+8). To confirm the nature of this activity, an oligonucleotide mimicking the last 10 nucleotides of the SARS-CoV genome was first blocked with cordycepin (3'deoxyadenosine triphosphate, 3'dATP) using a commercially available polyA polymerase and subsequently incubated with purified SARS-CoV nsp(7+8) and [ $\alpha$ - $^{32}$ P]ATP. As shown in Fig. 2A,  $^{32}$ P-labelling of 3'-terminally blocked template did not occur, whereas tailing was still observed for the unblocked control substrate (Fig. 2A).

We next explored the nucleotide preference of SARS-CoV nsp(7+8) for 3' extension of the SARS-CoV genome and titrated ~30-, ~300- or ~3000-fold excess of either ATP, GTP, CTP or UTP in a terminal transferase assay containing 0.17  $\mu$ M [ $\alpha$ - $^{32}$ P] ATP, to assess their ability to compete for incorporation. In line with the preference of poliovirus 3D<sup>pol</sup> [247], unlabelled ATP was the most effective competitor of [ $\alpha$ - $^{32}$ P] ATP, suggesting that nsp(7+8) preferentially synthesises polyA tails (Fig. 2B, Fig. S2). At larger excesses, pyrimidines were also efficiently incorporated into the tails, while semi-efficient GMP incorporation was observed at a 3000-fold excess over [ $\alpha$ - $^{32}$ P]AMP as was evident from the 'G-jumps' in the tailing pattern (Fig. 2B, Fig. S2). This thus leads us to suggest that the order of nucleotide preference for nsp(7+8) is ATP>CTP>UTP>GTP.

Following up on the above finding and our observation that mutation of the 3' sequence of the SARS-CoV genome influenced terminal transferase activity (Fig. 1B), we investigated the relation between the nucleotide substrates present and the template sequence. Interestingly, when nsp(7+8) was incubated with [ $\alpha$ - $^{32}$ P]AMP in absence of other nucleotides, nsp(7+8) preferentially tailed our control oligonucleotide consisting of a polyU tail followed by a 3'-terminal 5'-CCAAAA<sup>OH</sup> sequence and disregarded the templates that were based on the genomic 3' end (Fig. 2C, left panel). This picture completely changed, however, upon the addition of GTP, which significantly shifted the tailing preference to the substrates matching the viral 3'-terminal sequence (Fig. 2C, compare left and middle panel). This shift in template selection also included an equal preference for the mutant 3' end oligonucleotides, in the presence of either GTP only or all four NTPs (Fig. 2C, middle and right panel), suggesting that although GTP plays a role in template selection, nsp(7+8) likely uses additional template-enzyme contacts for fine-tuning and to achieve the specificity observed in Fig. 1B.

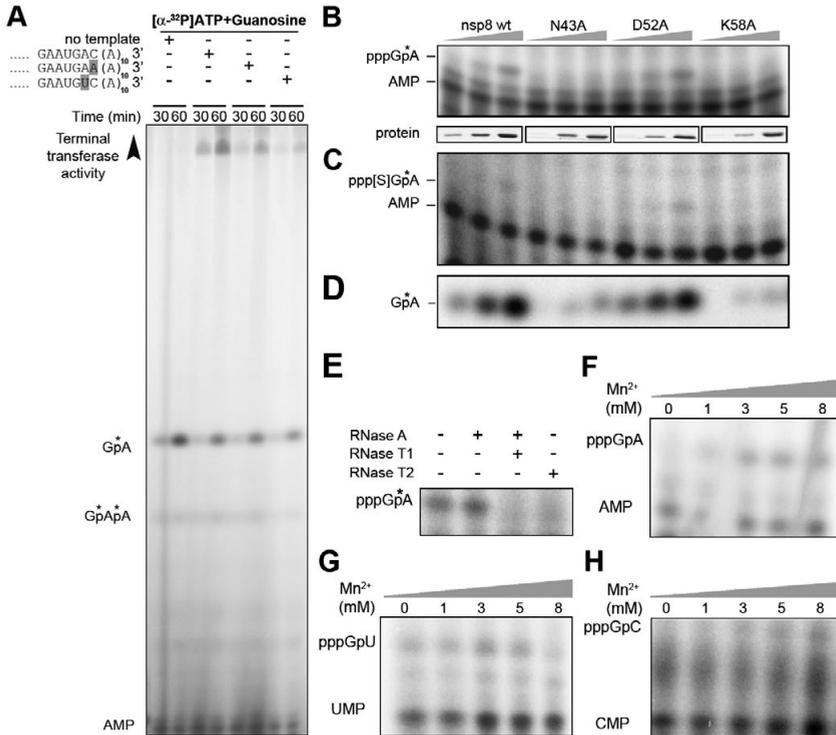
### SARS-CoV nsp(7+8) can synthesise dinucleotides in the absence of an RNA template

To investigate nsp(7+8)'s activity on the 3' end during *de novo* RNA synthesis and to study in particular the first step leading to dinucleotide formation - which was previ-



**Figure 2. Terminal transferase specificity of SARS-CoVnsp(7+8).** (A) No terminal transferase activity of nsp(7+8) was observed when the 3' end of an oligoribonucleotide representing the 3' end of the SARS-CoV genome (oligo t3, see Fig. 2C) was blocked with cordycepin (3'dA) and polyA polymerase. Some residual labelling was visible due to the incomplete blocking of the oligo. (B) Using [α-<sup>32</sup>P]ATP as readout, we tested the competition for other NTPs with the 0.17 μM labelled nucleotide for their use in the terminal transferase activity of nsp(7+8). Ratios were 1:30, 1:300 and 1:3000. For original gel data see Fig. S2. (C) To further explore the requirements for the nsp(7+8) terminal transferase activity, we reduced the RNA template length and removed all nucleotides except the [α-<sup>32</sup>P]ATP readout. Interestingly, this resulted in a general change in terminal transferase specificity (left panel) and a preference for the control templates t1 and t2. The effect could be reversed via the addition of 500 μM GTP alone (middle panel), or through the addition of 500 μM GTP and 100 μM CTP and UTP (right panel). No unlabelled ATP was added to prevent competition for the radioactive signal. All reactions were incubated for 30 min at 30 °C and analysed on 20% 7M Urea PAGE gels.

ously postulated to be pppGpA synthesis [155] - we incubated nsp(7+8) with either wild-type or mutant 3' template, [α-<sup>32</sup>P] ATP and guanosine. This latter component was added to replace GTP in the reaction and to exclude the incorporation of GMP at the +2 position. Strikingly, dinucleotide products were observed in all reactions and their formation appeared to be unabated by mutation of the penultimate 3' C or substitution of the upstream A with U (Fig. 3A). The nps(7+8)-dependent terminal transferase activity was affected, however, in line with the results presented in Fig. 1B. These results therefore suggested that nsp(7+8) was able to identify the template correctly under these conditions and that dinucleotide synthesis had taken place in a template-independent manner. Indeed, GpA formation was also observed in the absence of an RNA template (Fig. 3A).



**Figure 3. Initiation of RNA synthesis by SARS-CoV nsp(7+8) requires GTP and is non-templated.**

(A) To test whether GTP was used for primer synthesis nsp(7+8), we incubated purified nsp(7+8) with 111 nt long SARS-CoV 3' end templates, 500  $\mu$ M guanosine and [ $\alpha$ - $^{32}$ P]ATP. The guanosine was used to better separate the formed dinucleotide from the unincorporated label, distinguish it from putative pppApA products and prevent nsp(7+8) from forming pppGpG products. Interestingly, a GpA dinucleotide was also formed in absence of template. Asterisk indicates position of the [ $\alpha$ - $^{32}$ P]-group. The AMP contaminant present in the radioactive label is indicated as internal loading control. (B) Control reactions showing that dinucleotides can be efficiently formed using GTP, (C) alpha-S-GTP or (D) guanosine at the +1 position in absence of template. In addition, these figures show that the activity is nsp8-specific using nsp8 mutant N43A, D52A and K58A. Reactions contained 0.1, 0.5 or 1  $\mu$ M of wild-type nsp8, mutant D52A or mutant K58A. Mutation of conserved residues N43 and K58 to alanine results in a complete loss (<1%) of *de novo* dinucleotide synthesis in the presence of GTP. Similar to the terminal transferase assay, the activity of D52A was not significantly altered in comparison to the wild-type protein. Lower panels show the SDS-PAGE analysis of the diluted protein stocks of wild-type or mutant nsp8 used in the assay. (E) Synthesis of the pppGpA dinucleotide product by nsp8 in the presence or absence of RNases A, T1 and T2. No template was added to these reactions. (F) Comparative analysis of [ $\alpha$ - $^{32}$ P]ATP, (G) [ $\alpha$ - $^{32}$ P]UTP, and (H) [ $\alpha$ - $^{32}$ P]CTP incorporation at the +2 site under different Mn<sup>2+</sup> concentrations. The ANP contaminants present in the radioactive labels are indicated as internal loading controls.

To exclude that this effect was the result of a contamination or the absence of the triphosphate on the guanosine, we performed dinucleotide assays in the presence of GTP, the non-hydrolysable GTP analogue GTP-alpha-S (guanosine 5'-[alpha-thiotriphosphate];

ppp[S]G) and guanosine, and compared the *de novo* activity of wild-type nsp(7+8) and nsp8 mutants N43A, D52A and K58A. As shown in Fig. 3B and 3C, nsp(7+8) was able to synthesise equal amounts of pppGpA and ppp[S]GpA dinucleotide products, which not only confirmed the important role of the guanosine at the +1 position, but also suggested that it was not significantly competing with [ $\alpha$ - $^{32}$ P]ATP for incorporation at the +2 site. In addition, this result also implied that hydrolysis of the GTP's triphosphate was not required to drive the reaction, which is in contrast to its role as energy donor for the covalent uridylylation of the poliovirus VPg protein primer [206]. Of all nsp8 mutations tested, K58A had the strongest effect on dinucleotide synthesis in all control reactions, thus confirming the specificity of the assay (Fig. 3B-D). Mutation D52A was not able to impair formation of pppGpA, in stark contrast to its role in the terminal transferase assay (Fig. 1).

To exclude that the observed reactions were templated by inadvertently co-purified RNAs, we supplemented the template-free dinucleotide reactions with RNase A (Fig. 3E). Known for its ability to hydrolyse RNA 3' of C and U residues, this RNase could thus specifically degrade any RNAs containing sequences that may have templated pppGpA synthesis. However, the presence of RNase A did clearly not affect the synthesis of a pppGpA product at all, whereas all product formation was lost in negative control reactions that included RNase T1 or T2 to cleave ssRNA products 3' of G or N, respectively (Fig. 3F). These results thus suggest that dinucleotide synthesis by nsp(7+8) is insensitive to both the absence of template and the presence of nucleases capable of degrading all RNA templates complementary to the product. Moreover, together with our other observations, they provide a strong indication that the observed pppGpA dinucleotide product was synthesised *de novo* without the use of an RNA template.

### **SARS-CoV nsp(7+8) preferentially synthesises a pppGpU dinucleotide**

Up to this point, all dinucleotide activity experiments presented in Fig. 3 had been performed in the presence of  $Mn^{2+}$ , given the reported dependence of His-nsp8's *de novo* activity on this cation [155]. In light of the effect of  $Mn^{2+}$  on the SARS-CoV nsp12 and poliovirus 3D<sup>pol</sup> fidelity [154,279], we tested whether the presence of  $Mn^{2+}$  had the ability to alter the nucleotide specificity for nucleotide incorporation at the +2 position. To this end, we titrated  $Mn^{2+}$  in dinucleotide synthesis assays containing either [ $\alpha$ - $^{32}$ P]AMP, [ $\alpha$ - $^{32}$ P]UMP or [ $\alpha$ - $^{32}$ P]CMP as readout. Interestingly, the incorporation of both AMP and CMP required millimolar concentrations of  $Mn^{2+}$  in addition to  $Mg^{2+}$  (Fig. 3G and 3H), whereas the incorporation of UMP only required  $Mg^{2+}$  (Fig. 3I). This thus leads us to suggest that the preferred nsp(7+8) dinucleotide product is pppGpU, which is strikingly complementary to the 3'-terminal sequence immediately upstream of the polyA tail in the genome of SARS-CoV and all closely related betacoronaviruses (Fig. S1). In addition, this result explains the absence of a significant dinucleotide product in the reactions

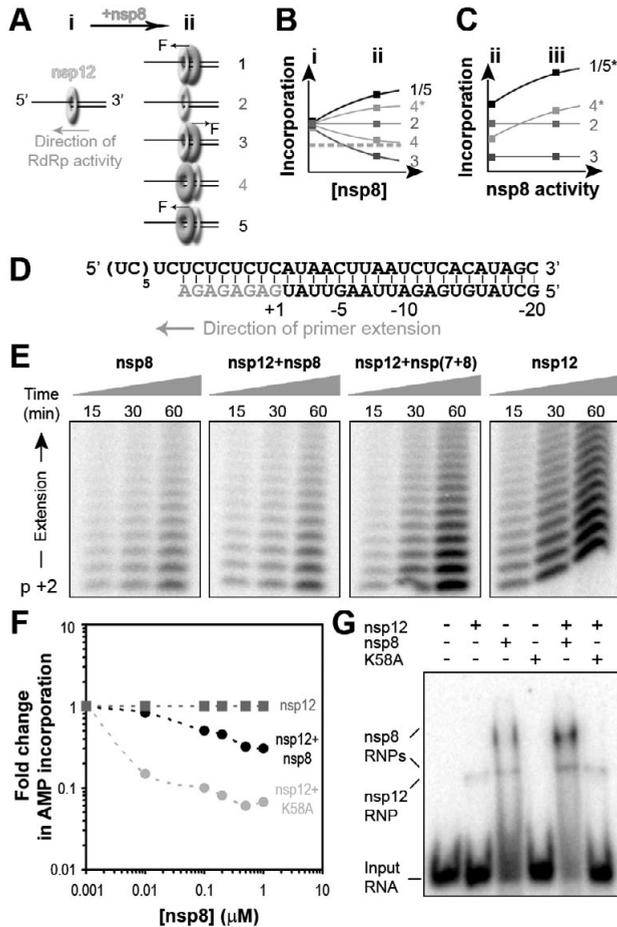
presented in Fig. 1B, as these were visualised through [ $\alpha$ - $^{32}$ P]AMP incorporation rather than [ $\alpha$ - $^{32}$ P]UMP.

### A two-way regulation of SARS-CoV nsp12 and nsp8 activity

With the previously proposed primase hypothesis in mind [155], we next sought to investigate whether the products produced by nsp(7+8) under the conditions in Fig. 1 could be extended by nsp12. Unfortunately, however, no change in the position of the products or product accumulation was observed, both when nsp12 was added after a preincubation with nsp8 or nsp(7+8) and when nsp12 was present from the start. Given that three studies have independently shown that SARS-CoV nsp12 and nsp8 are able to interact [85,289,423], we hypothesised that one of the enzymes was able to modulate the activity of the other and prevent the reaction from progressing. To verify the existence of such an effect, we designed an experiment to model the hypothesised nascent strand hand-off between nsp(7+8) and nsp12 (Fig. 4). Specifically, we employed a primed-template that consisted of a duplex derived from an internal sequence of the 3' UTR of the genome [325] and a single-stranded template sequence with low secondary structure (Fig. 4D) on which both enzymes were active (Fig. S3) [154,325].

To minimise the number of potential outcomes of this experiment and simplify interpretation we decided to fix the nsp12 concentration and vary the amount of nsp8 or nsp(7+8) added to the reaction. As presented in Fig. 4A and 4B, putative outcomes were likely to be qualitatively similar (multiple were either inhibiting or stimulating), but given that nsp12 is a stronger polymerase and has a lower RNA binding affinity than nsp8 and nsp(7+8) [156], we expected the differences in signal to be significant enough to postulate whether nsp8 inhibited nsp12, nsp12 inhibited nsp8, or whether the inhibition was reciprocal (Fig. 4A and 4B). In addition, we hoped to achieve further discrimination between the outcomes by performing a titration experiment under conditions where nsp8 had gained or lost activity, since this could either negate part of the effect or have no effect at, depending on whether the influence of the two enzymes was mediated through the template or via a direct protein-protein interaction (compare Fig. 4B and C).

Interestingly, when a fixed concentration of 0.1  $\mu$ M wild-type nsp12 was titrated with either wild-type nsp8 or the nsp8 template-binding mutant K58A in the presence of an excess of RNA template and NTPs, we observed a steep drop in the overall polymerase activity (Fig. 4E and F). The inhibitory effect of K58A was significantly stronger than that of wild-type nsp8, however, and essentially resulted in a reduction to the lower detection limits of the assay. A smaller effect on the initial nsp12 activity was observed when nsp(7+8) was added to the reaction (Fig. 4E panel 3, Fig. S3). To confirm that the highest concentrations used did not result in competition for template binding, we analysed the reactions by native PAGE. As shown in Fig. 4G, mutant K58A was unable to bind substrate and inhibit nsp12 template binding, whereas the presence of wild-type nsp8



**Figure 4: SARS-CoV nsp8 inhibits nsp12-dependent primer-extension activity.** (A) Addition of nsp8 to a model primer-extension reaction by the main SARS-CoV RdRp nsp12 may result in four potential scenarios: 1) nsp8 has a stimulatory effect on nsp12 (indicated by F in the direction of RNA synthesis) through binding to nsp12, while we assume that the observed activity is dependent on nsp12's interaction with the primer; 2) nsp8 has no effect and is outcompeted by nsp12 for access to the nascent strand; 3) nsp8 has an inhibitory effect on nsp12 (indicated by F in the direction antipodal to RNA synthesis) by binding to nsp12, while catalysis is performed by nsp12; 4) nsp8 is active rather than nsp12, prevent nsp12 from accessing the 3'OH of the nascent strand and slows down the reaction as a results of its lower activity; 5) nsp8 is active rather than nsp12 and is stimulated by nsp12 through an allosteric effect, indicated with F in the direction of RNA synthesis. The dark grey dashed line indicates the basic nsp8 activity level. (B) The addition of nsp8 to nsp12 results in an overall activity that is one of five hypothetical functions of the nsp8 concentration. The effect indicated by 4\* is presented to portray the difference with Fig. 4C. (C) In theory, the combined activity of nsp12 and nsp8 can also be studied as a function of the nsp8 activity, while keeping the nsp8 concentration constant. (D) RNA template used to study the primer extension activity of nsp12 and nsp8. (E) Side-by-side comparison of the incorporation activities of nsp8 alone, nsp12+nsp8, nsp12+nsp(7+8), and nsp12 alone in time. (F) Average change in AMP incorporation. In the presence of nsp8, we observe a decrease in the total activity (black circles).

**Figure 4** continued

A much stronger decrease in activity was observed when we combined nsp12-RdRp with nsp8 template binding mutant K58A (grey circles). **(G)** Electromobility shift assays showing that the presence of nsp8 K58A does not negatively affect nsp12 template binding. In fact, when both nsp8 and nsp12 were present the total amount of bound template was higher. This was particularly noticeable at the height of the upper nsp8 RNP complex.

resulted in an overall increase in template binding that was not only additive due to the fact that both enzymes bound RNA, but also evident as a ~2 fold enhanced complex formation with the largest wild-type nsp8 complex (Fig. 4G). Given that the smaller migrating nsp8 complex runs at a similar position as the nsp12 enzyme (106 kDa), it is likely that the smaller complex represents the nsp8 tetramer, while the larger complex would represent the nsp8 octamer (8x 22 kDa) bound to RNA [153,156].

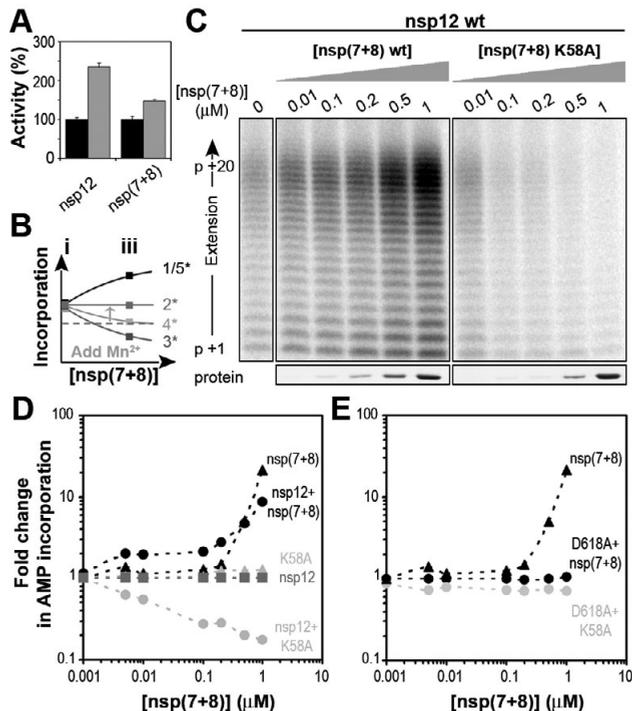
**Increased nsp(7+8) activity overcomes nsp12 inhibition**

To further untwine the interplay between both enzymes and differentiate between the possible interactions of nsp8 and nsp12, we aimed to stimulate the activity of nsp(7+8) relative to nsp12 (Fig. 4C). Specifically, we hoped to find whether the effect of nsp8 was interaction- or activity-dependent. Presently, however, no chemical has been identified that can specifically inhibit or stimulate one SARS-CoV RdRp relative to the other. We therefore attempted to raise the overall activity level of the assay through the addition of  $Mn^{2+}$  and a reduction of the  $Mg^{2+}$  concentration, thereby stimulating the activity of both nsp12 and nsp8 (Fig 5A). To negate the effect of these conditions on incorporation fidelity, we simplified the template to  $U_{20}$ . Both enzymes are active on this template as shown in Fig. S3 and elsewhere [154,156].

As shown in Fig. 5, the more active nsp(7+8) now stimulated the overall incorporation level of the assay, while the inactive mutant K58A showed an unaltered inhibitory effect in line with its inability to benefit from the stimulating reaction conditions (Fig. 5C and D). Lastly and as a further control measure, we also titrated the nsp12 active site mutant D618A in reactions containing either wild type nsp8 or nsp8 mutant K58A. As shown in Fig. 5E, nsp12 was able to fully inhibit nsp8 activity and thus essentially behaved as a permanent 'brake' on the nsp8 activity (Fig. 5E). Overall, these results thus lead us to conclude that on the one hand, nsp8 must be able to exert an influence on nsp12 via a direct interaction given the effect of the template-binding mutant, but that the interplay is also regulated by the interplay on the template on the other hand.

**Discussion**

The complex replication and transcription process that coronaviruses initiate upon infection involves up to sixteen viral nsps and at least one host factor [51,79,290]. In spite of increased effort to untwine their interactions since the outbreak of SARS in 2003, the



**Figure 5: Stimulated SARS-CoV nsp8 can stimulate nsp12 in a concentration-dependent manner, whereas inactive nsp12 blocks nsp8 activity.** (A) To stimulate the activity of SARS-CoV nsp8 we reduced the Mg<sup>2+</sup> concentration from 6 mM Mg<sup>2+</sup> (black bars) to 4 mM and added 1 mM Mn<sup>2+</sup> (green bars). This also resulted in a larger increase in activity of nsp12, albeit at a previously observed loss of fidelity. (B) The increase in nsp12-RdRp activity due to Mn<sup>2+</sup> raised the basal nsp12 activity level in the assay (green arrow). Furthermore, using the higher nsp8 activity we were now able to explore its corresponding effects as proposed in Fig. 4B and 4C by varying the nsp8 concentration. (C) Fixed amounts of 0.1 μM nsp12-RdRp and 1 μM of U<sub>20</sub> template were titrated with either wild-type nsp8 or nsp8 template-binding mutant K58A. Both nsp8 and nsp(7+8) were active on this template (see Fig. S3). Clearly, wild type nsp8 now stimulates the total RdRp activity, whereas the K58A mutant, which cannot benefit from the stimulating reaction conditions, still inhibits activity. (D) Quantification of the nucleotide incorporation activity relative to the activity of nsp12 alone (red squares). In absence of nsp12, the nsp8 activity increases  $21 \pm 0.5$  fold over the course of the curve (black triangles), whereas in the presence of nsp12, this increase is only  $8 \pm 1$  fold ( $n = 3$ ) (black circles). Addition of nsp8 mutant K58A reduces the total AMP incorporation  $9 \pm 1$  fold (grey circles). (E) To test whether nsp8-dependent nucleotide incorporation was still observed in the presence of inactive nsp12, we titrated a fixed amount of nsp12 D618A with either wild type nsp8 (black circles) or nsp8 mutant K58A (grey circles). In both cases no activity was observed. For reference, the activity of wild type nsp8 in absence of nsp12 is shown as black triangles.

exact mechanism that these enzymes use to catalyse RNA synthesis in the CoV RTC is still poorly understood. Moreover, the replication strategy is likely more complex than proposed for other +RNA viruses, since, uniquely among RNA viruses, CoVs employ two viral RdRps to catalyse CoV RNA synthesis [154,155,156]. The prevailing hypothesis to explain

this unique strategy suggests that these enzymes are functioning closely together and form a primase-replicative polymerase pair [155], putatively similar to the eukaryotic replisome, while one alternative hypothesis suggests that their function is separable and that their activity may be strictly regulated in space by other viral proteins [156].

Following up on the description of the polymerase activity of nsp8 and nsp(7+8) [155,156], we here demonstrate that the nsp(7+8) polymerase is capable of recognising the genomic 3' end in order to perform two distinct activities: *de novo* RNA synthesis and terminal transferase activity (Fig. 1B and 6A). We further show that this enzyme can preferentially synthesise a pppGpU product (Fig. 3), which is complementary to the 3' terminal 5'-GAC<sup>OH</sup> of the SARS-CoV genome and may assist in recognition of the template and regulation of the two activities (Fig. 1 and 2). Moreover and significant for our understanding of CoV RNA synthesis, the recognition of the genomic 3' end facilitated the synthesis of 8-14 nt short RNA products (Fig. 1B), which is in line with previous observations implicating 3' terminal sequences and their interaction with nsp8 in the initiation of CoV replication [334,424]. Subsequent mutational analysis of the conserved 3' terminal 5'-GAC<sup>OH</sup> signature abolished the synthesis of these RNAs (Fig. 1B), similar to the observed effect of the mutational analysis of the 3'-terminal 5'-CC<sup>OH</sup> for rotavirus and arterivirus RNA synthesis *in vitro* and *in vivo* [196,407].

Together, these findings suggest that although nsp(7+8) can initiate nucleotide condensation in the absence of an RNA template, it likely prefers an RNA template with a sequence complementary to the pre-formed dinucleotide to catalyse further incorporations (Fig. 6A). We assume therefore that the initiation of -RNA synthesis depends on the correct substrate stoichiometry as well as base pairing and interactions with secondary structures in the template, as also suggested by Züst *et al.* [334]. Future experiments should investigate these steps in more detail. Of particular interest in that regard is the mutational analysis of the nsp8 subunit presented here. As shown in Fig. 1D, this analysis revealed that contrary to D50's involvement in primer-extension [156], this aspartate was not required for terminal transferase activity. Similarly, the for primer-extension crucial D52 was not involved in pppGpA formation (Fig. 1D) [156]. These observations suggest that the various activities noted for nsp(7+8) thus far can be separated *in vitro*, which, for instance, opens up avenues to complementation experiments.

In addition, we also investigated the interplay between nsp8 or nsp(7+8) and nsp12 on a primed template, and observed in Fig. 4 and Fig. 5 that nsp8 has an inhibitory effect on nsp12 activity. Specifically, we found that this effect was dependent on the nsp(7+8) concentration and activity, suggesting that i) nsp8 is allosterically limiting nsp12 activity (Fig. 4A, model 3) or ii) preventing it from accessing the 3'-OH of the primer (Fig. 4A, model 4). In the latter case, the observed activities would become solely nsp8-driven at the higher concentrations, which in turn would postulate a requirement for additional factors to induce nsp(7+8)'s dissociation from the primer. Unfortunately, we can pres-

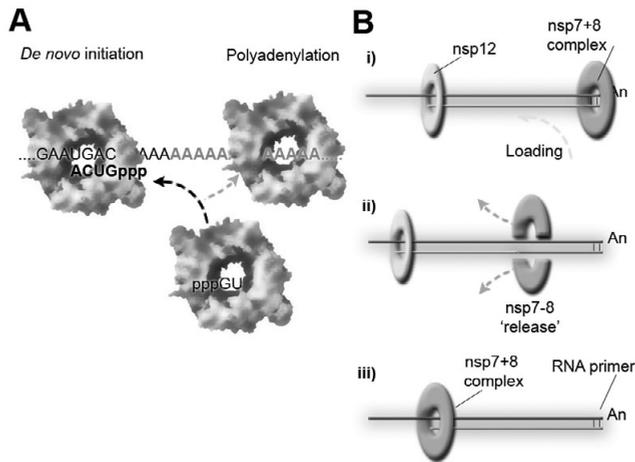
ently not ascribe a conclusive preference to either model, since the results obtained with nsp8 mutant K58A suggest that part of the influence must be mediated through an interaction, whereas the complete inhibitory effect of inactive nsp12 on active nsp8 suggests that the interplay may be organised through the template. Alternatively, it is a distinct possibility that the regulation between the two enzymes requires both direct (protein-protein) and indirect (substrate) influences, and possibly even additional viral proteins.

In summary, our results explain a number of ill-understood observations and touch upon at least three important questions regarding CoV replication, including the genetic linkage between mutations in the SARS-CoV genomic 3' end and the nsp8-coding sequence, and the conservation of the 3' terminal residue (Fig. S1). In light of the above two subjects, we demonstrated here that the nsp(7+8) complex can synthesise 8-14 nt short RNA products on RNA templates representing the 3'-terminal domain of the SARS-CoV genome and that these templates appeared to stimulate the RdRp activity compared to non-biological sequences in the presence of NTPs (Fig. 1B and 2). These results make it tempting to envision a mechanism in which these products are utilized by an elongation complex containing the putative main primer-dependent CoV RdRp, nsp12, as its key enzyme [155] (Fig. 6B, panel i). This brings us to a third question, however, because based on the activity of nps(7+8) on non-structured RNA templates [156] (Fig. S3), we cannot formally exclude the possibility that nsp(7+8) may synthesise substantially longer products *in vivo* (Fig. 6B, panel iii), possibly stimulated by the presence of additional viral protein factors. Moreover, the observed and apparently complex inhibitory effect of nsp8 and nsp(7+8) on the nsp12-dependent primer-extension activity (Fig. 4 and 5) asks whether i) additional factors are required to regulate their interaction and the nascent strand hand-off, possibly even through the release of nsp(7+8) from the template as proposed by Li *et al.* [425] (Fig. 6B, panel ii), or ii) that these two enzymes simply operate in separated complexes that each have their own dedicated RdRp and function in viral +RNA or -RNA synthesis in the infected cell. It will therefore become now particularly crucial to study whether these different polymerases associate *in vivo* and to what extent they influence each other's activity there.

## Material and methods

### Nucleotides and RNA modification enzymes

RNAases RNase A, T1 and T2 were purchased from Ambion. PolyA polymerase was purchased from USB, cordycepin 5'-triphosphate (3'dATP) from Sigma, and T4 kinase and Shrimp Alkaline Phosphatase (SAP) from Invitrogen. Radiolabelled nucleotides were purchased from Perkin-Elmer, marker dinucleotides GpA, GpU and ApA, and all RNA and DNA oligonucleotides from Eurogentec, nucleotide analog GTP-alpha-S from Jena Bi-



**Figure 6: Models for nsp8 activity and the initiation of negative strand RNA synthesis.** (A) Schematic presentation of the activities on SARS-CoVnsp(7+8). We hypothesise nsp(7+8) preferentially forms a pppGpU dinucleotide in absence of template that can subsequently be extended in a template dependent manner (black arrow). Additionally, nsp(7+8) may use the preformed product to align itself with the genomic end and catalyse the formation of a polyA tail (dark grey arrow). It is unknown whether in absence of *de novo* initiation, the dinucleotide product is release or whether it remains associated with nsp(7+8). (B) Panel i) shows a schematic presentation of the 'primase' hypothesis first proposed by Imbert *et al*, in which nsp8 or nsp(7+8) synthesises a primer that can be extended by SARS-CoV nsp12. Based on the data present in Fig. 1, the initiation site for nsp(7+8) may be on the genomic 3' end. Panel ii) demonstrates an extension of the 'primase' hypothesis, in which SARS-CoVnsp(7+8) must first release its primer before nsp12 can commence extension of the primer. In panel iii) a model is presented in which no primer-hand off between nsp(7+8) and nsp12 is required at all and nsp(7+8) is assumed to fully catalyse minus strand formation.

osience, whereas guanosine and all standard nucleoside triphosphates were purchased from Sigma. For T7 transcription, an Ambion MegaShortScript kit was used according to the manufacturer's instructions.

### Mutagenesis, protein purification and polymerase assays

All described nsp8 mutants were engineered via site-directed mutagenesis according to the QuikChange protocol (Stratagene) using the primers listed in table S1. All proteins were expressed and analysed as described previously [156]. SARS-CoV nsp7+8 assays were essentially performed as described elsewhere [156], with modifications. Specifically, nsp8 dinucleotide assays were performed either in the presence or absence of 1  $\mu$ M synthetic oligoribonucleotides as listed in Fig. 2 or the presence or absence of SARS-CoV genomic 3' end templates (see section below). Per nsp8 *de novo* reaction, typically 1  $\mu$ M monomeric nsp8 or nsp8 mutant and 1  $\mu$ M monomeric nsp7 was incubated with 4 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, 1 mM GTP, 10  $\mu$ M ATP, 0.17  $\mu$ M [ $\alpha$ -<sup>32</sup>P]ATP, 1mM DTT, 10 mM

KCl and 20 mM Tris (pH 8.0). ATP and GTP were substituted with other nucleotides or nucleosides as indicated in the main text. Note that we performed the activity assays at pH 8.0 instead of the for nsp8 optimal pH of 9.5 [156] to better preserve RNA integrity. For reactions containing guanosine, we compensated for the addition of NaOH that was used to dissolve guanosine by lowering the pH of the reaction buffer to 6.5 prior to the addition of guanosine and enzyme.

### RNA templates

Oligoribonucleotides for polymerase assays were excised as a single band from 7 M urea/15% polyacrylamide gels, eluted overnight in deionized water and desalted using NAP-10 columns (GE healthcare) that were equilibrated with deionized water. To anneal RNA duplexes, oligonucleotide mixtures in annealing buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl and 5 mM EDTA) were heat denatured and slowly cooled to room temperature.

Genomic 3' end templates of SARS-CoV Frankfurt-1 were amplified via *pfu*-based PCR (Fermentas) using forward primers containing a T7 promoter region (see table S2). After the PCR reaction, the cDNA template was digested with *DpnI* for 1 hour at 37 °C and the product purified via gel-extraction. Subsequently, T7 transcriptions were performed and the RNA products extracted with Trizol (Invitrogen), and glycogen (Roche) co-precipitation in ethanol.

### Sequence alignment

Alignments of the genomic 3' terminal residues were made using Muscle [327] and visualized with Weblogo 3.0. Sequences used included the alpha coronaviruses human CoV 229E (NC\_002645), rat CoV (NC\_012936), and bat CoV HKU8 (NC\_010438); the beta coronaviruses SARS-CoV (AY291315), mouse hepatitis virus (NC\_001849), and human CoV OC43 (NC\_005147); and the gamma coronaviruses beluga whale CoV (NC010646), bulbul CoV (FJ376620) and avian infectious bronchitis virus (IBV, AJ311317).

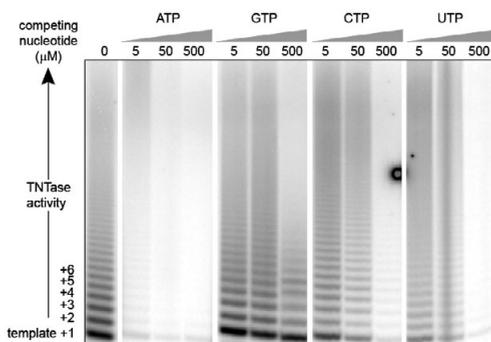
### Acknowledgements

The authors thank Dr. Danny Nedialkova, Dr. Clara Posthuma and Lorenzo Subissi for stimulating discussions. This work was supported by the Netherlands Organization for Scientific Research (NWO) through Toptalent grant 021.001.037 and ECHO grant 700.55.002 from the Council for Chemical Sciences (NWO-CW).

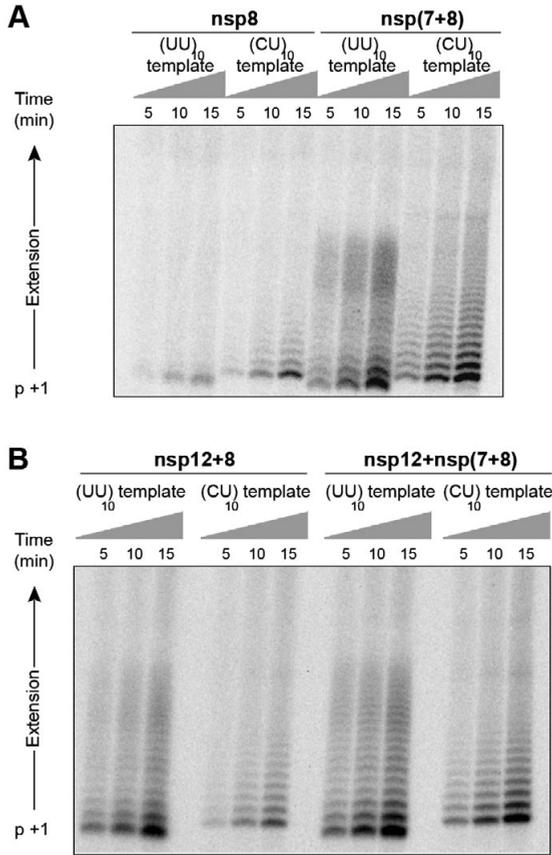
## CHAPTER 8 - SUPPLEMENTAL INFORMATION

			. — polyA —
α	RAT-CoV	UGAAUGAAGUUGAUC AUGG	CAAAAAAAAAA
α	HCoV-229E	CGGUUUCGAUAUGGAUACA	CAAAAAAAAAA
α	Bat-CoV	UAGUUUUGAUAGGGAUUCA	CAAAAAAAAAA
β	MHV-A59	UGGCCAAUUGGAAGAAUCA	CAAAAAAAAAA
β	SARS-CoV	UAGCUUCUUGAGGAAUGA	CAAAAAAAAAA
β	HCoV-OC43	UGGCCAAUUGGAAGAAUCA	CAAAAAAAAAA
γ	IBV-B	GGCUAGUAUAGAGUUAGAG	CAAAAAAAAAA
γ	BW-CoV	GACUAUAGGUAAUUGUUAGC	CAAAAAAAAAA
γ	Turkey-CoV	GGCTAGTATAGAGTTAGAG	CAAAAAAAAAA
δ	Bulbul-CoV	GCUUAAAUGGGGAGGGAG	CAAAAAAAAAA

**Figure S1. Sequence alignment of genomic 3' terminal sequences of alpha, beta, gamma and delta coronavirinae.** Sequences used included the alpha coronaviruses Rat Parker CoV (NC\_012936), Human CoV 229E (NC\_002645), and Bat CoV HKU8 (NC\_010438); the beta coronaviruses MHV A59 (NC\_001849), SARS-CoV Fr-1 (AY291315), and Human CoV OC43 (NC\_005147); the gamma coronaviruses avian infectious bronchitis virus (IBV, AJ311317), Beluga Whale CoV (NC010646), and Turkey CoV (NC\_010800); and the delta coronavirus Bulbul CoV (FJ376620). The conserved 3' cytosine is marked with an asterisk.



**Figure S2: Nucleotide competition during terminal transferase activity.** A SARS-CoV 3' end template was incubated with SARS-CoVnsp(7+8) and [ $\alpha$ - $^{32}$ P]ATP (lane 1). In parallel, 5, 50 or 500  $\mu$ M concentrations of either ATP, GTP, CTP and UTP were added to this mixture to study their potential to compete with [ $\alpha$ - $^{32}$ P]ATP for incorporation into the template RNA. Note that although the incorporation preference for GTP is lowest, it is effectively incorporated when present in ~3000 fold excess over the [ $\alpha$ - $^{32}$ P]ATP as is evident of the 'G-jumps' in the migration pattern on 20% 7M Urea PAGE.



**Figure S3: Primer-extension activities on various templates.** RdRp reactions containing either the U20 or the UC10 template, SARS-CoV polymerases nsp8, nsp(7+8) or nsp12 were resolved by 20% 7M Urea PAGE. Since extension of the UC10 template requires a GTP for position +1 and the readout of this assay is [ $\alpha$ -<sup>32</sup>P]ATP, the first visible product is p+2.

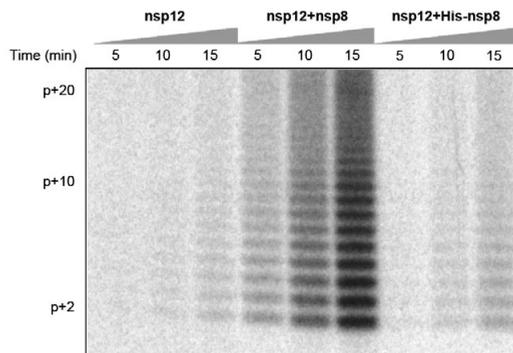


Table S1

nsp8 mutation	PCR primers	Sequence
<b>N43A</b>	SAV570	5'-TTAAAGAAATCTTTGGCTGTGGCTAAATCTGAG-3'
	SAV571	5'-CTCAGATTTAGCCACACCAAAGATTTCTTTAA-3'
<b>D50A</b>	SAV574	5'-GCTAAATCTGAGTTTGCCCGTGATGCTGCCATG-3'
	SAV575	5'-CATGGCAGCATCACGGCAAACCTCAGATTTAGC-3'
<b>D52A</b>	SAV590	5'-TCTGAGTTTGACCGTGCTGCTGCCATGCAACGC-3'
	SAV591	5'-GCGTTGCATGGCAGCAGCACGGTCAAACCTCAGT-3'
<b>K58A</b>	SAV402	5'-GCCATGCAACGCGCTTTGGAAAAGATGG-3'
	SAV403	5'-CCATCTTTTCCAAAGCGCGTTGCATGGC-3'
<b>D64A</b>	SAV580	5'-TTGGAAAAGATGGCAGCTCAGGCTATGACCCAA-3'
	SAV581	5'-TTGGGTCATAGCCTGAGCTGCCATCTTTTCCAA-3'

Table S2

3' terminal sequence	PCR primers	Sequence
--- <b>GACpA</b>	SAV453	TTTTTTTTTT <b>GTC</b> ATTCTCCTAAGAAGCTATTA
	SAV209	TAATACGACTCACTATAGGGAGAGCTGCCTATATGGAAGAGC
--- <b>GAC</b>	SAV212	<b>GTC</b> ATTCTCCTAAGAAGCTATTAATCAACATGGGGATAGC
	SAV209	TAATACGACTCACTATAGGGAGAGCTGCCTATATGGAAGAGC
--- <b>GAGpA</b>	SAV489	TTTTTTTTTT <b>CTC</b> ATTCTCCTAAGAAGCTATTA
	SAV209	TAATACGACTCACTATAGGGAGAGCTGCCTATATGGAAGAGC
--- <b>GAApA</b>	SAV483	TTTTTTTTTT <b>TTC</b> ATTCTCCTAAGAAGCTATTA
	SAV209	TAATACGACTCACTATAGGGAGAGCTGCCTATATGGAAGAGC
--- <b>GAUpA</b>	SAV490	TTTTTTTTTT <b>ATC</b> ATTCTCCTAAGAAGCTATTA
	SAV209	TAATACGACTCACTATAGGGAGAGCTGCCTATATGGAAGAGC
--- <b>GUCpA</b>	SAV486	TTTTTTTTTT <b>GAC</b> ATTCTCCTAAGAAGCTATTA
	SAV209	TAATACGACTCACTATAGGGAGAGCTGCCTATATGGAAGAGC



# Chapter 9

Ribavirin triphosphate-  
and mismatch-  
stimulated 3'-to-5'  
exonuclease activity of  
the SARS coronavirus  
RNA polymerase nsp12

Aartjan J.W. te Velthuis and Eric J.  
Snijder

Molecular Virology laboratory,  
Department of Medical  
Microbiology, Center of Infectious  
Diseases, Leiden University  
Medical Center, PO Box 9600,  
2300RC Leiden, The Netherlands.

## Abstract

Ribavirin is a purine analogue that can be efficiently incorporated into many viral RNA genomes. Once present, it induces the accumulation of transition mutations, which subsequently lead to error catastrophe and the inhibition of virus propagation. Furthermore, these effects may be more pronounced in larger viral RNA genomes. However, in spite of having RNA genomes that are ~3-fold larger than the typical RNA virus genome, the inhibitory effect of ribavirin on the replication of coronaviruses like SARS coronavirus (SARS-CoV) has been shown to be remarkably small. It was therefore hypothesised that CoVs may encode a rudimentary mechanism to prevent or correct misincorporation, and thus to avoid error catastrophe. Using *in vitro* assays and recombinant SARS-CoV nonstructural protein (nsp) 12, we show here that a 3'-to-5' exonuclease activity is associated with this coronavirus polymerase. Importantly, this activity was stimulated by the presence of either ribavirin triphosphate (RTP) or an unpaired residue in the primer-template duplex, whereas mutations of RdRp motif B and residues in the nsp12 N-terminal domain altered these effects. Overall, these results lead us to suggest that SARS-CoV nsp12 may indeed be a viral RdRp that is capable of reducing its own error rate.

## Introduction

The nucleoside ribavirin (1- $\beta$ -D-ribofuranosyl-1,2,4-triazole-3-carboxamide) is a purine analogue that can enter cells as nucleotide precursor. Once converted to ribavirin triphosphate (RTP), it can be incorporated into RNA virus genomes by their cognate RNA-dependent RNA polymerases (RdRps) with efficiencies similar to those of ATP and GTP [180,276,320,321,426]. This property derives from the fact that ribavirin's carboxamide group can rotate and thereby alternate ribavirin's base pairing between uracil and cytosine. The consequence of this property is that ribavirin monophosphate (RMP) incorporated into the viral RNA can randomly induce transition mutations (G $\leftrightarrow$ A; and C $\leftrightarrow$ U in the opposite strand). In turn, this effect increases the overall mutation frequency of the virus and may thus rapidly result in an inhibition of viral replication and transcription [39,56,244,427,428,429]. Ribavirin is therefore a widely administered antiviral drug - alone or in combination with other compounds - for treatment of infections with viruses as diverse as Lassa virus, respiratory syncytial virus, and hepatitis C virus [92,93,94,95]. Furthermore, the analogue can effectively inhibit the replication of picornaviruses, orthomyxoviruses [96,97], hantaviruses [98,99], vaccinia virus [100], and reoviruses [101] in cell culture.

The sensitivity to ribavirin is highly variable among viruses, however. For example, to prevent plaque formation on Vero E6 cells, 1-25  $\mu$ g/ml ribavirin is needed for the inhibition of the influenza A virus, while 25-100  $\mu$ g/ml is required to inhibit HIV and other retroviruses [430]. Interestingly, to prevent cytopathology of severe acute respiratory syndrome coronavirus (SARS-CoV) on the same cell type, ribavirin concentrations of 0.5-5 mg/ml are required, a dose range that overlaps with the 50% cytotoxic dose (CD<sub>50</sub>) for Vero E6 cell (generally ~0.2-1 mg/ml) [102,103,308,309]. Inhibiting concentrations close to the cytotoxic concentration (CC<sub>50</sub>) of ribavirin on ST cells were also reported for two related swine CoVs, porcine respiratory CoV (PRCV) and transmissible gastroenteritis virus (TGEV) [431]. In line with these observations, ribavirin treatment of SARS-CoV-infected humans did not result in convincing antiviral effects during the 2003 outbreak [103], and similar findings were reported in mice infected with a mouse-adapted SARS-CoV strain [432].

With a nonsegmented, positive-stranded RNA (+RNA) genome of ~30 kb, approximately 3 fold larger than the average +RNA viral genome, SARS-CoV has one of the largest genomes among RNA viruses [58]. Theoretically, this unprecedented genome size should thus make it more sensitive to mutagen-induced error catastrophe, provided that we assume that the fidelity of the SARS-CoV replication and transcription complex (RTC) is similar to that of other RNA virus replicases. It has therefore been hypothesised that the surprisingly limited effect of ribavirin on CoV infections must be due to i) inefficient incorporation of RTP by one of the two SARS-CoV RdRps (*i.e.*, nonstructural proteins [nsps] 12 and 8) due to structural constraints in the nucleotide selection pocket

[89,154,156]; ii) negation of erroneous nucleotide incorporation through a correction mechanism involving a proofreading activity residing in one of the RdRps; iii) proofreading through a mechanism involving the nsp14 DEDD-type exonuclease [35,58], a viral enzyme whose inactivation was previously linked with the accumulation of viral mutations [105,106,107]; or iv) increased recombination rates between correct genome segments.

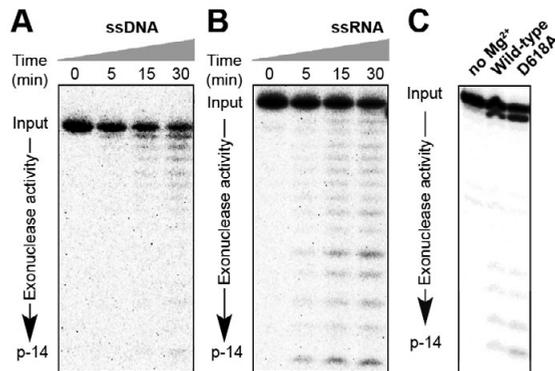
In light of the above hypotheses [35,89], we here document the observation that RTP and misaligned primer-template duplexes can stimulate a 3'-to-5' exonuclease activity that is associated with the SARS-CoV RNA polymerase nsp12. Mutation of the polymerase motif B altered the efficiency of mismatch recognition, whereas single mutations in the unique N-terminal domain allowed us to impair exonuclease activity. These results thus lead us to suggest that SARS-CoV nsp12 possesses 3'-to-5' exonuclease activity that is stimulated by an ability to recognise the nucleotide analogue RTP and errors in the template-primer duplex.

## Results and discussion

### SARS-CoV nsp12 has 3'-to-5' exonuclease activity

Comparative sequence analysis has shown that the conserved, canonical RdRp domain of CoVs resides in the C-terminal two thirds of the 106-kDa nsp12 protein [274]. Subsequent biochemical studies have demonstrated the ability of this replicase subunit to incorporate nucleotides using both RNA and DNA templates [89,154]. Interestingly, during our analysis of the nsp12 primer-extension activity in the presence of ATP [154], we also noted a weak degradation of 5' <sup>32</sup>P-labelled RNA oligos that was more pronounced in reactions containing the RdRp active site mutant D618A than in reactions containing wild-type nsp12. To investigate this activity in more detail and its significance in view of the hypothesised proofreading ability of CoV RdRps, we purified a recombinant SARS-CoV nsp12-RdRp as described previously [154] and incubated this enzyme with 5' <sup>32</sup>P-labelled single-stranded DNA (ssDNA) and ssRNA oligonucleotides. Analysis of the reaction by 20% denaturing PAGE and autoradiography demonstrated that these nsp12 preparations were indeed able to slowly degrade both DNA and RNA (Fig. 1A and 1B). Interestingly, we previously observed that nsp12 has polymerase activity on both templates as well [154]. Given that the 5' <sup>32</sup>P-labelled substrates were gradually hydrolysed to increasingly smaller labelled products, we conclude that this exonuclease activity has a 3'-to-5' directionality.

Although the exonuclease activity for most optimal substrate for the exonuclease activity of DNA-dependent DNA polymerase (DdDps) is ssDNA, the physiologic relevant one is a primer-template duplex, whose primer strand must be melted in order to transfer the 3'-end of the primer to the exonuclease active site [54,433]. To study the



**Figure 1. Nsp12 exonuclease activity has 3'-to-5' directionality on both ssDNA and ssRNA templates.** (A) To investigate the directionality of the SARS-CoV exonuclease activity,  $5^{32}\text{P}$ -labelled DNA or (B) RNA oligos were incubated in the presence of 6 mM  $\text{Mg}^{2+}$  and purified SARS-CoV nsp12. Samples were stopped at the indicated time points and resolved by 20% PAGE/7M urea gels. (C) SARS-CoV nsp12 is able to cleave a primer that was annealed to template RNA in the presence of 6 mM  $\text{Mg}^{2+}$ . A similar activity was observed for RdRp active site mutant D618A. Assays were stopped after 60 min and resolved with 20% PAGE/7M urea gels.

activity of nsp12 on such a template, we hybridised the 5'  $^{32}\text{P}$ -labelled RNA primer to a 40-mer template that was partially based on the 3' untranslated region (UTR) of the SARS-CoV genome (residues 29512-29532) [154]. As shown in Fig. 1C, both nsp12 and the polymerase active site mutant D618A both degraded the primer in this 20-mer/40-mer primer-template configuration, albeit with very limited processivity. In absence of magnesium ions, no significant exonuclease activity was observed (Fig. 1C).

### SARS-CoV nsp12 prefers magnesium ions for activity

To characterise the influence of divalent metal ions on the nsp12 exonuclease activity, we next substituted the optimal  $\text{Mg}^{2+}$  concentration for primer extension (6 mM; [154]) with different concentrations of  $\text{Mg}^{2+}$  or other divalent metals. As shown in Fig. S1, nsp12 was able to hydrolyse the 3' end of ssRNAs under all  $\text{Mg}^{2+}$  concentrations and a low mM  $\text{Mn}^{2+}$  concentrations. The presence of  $\text{Zn}^{2+}$ , in line with its negative effects on the template binding affinity of SARS-CoV nsp12 [325], or high concentrations of  $\text{Mn}^{2+}$  significantly reduced nsp12's exonuclease activity (Fig. S1).

### Sequence analysis of the coronavirus nsp12 N-terminal domain identifies residues involved in exonuclease activity

Presently, the active site residues involved in the exonuclease and proofreading abilities of viral polymerases have only been identified in DNA virus polymerases [166,434]. Similar to the Klenow fragment, these polymerases employ a two-metal ion mechanism to bind an attacking hydroxide ion and stabilise the leaving 3'-OH group [54,168,435,436].

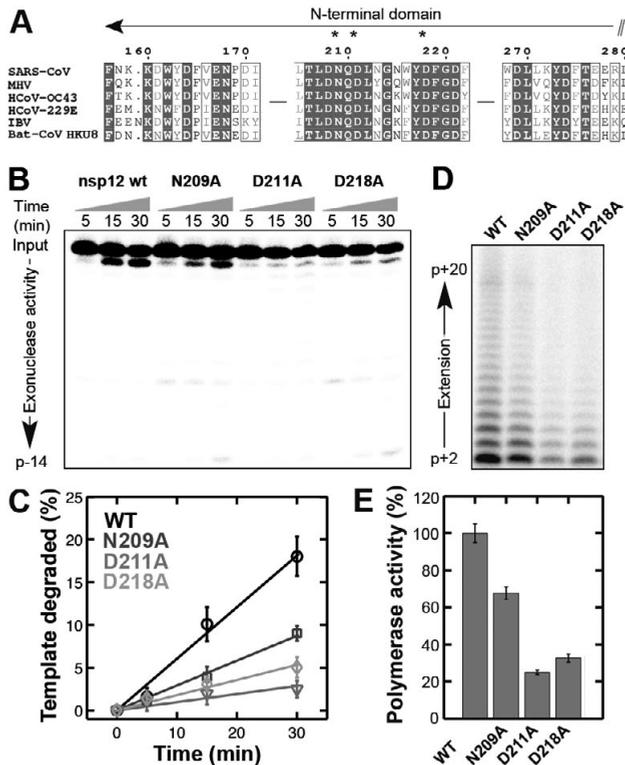
Typically the exonuclease domain is a ~150-amino acid structure in the DdDp N-terminus and dependent on motif I residues DxE, motif II residues Nx<sub>2-4</sub>F/YD, and motif III residues Yx<sub>3</sub>D for activity [54,169,433,434]. Metal ion coordination in this structure is performed by the invariant acidic amino acids in the three motifs, whereas the binding and positioning of the single stranded part of the melded nascent strand requires the strictly conserved asparagine and aromatic residues (*i.e.*, Phe or Tyr) [54,169,433]. Using these well-characterised motifs as reference, we scanned the nsp12 sequence for putative exonuclease active site residues and identified a set of conserved aspartic acid and asparagine motifs positioned within ~120 amino acids of each other. As shown in Fig. 2A, these residues are spaced according to conserved Dx<sub>2</sub>EN, DNx<sub>2</sub>Dx<sub>3</sub>YD and Dx<sub>3</sub>YD motifs and reside in the N-terminal domain, a domain that is unique for CoVs and presently without assigned enzymatic function [58,89].

To investigate the involvement of the N-terminal domain in nsp12's ability to degrade an annealed RNA primer, we engineered three alanine substitutions in the most conserved N-terminal nsp12 motif (*i.e.*, motif II), targeting asparagine-209, aspartate-211 and aspartate-218 (*i.e.*, N209, D211 and D218). All mutants were purified side-by-side using the same expression and purification protocol [154]. When we next examined the ability of these three mutants to degrade 5' <sup>32</sup>P-labelled RNA, we found that the exonuclease activity of N209A was impaired by ~50% compared to the wild-type protein, whereas D211A and D218A demonstrated ~86% and ~72% reduced exonuclease activities, respectively (Fig. 2B and C).

To test the impact of the mutations on the basic polymerase activity of nsp12, we assessed their ability to extend the 20-nt primer used above as readout for the exonuclease assay through the incorporation of ATP and GTP, as described previously [154]. As shown in Fig. 2D and 2E, all mutations had less than wild-type primer extension activity on this template, suggesting that the C-terminal polymerase domain and the N-terminal domain are tightly integrated into the nsp12 structure. Such an integration of the two active sites and a correspondingly reciprocal effect of exonuclease active site mutations has also been observed for a number of viral DdDps [434,435].

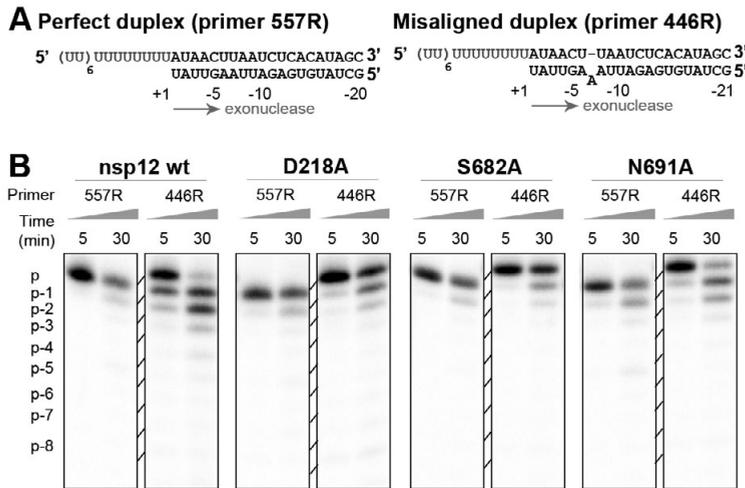
### **An unpaired residue in the primer-template duplex stimulates SARS-CoV nsp12's 3'-to-5' exonuclease activity.**

Together the results presented in Fig. 1 and 2 suggested that nsp12 is able to hydrolyse RNA molecules with a 3'-to-5' polarity and that this activity involves residues in the N-terminal domain. To investigate the significance of this activity for the proposed proof-reading hypothesis, we incubated nsp12 with a primer-template duplex containing an unpaired residue at position -6-8 of the primer (Fig. 3A, primer 446). As was argued previously for DdDps [436], the recognition of such internally mismatched residues critically is important for the overall fidelity of the polymerase reaction, as they can then induce a



**Figure 2. Conserved residues in the coronavirus nsp12 N-terminal domain influence exonuclease and RdRp activity.** (A) Partial alignment of the N-terminal domain of CoV nsp12. Fully conserved residues are shaded dark grey, while partially conserved residues are boxed. The residues targeted by mutagenesis are indicated with asterisks. Mutants of aspartic acid and asparagine residues were chosen based on their conservation among CoV nsp12 proteins and their relatively similar spacing compared to the exonuclease domains of viral DNA polymerases. (B) Analysis of the effect of N-terminal domain mutations on SARS-CoV nsp12's ability to degrade a 5' <sup>32</sup>P-labelled primer in a dsRNA substrate. (C) Quantitation of the exonuclease activity. Percentages of degraded input RNA were plotted as function of time and fitted with a single polynomial. Error bars present standard deviation. (D) Analysis of the primer extension activity of wild-type SARS-CoV nsp12 and the nsp12 mutants on a heteromeric template. (E) Quantitation of the primer extension activity of SARS-CoV nsp12 mutants expressed as percentage of activity of the wild-type protein. Error bars represent standard deviations (n = 3).

global conformational change in the polymerase ground-state, which stalls polymerisation and allows the single stranded nascent strand to travel to the exonuclease site for mismatch removal. Indeed, when we incubated wild-type nsp12 with the misaligned template, we found that the exonuclease activity increased relative to the fully Watson-Crick base-paired template (Fig. 3B, left panel). In contrast, no significant stimulation of exonuclease activity was observed for mutant D218 in line with our observations in Fig. 2.



**Figure 3. The 3'-to-5' exonuclease activity of SARS-CoV nsp12 mutants is stimulated by mutation of conserved residues in nsp12.** (A) To study the sensitivity of the nsp12 exonuclease activity to misalignments in the primer-template duplex, two templates were used. The first consisted of a perfect primer-duplex (primer 557), while the second used a primer (primer 446) with an additional adenosine, thus creating a primer-duplex containing a non-Watson-Crick paired residue at position -6 of the primer. (B) Wild type nsp12 and mutants D218A, S682A and N691A were incubated with 5' <sup>32</sup>P-labelled primers hybridised to the template. Samples were taken after 5 and 30 minutes and resolved with 20% PAGE/7M urea gels.

In the RdRp structure, motif B residues are generally involved in nucleotide selection and the recognition of the primer-template duplex [89,132]. To explore the involvement of nsp12's motif B residues in exonuclease stimulation, we performed the above assay with nsp12 mutants containing alanine substitutions at the strictly conserved positions serine-682 and asparagine-691 (*i.e.*, S682 and N691; Fig. S1). As shown in Fig. 3C, we observed that mutation S682A failed to degrade the two templates with a significant difference, suggesting that this residue may be involved in the contact formation between the RdRp and the primer-template duplex. Mutant N691A on the other hand, was effectively stimulated by the presence of the mutation and degraded both primers with similar efficiency as the wild-type enzyme.

### Ribavirin triphosphate stimulates exonuclease and re-incorporation activity

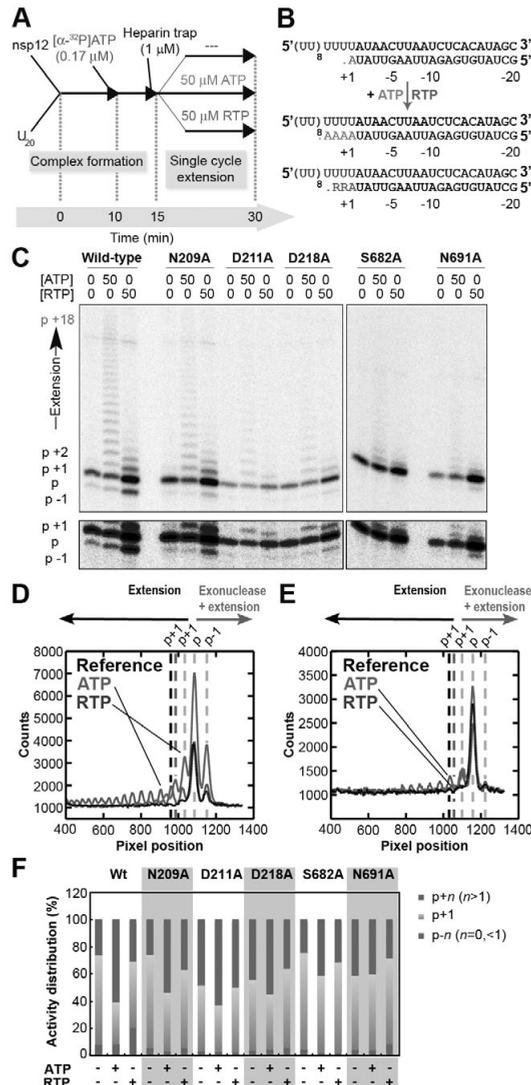
As was noted above, the commonly used antiviral drug and nucleotide analogue ribavirin has been deemed relatively ineffective against SARS-CoV infections [103,432]. One hypothesis to explain this phenomenon suggests that one or both of the SARS-CoV RdRps may be able to select against RTP incorporation or remove RMP moieties from the nascent strand [89,156]. To analyse the relevance of the observed SARS-CoV nsp12 exonuclease activity for RTP incorporation, we next incubated wild-type and mutant

nsp12 enzymes with either ATP or RTP and tested their ability to extend a 20-nt primer under these conditions. To ensure that only single extension cycles could be performed, the RdRp assays were done in the presence of sufficient heparin to trap unbound polymerases (Fig. 4A). As outlined in Fig. 4, complex formation and 3'-labelling of the primer was achieved by incubating nsp12 with the partly double-stranded RNA template and a limiting concentration of [ $\alpha$ - $^{32}$ P]ATP (Fig. 4B, grey A residue). Next, heparin was added in addition to a 300-fold excess of non-radioactive ATP or RTP to facilitate further elongation of the primer (Fig. 4B).

As shown in Fig. 4C, the reference lane of wild-type nsp12, which had not been stimulated by addition of ATP or RTP, yielded the expected 3'-radiolabelled primer (indicated as 'p+1') as well as a minor product that was equal to the primer length (labelled 'p') (Fig. 4C, leftmost lane). This radioactive pattern suggested that a small fraction of the wild-type nsp12 had first hydrolysed the 3' end of the primer and subsequently extended it by one nucleotide using the only available substrate in the reaction (*i.e.*, [ $\alpha$ - $^{32}$ P]ATP) before becoming inactive (Fig. 4C, Fig. S2). We observed significantly less p product in reactions containing nsp12 mutants D211A and S682A (Fig. 4C), in line with their reduced ability to attack the 3' terminal residue of the primer in the exonuclease assays.

In the reactions containing wild-type nsp12, the subsequent addition of an excess of unlabelled ATP resulted in a distributive extension of the primer up to a dominant p+18 product (Fig. 4C and 4D), which is in line with previous observations [154]. A more detailed analysis of the lane profiles confirmed that the spacing between the p, p+1 and p+2 signals was identical and that the primer signal must have been the result of an ATP incorporation event (Fig. 4D). No decrease in the p signal was observed upon addition of an excess of unlabelled ATP, suggesting that the RdRps responsible for this product had indeed either dissociated from the template or had become inactive. In addition, this approach allowed us to verify that the p product in the D211A-reaction was indeed significantly reduced (Fig. 4E).

In contrast to the ATP reactions, the addition of RTP predominantly resulted in an increase of the p and p+1 signals (Fig. 4C). This was particularly evident in the lane profile analysis, in which the p+1 signal at pixel position 1075 showed a dramatic increase over the p+1 peak in the ATP reaction. Strikingly, a weak band corresponding to a p-1 signal, *i.e.* a product that was shorter than the original RNA primer, was observed as well (Fig. 4C, lower panel), suggesting that the RTP had stimulated the 3' attack of the primer and (re)incorporation of the still present [ $\alpha$ - $^{32}$ P]ATP. Lane profile analysis confirmed that ATP had indeed been incorporated in all these three peaks, whereas RTP had clearly been used for incorporation in the p+2 product, as was evident from the migration shift of these bands (Fig. 4D). This result is in line with the 23.2 g/mol molecular weight (MW) difference between adenosine and ribavirin ( $MW_{\text{ATP}} = 507.2$  g/mol;  $MW_{\text{RTP}} = 484.0$  g/mol).



**Figure 4. Interplay between the SARS-CoV nsp12 polymerase and exonuclease activities.**

(A) Schematic of nsp12 primer-extension experiments. Recombinant SARS-CoV nsp12 was first preincubated with an unlabelled primer-template complex and a low concentration of  $[\alpha\text{-}^{32}\text{P}]\text{ATP}$  to establish complex formation and 3' radiolabelling of the primer. Heparin was subsequently added to a final concentration of 1  $\mu\text{M}$  to trap free polymerase before reactions were split in 3 aliquots. To allow extension, 50  $\mu\text{M}$  non-radioactive ATP or RTP was then added to two of these samples, followed by a 30-min incubation. (B) Overview of expected radiolabelled products on the provided template assuming that nsp12 only has polymerase activity. (C) Activity of SARS-CoV nsp12 in the presence of an excess of ATP or RTP. For wild-type nsp12, elevated p-1, primer and p+1 signals are evident in the presence of RTP relative to reactions performed in the presence of ATP. Interestingly, this behaviour changed upon mutation of conserved nsp12 residues to alanine, most notably through mutation D211A. Samples were resolved with 20% PAGE/7M urea gels. The lower

**Figure 4** continued

panel is a longer exposure of the upper panel. **(D)** Lane profile analysis can be used to distinguish between the incorporation of ATP and RTP in the ATP- and RTP-containing lanes given their different migration in the 20% PAGE gels. These differences are only visible for the p+2 and longer products (compare black dashed line [ATP] with dark grey dashed line [RTP]), whereas the p and p-1 products are all separated with an AMP-like interval (light grey dashed lines), suggesting that they result from hydrolysis of the 3' terminal nucleotide of the primer and the (mis)incorporation of [ $\alpha$ - $^{32}$ P]ATP. **(E)** The p and p-1 products are significantly reduced compared to the reference signalin mutant D211A. **(F)** Normalised distributions of the SARS-CoV nsp12 polymerase activity.

To obtain a better estimate of the effects of ATP and RTP on the overall polymerase reaction, we calculated the normalised distribution of the primer extension signal (all products longer than p+1, *i.e.*, p+n with n>1), the initial 3' labelling (p+1), and all products that were shorter than any extension of the primer (p+n with n=0 or n<1). As shown in Fig. 4F, this analysis supported our previous conclusions and showed that only the wild-type nsp12 and mutant N691 displayed a >2 fold increase of the activity that required RTP-induced exonuclease active site and a reincorporation of [ $\alpha$ - $^{32}$ P]ATP (Fig. 4F and Fig. S2). This signal was reduced in all other mutants, in line the significantly impaired exonuclease activity of the N-terminal domain mutants, and mutant S682A's inability to stimulate exonuclease activity in the presence of a template-primer mismatch. Interestingly, all extension signals were affected by the presence of RTP (Fig. 4F), suggesting that nsp12 is able to discriminate between the ATP and RTP, but that none of the residues targeted by our mutagenesis is essential for this discrimination during incorporation.

### Concluding remarks

With genomes of ~30 kb, CoVs are a unique family of +RNA viruses that have evolved the largest nonsegmented RNA genomes known to date. In addition to genome replication, expression of these giant genomes requires the synthesis of subgenomic (sg) mRNAs, which have a common leader sequence at their 5' end and an identical 3'-terminal segment [323]. Current evidence suggests that these products are formed during a process that starts with discontinuous minus-strand RNA (-RNA) synthesis and produces a nested set of sg-length -RNAs that can be used as template for the production of the sg mRNAs [51,266,416]. It is believed that as CoVs evolved these features, genome expansion and the development of a more complex enzymology than other +RNA viruses went hand in hand. To minimise the number of mutations per replicated genome, these enzymes likely included proteins capable of performing RNA-based proofreading or some other form of error prevention [35,58,323,437]. Interestingly, previous deep-sequencing analysis of the progeny of a SARS-CoV mutant containing an inactivated nsp14 exonuclease domain did indeed demonstrate a significantly increased mutation frequency compared to the-wild type control [106,107]. This virus mutant was therefore denoted as "infidelity

mutant" and nsp14 was proposed to be a key component of a putative coronavirus proofreading mechanism. A direct observation of mutation-recognition by nsp14 has so far not been made, however, and mechanistic information about its possible mode of action in the viral RTC is still lacking.

Interesting in the light of the proposed proofreading hypothesis, we here describe that the SARS-CoV polymerase nsp12 has 3'-to-5' exonuclease activity as well. Furthermore, guided by the well-established of the DdDp exonuclease active site and a model structure based on comparative sequence analysis between the C-terminal two-thirds of SARS-CoV nsp12 and many well-studied RdRps [89,166,434], we were able to engineer nsp12 mutants that could alter its exonuclease activity (Fig. 2) and its ability to recognise primer-template mismatches (Fig. 3). Together, these observations lead us to suggest that one of the SARS-CoV RdRps has a proofreading ability that is tightly interwoven with its polymerase function. It will be of importance to investigate the involvement of nsp12 in the proposed CoV proofreading mechanism and to test whether it may function in conjunction with the multifunctional exonuclease-methyltransferase enzyme nsp14 [105,107].

## Materials and methods

### SARS-CoV nsp12 purification and mutagenesis

All proteins used were expressed and purified as described previously [154]. The nsp12 mutants listed in Table S1 were engineered in the pASK3-Ub-nsp12-CHis<sub>6</sub> expression vector [154] via site-directed mutagenesis according to the QuikChange protocol (Stratagene) and the in Table S1 listed primers.

### Polymerase and exonuclease experiments

The oligoribonucleotide substrates used for polymerase assays indicated in Fig. 2 and 4 were prepared as described previously [154]. Primer-extension assays were essentially performed as described elsewhere [154], using 50  $\mu$ M ATP (Roche) or RTP (Jena Bioscience) for extension. For exonuclease assays, 0.25  $\mu$ M of purified nsp12 was typically incubated with 1 nM of 5' <sup>32</sup>P-labelled ssDNA, ssRNA or duplex RNA and 6 mM Mg<sup>2+</sup> at 20 °C. Reaction times are indicated in the figures, while all reaction buffers contained 1mM DTT, 0.1% Triton X-100, 10 mM KCl, 10 mM NaCl, 5% glycerol and 20 mM Tris (pH 8.0). Gels were dried on Whatman filter paper and bands were quantified by phosphorimaging using a Typhoon variable mode scanner (GE Healthcare) and ImageQuant TL 7.0 software (GE Healthcare). Lane profile densities were exported from ImageQuant TL 7.0 and aligned by the p+1 density using Matlab 2009a.

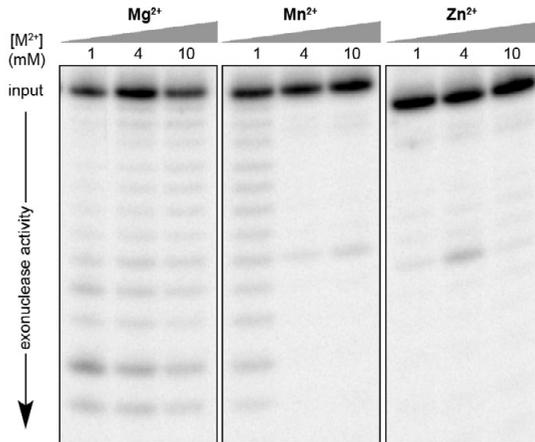
### **Sequence alignment**

Alignments of nsp12 sequences were made using Muscle [327]. Sequences used included the alphacoronaviruses human CoV 229E (NC\_002645); the betacoronaviruses SARS-CoV Frankfurt-1 (AY291315), mouse hepatitis virus A59 (MHV, NC\_001849), and human CoV OC43 (NC\_005147); and the gammacoronavirus avian infectious bronchitis virus (IBV, AJ311317).

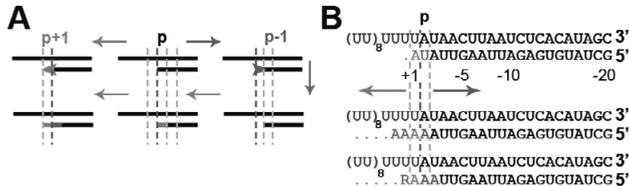
### **Acknowledgements**

We thank Dr Alexander Gorbalenya, Dr Craig Cameron, and Dr Clara Posthuma for stimulating discussions. This work was supported by the Netherlands Organization for Scientific Research (NWO) through Toptalent grant (021.001.037), a grant from the Council for Chemical Sciences (700.55.002), and the European Union Seventh Framework Programme (FP7/2007-2013) under SILVER grant agreement n° 260644.

## CHAPTER 9 - SUPPLEMENTAL INFORMATION



**Figure S1: Influence of divalent metal ions on the SARS-CoV nsp12 activity.** Wild-type nsp12 was incubated with different concentrations of  $Mg^{2+}$ ,  $Mn^{2+}$  and  $Zn^{2+}$  and assessed for its ability to degrade a 5'  $^{32}P$ -labelled ssRNA under these conditions. Samples were resolved by 20%/7M Urea PAGE after 60 min.



**Figure S2: Schematic presentation of interplay between the SARS-CoV nsp12 primer-extension and exonuclease activities.** (A) In the presence of low concentrations of  $[\alpha\text{-}^{32}P]\text{ATP}$ , wild type SARS-CoV nsp12 can extend the 3' end of the primer (short black line) with a radioactively labelled AMP (indicated in light grey), but also perform a 3'→5' digestion-extension reaction (indicated in dark grey). Given that only  $[\alpha\text{-}^{32}P]\text{AMP}$  was provided during the preincubation, this can result in the synthesis of p and p+1 products through the incorporation of  $[\alpha\text{-}^{32}P]\text{AMP}$  and thus two radioactive products. (B) Effectively the radiolabelling of p is a misincorporation (indicated as grey A) and thus results in the replacement of the 3' U with a  $[\alpha\text{-}^{32}P]\text{AMP}$  in a small population of the templates. When an excess of ATP is subsequently added, further extension of the primer can be observed (indicated with green A residues). In the presence of RTP, however, the frequency of the 3' exonuclease-extension reactions increases, which results in  $[\alpha\text{-}^{32}P]\text{AMP}$ -labelling of the p-1 position.

**Table S1: primers for nsp12 mutagenesis**

<b>Mutant</b>	<b>Primer</b>	<b>Sequence (5'-to-3')</b>
N209A	SAV547	GTACTGACATTAGATGCTCAGGATCTTAATGGG
	SAV548	CCCATTAAGATCCTGAGCATCTAATGTCAGTAC
D211A	SAV549	ACATTAGATAATCAGGCTCTTAATGGGAACTGG
	SAV550	CCAGTCCCATTAAGAGCCTGATTATCTAATGA
D218A	SAV551	AATGGGAACTGGTACGCTTTCGGTGATTCGTA
	SAV552	TACGAAATCACCGAAAGCGTACCAGTTCCTCCATT
S682A	SAV618	CAGGTGGAACATCAGCCGGTGATGCTACAACCT
	SAV619	AGTTGTAGCATCACCGCTGATGTCCACCTG
N691A	SAV622	ACAACCTGCTTATGCTGCTAGTGCTTTAACATT
	SAV623	AATGTTAAAGACACTAGCAGCATAAGCAGTTGT
S795A	SAV626	AATAATGTGTTTCATGGCTGAGGCAAAATGTTGG
	SAV627	CCAACATTTGCCTCAGCCATGAACACATTATT



# Chapter 10

A discussion of  
nidovirus RNA synthesis

---

## Introduction

Daily, numerous RNA molecules are created inside the eukaryotic cell. In the nucleus for example, RNAs of several hundred to thousand bases long are synthesised by RNA polymerases operating at local concentrations of  $\sim 1 \mu\text{M}$  [130,131,438]. However, most of these RNAs need to be shortened through internal splicing events or extended at their 3' end by polyA polymerases before they can become functional [415,439]. RNA modifications are also made in the cytoplasm, where various protein complexes regulate, for instance, i) the stability of miRNAs through the 3' addition of adenosines [440], ii) the degradation of miRNAs by synthesising polyU tails on their 3' ends [441], iii) the efficiency of mRNA translation by restoring or extending their polyA tails [442], or iv) the turnover of ribosomal and mitochondrial RNAs through their ability to add polyadenylyl or heteromeric tails to these molecules [443]. In addition, many enzymes are also involved in the cleavage of RNA molecules, such as exo- and endonucleases, which may be involved in RNA metabolism in general, the maturation of Okazaki fragments or defences against pathogens with RNA genomes [444,445]. Furthermore, motor proteins such as helicases play a crucial role in regulating the dynamics of nucleic acid substrates and the molecules that bind them, including some RNA polymerases that need to switch from initiation to extension, *i.e.*, 'escape' their promoter.

Strikingly and unique for RNA pathogens, many of these functions were also predicted by comparative sequence analysis or found via screening methods in the nidovirus genomes [58,84,273]. This virus group consists of enveloped positive stranded RNA (+RNA) viruses and unites the *Corona-*, *Roni-* and *Arteriviridae* based on comparative sequence analyses, phylogenetic analyses and a similar genome organisation [58,84]. So far, the existence of a large number of these enzymatic functions has been verified using *in vitro* methods, while others were added by accidental discoveries. Together they demonstrate that nidoviruses encode two different RNA polymerase activities [154,156,196] (chapter 3-5), an RNA helicase [109,110], two separate exonuclease activities [105] (chapter 9), an endonuclease [269,446], single strand RNA (ssRNA) binding proteins [270,447], and two methyltransferases [272,273]. Although the identification and characterisation of these RNA proteins is of medical interest - partly since nidoviruses were the protagonists in a number of pandemics, including the 2003 outbreak of the lethal severe acute respiratory syndrome coronavirus (SARS-CoV) [58,65,437,448] and because the development of anti-nidoviral strategies greatly depends on our understanding of the function of these enzymes in the nidovirus replication cycle - they may one day also give us an answer to the question "What constitutes a nidovirus replicase and why do nidoviruses need this enzymatic diversity, while other RNA viruses do not?"

### Nidovirus RNA synthesis, a little background...

All nidovirus replicase enzymes, putatively with the exception of at least one actively or passively contributing host factor [290,313], are encoded by the nidovirus genome [58,390]. To accommodate the large number of replicase enzymes, approximately two-thirds of the nidovirus genome is dedicated to encoding two large polyproteins that contain all the mature viral replicative enzymes or non-structural proteins (nsps). Nidovirus genomes are further marked by their polycistronic nature, meaning that they encompass multiple open reading frames (ORFs). To express these ORFs, which are all situated downstream of the replicase gene and encode the structural and accessory proteins required for, *e.g.*, virion formation, nidoviruses also employ a unique discontinuous RNA synthesis process. Quintessentially, this process involves the synthesis and subsequent transcription of a nested set of subgenome (sg) -RNA molecules [266,416,449]. These types of -RNA molecules are both 3'- and 5'-co-terminal with each other and the genome-length -RNA - also referred to as the replicative intermediate when associated with a complementary genomic +RNA and replicase enzymes -, but by definition shorter in length as they progressively lack internal ORFs of the genome. Interestingly, various lines of research have indicated that the replicative intermediate must be derived from a continuous mode of RNA synthesis, whereas a discontinuous mode, including a strand transfer step, is required to produce sg -RNAs [265,416,449]. To produce capped and polyadenylated subgenome and genomic +RNAs, both types of -RNA molecules need to be transcribed [51,265,449].

In spite of having similar genomic structures, the replicase gene is significantly larger in CoVs than in arteriviruses (*i.e.*, encoding ~4000 compared to ~6500 amino acids). Correspondingly, the CoV, torovirus and ronivirus genomes are dramatically larger than those of arteriviruses and have been found to range up to 31 kb [35]. Regardless of these size differences, the replication and transcription of the nidoviral genomes has long been believed to be catalysed by the same conserved viral RNA-dependent RNA polymerase (RdRp) activity [58,84,196,274]. More recent evidence and the observations from chapter 4 and 5 suggest, however, that the CoV and arterivirus RdRps are different and that a second RdRp is involved in CoV RNA synthesis as well [154,155,156]. Whether this second RdRp is crucial for the replication of these large genomes remains to be investigated in detail, but initial experiments suggest that the lysine to alanine mutation at position 58 in nsp8 (*i.e.*, K58A) is lethal for SARS-CoV (Posthuma, Zevenhoven-Dobbe, te Velhuis and Snijder, unpublished results).

The initial expression of the viral replicase enzymes and the RdRp(s) as large polyproteins and their subsequent release from these polyproteins by viral proteases is vital for nidovirus replication. Furthermore, the maturation of nsps appears to follow a conserved pattern across the nidovirus order [44,323]. Although various explanations can be offered to explain this expression strategy, it is presently believed that the above

process chiefly evolved to establish control over replicase assembly and the activation of the enzymes that are part of it. Additional fine-tuning of the molecular ratios between the nidoviral replicase proteins - and thus the regulation of activities in the replication and transcription complex (RTC) - is likely provided by a ribosomal frameshift signal encoded inside the RdRp-coding region. During translation, this structure can interact with and regulate the ribosome and thereby establish a down-regulated expression of the nsps downstream of the frameshift (including the conserved viral polymerase and helicase functions) relative to those encoded upstream [45,323]. Whether this structure also influences (*e.g.*, stalls) the viral polymerase or helicase is presently unknown.

A much overlooked element that is present on all nidoviral mRNAs and required for efficient translation is the 3' polyA tail [418]. It is presently unknown at which stage polyadenylation - which as we described in chapter 8 may be performed by SARS-CoV nsp(7+8) - is initiated on the 3' untranslated region (UTR) present in each sg mRNA and the genome. In fact, it first needs to be established whether the same RdRp that catalyses +RNA synthesis can also perform this activity - in which case polyadenylation would be continuous with extension - or whether two different RdRps, in line with polyadenylation in eukaryotic cells [415,450], are required to complete +RNA synthesis and termination. In this light, it is interesting to note that no clear CoV nsp8 homologue is present in the arterivirus genome, suggesting that both functions may be united in the arterivirus RdRp nsp9, but physically separated in CoVs. Interestingly, if we observe the RNA products in Fig. S1 of chapter 7, we may indeed reach the conclusion that EAV nsp9 is able to perform terminal transferase activity as well as copy an RNA template.

The experiments in chapter 8 suggest SARS-CoV nsp8 may initiate -RNA synthesis on the 3' terminal cytosine of the genome in addition to polyadenylation. The former activity would rule out the attachment of a polyU tail at the 5' ends of the -RNAs. However, the presence of such a tail was reported for bovine CoV -RNAs in 1991 [451] and not necessarily excluded in poly(dT) binding experiments of isolated viral dsRNAs [418], suggesting that -RNAs may become polyU-tailed after the initiation of -RNA synthesis or that -RNA synthesis is significantly different *in vivo* (due to additional protein factors or regulating signals) than could currently be assessed *in vitro*.

## Proofreading

Ranking first in genome size among all other RNA viruses, the large genomes of the corona, bafini-, toro- and roniviruses are regarded as unique products of +RNA virus evolution [58,323,437]. As mentioned in chapter 9 and discussed by Gorbalenya *et al.* [35], the faithful replication of such gigantic RNA sequences is a precarious undertaking, particularly if one assumes that their viral polymerases have the same incorporation fidelity as other viral RdRps (~1 error in every  $10^4$  bases). Interestingly, it was argued that the 3'-to-5' exonuclease function encoded in these large genomes may have facilitated

nidovirus genome expansion by 'adding' a new level of error control to the nidoviral RTC [35,437], an observation that appears to be corroborated by the difference in mutation frequency between wild-type and nsp14 exonuclease knockout CoVs [106,107].

If we consider just their basic function, replicases are already truly remarkable enzymes. Whereas most enzymes either have a relaxed substrate specificity in order to utilise a wide range of relatively similar substrates or a very strict specificity to select a single substrate from a pool of (near) homologues, polymerases essentially do both and are able to adapt their substrate specificity as they translocate along the nucleic acid template [436]. To appreciate this, one merely has to observe the following: the free energy difference in solution between the formation of Watson-Crick base pairs and non-Watson-Crick base pairs is 1-5 kcal/mol (2-8  $K_bT$ ) [452,453]. That isn't much, since the incorporation of an NTP already gives us ~11 kcal/mol (~18  $K_bT$ ). Further, if we use the above values to calculate the bare error frequency using  $\Delta\Delta G = RT\ln(k_c/k_i)$  - where  $k$  is the binding constant for correct ( $k_c$ ) and incorrect bases ( $k_i$ ),  $T$  the temperature, and  $R$  molar gas constant - we get an estimate of the basic polymerase error that is around one misincorporation in every 5-150 bases [436]. If we compare this to the observed error rate in RNA viruses (~ $10^{-3}$ - $10^{-4}$ ) or eukaryotes ( $10^{-5}$ - $10^{-9}$ ), it is clear that base pairing in itself cannot account for the selectivity in the polymerase reaction.

The polymerase is thus able to enhance the free energy differences between correct and incorrect base incorporations, a value that can be obtained from the relative rate of incorporating correct and incorrect base pairs, weighted by the concentration of each base pair [436,454]:

$$\text{relative rate} = \frac{(k_{cat} \cdot K_m)_{correct} \cdot [\text{correct base}]}{(k_{cat} \cdot K_m)_{incorrect} \cdot [\text{incorrect base}]}$$

Although these values have been obtained for the well-studied poliovirus RdRp [56,180], they are currently unknown for the nidovirus polymerases and we are thus unable to put a quantitative measure on the fidelity of the nidovirus RNA polymerase. In fact, we only have the qualitative evidence from chapter 9 and the deep-sequencing of passaged virus genomes by Eckerle *et al.* [106,107] for arguing that the CoVs encode a mechanism that has the ability to recognise and correct mismatches in the nascent RNA strand.

Given that two exonucleases have been discovered in the CoV genome and only one has so far been shown to be capable of recognising mismatches, it is of course tempting to hypothesise that the nidovirus proofreading mechanism combines both enzymes to improve the fidelity. However, one might just as well argue that they should work in separate complexes, if only to achieve a higher overall replication rate and efficiency. Here, the latter simply follows from the fact that an interplay between nsp12 and nsp14

would not only require the canonical transfer of the nascent strand from the polymerase active site ( $E_{pol}$ ) to the exonuclease active site of the RdRp ( $E_{exo}$ ) [52,436], but it would also entail a subsequent transfer from  $E_{exo}$  to the active site of the dedicated exonuclease nsp14 ( $E_{ExoN}$ ). Clearly, such a multi-step correction process would significantly slow down RNA synthesis and likely be far too elaborate and 'costly' to support just a three-fold larger genome. Interestingly, if we analyse the recently published pair-wise interaction studies between the SARS-CoV nsp8, we find that nsp14 only binds to SARS-CoV's second polymerase, nsp8, and not to nsp12 [85,289,423]. In turn, this suggests that each polymerase may have its own exonuclease to improve the overall fidelity of RNA synthesis that may itself consist of two or more separated processes. However, until we observe a difference in the relative incorporation rate between a CoV nsp12  $E_{exo}$  mutant or a wild-type CoV nsp12 in the presence of a wild-type nsp14 relative to wild-type CoV nsp12 alone, multiple explanations for the observed phenomena are still possible.

### Nidovirus RNA products

Regardless of the nidovirus replicase composition and whether it can exist in two or more molecularly and functionally different entities, the polymerases need to direct the catalysis of a 3'- and 5'-coterminally nested set of sg mRNAs. Most of these molecules serve as templates for the translation of only their 5'-proximal ORF, although some are functionally polycistronic and can thus be translated into more than one protein [58,390]. As outlined above, it is now understood that the production of these molecules is orchestrated during -RNA synthesis [265,266,416]. In addition, it is clear that it must involve both a discontinuous step, which produces sg -RNAs, and at least two continuous processes that yield i) the full-length anti-genome template for replication and ii) each type of viral mRNAs [51,265,449].

Crucial parameters that influence and ultimately characterise discontinuous RNA synthesis are the pausing/dissociation frequency of the nidovirus RdRps, the base-pairing interactions between sense and antisense transcription-regulating RNA sequences (TRSs) [455,456], and various protein factors that may switch the replicase from a continuous to a discontinuous mode or just generally stimulate template switching [457,458]. Furthermore, research into the TRSs in arteri- and coronavirus genomes revealed that the genomic sequence surrounding the TRS and the proximity of the TRS to the genomic 3' end might play a role in discontinuous RNA synthesis as well [265,267,422,459,460]. Presently, however, we do not have an estimate of the RdRp processivity, as we lack knowledge of basic parameters like the nucleotide incorporation rate under single cycle conditions (in nt/s) and the RdRp dissociation rate (in  $s^{-1}$ ). The ratio of these two parameters would easily provide us with an estimate of how far a typical nidovirus polymerase can extend a given RNA molecule before it becomes prone to dissociation, and thus how processive it will be on average. Furthermore, we could use one or both of these

relatively simple parameters to screen the influence of other nsps on the polymerase activity and thereby build up a much wider knowledgebase that could help explain the interplay and composition of the replicase. Of course, the apparent disparity between the continuous and discontinuous processes does not necessarily depend on the regulation of an inherently highly processive RdRp. In fact, the replicase may be able to synthesise long RNAs just as well by using a non-processive RdRp if the polymerase is frequently replaced with 'fresh' polymerase subunits, putatively in a fashion that resembles the polymerase turn-over of the DNA replication machinery [461].

A more direct study estimate of the replicase activity may be obtained with single-molecule force-spectroscopy studies, such as those shown in chapters 6 and 7. These experiments can provide insight into the sequence and force dependency of the enzyme under study, and the enzyme's processivity under various conditions. The data presented in chapter 7 already provide such information for the EAV helicase nsp10 - a component of the EAV replicase and involved in discontinuous RNA synthesis [403] - and demonstrate the effect of co-factors and the local sequence on nsp10's tendency to pause. It is tempting to speculate that this information gives us a glimpse of the processivity of the replicase as well if we assume that the RdRp nsp9 follows in the wake of nsp10. Indeed, given nsp9's inability to displace strands downstream of its polymerase direction itself (chapter 7, Fig. S1) it would be highly dependent on a helicase function and putatively forced to pause at regular intervals if nsp10 fails to unwind the dsRNA. On the other hand, the RdRp-helicase tandem may also be envisioned to be more processive and efficient than each is enzyme is on its own: 1) the helicase can unwind the dsRNA, thereby allowing the polymerase to use a locally single-stranded template and reach its optimal incorporation rate, while 2) the helicase can achieve a higher unwinding velocity since the polymerase may function as a moving roadblock behind the helicase and prevent it from translocating backwards and away from the unwinding fork [386,399].

### **Initiation of RNA synthesis and complex formation**

As mentioned above, all nidovirus replicase enzymes are initially part of large polyproteins. Interestingly, these polyproteins contain trans-membrane proteins that can associate the replicase with cellular membranes [58,72,294,390,462,463,464]. Although this process is often considered to be a strategy that establishes a vital (micro)-environment for viral RNA synthesis and a protection of viral replication intermediates and triphosphate-containing, uncapped RNAs from host defence mechanisms [185,228,235], it is in theory also a mechanism to control viral anti-host defence enzymes as they i) include the vital viral proteases required for polyproteins processing and the putative activation of enzymatic functions [44,88] and ii) may inadvertently disrupt cellular regulatory processes that depend on ubiquitination such as organelle biogenesis, ribosome biosynthesis and cellular transcription [465,466,467]. Furthermore, it allows the polyproteins

to be a vehicle that not only contributes to the regulation of the RdRp activity and the assembly of the replicase, it theoretically also enables them to control the initiation of RNA synthesis and thus assist in regulating the RdRp. I will discuss this in the next two sections below.

If we assume that a limited multiplicity of infection (MOI) facilitates, on average, only one infection event per cell, only a single viral genome will be released per cell. Under such conditions, the initial level of viral protein synthesis will thus be limited and the starting concentration of viral proteins correspondingly low. Consequently, without any build-in strategy to ensure that translation is immediately followed by the association of the newly synthesised viral RdRp with this singular viral genome, the likelihood of a chance encounter - even if we account for microdomain formation in the crowded cytoplasm [468,469] - is likely minute. Moreover, given the extensive protein network that surrounds the ER membrane at cellular homeostasis, it seems improbable that the effects of a single translational event and the concomitant insertion of the viral transmembrane domains would suffice to induce sufficient membrane-pairing and curvature to, *e.g.*, produce an invagination that could confine the viral genome. Furthermore, the activities of the viral replicase likely depend on several more protein factors than just the RdRp [111,457] (chapter 7), so in order to support a rapid initiation of viral RNA synthesis, the viral genome must encode signals that establish a long enough association with the nascent polyprotein to allow the formation of the most optimal platform for -RNA synthesis.

Taking the above, the polyprotein-based expression of the viral replicase proteins, and the conserved membrane-association strategies into consideration, the association of the viral genome with the RdRp or other replicase proteins may thus be achieved as follows: i) a direct integration of the transmembrane proteins into the ER membrane by and near the ER translocon (a protein complex that resides in the ER membrane and uses the surrounding phospholipids as medium for protein insertion [470,471]), ii) the production of the viral RdRp and its immediate association with the scaffold of transmembrane proteins (likely via the polyprotein), iii) and the continued association of the genome with the site of transmembrane insertion, possibly due to the presence of multiple translating ribosomes on the genome (*i.e.*, through polysome formation [472,473]). Interesting though inconclusive in light of point iii) is the observation by Sawicki *et al.* that -RNA synthesis is four times more sensitive to translation inhibition than +RNA synthesis [474].

Although association of the RdRp with the membrane scaffold is likely achieved through the covalent linkage of the RdRp to the membrane-spanning subunits in the polyproteins, activation of the RdRp may require its release from the polyproteins, particularly given the importance of the SARS-CoV N-terminal RdRp domains for activity (chapters 3 and 5). It was therefore intriguing to note that nsp8, which is expressed

at equal ratio's with the trans-membrane subunits of the replicase, is able to recognise the 3' UTR of the SARS-CoV genome and use it as template for *de novo* RNA synthesis (chapter 8).

### Initiation, continued...

The low viral protein levels that characterise the early stages of infection may also have selected for an RdRp initiation mechanism that requires as few cofactors as possible. This may, for instance, explain why the RdRp subunit of most +RNA viruses studied to date is sufficient to catalyse the condensation of ribonucleoside triphosphates *in vitro* [134,277]. As demonstrated in chapters 3 and 5, both SARS-CoV polymerases nsp12 and nsp(7+8) are active under such conditions, although it is presently not clear how well this activity reflects the incorporation rate *in vivo*. Interestingly, the hypothesised selection for a rapid initiation mechanism may also clarify why SARS-CoV nsp8 initiates dinucleotide formation in absence of a template, even as part of polyproteins nsp7-8 and nsp7-10 (chapter 5, te Velhuis and Snijder, unpublished observations). Such a feature may stabilise its binding to the genomic 3' end and thus facilitate a more rapid transition from translation to the initiation of -RNA synthesis. In addition, such a system predicts that nsp8 would function primarily *in cis*, a hypothesis that is attractively well in line with *in vivo* observations for MHV nsp8 [475].

However, among +RNA viruses like the picornaviruses and the flaviviruses, there is a tendency to use and initiate on RNA structures encoded in the 5' UTR, and not to utilise a multimeric, additional polymerase like nsp8. Poliovirus 3D<sup>pol</sup>, for instance, initiates RNA synthesis by adding uridylyl moieties to the viral protein VPg using an internal genomic region as template [197], but requires structures in the 5' UTR to coordinate its association with the polyA tail and the initiation of -RNA synthesis [197,206,207]. Flavivirus initiation by NS5 on the other hand immediately starts -RNA synthesis on a 5' UTR promoter, but then similarly relocates to the 3' end of the genome to produce a replicative intermediate [205,223,226].

A factor contributing to the differences between nidoviruses and other +RNA viruses may be the formation of ribonucleoprotein (RNP) complex that associates with the ends of the viral genome, a process that is required for proper flavivirus and picornavirus replication [197,199,223,406]. However, (RNP-based) circularisation of nidovirus genomes has been postulated as well [476] and, together with a sequence-induced folding of the genome [459], provides an attractive mechanism that brings together the body and leader TRSs in order to facilitate strand transfer during discontinuous RNA synthesis.

Presently, there is insufficient evidence to conclude that circularisation of the genome and 3' UTR-based initiation of RNA synthesis by nsp8 are mutually exclusive. A further analysis of the various protein-protein and RNA-protein interactions will, however, be a daunting, but likely also a highly rewarding task. One can think of *in vitro* studies in

which nsp8 RNA synthesis is studied in the presence of both genomic ends and one or more of the various protein co-factors that were found to interact with nsp8 in pair-wise protein-protein interaction studies [289,477]. In addition, recent advances in single-molecule FRET technologies now allow investigators to use multiple FRET pairs, which can facilitate studies of the interactions of multiple biological molecules at the same time [439]. Consequently, such an approach would provide a quantitative and real-time measure of how well RNA synthesis is performed by nsp8 [or nsp(7+8), 7-8 and 7-10] and whether the dissociation constants and incorporation rates are altered in the presence of both genomic ends or co-factors like ssRNA binding proteins, the helicase, or the TRS-unwinding nucleocapsid protein [447].

### **Finding the replicase in the membrane stacks...**

After initiation of -RNA synthesis, nidovirus RNA synthesis is generally assumed to proceed rapidly, since virus-specific radio-active signals in metabolic labelling experiments can already be detected within minutes [418]. Interestingly, as nidovirus infection progresses, the membrane-bound complexes also start to induce various membrane structures, including double membranes, CMs and DMVs [67,68,381,478]. Due to their characteristic morphology and association with nidovirus infections, they have often been used as a signature readout for the efficiency of nidovirus infections and putative sites of replication, and have therefore drawn substantial attention and research investments [70,479]. Interestingly, the involvement of DMVs in nidovirus replication has also been subject to much debate ever since electron tomography failed to visualise clear connections between the inside of SARS-CoV-induced DMVs and the cytosol [68], an observation that is in contrast with the replication vesicles of most other +RNA viruses [75,228]. Furthermore, immunolabelling of CoV-infected cells, showed that the viral nsps, including the second RdRp nsp8, preferentially co-localise with CMs instead of DMVs, suggesting that CMs may be the actual sites of RNA synthesis [68].

Conclusive evidence that establishes that CMs are the sites of RdRp activity in the infected cell is presently unavailable, however. For instance, viral dsRNA, putatively representing the replicative intermediate, has so far been mainly found inside DMVs [68]. In addition, the RNA-synthesising activity of membrane-associated RTCs isolated from infected cells has been shown to be insensitive to both nuclease and protease digestions in absence of detergents, suggesting that active RTCs are protected by their membranes and, serving as a reality check, that membranes, active RdRps and viral nucleic acid are associated with one another [290,313].

Several explanations have been offered to justify the above, ostensibly paradoxical observations for the non-overlapping locations of the viral RNA and the replicase enzymes, including analyses made for flaviviruses in which only a fraction of the replicase proteins was found to be actively contributing to RNA synthesis [68,480]. Still, if we then

assume that DMVs are indeed sealed and that they give a correct indication of the location of active nidovirus RTCs, how do we ratify the impossible diffusion of nucleoside triphosphates into these vesicles and the possibly even more challenging export of viral RNAs out of them and into the cytoplasm for translation or packaging? Furthermore, the increased stability of dsRNA molecules inside these vesicles seems at odds with the reported high turnover of negative strands in MHV infections [481].

### Concluding remarks

The outbreak of SARS caused significant economical damages and many human casualties [65,448]. However, in the wake of the 2003 pandemic, the efforts to understand CoV infection and replication increased, and the expansion of our knowledge of nidoviruses has certainly followed suit. Unfortunately, still many nidovirus secrets remain well hidden under wraps and our present understanding of the viral replicase continues to be significantly fragmented.

It is likely that future studies will strongly depend on the outcomes of the sustained development of nidovirus *in vitro* and surrogate systems. The latter may of course utilise the fact that membrane modifications can be induced by the membrane nsps alone and may offer valuable insight into the (micro)-environment that nurtures viral RNA synthesis *in vivo*. Indeed, it is presently largely ambiguous what this environment is and what the RdRp requires to perform all its documented functions *in vitro*. The studies presented in chapters 3-5 and 7-9, should therefore be interpreted with caution and only be regarded as the first steps towards the reconstitution of an active nidoviral replicase *in vitro* given that they still require significantly more quantitative analyses and extensive comparisons with *in vivo* data.

Of course, the RdRp itself should be part of further future scrutiny as well, if only to explain the integration of the exonuclease and polymerase activities in SARS-CoV nsp12 and how this affects the fidelity of the virus in cell culture. Lastly, a better understanding and identification of the nidovirus RNA structures in the 3' and 5' UTRs will likely become important as well, because they appear to play a crucial role in various activities of the viral replicase. Together, the insights that ensue from all these future endeavours will likely allow us to better comprehend what components the nidovirus replicase is made of and how they integrate to give the wide array of phenomena that we can see to date.



## REFERENCES

1. Lai MM (1995) The molecular biology of hepatitis delta virus. *Annu Rev Biochem* 64: 259-286.
2. Pelchat M, Grenier C, Perrault J (2002) Characterization of a viroid-derived RNA promoter for the DNA-dependent RNA polymerase from *Escherichia coli*. *Biochemistry* 41: 6561-6571.
3. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, *et al.* (2008) The virophage as a unique parasite of the giant mimivirus. *Nature* 455: 100-104.
4. Xiao C, Kuznetsov YG, Sun S, Hafenstein S, Kostyuchenko V, *et al.* (2009) Structural studies of the giant mimivirus. *PLoS Biol* 7: e92.
5. Attwater J, Wochner A, Pinheiro V, Coulson A, Holliger P (2010) Ice as a protocellular medium for RNA replication. *Nat Commun* 1: doi:10.1038/ncomms1076.
6. Kanavarioti A, Monnard P, Deamer D (2001) Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology* 1: 271-281.
7. Trinks H, Schroder W, Biebricher C (2005) Ice and the origin of life. *Orig Life Evol Biosph* 35: 429-445.
8. Gilbert W (1986) The RNA world. *Nature* 319: 618.
9. Cech TR (1986) A model for the RNA-catalyzed replication of RNA. *Proc Natl Acad Sci U S A* 83: 4360-4363.
10. Kuhn H, Waser J (1982) Evolution of early mechanisms of translation of genetic information into polypeptides. *Nature* 298: 585-586.
11. Kuhn H (2001) Computer-modeling origin of a simple genetic apparatus. *Proc Natl Acad Sci U S A* 98: 8620-8625.
12. Böhler C, Nielsen P, Orgel L (1995) Template switching between PNA and RNA oligonucleotides. *Nature* 376: 578-581.
13. Crick F, Orgel L (1973) Directed Panspermia. *Icarus* 19: 341-346.
14. Coogan M (2009) *A Brief Introduction to the Old Testament: The Hebrew Bible in its Context*. New York: Oxford University Press.
15. Bommarito S, Peyret N, SantaLucia J (2000) Thermodynamic paramets for DNA sequences with dangling ends. *Nucleic Acid Res* 28: 1929-1934.
16. Hammann C, Lilley D (2002) Folding and activity of the hammerhead ribozyme. *Chembiochem* 3: 690-700.
17. Scott W, Finch J, Klug A (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* 81: 991-1002.
18. Bieling P, Beringer M, Adio S, Rodnina M (2006) Peptide bond formation does not involve acid-base catalysis by ribosomal residues. *Nat Struct Mol Biol* 13: 423-428.
19. Steitz TA (2008) A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* 9: 242-253.
20. Lincoln T, Joyce GF (2009) Self-sustained replication of an RNA enzyme. *Science* 323: 1229-3122.
21. Paul N, Joyce G (2002) A self-replicating ligase ribozyme. *Proc Natl Acad Sci U S A* 99: 12733-12740.
22. Shechner DM, Grant RA, Bagby SC, Koldobskaya Y, Piccirilli JA, *et al.* (2009) Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* 326: 1271-1275.
23. Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* 70: 39-80.
24. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35: 849-857.

25. Kruger K, Grabowski PJ, Zaugg AJ, Sands J, Gottschling D, *et al.* (1989) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31: 147-157.
26. Dai L, Chai D, Gu S, Gabel J, Noskov S, *et al.* (2008) A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol Cell* 30: 472-485.
27. Bagby SC, Bergman N, Shechner D, Yen C, Bartel D (2009) A class I ligase ribozyme with reduced Mg<sup>2+</sup> dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA* 15: 2129-2146.
28. Castro C, Smidansky ED, Arnold JJ, Maksimchuk KR, Moustafa I, *et al.* (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol* 16: 212-218.
29. Bergman N, Johnston W, Bartel D (2000) Kinetic framework of ligation by an efficient RNA ligase ribozyme. *Biochemistry* 39: 3115-3123.
30. Eklund EH, Bartel DP (1996) RNA-catalysed RNA polymerization using nucleotide triphosphates. *Nature* 382: 373-376.
31. Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261: 1411-1418.
32. Zaher H, Unrau P (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* 13: 1017-1026.
33. Grosjean H (2009) DNA and RNA modifying enzymes: structure, mechanism, function and evolution; Grosjean H, editor: Landes Bioscience.
34. Monnard P, Kanavarioti A, Deamer D (2003) Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *J Am Chem Soc* 125: 13734-13740.
35. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res* 117: 17-37.
36. Lehmann E, Bruekner F, Cramer P (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* 450: 445-449.
37. Gorbalenya AE, Koonin E (1989) Viral proteins containing the NTP-binding sequence pattern. *Nucleic Acids Res* 17: 8413-8440.
38. Holmes E, Rambaut A (2004) Viral evolution and the emergence of SARS coronavirus. *Philos Trans R Soc Lond B Biol Sci* 359: 1059-1065.
39. Crotty S, Cameron CE, Andino R (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci U S A* 98: 6895-6900.
40. Beerens N, Snijder EJ (2007) An RNA Pseudoknot in the 3' End of the Arterivirus Genome Has a Critical Role in Regulating Viral RNA Synthesis. *J Virol* 81: 9426-9436.
41. Narayanan K, Makino S (2001) Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging. *J Virol* 75: 9059-9067.
42. Goebel SJ, Miller TB, Bennett CJ, Bernard KA, Masters PS (2007) A Hypervariable Region within the 3' cis-Acting Element of the Murine Coronavirus Genome Is Nonessential for RNA Synthesis but Affects Pathogenesis. *J Virol* 81: 1274-1287.
43. Proshkin S, Rahmouni AR, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328: 504-508.
44. Ziebuhr J, Snijder EJ, Gorbalenya AE (2000) Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol* 81: 853-879.
45. Brierley I, Dos Ramos FJ (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res* 119: 29-42.

46. Duffy S, Shackleton L, Holmes E (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9: 267-276.
47. Domingo E, Holland J (1997) RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 51: 151-178.
48. Walsh C, Xu G (2006) Cytosine methylation and DNA repair. *Curr Topics Microbiol Immunol* 201: 283-315.
49. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, *et al.* (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424: 99-103.
50. Boni M, Zhou Y, Taubenberger J, Holmes E (2008) Homologous recombination is very rare or absent in human influenza A virus. *J Virol* 82: 4807-4811.
51. Pasternak AO, Spaan WJ, Snijder EJ (2006) Nidovirus transcription: how to make sense...? *J Gen Virol* 80: 1403-1421.
52. Steitz T, Yin Y (2004) Accuracy, lesion bypass, strand displacement and translocation by DNA polymerases. *Philos Trans R Soc Lond B Biol Sci* 359: 17-23.
53. Beese L, Steitz T (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J* 10: 25-33.
54. Derbyshire V, Freemont P, Sanderson M, Beese L, Friedman J, *et al.* (1988) Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* 240: 199-201.
55. Mizrahi V, Benkovic P, Benkovic S (1986) Mechanism of DNA polymerase I: exonuclease/polymerase activity switch and DNA sequence dependence of pyrophosphorolysis and misincorporation reactions. *Proc Natl Acad Sci U S A* 83: 5769-5773.
56. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, *et al.* (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat Med* 6: 1375-1379.
57. Vignuzzi M, Stone J, Arnold J, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439: 344-348.
58. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, *et al.* (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 331: 991-1004.
59. Zhao Z, Li H, Wu X, Zhong Y, Zhang K, *et al.* (2004) Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol* 4: 21.
60. Yeh S, Wang H, Tsai C, Kao C, Yang J, *et al.* (2004) Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. *Proc Natl Acad Sci U S A* 101: 2542-2547.
61. Ziebuhr J (2004) Molecular biology of the severe acute respiratory syndrome coronavirus. *Current Opinion in Microbiology* 7: 412-419.
62. Ma Y, Feng Y, Liu D, Gao G (2009) Avian influenza virus, *Streptococcus suis* serotype 2, severe acute respiratory syndrome-coronavirus and beyond: molecular epidemiology, ecology and the situation in China. *Philos Trans R Soc Lond B Biol Sci* 364: 2725-2737.
63. Becker M, Graham R, Donaldson E, Rockx B, Sims A, *et al.* (2008) Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A* 105: 19944-19949.
64. Sheahan T, Rockx B, Donaldson E, Sims A, Pickles R, *et al.* (2008) Mechanisms of zoonotic severe acute respiratory syndrome coronavirus host range expansion in human airway epithelium. *J Virol* 82: 2274-2285.
65. Perlman S, Netland J (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Micro* 7: 439-450.

66. Flegel T (1997) Major viral diseases of the black tigerprawn (*Penaeus monodon*) in Thailand. *World J Microbiol Biotechnol* 13: 433-442.
67. Snijder EJ, van der Meer Y, Zevenhoven-Dobbe J, Onderwater JJM, van der Meulen J, *et al.* (2006) Ultrastructure and Origin of Membrane Vesicles Associated with the Severe Acute Respiratory Syndrome Coronavirus Replication Complex. *J Virol* 80: 5927-5940.
68. Knoops Kv, Kikkert M, Worm SHEvd, Zevenhoven-Dobbe JC, van der Meer Y, *et al.* (2008) SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLoS Biol* 6: e226.
69. Gosert R, Kanjanahaluethai A, Egger D, Bienz K, Baker SC (2002) RNA Replication of Mouse Hepatitis Virus Takes Place at Double-Membrane Vesicles. *J Virol* 76: 3697-3708.
70. Pedersen KW, van der Meer Y, Roos N, Snijder EJ (1999) Open Reading Frame 1a-Encoded Subunits of the Arterivirus Replicase Induce Endoplasmic Reticulum-Derived Double-Membrane Vesicles Which Carry the Viral Replication Complex. *J Virol* 73: 2016-2026.
71. Snijder EJ, van Tol H, Roos N, Pedersen KW (2001) Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex. *J Gen Virol* 82: 985-994.
72. van der Meer Y, Snijder EJ, Dobbe JC, Schleich S, Denison MR, *et al.* (1999) Localization of Mouse Hepatitis Virus Nonstructural Proteins and RNA Synthesis Indicates a Role for Late Endosomes in Viral Replication. *J Virol* 73: 7641-7657.
73. Prentice E, Jerome WG, Yoshimori T, Mizushima N, Denison MR (2004) Coronavirus Replication Complex Formation Utilizes Components of Cellular Autophagy. *J Biol Chem* 279: 10136-10141.
74. Snijder EJ, van der Meer Y, Zevenhoven-Dobbe J, Onderwater JJM, van der Meulen J, *et al.* (2006) Ultrastructure and Origin of Membrane Vesicles Associated with the Severe Acute Respiratory Syndrome Coronavirus Replication Complex. *J Virol* 80: 5927-5940.
75. Miller DJ, Schwartz MD, Ahlquist P (2001) Flock House Virus RNA Replicates on Outer Mitochondrial Membranes in *Drosophila* Cells. *J Virol* 75: 11664-11676.
76. Burgyan J, Rubino L, Russo M (1996) The 5'-terminal region of a tobusvirus genome determines the origin of multivesicular bodies. *J Gen Virol* 77: 1967-1974.
77. Kujala P, Ikaheimonen A, Ehsani N, Vihinen H, Auvinen P, *et al.* (2001) Biogenesis of the Semliki Forest Virus RNA Replication Complex. *J Virol* 75: 3873-3884.
78. Salonen A, Vasiljeva L, Merits A, Magden J, Jokitalo E, *et al.* (2003) Properly Folded Nonstructural Polyprotein Directs the Semliki Forest Virus Replication Complex to the Endosomal Compartment. *J Virol* 77: 1691-1702.
79. Ziebuhr J (2006) The coronavirus replicase: insights into a sophisticated enzyme machinery. *Adv Exp Med Biol* 581: 3-11.
80. Sawicki SG, Sawicki DL, Siddell SG (2007) A Contemporary View of Coronavirus Transcription. *J Virol* 81: 20-29.
81. Namy O, Moran SJ, Stuart DI, Gilbert RJC, Brierley I (2006) A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* 441: 244-247.
82. Plant EP, eacute, rez-Alvarado GC, Jacobs JL, Mukhopadhyay B, *et al.* (2005) A Three-Stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLoS Biology* 3: e172.
83. Baretto N, Jukneliene D, Ratia K, Chen Z, Mesecar A, *et al.* (2006) Deubiquitinating activity of the SARS-CoV papain-like protease. *Adv Exp Med Biol* 581: 43-48.
84. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1989) Coronavirus genome: prediction of putative functional domain in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res* 17: 4847-4861.

85. Imbert I, Snijder EJ, Dimitrova M, Guillemot J-C, Lécine P, *et al.* (2008) The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. *Virus Res* 133: 136-148.
86. Wojdyla J, Manolaridis I, van Kasteren P, Kikkert M, Snijder E, *et al.* (2010) Papain-like protease 1 from transmissible gastroenteritis virus: crystal structure and enzymatic activity toward viral and cellular substrates. *J Virol* 84: 10063-10073.
87. Barretto N, Jukneliene D, Ratia K, Chen Z, Mesecar A, *et al.* (2005) The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *J Virol* 79: 15189-15198.
88. Frias-Staheli N, Giannakopoulos NV, Kikkert M, Taylor SL, Bridgen A, *et al.* (2007) Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe* 2: 404-416.
89. Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, *et al.* (2003) Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res* 31: 7117-7130.
90. Clercq ED (2004) Antivirals and antiviral strategies. *Nature Reviews Microbiology* 2: 704-720.
91. Tsai C-H, Lee P-Y, Stollar V, Li M-L (2006) Antiviral Therapy Targeting Viral Polymerase. *Current Pharmaceutical Design* 12: 1339-1355.
92. Webster DP, Kleneman P, Collier J, Jeffery KJ (2009) Development of novel treatments for hepatitis C. *Lancet Infect Dis* 9: 108-117.
93. Soares MM, King SW, Thorpe PE (2008) Targeting inside-out phosphatidylserine as a therapeutic strategy for viral diseases. *Nat Med* 14: 1357-1362.
94. Graci JD, Cameron CE (2005) Mechanisms of action of ribavirin against distinct viruses. *Rev Med Virol* 16: 37-48.
95. Hall CB, Walsh EE, Hruska JF, Betts RF, Hall WJ (1983) Ribavirin treatment of experimental respiratory syncytial virus infection. A controlled double-blind study in young adults. *JAMA* 249: 2666-2670.
96. Oxford JS (1975) Inhibition of the replication of influenza A and B viruses by a nucleoside analogue (Ribavirin). *J Gen Virol* 28: 409-414.
97. Eriksson B, Helgstrand E, Johansson NG, Larsson A, Misiorny A, *et al.* (1977) Inhibition of influenza virus ribonucleic acid polymerase by ribavirin triphosphate. *Antimicrob Agents Chemother* 11: 946-951.
98. Huggins JW, Hsiang CM, Cosgriff TM, Guang MY, Smith JI, *et al.* (1991) Prospective, double-blind, concurrent, placebo-controlled clinical trial of intravenous ribavirin therapy of hemorrhagic fever with renal syndrome. *J Infect Dis* 164: 1119-1127.
99. Chung DH, Sun Y, Perker W, Arterburn J, Bartolucci A, *et al.* (2007) Ribavirin reveals a lethal threshold of allowable mutation frequency for hantaan virus. *J Virol* 81: 11722-11729.
100. Katz E, Margalith E, Winer B (1976) Inhibition of Vaccinia Virus Growth by the Nucleoside Analogue 1- $\beta$ -D-Ribofuranosyl-1,2,4-Triazole-3-Carboxamide (Virazole, Ribavirin). *J Gen Virol* 32: 327-330.
101. Rankin JT, Eppes SB, Antczak JB, Joklik WK (1989) Studies on the mechanism of the antiviral activity of ribavirin against reovirus. *Virology* 168: 147-158.
102. Tan EL, Ooi EE, Lin CY, Tan HC, Ling AE, *et al.* (2004) Inhibition of SARS coronavirus infection in vitro with clinically approved drugs. *Emerg Infect Dis* 10: 581-586.
103. Stockman LJ, Bellamy R, Garner P (2006) SARS: systematic review of treatment effects. *PLoS Med* 3: e343.
104. Imbert I, Guillemot J, Bourhis J, Bussetta C, Coutard B, *et al.* (2006) A second, non-canonical RNA-dependent RNA polymerase in SARS Coronavirus. *EMBO* 25: 4933-4942.

105. Minskaia E, Hertzog T, Gorbalenya AE, Campanacci V, Cambillau C, *et al.* (2006) Discovery of an RNA virus 3' to 5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci USA* 103: 5108-5113.
106. Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR (2007) High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol* 81: 12135-12144.
107. Eckerle L, Becker M, Halpin R, Li K, Venter E, *et al.* (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog* 6: e1000896.
108. Ivanov KA, Thiel V, Dobbe JC, van der Meer Y, Snijder EJ, *et al.* (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol* 78: 5619-5632.
109. Seybert A, Hegyi A, Sidell S, Ziebuhr J (2000) The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA* 6: 1056-1068.
110. Seybert A, van Dinten L, Snijder E, Ziebuhr J (2000) Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. *J Virol* 74: 9586-9593.
111. van Dinten LC, van Tol H, Gorbalenya AE, Snijder EJ (2000) The Predicted Metal-Binding Region of the Arterivirus Helicase Protein Is Involved in Subgenomic mRNA Synthesis, Genome Replication, and Virion Biogenesis. *J Virol* 74: 5213-5223.
112. Niehl A, Heinlein M (2009) Impact of RNA virus infection on plant cell function and evolution. *Ann N Y Acad Sci* 1178: 120-128.
113. Sharp M (2002) Origins of human virus diversity. *Cell* 108: 305-312.
114. Sverdlov ED (2000) Retroviruses and primate evolution. *Bioessays* 22: 161-171.
115. Taylor DJ, Bruenn JA (2009) The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC Biol* 18: 88.
116. Arnaud F, Varela M, Spencer T, Palmarini M (2008) Coevolution of endogenous betaretroviruses of sheep and their host. *Cell Mol Life Sci* 65: 3422-3432.
117. Dunlap K, Palmarini M, Varela M, Burghardt R, Hayashi K, *et al.* (2006) Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc Natl Acad Sci U S A* 103: 14390-14395.
118. Pijlman GP, Suhrbier A, Khromykh AA (2006) Kunjin virus replicons: an RNA-based, non-cytopathic viral vector system for protein production, vaccine and gene therapy applications. *Expert Opin Biol Ther* 6: 135-145.
119. Russell SJ (2002) RNA viruses as virotherapy agents. *Cancer Gene Ther* 9: 961-966.
120. Palese P, Zheng H, Engelhardt OG, Pleschka S, García-Sastre A (1996) Negative-strand RNA viruses: Genetic engineering and applications. *Proc Natl Acad Sci U S A* 93: 11354-11358.
121. Cavazzana-Calvo M, Hacein-Bey S, Basile G, et al. (2000) Gene Therapy of Human Severe Combined Immunodeficiency (SCID)-X1 Disease. *Science* 288: 669-672.
122. Baltimore D (1971) Expression of animal virus genomes. *Bacteriol Rev* 35: 235-241.
123. Fraaij PL, Bodewes R, Osterhaus AD, Rimmelzwaan GF (2011) The ins and outs of universal childhood influenza vaccination. *Future Microbiol* 6: 1171-1184.
124. Medina RA, García-Sastre A (2011) Influenza A viruses: new research developments. *Nat Rev Microbiol* 9: 590-603.
125. Bruenn JA (2003) A structural and primary sequence comparison of the viral RNA-dependent RNA polymerase. *Nucleic Acids Res* 31: 1821-1829.
126. Bruenn JA (1991) Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res* 19: 217-226.

127. Meselson M, Stahl FW (1958) The replication of DNA in *Escherichia coli*. Proc Natl Acad Sci U S A 44: 671-682.
128. Richardson CC (1983) Bacteriophage T7: minimal requirements for the replication of a duplex DNA molecule. Cell 33: 315-317.
129. Erhard KJ, Stonaker J, Parkinson S, Lim J, Hale C, et al. (2009) RNA polymerase IV functions in paramutation in *Zea mays*. Science 323: 1201-1205.
130. Werner F, Grohman D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Micro 9: 85-98.
131. Wierzbicki A, Haag J, Pikaard C (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135: 635-648.
132. Gorbalenya AE, Pringle FM, Zeddam J, Luke BT, Cameron CE, et al. (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J Mol Biol 324: 47-62.
133. Lescar J, Canard B (2009) RNA-dependent RNA polymerases from flaviviruses and Picornaviridae. Curr Opin Struct Biol 19: 759-767.
134. van Dijk AA, Makeyev EV, Bamford DH (2004) Initiation of viral RNA-dependent RNA polymerization. J Gen Virol 85: 1077-1093.
135. Ng KK, Arnold JJ, Cameron CE (2008) Structure-function relationships among RNA-dependent polymerases. Curr Top Microbiol Immunol 320: 137-156.
136. Delarue M, Poch O, Tordo N, Moras D, Argos P (1990) An attempt to unify the structure of polymerases. Protein Eng 3: 461-467.
137. Poch O, Sauvaget I, Delarue M, Tordo N (1989) Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J 8: 3867-3874.
138. Butcher SJ, Grimes JM, Makeyev EV, Bamford DH, Stuart DI (2001) A mechanism for initiating RNA-dependent RNA polymerization. Nature 410: 235-240.
139. Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, et al. (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nat Struct Biol 6: 937-943.
140. Braithwaite DK, Ito J (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res 21: 787-802.
141. Ollis DL, Brick P, Hamlin R, Xuong NG, Steitz TA (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature 313: 762-766.
142. Steitz TA (1999) DNA polymerases: structural diversity and common mechanisms. J Biol Chem 274: 17395-17398.
143. Wang J, Sattar AK, Wang CC, Karam JD, Konigsberg WH, et al. (1997) Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. Cell 89: 1087-1099.
144. Gohara DW, Crotty S, Arnold JJ, Yoder JD, Andino R, et al. (2000) Poliovirus RNA-dependent RNA polymerase (3Dpol): structural, biochemical, and biological analysis of conserved structural motifs A and B. J Biol Chem 275: 25523-25532.
145. Yap TL, Xu T, Chen Y-L, Malet H, Egloff M-P, et al. (2007) Crystal Structure of the Dengue Virus RNA-Dependent RNA Polymerase Catalytic Domain at 1.85-Angstrom Resolution. J Virol 81: 4753-4765.
146. Choi KH, Rossmann MG (2009) RNA-dependent RNA polymerases from *Flaviviridae*. Curr Opin Struct Biol 19: 746-751.

147. Chinnaswamy S, Yarbrough I, Palaninathan S, Kumar CT, Vijayaraghavan V, *et al.* (2008) A locking mechanism regulates RNA synthesis and host protein interaction by the hepatitis C virus polymerase. *J Biol Chem* 283: 20535–20546.
148. Chinnaswamy S, Murali A, Li P, Fujisaki K, Kao CC (2010) Regulation of *De Novo*-Initiated RNA Synthesis in Hepatitis C Virus RNA-Dependent RNA Polymerase by Intermolecular Interactions. *J Virol* 84: 5923–5935.
149. McDonald WF, Klemperer N, Traktman P (1997) Characterization of a processive form of the vaccinia virus DNA polymerase. *Virology* 234: 168–175.
150. Boehmer PE, Lehman IR (1997) Herpes simplex virus DNA replication. *Annu Rev Biochem* 66: 347–383.
151. Olson MW, Wang Y, Elder RH, Kaguni LS (1995) Subunit structure of mitochondrial DNA polymerase from *Drosophila* embryos. Physical and immunological studies. *J Biol Chem* 270: 28931–28937.
152. Kelman Z, Hurwitz J, O'Donnell M (1998) Processivity of DNA polymerases: two mechanisms, one goal. *Structure* 6: 121–125.
153. Zhai Y, Sun F, Li X, Pang H, Xu X, *et al.* (2005) Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat Struct Mol Biol* 12: 980–986.
154. te Velthuis AJ, Arnold JJ, Cameron CE, van den Worm SH, Snijder EJ (2010) The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res* 38: 203–214.
155. Imbert I, Guillemot J, Bourhis J, Bussetta C, Coutard B, *et al.* (2006) A second, non-canonical RNA-dependent RNA polymerase in SARS Coronavirus. *EMBO J* 25: 4933–4942.
156. te Velthuis AJ, van den Worm SH, Snijder EJ (2011) The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both *de novo* initiation and primer extension. *Nucleic Acids Res*: doi:10.1093/nar/gkr1893.
157. Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, *et al.* (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* 458: 909–913.
158. Ruigrok RW, Crepin T, Hart DJ, Cusack S (2010) Towards an atomic resolution understanding of the influenza virus replication machinery. *Curr Opin Struct Biol* 20: 104–113.
159. Kawaguchi A, Momose F, Nagata K (2011) Replication-Coupled and Host Factor-Mediated Encapsidation of the Influenza Virus Genome by Viral Nucleoprotein. *J Virol* 85: 6197–6204.
160. Torreira E, Schoehn G, Fernandez Y, Jorba N, Ruigrok RW, *et al.* (2007) Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic Acid Res* 35: 3774–3783.
161. Lai VC, Kao CC, Ferrari E, Park J, Uss AS, *et al.* (1999) Mutational analysis of bovine viral diarrhea virus RNA-dependent RNA polymerase. *J Virol* 73: 10129–10136.
162. Lohmann V, Korner F, Herian U, Bartenschlager R (1997) Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J Virol* 71: 8416–8428.
163. Eglhoff MP, Benarroch D, Selisko B, Romette JL, Canard B (2002) An RNA cap (nucleoside-2'-O)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J* 21: 2757–2768.
164. Lee HR, Helquist SA, Kool ET, Johnson KA (2007) Base pair hydrogen bonds are essential for proof-reading selectivity by the human mitochondrial DNA polymerase. *J Biol Chem* 283: 14411–14416.
165. Ibarra B, Chemla YR, Plyasunov S, Smith SB, Lázaro JM, *et al.* (2009) Proofreading dynamics of a processive DNA polymerase. *EMBO J* 28: 2794–2802.
166. Blanco L, Salas M (1985) Characterization of a 3'→5' exonuclease activity in the phage phi 29-encoded DNA polymerase. *Nucleic Acids Res* 13: 1239–1249.

167. Soengas MS, Esteban JA, Lázaro JM, Bernad A, Blasco MA, *et al.* (1992) Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. *EMBO J* 11: 4227-4237.
168. Bernad A, Blanco L, Lazaro JM, Martin P, Salas M (1989) A conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* 59: 219-228.
169. Derbyshire V, Grindley ND, Joyce CM (1991) The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid at the active site to the reaction. *EMBO J* 10: 17-24.
170. Drake J, Holland J (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* 96: 13910-13913.
171. Hostomsky Z, Hostomska Z, Fu TB, Taylor J (1992) Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. *J Virol* 66: 3179-3182.
172. Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, *et al.* (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 458: 914-918.
173. Liu S, Abbondanzieri EA, Rausch JW, Le Grice SF, Zhuang X (2008) Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science* 322: 1092-1097.
174. Abbondanzieri EA, Bokinsky G, Rausch JW, Zhang JX, Le Grice SF, *et al.* (2008) Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* 453: 184-189.
175. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, *et al.* (2009) Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition. *J Mol Biol* 385: 693-713.
176. Joyce CM (1989) How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow Fragment). *J Biol Chem* 264: 10858-10866.
177. Moustafa IM, Shen H, Morton B, Colina CM, Cameron CE (2011) Molecular dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity. *J Mol Biol* 410: 159-181.
178. Ren Z, Wang H, Ghose R (2010) Dynamics on multiple timescales in the RNA-directed RNA polymerase from the cystovirus  $\phi$ 6. *Nucleic Acids Res* 38: 5105-5118.
179. Patel SS, Wong I, Johnson KA (1991) Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* 30: 511-525.
180. Arnold JJ, Cameron CE (2004) Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg<sup>2+</sup>. *Biochemistry* 43: 5126-5137.
181. Xiong Y, Deng J, Sudarsanakumar C, Sundaralingam M (2001) Crystal structure of an RNA duplex r(gugucgac)(2) with uridine bulges. *J Mol Biol* 313: 573-582.
182. Nomoto A, Lee YF, Wimmer E (1976) The 5' end of poliovirus mRNA is not capped with m<sup>7</sup>G(5')ppp(5')Np. *Proc Natl Acad Sci U S A* 73: 375-380.
183. Yogo Y, Wimmer E (1972) Polyadenylic acid at the 3'-terminus of poliovirus RNA. *Proc Natl Acad Sci U S A* 69: 1877-1882.
184. Thurner C, Witwer C, Hofacker IL, Stadler PF (2004) Conserved RNA secondary structures in Flaviviridae genomes. *J Gen Virol* 85: 1113-1124.
185. Ahlquist P (2002) RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science* 296: 1270-1273.
186. Bartenschlager R, Lohmann V (2000) Replication of hepatitis C virus. *J Gen Virol* 81: 1631-1648.

187. Tchatalbachev S, Flick R, Hobom G (2001) The packaging signal of influenza viral RNA molecules. *RNA* 7: 979-989.
188. Flick R, Hobom G (1999) Interaction of influenza virus polymerase with viral RNA in the 'corkscrew' conformation. *J Gen Virol* 80: 2565-2572.
189. Hsu MT, Parvin JD, Gupta S, Krystal M, Palese P (1987) Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc Natl Acad Sci U S A* 84: 8140-8144.
190. Deng T, Vreede F, Brownlee GG (2006) Different *de novo* initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J Virol* 80: 2337-2348.
191. Flick R, Neumann G, Hoffmann E, Neumeier E, Hobom G (1996) Promoter elements in the influenza vRNA terminal structure. *RNA* 2: 1046-1057.
192. Kainov DE, Lisal J, Bamford DH, Tuma R (2004) Packaging motor from double-stranded RNA bacteriophage  $\phi$ 12 acts as an obligatory passive conduit during transcription. *Nucleic Acid Res* 32: 3515-3521.
193. Poranen MM, Tuma R (2004) Self-assembly of double-stranded RNA bacteriophages. *Virus Res* 101: 93-100.
194. Pirttimaa MJ, Paatero AO, Frilander MJ, Bamford DH (2002) Nonspecific nucleoside triphosphatase P4 of double-stranded RNA bacteriophage  $\phi$ 6 is required for single-stranded RNA packaging and transcription. *J Virol* 72: 10122-10127.
195. Makeyev EV, Bamford DH (2000) Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage  $\phi$ i6. *EMBO J* 19: 124-133.
196. Beerens N, Selisko B, Ricagno S, Imbert I, van der Zanden L, *et al.* (2007) *De Novo* Initiation of RNA Synthesis by the Arterivirus RNA-Dependent RNA Polymerase. *J Virol* 81: 8384-8395.
197. Vogt DA, Andino R (2010) An RNA element at the 5'-end of the poliovirus genome functions as a general promoter for RNA synthesis. *PLoS Pathog* 6: e1000936.
198. Gamarnik AV, Andino R (1997) Two functional complexes formed by KH domain containing proteins with the 5' noncoding region of poliovirus RNA. *RNA* 3: 882-892.
199. Andino R, Rieckhof GE, Achacoso PL, Baltimore D (1993) Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *EMBO J* 12: 3587-3598.
200. Andino R, Rieckhof GE, Baltimore D (1990) A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell* 63: 369-380.
201. Herold J, Andino R (2001) Poliovirus RNA replication requires genome circulization through a protein bridge. *Mol Cell* 7: 581-591.
202. Lemay JF, Lemieux C, St-André O, Bachand F (2010) Crossing the borders: poly(A)-binding proteins working on both sides of the fence. *RNA Biol* 7: 291-295.
203. Mangus DA, Evans MC, Jacobson A (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* 4: 223.
204. Steil BP, Barton DJ (2009) Conversion of VPg into VPgUpUOH before and during poliovirus negative-strand RNA synthesis. *J Virol* 83: 12660-12670.
205. Paul AV, Yin J, Mugavero J, Rieder E, Liu Y, *et al.* (2003) A "slide-back" mechanism for the initiation of protein-primed RNA synthesis by the RNA polymerase of poliovirus. *J Biol Chem* 278: 43951-43960.
206. Paul AV, van Boom JH, Filippov D, Wimmer E (1998) Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature* 393: 280-284.

207. Paul AV, Rieder E, Kim DW, van Boom JH, Wimmer E (2000) Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *J Virol* 74: 10359-10370.
208. Sharma N, O'Donnell BJ, Flanagan JB (2005) 3'-Terminal sequence in poliovirus negative-strand templates is the primary cis-acting element required for VPgUpU-primed positive-strand initiation. *J Virol* 79: 3565-3577.
209. Morasco BJ, Sharma N, Parilla J, Flanagan JB (2003) Poliovirus cre(2C)-dependent synthesis of VPgUpU is required for positive- but not negative-strand RNA synthesis. *J Virol* 77: 5136-5144.
210. Ferrer-Orta C, Arias A, Agudo R, Pérez-Luque R, Escarmís C, *et al.* (2006) The structure of a protein primer-polymerase complex in the initiation of genome replication. *EMBO J* 25: 880-888.
211. Teterina N, Kean KM, Gorbalenya AE, Agol VI, Girard M (1992) Analysis of the functional significance of amino acid residues in the putative NTP-binding pattern of the poliovirus 2C protein. *J Gen Virol* 73: 1977-1986.
212. Banerjee R, Tsai W, Kim W, Dasgupta A (2001) Interaction of poliovirus-encoded 2C/2BC polypeptides with the 3' ternegative negative-strand cloverleaf requires an intact stem-loop b. *Virology* 280: 41-51.
213. Banerjee R, Echeverri A, Dasgupta A (1997) Poliovirus-encoded 2C polypeptide specifically binds to the 3'-terminal sequences of viral negative-strand RNA. *J Virol* 71: 9570-9578.
214. Adams P, Kandiah E, Effantin G, Steven AC, Ehrenfeld E (2009) Poliovirus 2C protein forms homooligomeric structures required for ATPase activity. *J Biol Chem* 284: 22012-22221.
215. Teterina NL, Gorbalenya AE, Egger D, Bienz K, Ehrenfeld E (1997) Poliovirus 2C protein determinants of membrane binding and rearrangements in mammalian cells. *J Virol* 71: 8962-8972.
216. Echeverri AC, Dasgupta A (1995) Amino terminal regions of poliovirus 2C protein mediate membrane binding. *Virology* 208: 540-553.
217. Hardy CD, Cozzarelli NR (2005) A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol Microbiol* 57: 1636-1652.
218. Travers A, Muskhelishvili G (2007) A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep* 8: 147-151.
219. Schaarschmidt D, Baltin J, Stehle IM, Lipps HJ, Knippers R (2004) An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J* 23: 191-201.
220. Brinton MA, Fernandez AV, Dispoto JH (1986) The 3'-nucleotides of flavivirus genomic RNA form a conserved secondary structure. *Virology* 153: 113-121.
221. Proutski V, Gould EA, Holmes EC (1997) Secondary structure of the 3' untranslated region of flaviviruses: similarities and differences. *Nucleic Acid Res* 25: 1194-1202.
222. Khromykh AA, Kondratieva N, Sgro Y, Palmenberg A, Westaway EG (2003) Significance in Replication of the Terminal Nucleotides of the Flavivirus Genome. *J Virol* 77: 10623-10629.
223. Filomatori CV, Lodeiro MF, Alvarez DE, Samsa MM, Pietrasanta L, *et al.* (2006) A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev* 20: 2238-2249.
224. Nomaguchi M, Ackermann M, Yon C, You S, Padmanabhan R (2003) *De novo* synthesis of negative-strand RNA by Dengue virus RNA-dependent RNA polymerase in vitro: nucleotide, primer, and template parameters. *J Virol* 77: 8831-8842.
225. Villordo SM, Alvarez DE, Gamarnik AV (2010) A balance between circular and linear forms of the dengue virus genome is crucial for viral replication. *RNA* 16: 2325-2335.
226. Filomatori CV, Iglesias NG, Villordo SM, Alvarez DE, Gamarnik AV (2011) RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. *J Biol Chem* 286: 6929-6939.

227. Engelhardt OG, Fodor E (2006) Functional association between viral and cellular transcription during influenza virus infection. *Rev Med Virol* 16: 329-345.
228. Salonen A, Ahola T, Kaariainen L (2005) Viral RNA replication in association with cellular membranes. *Curr Top Microbiol Immunol* 285: 139-173.
229. Miller DJ, Schwartz MD, Dye BT, Ahlquist P (2003) Engineered retargeting of viral RNA replication complexes to an alternative intracellular membrane. *J Virol* 77: 12193-12202.
230. Knoops K, Swett-Tapia C, van den Worm SH, Te Velthuis AJ, Koster AJ, *et al.* (2010) Integrity of the early secretory pathway promotes, but is not required for, severe acute respiratory syndrome coronavirus RNA synthesis and virus-induced remodeling of endoplasmic reticulum membranes. *Journal of virology* 84: 833-846.
231. Matto M, Sklan EH, David N, Melamed-Book N, Casanova JE, *et al.* (2011) Role for ADP ribosylation factor 1 in the regulation of hepatitis C virus replication. *J Virol* 85: 946-956.
232. Belov GA, Feng Q, Nikovics K, Jackson CL, Ehrenfeld E (2008) A critical role of a cellular membrane traffic protein in poliovirus RNA replication. *PLoS Pathog* 4: e1000216.
233. Posthuma CC, Pedersen KW, Lu Z, Joosten RG, Roos N, *et al.* (2008) Formation of the arterivirus replication/transcription complex: a key role for nonstructural protein 3 in the remodeling of intracellular membranes. *J Virol* 82: 4480-4491.
234. Paul D, Romero-Brey I, Gouttenoire J, Stoitsova S, Krijnse-Locker J, *et al.* (2011) NS4B self-interaction through conserved C-terminal elements is required for the establishment of functional hepatitis C virus replication complexes. *J Virol* 85: 6963-6976.
235. Ahlquist P, Noueir AO, Lee W-M, Kushner DB, Dye BT (2003) Host Factors in Positive-Strand RNA Virus Genome Replication. *J Virol* 77: 8181-8186.
236. Richards OC, Ehrenfeld E (1998) Effects of Poliovirus 3AB Protein on 3D Polymerase-catalyzed Reaction. *J Biol Chem* 273: 12832-12840.
237. Hsu NY, Ilnytska O, Belov G, Santiana M, Chen YH, *et al.* (2010) Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. *Cell* 141: 799-811.
238. Kopek BG, Perkins G, Miller DJ, Ellisman MH, Ahlquist P (2007) Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. *PLoS Biol* 5: e220.
239. Perera R, Daijogo S, Walter BL, Nguyen JH, Semler BL (2007) Cellular protein modification by poliovirus: the two faces of poly(rC)-binding protein. *J Virol* 81: 8919-8932.
240. Kuyumcu-Martinez NM, Joachims M, Lloyd RE (2002) Efficient cleavage of ribosome-associated poly(A)-binding protein by enterovirus 3C protease. *J Virol* 76: 2062-2074.
241. Gamarnik AV, Andino R (1998) Switch from translation to RNA replication in a positive-stranded RNA virus. *Genes Dev* 12: 229-2304.
242. Eckert KA, Kunkel TA (1993) Fidelity of DNA synthesis catalyzed by human DNA polymerase alpha and HIV-1 reverse transcriptase: effect of reaction pH. *Nucleic acids research* 21: 5212-5220.
243. Arnold JJ, Ghosh SK, Cameron CE (1999) Poliovirus RNA-dependent RNA Polymerase (3Dpol). Divalent cation modulation of primer, template and nucleotide selection. *J Biol Chem* 274: 37060-37069.
244. Crotty S, Cameron CE, Andino R (2002) Ribavirin's antiviral mechanism of action: lethal mutagenesis? *J Mol Med* 80: 86-95.
245. Guan H, Simon AE (2000) Polymerization of nontemplate bases before transcription initiation at the 3' ends of templates by an RNA-dependent RNA polymerase: an activity involved in 3' end repair of viral RNAs. *Proc Natl Acad Sci U S A* 97: 12451-12456.
246. Teramoto T, Kohno Y, Mattoo P, Markoff L, Falgout B, *et al.* (2008) Genome 3'-end repair in dengue virus type 2. *RNA* 14: 2645-2656.

247. Neufeld KL, Galarza JM, Richards OC, Summers DF, Ehrenfeld E (1994) Identification of terminal adenyllyl transferase activity of the poliovirus polymerase 3Dpol. *J Virol* 68: 5811-5818.
248. Poranen MM, Koivunen MR, Bamford DH (2008) Nontemplated terminal nucleotidyltransferase activity of double-stranded RNA bacteriophage phi6 RNA-dependent RNA polymerase. *J Virol* 82: 9254-9264.
249. Graham SC, Sarin LP, Bahar MW, Myers RA, Stuart DI, *et al.* (2011) The N-terminus of the RNA polymerase from infectious pancreatic necrosis virus is the determinant of genome attachment. *PLoS Pathog* 7: e1002085.
250. Tomar S, Hardy RW, Smith JL, Kuhn RJ (2006) Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenyllyltransferase activity. *J Virol* 80: 9962-9969.
251. van Ooij MJ, Polacek C, Glaudemans DH, Kuijpers J, van Kuppeveld FJ, *et al.* (2006) Polyadenylation of genomic RNA and initiation of antigenomic RNA in a positive-strand RNA virus are controlled by the same cis-element. *Nucleic Acids Res* 34: 2953-2965.
252. Li X, Palese P (1994) Characterization of the polyadenylation signal of influenza virus RNA. *J Virol* 68: 1245-1249.
253. Luo GX, Luytjes W, Enami M, Palese P (1991) The polyadenylation signal of influenza virus RNA involves a stretch of uridines followed by the RNA duplex of the panhandle structure. *J Virol* 65: 2861-2867.
254. Kunkel T, Burgers P (2008) Dividing the workload of at the eukaryotic replication fork. *Mol Cell* 10: 521-527.
255. Yang W (2003) Damage repair DNA polymerases. *Curr Opin Struct Biol* 13: 23-30.
256. Prakash S, Johnson R, Prakash L (2005) Eukaryotic translesion synthesis. *Annu Rev Biochem* 74: 317-353.
257. Lovett ST (2007) Polymerase switching in DNA replication. *Mol Cell* 27: 523-526.
258. Spagnolo JF, Rossignol E, Bullitt E, Kirkegaard K (2010) Enzymatic and nonenzymatic functions of viral RNA-dependent RNA polymerases within oligomeric arrays. *RNA* 16: 382-393.
259. Lyle JM, Bullitt E, Bienz K, Kirkegaard K (2002) Visualization and functional analysis of RNA-dependent RNA polymerase lattices. *Science* 296: 2218-2222.
260. Högbom M, Jäger K, Robel I, Unge T, Rohayem J (2009) The active form of the norovirus RNA-dependent RNA polymerase is a homodimer with cooperative activity. *J Gen Virol* 90: 281-291.
261. Wang J, Smerdon SJ, Jäger J, Kohlstaedt LA, Rice PA, *et al.* (1991) Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. *Proc Natl Acad Sci U S A* 91: 7242-7246.
262. Jorba N, Coloma R, Ortin J (2009) Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog* 5: e1000462.
263. Luo G, Hamatake RK, Mathis DM, Racela J, Rigat KL, *et al.* (2000) *De Novo* Initiation of RNA Synthesis by the RNA-Dependent RNA Polymerase (NS5B) of Hepatitis C Virus. *J Virol* 74: 851-863.
264. Miller WA, Koev G (2000) Synthesis of subgenomic RNAs by positive-strand RNA viruses. *Virology* 273: 1-8.
265. Sawicki SG, Sawicki DL, Siddell SG (2007) A Contemporary View of Coronavirus Transcription. *J Virol* 81: 20-29.
266. Sawicki SG, Sawicki DL (1995) Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv Exp Med Biol* 380: 499-506.
267. van Marle G, Dobbe JC, Gultyaev AP, Luytjes W, Spaan WJM, *et al.* (1999) Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proc Natl Acad Sci U S A* 96: 12056-12061.

268. Bartlam M, Yang H, Rao Z (2006) Structural insights into SARS coronavirus proteins. *Curr Opin Struct Biol* 15: 664-672.
269. Ivanov KA, Hertzog T, Rozanov M, Bayer S, Thiel V, *et al.* (2004) Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc Natl Acad Sci USA* 101: 12694-12699.
270. Egloff MP, Ferron F, Campanacci V, Longhi S, Rancurel C, *et al.* (2004) The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc Natl Acad Sci USA* 101: 3792-3796.
271. Su D, Lou Z, Sun F, Zhai Y, Yang H, *et al.* (2006) Dodecamer Structure of Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein nsp10. *J Virol* 80: 7902-7908.
272. Decroly E, Imbert I, Coutard B, Bouvet M, Selisko B, *et al.* (2008) Coronavirus Nonstructural Protein 16 Is a Cap-0 Binding Enzyme Possessing (Nucleoside-2'O)-Methyltransferase Activity. *J Virol* 82: 8071-8084.
273. Chen Y, Cai H, Pan Ja, Xiang N, Tien P, *et al.* (2009) Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci USA* 106: 3484-3489.
274. Bournsnell MEG, Brown TD, Foulds IJ, Green PF, Tomley FM, *et al.* (1987) Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J Gen Virol* 68: 57-77.
275. Cheng A, Zhang W, Xie Y, Jiang W, Arnold E, *et al.* (2005) Expression, purification, and characterization of SARS coronavirus RNA polymerase. *Virology* 335: 165-176.
276. De Clercq E (2004) Antivirals and antiviral strategies. *Nat Rev Microbiol* 2: 704-720.
277. Gohara D, Ha C, Kumar S, Ghosh B, Arnold J, *et al.* (1999) Production of "authentic" poliovirus RNA-dependent RNA polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in *Escherichia coli*. *Protein Expr Purif* 17: 128-138.
278. Thompson AA, Peersen OB (2004) Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. *EMBO J* 23: 3462-3471.
279. Castro C, Smidansky E, Maksimchuk KR, Arnold JJ, Korneeva VS, *et al.* (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci USA* 104: 4267-4272.
280. Yang W, Lee JY, Nowotny M (2006) Making and Breaking Nucleic Acids: Two-Mg<sup>2+</sup>-Ion Catalysis and Substrate Specificity. *Mol Cell* 22: 5-13.
281. Jablonski SA, Morrow CD (1995) Mutation of the aspartic acid residues of the GDD sequence motif of poliovirus RNA-dependent RNA polymerase results in enzymes with altered metal ion requirements for activity. *J Virol* 69: 1532-1539.
282. Vazquez AL, Alonso JMM, Parra F (2000) Mutation Analysis of the GDD Sequence Motif of a Calicivirus RNA-Dependent RNA Polymerase. *J Virol* 74: 3888-3891.
283. Deng T, Sharps J, Fodor E, Brownlee GG (2005) In Vitro Assembly of PB2 with a PB1-PA Dimer Supports a New Model of Assembly of Influenza A Virus Polymerase Subunits into a Functional Trimeric Complex. *J Virol* 79: 8669-8674.
284. Arnold JJ, Cameron CE (1999) Poliovirus RNA-dependent RNA Polymerase (3Dpol) Is Sufficient for Template Switching in Vitro. *J Biol Chem* 274: 2706-2716.
285. Deming DJ, Graham RL, Denison MR, Baric RS (2007) Processing of Open Reading Frame 1a Replicase Proteins nsp7 to nsp10 in Murine Hepatitis Virus Strain A59 Replication. *J Virol* 81: 10280-10291.
286. Beckman MTL, Kirkegaard K (1998) Site Size of Cooperative Single-stranded RNA Binding by Poliovirus RNA-dependent RNA Polymerase. *J Biol Chem* 273: 6724-6730.

287. Liu Z, Robida JM, Chinnaswamy S, Yi G, Robotham JM, *et al.* (2009) Mutations in the hepatitis C virus polymerase that increase RNA binding can confer resistance to cyclosporine A. *Hepatology* 50: 25-33.
288. McDonald SM, Aguayo D, Gonzalez-Nilo FD, Patton JT (2009) Shared and Group-Specific Features of the Rotavirus RNA Polymerase Reveal Potential Determinants of Gene Reassortment Restriction. *J Virol* 83: 6135-6148.
289. Pan JA, Peng X, Gao Y, Li Z, Lu X, *et al.* (2008) Genome-Wide Analysis of Protein-Protein Interactions and Involvement of Viral Proteins in SARS-CoV Replication. *PLoS ONE* 3: e3299.
290. van Hemert MJ, van den Worm SHE, Knoop K, Mommaas AM, Gorbalenya AE, *et al.* (2008) SARS-Coronavirus Replication/Transcription Complexes Are Membrane-Protected and Need a Host Factor for Activity In Vitro. *PLoS Pathog* 4: e1000054.
291. Labonte P, Axelrod V, Agarwal A, Aulabaugh A, Amin A, *et al.* (2002) Modulation of Hepatitis C Virus RNA-dependent RNA Polymerase Activity by Structure-based Site-directed Mutagenesis. *J Biol Chem* 277: 38838-38846.
292. Studier F (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41: 207-234.
293. Laemmli U (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685.
294. Prentice E, McAuliffe J, Lu X, Subbarao K, Denison MR (2004) Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J Virol* 78: 9977-9986.
295. Lazarczyk M, Favre M (2008) Role of Zn<sup>2+</sup> ions in host-virus interactions. *J Virol* 82: 11486-11494.
296. Frederickson CJ, Koh JY, Bush AI (2005) Neurobiology of zinc in health and disease. *Nat Rev Neurosci* 6: 449-462.
297. Alirezaei M, Nairn AC, Glowinski J, Premont J, Marin P (1999) Zinc inhibits protein synthesis in neurons: potential role of phosphorylation of translation initiation factor-2a. *J Biol Chem* 274: 32433-32438.
298. Uchida N, Ohyama K, Bessho T, Yuan B, Yamakawa T (2002) Effect of antioxidants on apoptosis induced by influenza virus infection: inhibition of viral gene replication and transcription with pyrrolidine dithiocarbamate. *Antiviral Res* 56: 207-217.
299. Suara RO, Crowe JEJ (2004) Effect of zinc salts on respiratory syncytial virus replication. *Antimicrob Agents Chemother* 48: 783-790.
300. Gaudernak E, Seipelt J, Triendl A, Grassauer A, Kuechler E (2002) Antiviral Effects of Pyrrolidine Dithiocarbamate on Human Rhinoviruses. *J Virol* 76: 6004-6015.
301. Si X, McManus BM, Zhang J, Yuan J, Cheung C, *et al.* (2005) Pyrrolidine Dithiocarbamate Reduces Coxsackievirus B3 Replication through Inhibition of the Ubiquitin-Proteasome Pathway. *J Virol* 79: 8014-8023.
302. Korant BD, Kauer JC, Butterworth BE (1974) Zinc ions inhibit replication of rhinoviruses. *Nature* 248: 588-590.
303. Polatnick J, Bachrach HL (1978) Effect of zinc and other chemical agents on foot-and-mouth-disease virus replication. *Antimicrob Agents Chemother* 13: 731-734.
304. Lanke K, Krenn BM, Melchers WJG, Seipelt J, van Kuppeveld FJM (2007) PDTC inhibits picornavirus polyprotein processing and RNA replication by transporting zinc ions into cells. *J Gen Virol* 88: 1206-1217.
305. Krenn BM, Gaudernak E, Holzer B, Lanke K, Van Kuppeveld FJM, *et al.* (2009) Antiviral Activity of the Zinc Ionophores Pyrithione and Hinokitiol against Picornavirus Infections. *J Virol* 83: 58-64.

306. Zalewski PD, Forbes IJ, Betts WH (1993) Correlation of apoptosis with change in intracellular labile Zn(II) using Zinquin [(2-methyl-8-p-toluenesulphonamide-6-quinolyloxy)acetic acid], a new specific fluorescent probe for Zn(II). *Biochem J* 296: 403-408.
307. Baum EZ, Bebernitz GA, Palant O, Mueller T, Plotch SJ (1991) Purification, properties, and mutagenesis of poliovirus 3C protease. *Virology* 165: 140-150.
308. Chen F, Chan KH, Jiang Y, Kao RY, Lu HT, *et al.* (2004) In vitro susceptibility of 10 clinical isolates of SARS coronavirus to selected antiviral compounds. *J Clin Virol* 31: 69-75.
309. Cinatl J, Morgenstern B, Bauer G, Chandra P, Rabenau H, *et al.* (2003) Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus. *Lancet* 361: 2045-2046.
310. Butterworth BE, Korant BD (1974) Characterization of the large picornaviral polypeptides produced in the presence of zinc ion. *J Virol* 14: 282-291.
311. Denison MR, Perlman S (1986) Translation and processing of mouse hepatitis virus virion RNA in a cell-free system. *J Virol* 60: 12-18.
312. Denison MR, Zoltick PW, Hughes SA, Giangreco B, Olson AL, *et al.* (1992) Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology* 189: 274-284.
313. van Hemert MJ, de Wilde AH, Gorbalenya AE, Snijder EJ (2008) The in Vitro RNA Synthesizing Activity of the Isolated Arterivirus Replication/Transcription Complex Is Dependent on a Host Factor. *J Biol Chem* 283: 16525-16536.
314. van den Born E, Posthuma CC, Knoops K, Snijder EJ (2007) An infectious recombinant equine arteritis virus expressing green fluorescent protein from its replicase gene. *J Gen Virol* 88: 1196-1205.
315. Sims AC, Burkett SE, Yount B, Pickles RJ (2008) SARS-CoV replication and pathogenesis in an in vitro model of the human conducting airway epithelium. *Virus Res* 133: 33-44.
316. Stockman LJ, Bellamy R, Garner P (2006) SARS: Systematic Review of Treatment Effects. *PLoS Med* 3: e343.
317. Thompson A, Patel K, Tillman H, McHutchison JG (2009) Directly acting antivirals for the treatment of patients with hepatitis C infection: A clinical development update addressing key future challenges. *J Hepatol* 50: 184-194.
318. Thompson AJV, McHutchison JG (2009) Antiviral resistance and specifically targeted therapy for HCV (STAT-C). *J Viral Hepat* 16: 377-387.
319. Günther S, Asper M, Röser C, Luna LK, Drosten C, *et al.* (2004) Application of real-time PCR for testing antiviral compounds against Lassa virus, SARS coronavirus and Ebola virus in vitro. *Antiviral Res* 63: 209-215.
320. Hakimelahi GH, Ly TW, Moosavi-Movahedi AA, Jain ML, Zakerinia M, *et al.* (2001) Design, synthesis, and biological evaluation of novel nucleoside and nucleotide analogues as agents against DNA viruses and/or retroviruses. *J Med Chem* 44: 3710-3720.
321. Miller JP, Kigwana LJ, Streeter DG, Robins RK, Simon LN, *et al.* (1977) The relationship between the metabolism of ribavirin and its proposed mechanism of action. *Ann N Y Acad Sci* 284: 211-229.
322. Chen HM, Hosmane RS (2001) Acyclic nucleoside/nucleotide analogues with an imidazole ring skeleton. *Nucleosides Nucleotides Nucleic Acids* 20: 1599-1614.
323. Ziebuhr J, Snijder EJ (2006) *Coronaviruses: Molecular Biology and Diseases*; Thiel V, editor: Horizon Scientific Press, Norwich UK.
324. Chan-Yeung M, Xu RH (2003) SARS: epidemiology. *Respirology* 8: S9-14.

325. te Velthuis AJ, van den Worm SH, Sims AC, Baric RS, Snijder EJ, *et al.* (2010) Zn<sup>2+</sup> inhibits coronavirus and arterivirus RNA polymerase activity in vitro and zinc ionophores block the replication of these viruses in cell culture. *PLoS Pathog* 6: e1001176.
326. Kaeppler U, Stiefl N, Schiller M, Vicik R, Breuning A, *et al.* (2005) A new lead for nonpeptidic active-site-directed inhibitors of the severe acute respiratory syndrome coronavirus main protease discovered by a combination of screening and docking methods. *J Med Chem* 48: 6832-6842.
327. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32.
328. Iyer LM, Koonin EV, Leippe DD, Aravind L (2005) Origin and evolution of the archeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucl Acid Res* 33: 3875-3896.
329. Saijo M, Morikawa S, Fukushi S, Mizutani T, Hasegawa H, *et al.* (2005) Inhibitory effect of mizoribine and ribavirin on the replication of severe acute respiratory syndrome (SARS)-associated coronavirus. *Antiviral Research* 66: 159-163.
330. Morgenstern B, Michaelis M, Baer PC, Doerr HW, Cinatl J, Jr. (2005) Ribavirin and interferon-beta synergistically inhibit SARS-associated coronavirus replication in animal and human cell lines. *Biochemical and biophysical research communications* 326: 905-908.
331. Marcotte LL, Wass AB, Gohara DW, Pathak HB, Arnold JJ, *et al.* (2007) Crystal structure of poliovirus 3CD protein: virally encoded protease and precursor to the RNA-dependent RNA polymerase. *J Virol* 81: 3583-3596.
332. Ratia K, Pegan S, Takayama J, Sleeman K, Coughlin M, *et al.* (2008) A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc Natl Acad Sci U S A* 105: 16119-16124.
333. Bouvet M, Debarnot C, Imbert I, Selisko B, Snijder EJ, *et al.* (2010) In vitro reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog* 6: e1000863.
334. Züst R, Miller TB, Goebel SJ, Thiel V, Masters PS (2008) Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *J Virol* 82: 1214-1228.
335. Tinoco Jr I, Pan TXL, Bustamante C (2006) Determination of thermodynamics and kinetics of RNA reaction by force. *Q Rev Biophys* 39: 325-360.
336. Neuman KC, Nagy A (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods* 5: 491-505.
337. Strick TR, Allemand JF, Bensimon D, Bensimon A, Croquette V (1996) The elasticity of a single supercoiled DNA molecule. *Science* 271: 1835-1837.
338. Danilowicz C, Lee CH, Kim K, Hatch K, Coljee VW, *et al.* (2009) Single molecule detection of direct, homologous, DNA/DNA pairing. *Proc Natl Acad Sci USA* 106: 19824-19829.
339. Kruithof M, Chien FT, Routh A, Logie C, Rhodes D, *et al.* (2009) Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat Struct Mol Biol* 16: 534-540.
340. Kruithof M, Chien F, de Jager M, van Noort J (2008) Subpiconewton dynamic force spectroscopy using magnetic tweezers. *Biophys J* 94: 2343-2348.
341. Lipfert J, Hao X, Dekker N (2009) Quantitative modeling and optimization of magnetic tweezers. *Biophys J* 96: 5040-5049.
342. Ribbeck N, Saleh OA (2008) Multiplexed single-molecule measurements with magnetic tweezers. *Rev Sci Instrum* 79: 094301.

343. Ajjan R, LIM BCB, Standeven KF, Harrand R, Dolling S, *et al.* (2008) Common variation in the C-terminal region of the fibrinogen beta-chain: effects on fibrin structure, fibrinolysis and clot rigidity. *Blood* 111: 643-650.
344. Mierke CT, Kollmannsberger P, Zitterbart DP, Smith J, Fabry B, *et al.* (2008) Mechano-coupling and regulation of contractility by the vinculin tail domain. *Biophys J* 94: 661-670.
345. Strick TR, Croquette V, Bensimon D (1998) Homologous pairing in stretched supercoiled DNA. *Proc Natl Acad Sci USA* 95: 10579-10583.
346. Celedon A, Nodelman IM, Wildt B, Dewan R, Searson P, *et al.* (2009) Magnetic tweezers measurement of single molecule torque. *Nano Lett* 9: 1720-1725.
347. Abels JA, Moreno-Herrero F, van der Heijden T, Dekker C, Dekker NH (2005) Single-molecule measurements of the persistence length of double-stranded DNA. *Biophys J* 88: 2737-2744.
348. Mosconi F, Allemand JF, Bensimon D, Croquette V (2009) Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys Rev Lett* 102: 078301.
349. Koster DA, Palle K, Bot ES, Bjornsti MA, Dekker NH (2007) Antitumor drugs impede DNA uncoiling by topoisomerase I. *Nature* 448: 213-217.
350. Koster DA, Croquette V, Dekker C, Shuman S, Dekker NH (2005) Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature* 434: 671-674.
351. Leuba SH, Karymov MA, Tomschik M, Ramjit R, Smith P, *et al.* (2003) Assembly of single chromatin fibers depends on the tension in the DNA molecule: magnetic tweezers study. *Proc Natl Acad Sci USA* 100: 495-450.
352. Neuman KC, Charvin G, Bensimon A, Croquette V (2008) Mechanisms of chiral discrimination by topoisomerase IV. *Proc Natl Acad Sci USA* 106: 6986-6991.
353. Wong PW, Halvorsen K (2006) The effect of integration time on fluctuation measurements: calibrating an optical trap in the presence of motion blur. *Opt Express* 14: 12517-12531.
354. Berg-Sørensen K, Flyvbjerg H (2004) Power spectrum analysis for optical tweezers. *Rev Sci Instrum* 75: 594-612.
355. Savin T, Doyle PS (2005) Static and Dynamic Errors in Particle Tracking Microrheology. *Biophys J* 88: 623-638.
356. Lionnet T, Spiering M, Benkovic S, Bensimon D, Croquette V (2007) Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. *Proc Natl Acad Sci USA* 104: 19790-19975.
357. Wong WP, Halvorsen K (2009) Beyond the frame rate: measuring high-frequency fluctuations with light-intensity modulation. *Opt Lett* 34: 277-279.
358. Kim K, Saleh OA (2009) A high-resolution magnetic tweezer for single-molecule measurements. *Nucleic Acids Res* 37: e136.
359. Storm C, Nelson PC (2003) Theory of high-force DNA stretching and overstretching. *Phys Rev E stat Nonlin Soft Matter Phys* 67: 051906.
360. Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69: 1352-1360.
361. Vilfan ID, Lipfert J, Koster DA, Dekker NH (2009) Magnetic tweezers for single-molecule experiments. *Handbook of Single-Molecule Biophysics*: Springer.
362. Lipfert J, Koster DA, Vilfan ID, Hage S, Dekker NH (2009) Single-molecule magnetic tweezers studies of type IB topoisomerases. *Methods Mol Biol* 582: 71-89.
363. Crut A, Nair PA, Koster DA, Shuman S, Dekker NH (2008) Dynamics of phosphodiester synthesis by DNA ligase. *Proc Natl Acad Sci USA* 105: 6894-6899.

364. Wong PW (2006) Exploring single-molecule interactions through 3D optical trapping and tracking: from thermal noise to protein refolding [PhD Thesis]. Cambridge: Harvard University.
365. Beausang JF, Zurla C, Sullivan L, Finzi L, Nelson PC (2007) Elementary simulation of tethered Brownian motion. *Am J Phys* 75: 520-523.
366. Han L, Lui B, Blumberg S, Beausang JF, Nelson PC, *et al.* (2008) Calibration of tethered particle motion experiments: Springer.
367. Strick T (1999) Mechanical supercoiling of DNA and its relaxation by topoisomerases [PhD]. Paris: University of Paris VI.
368. van der Horst A, Forde NR (2010) Power spectral analysis for optical trap stiffness calibration from high-speed camera position detection with limited bandwidth. *Opt Express* 18: 7670-7677
369. Strick T (1999) Mechanical supercoiling of DNA and its relaxation by topoisomerases. Paris: University of Paris VI.
370. Schäffer E, Norrelykke SF, Howard J (2007) Surface forces and drag coefficients of microspheres near a plane surface measured with optical tweezers. *Langmuir* 23: 3654-3665.
371. Happel J, Brenner N (1965) Low Reynolds number hydrodynamics with special application to particulate media. N.J.: Prentice-Hall, Englewood Cliffs.
372. Frick DN (2007) The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target. *Curr Issues Mol Biol* 9: 1-20.
373. Fairman-Williams ME, Guenther UP, Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 20: 313-324.
374. Gorbalenya AE, Koonin EV (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* 3: 419-429.
375. Singleton MR, Wigley DB (2002) Modularity and specialization of superfamily 1 and 2 helicases. *J Bacteriol* 184: 1819-1826.
376. Dumont S, Cheng W, Serebrov V, Beran RK, Tinoco IJ, *et al.* (2006) RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. *Nature* 439: 105-108.
377. Cheng W, Arunajadai SG, Moffitt JR, Tinoco IJ, Bustamante C (2011) Single-base pair unwinding and asynchronous RNA release by the hepatitis C virus NS3 helicase. *Science* 333: 1746-1749.
378. Uchil PD, Satchidanandam V (2003) Architecture of the Flaviviral Replication Complex: protease, nuclease and detergents reveal encasement within double-layered membrane compartments. *J Biol Chem* 278: 24388-24398.
379. Westaway EG, Mackenzie JM, Kenney MT, Jones MK, Khromykh AA (1997) Ultrastructure of Kunjin virus-infected cells: colocalization of NS1 and NS3 with double-stranded RNA, and of NS2B with NS3, in virus-induced membrane structures. *J Virol* 71: 6650-6661.
380. Fang Y, Snijder EJ (2010) The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins. *Virus Res* 154: 61-76.
381. Snijder EJ (2001) Arterivirus RNA synthesis dissected. Nucleotides, membranes, amino acids, and a bit of zinc. *Adv Exp Med Biol* 494: 241-253.
382. Manosas M, Xi XG, Bensimon D, Croquette V (2010) Active and passive mechanisms of helicases. *Nucleic Acids Res* 38: 5518-5526.
383. Johnson DS, Bai L, Smith BY, Patel SS, Wang MD (2007) Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* 129: 1299-1309.
384. Thomsen ND, Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* 139: 523-534.

385. Kim JL, Morgenstern KA, Griffith JP, Dwyer MD, Thomson JA, *et al.* (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure* 6: 89-100.
386. Betterton MD, Jülicher F (2003) A motor that makes its own track: helicase unwinding of DNA. *Phys Rev Lett* 91: 258103.
387. Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* 37: 317-336.
388. Delagoutte E, von Hippel PH (2002) Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: Structures and properties of isolated helicases. *Q Rev Biophys* 35: 431-478.
389. Kool ET (2001) Active site tightness and substrate fit in DNA replication. *Annu Rev Biochem* 71: 191-219.
390. Snijder EJ, Meulenberg JJ (1998) The molecular biology of arteriviruses. *J Gen Virol* 79: 961-979.
391. Ahnert P, Patel SS (1997) Asymmetric interaction of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. *J Biol Chem* 272: 32267-32273.
392. Lee NR, Kwon HM, Park K, Oh S, Jeong YJ, *et al.* (2010) Cooperative translocation enhances the unwinding of duplex DNA by SARS coronavirus helicase nsP13. *Nucleic acids research* 38: 7626-7636.
393. te Velthuis AJ, Kerssemakers JW, Lipfert J, Dekker NH (2010) Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. *Biophys J* 99: 1292-1302.
394. Kim K, Saleh OA (2009) A high-resolution magnetic tweezer for single-molecule measurements. *Nucleic Acids Res* 37: 37.
395. Lionnet T, Spiering MM, Benkovic SJ, Bensimon D, Croquette V (2007) Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. *Proc Natl Acad Sci U S A* 104: 19790-19795.
396. Matson SW, Richardson CC (1983) DNA-dependent nucleoside 5' triphosphate activity of the gene 4 protein of bacteriophage T7. *J Biol Chem* 258: 14009-14016.
397. Sun B, Johnson DS, Patel G, Smith BY, Pandey M, *et al.* (2011) ATP-induced helicase slippage reveals highly coordinated subunits. *Nature* 478: 132-135.
398. Kim SH, Dallmann HG, McHenry CS, Marians KJ (1996) Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell* 84: 643-650.
399. Stano NM, Jeong YJ, Donmez I, Tummalapalli P, Levin MK, *et al.* (2005) DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature* 435: 370-373.
400. Sun B, Wei KJ, Zhang B, Zhang XH, Dou SX, *et al.* (2008) Impediment of *E. coli* UvrD by DNA-destabilizing force reveals a strained-inchworm mechanism of DNA unwinding. *EMBO J* 27: 3279-3287.
401. Gribble FM, Loussouarn G, Tucker SJ, Zhao C, Nichols CG, *et al.* (2000) A Novel Method for Measurement of Submembrane ATP Concentration. *J Biol Chem* 275: 30046-30049.
402. Steighardt J, Meyer K, Roos W (2000) Selective regulatory effects of purine and pyrimidine nucleotides on vacuolar transport of amino acids. *Biochim Biophys Acta* 1497: 321-327.
403. van Dinten LC, den Boon JA, Wassenaar AL, Spaan WJ, Snijder EJ (1997) An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc Natl Acad Sci U S A* 94: 991-996.
404. Wang Q, Arnold JJ, Uchida A, Raney KD, Cameron CE (2010) Phosphate release contributes to the rate-limiting step for unwinding by an RNA helicase. *Nucleic Acid Res* 38: 1312-1324.

405. Lipfert J, Hao X, Dekker NH (2009) Quantitative modeling and optimization of magnetic tweezers. *Biophys J* 96: 5040-5049.
406. Alvarez DE, Lodeiro MF, Ludueña SJ, Pietrasanta LI, Gamarnik AV (2005) Long-range RNA-RNA interactions circularize the dengue virus genome. *J Virol* 79: 6631-6643.
407. Chen D, Barros M, Spencer E, Patton J (2001) Features of the 3'-consensus sequence of rotavirus mRNAs critical to minus strand synthesis. *Virology* 282: 221-229.
408. Zoll J, Heus HA, van Kuppeveld FJ, Melchers WJ (2009) The structure-function relationship of the enterovirus 3'-UTR. *Virus Res* 139: 209-216.
409. Noton SL, Cowton VM, Zack CR, McGivern DR, Fearn R (2010) Evidence that the polymerase of respiratory syncytial virus initiates RNA replication in a nontemplated fashion. *Proc Natl Acad Sci U S A* 107: 10226-10231.
410. Chen D, Patton JT (2000) *De novo* synthesis of minus strand RNA by the rotavirus RNA polymerase in a cell-free system involves a novel mechanism of initiation. *RNA* 6: 1455-1467.
411. Drummond DA, McCrae MA, Colman A (1985) Stability and movement of mRNAs and their encoded proteins in *Xenopus* oocytes. *J Cell Biol* 100: 1148-1156.
412. Robertson JS, Schubert M, Lazzarini RA (1981) Polyadenylation sites for influenza virus mRNA. *J Virol* 38: 157-163.
413. Zheng H, Lee HA, Palese P, García-Sastre A (1999) Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication. *J Virol* 73: 5240-5243.
414. Laishram RS, Anderson RA (2010) The poly A polymerase Star-PAP controls 3'-end cleavage by promoting CPSF interaction and specificity toward the pre-mRNA. *EMBO J* 29: 4132-4145.
415. Wahle E (1992) The end of the message: 3'-end processing leading to polyadenylated messenger RNA. *Bioessays* 14: 113-118.
416. Sethna PB, Hung SL, Brian DA (1989) Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc Natl Acad Sci U S A* 86: 5626-5630.
417. Macnaughton MR, Madge MH (1978) The genome of human coronavirus strain 229E. *J Gen Virol* 39: 497-504.
418. Sawicki DL, Wang T, Sawicki SG (2001) The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. *J Gen Virol* 82: 385-396.
419. Lai MM, Patton CD, Stohlman SA (1982) Further characterization of mRNA's of mouse hepatitis virus: presence of common 5'-end nucleotides. *J Virol* 41: 557-565.
420. Lai MM, Stohlman SA (1981) Comparative analysis of RNA genomes of mouse hepatitis viruses. *J Virol* 38: 661-670.
421. Masters PS (2006) The molecular biology of coronaviruses. *Adv Virus Res* 66: 193-292.
422. Sola I, Mateos-Gomez PA, Almazan F, Zuñiga S, Enjuanes L (2011) RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol* 2.
423. von Brunn A, Teepe C, Simpson JC, Pepperkok R, Friedel CC, *et al.* (2007) Analysis of Intraviral Protein-Protein interactions of the SARS Coronavirus ORFome. *PLoS ONE* 2: e459.
424. Lai MM, Cavanagh D (1997) The molecular biology of coronaviruses. *Adv Virus Res* 48: 1-100.
425. Li S, Zhao Q, Zhang Y, Bartlam M, Xuemei L, *et al.* (2010) New nsp8 isoform suggests mechanism for tuning viral RNA synthesis. *Protein Cell* 1: 198-204.
426. Willis RC, Carson DA, Seegmiller JE (1978) Adenosine kinase initiates the major route of ribavirin activation in a cultured human cell line. *Proc Natl Acad Sci U S A* 75: 3041-3044.
427. Vo NV, Young KC, Lai MM (2003) Mutagenic and inhibitory effects of ribavirin on hepatitis C virus RNA polymerase. *Biochemistry* 42: 10462-10471.

428. Maag D, Castro C, Hong Z, Cameron CE (2001) Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin. *J Biol Chem* 276: 46094-46098.
429. Graci JD, Harki DA, Korneeva VS, Edathil JP, Too K, *et al.* (2007) Lethal mutagenesis of poliovirus mediated by a mutagenic pyrimidine analogue. *J Virol* 81: 11256-11266.
430. Nguyen JT, Hoopes JD, Le MH, Smee DF, Patick AK, *et al.* (2010) Triple Combination of Amantadine, Ribavirin, and Oseltamivir Is Highly Active and Synergistic against Drug Resistant Influenza Virus Strains In Vitro. *PLoS ONE* 5: e9332.
431. Choi HJ, Kim JH, Lee CH, Ahn YJ, Song JH, *et al.* (2009) Antiviral activity of quercetin 7-rhamnoside against porcine epidemic diarrhea virus. *Antiviral Res* 81: 77-81.
432. Day CW, Baric R, Cai SX, Frieman M, Kumaki Y, *et al.* (2009) A new mouse-adapted strain of SARS-CoV as a lethal model for evaluating antiviral agents in vitro and in vivo. *Virology* 395: 210-222.
433. de Vega M, Lazaro JM, Salas M, Blanco L (1998) Mutational analysis of phi29 DNA polymerase residues acting as ssDNA ligands for 3'-5' exonucleolysis. *J Mol Biol* 279: 807-822.
434. Knopf CW (1998) Evolution of viral DNA-dependent DNA polymerases. *Virus Genes* 16: 47-58.
435. Esteban JA, Soengas MS, Salas m, Blanco L (1994) 3'-5' exonuclease active site of phi29 DNA polymerase. Evidence favoring a metal ion-assisted reaction mechanism. *J Biol Chem* 244: 31946-31954.
436. Johnson KA (1993) Conformational coupling in DNA polymerase fidelity. *Annu Rev Biochem* 62: 685-713.
437. Nga PT, Parquet Mdel C, Lauber C, Parida M, Nabeshima T, *et al.* (2011) Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathog* 7: e1002215.
438. Cook PR (2010) A model for all genomes: the role of transcription factories. *J Mol Biol* 395: 1-10.
439. Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, *et al.* (2011) Ordered and dynamic assembly of single spliceosomes. *Science* 331: 1289-1295.
440. Kai ZS, Pasquinelli AE (2010) MicroRNA assassins: factors that regulate the disappearance of miRNAs. *Nat Struct Mol Biol* 17: 5-10.
441. Ibrahim F, Rymarquis LA, Kim EJ, Becker J, Balassa E, *et al.* (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *Proc Natl Acad Sci U S A* 107: 3906-3911.
442. Reinisch KM, Wolin SL (2007) Emerging themes in non-coding RNA quality control. *Curr Opin Struct Biol* 17: 209-214.
443. Slomovic S, Laufer D, Geiger D, Schuster G (2006) Polyadenylation of ribosomal RNA in human cells. *Nucleic Acid Res* 34: 2966-2975.
444. Dickson AM, Wilusz J (2011) Strategies for viral RNA stability: live long and prosper. *Trends Genet* 27: 286-293.
445. Rossi ML, Bambara RA (2006) Reconstituted Okazaki fragment processing indicates two pathways of primer removal. *J Biol Chem* 281: 26051-26061.
446. Nedialkova D, Ulferts R, van den Born E, Lauber C, Gorbalenya A, *et al.* (2009) Biochemical characterization of arterivirus nonstructural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease. *J Virol* 83: 5671-5682.
447. Grosseohme NE, Li L, Keane SC, Liu P, Dann CE, 3rd, *et al.* (2009) Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *Journal of Molecular Biology* 394: 544-557.
448. Drosten C, Günther S, Preiser W, van der Werf S, Brodt HR, *et al.* (2003) Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 348: 1967-1976.

449. Den Boon JA, Spaan WJ, Snijder EJ (1995) Equine arteritis virus subgenomic RNA transcription: UV inactivation and translation inhibition studies. *Virology* 213: 364-372.
450. Richard P, Manley JL (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev* 23: 1247-1269.
451. Hofmann MA, Brian DA (1991) The 5' end of coronavirus minus-strand RNAs contain a short poly(U) tract. *J Virol* 65: 6331-6333.
452. Donlin MJ, Patel S S, Johnson KA (1991) Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* 30: 538-546.
453. Petruska J, Goodman MF, Boosalis MS, Sowers LC, Cheong C, *et al.* (1988) Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc Natl Acad Sci U S A* 85: 6252-6256.
454. Benkovic SJ, Cameron CE (1995) Kinetic analysis of nucleotide incorporation and misincorporation by klenow fragment of *Escherichia coli* DNA polymerase I. *Methods Enzymols* 262: 257-269.
455. Pasternak AO, Spaan WJ, Snijder EJ (2004) Regulation of relative abundance of arterivirus subgenomic mRNAs. *J Virol* 78: 8102-8113.
456. Pasternak AO, van den Born E, Spaan WJ, Snijder EJ (2003) The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. *J Virol* 77: 1175-1183.
457. Nedialkova DD, Gorbalenya AE, Snijder EJ (2010) Arterivirus nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs. *PLoS Pathog* 6: e1000772.
458. Zuniga S, Cruz JL, Sola I, Mateos-Gomez PA, Palacio L, *et al.* (2010) Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *Journal of virology* 84: 2169-2175.
459. Zuniga S, Sola I, Alonso S, Enjuanes L (2004) Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J Virol* 78: 980-994.
460. van den Born E, Gultyaev AP, Snijder EJ (2004) Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. *RNA* 10: 424-437.
461. Loparo JJ, Kulczyk AW, Richardson CC, van Oijen AM (2011) Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proc Natl Acad Sci U S A*.
462. Brockway SM, Clay CT, Lu XT, Denison MR (2003) Characterization of the Expression, Intracellular Localization, and Replication Complex Association of the Putative Mouse Hepatitis Virus RNA-Dependent RNA Polymerase. *J Virol* 77: 10515-10527.
463. Brockway SM, Lu XT, Peters TR, Dermody TS, Denison MR (2004) Intracellular localization and protein interactions of the gene 1 protein p28 during mouse hepatitis virus replication. *J Virol* 78: 11551-11562.
464. Sims AC, Ostermann J, Denison MR (2000) Mouse Hepatitis Virus Replicase Proteins Associate with Two Distinct Populations of Intracellular Membranes. *J Virol* 74: 5647-5654.
465. Conaway RC, Brower CS, Conaway JW (2002) Emerging Roles of Ubiquitin in Transcription Regulation. *Science* 296: 1254-1258.
466. Kaiser P, Flick K, Wittenberg C, Reed SI (2000) Regulation of Transcription by Ubiquitination without Proteolysis. *Cell* 102: 303-314.
467. Stavreva DA, Kawasaki M, Dunder M, Koberna K, Müller WG, *et al.* (2006) Potential roles for ubiquitin and the proteasome during ribosome biogenesis. *Mol Cell Biol* 26: 5131-5145.

468. Luby-Phelps K, Taylor DL (1988) Subcellular compartmentalization by local differentiation of cytoplasmic structure. *Cell Motil Cytoskeleton* 10: 28-37.
469. Rizzuto R, Pinton P, Carrington W, Fay FS, Fogarty KE, *et al.* (1998) Close Contacts with the Endoplasmic Reticulum as Determinants of Mitochondrial Ca<sup>2+</sup> Responses. *Science* 280: 1763-1766.
470. Rapoport TA (2008) Protein transport across the endoplasmic reticulum membrane. *FEBS J* 275: 4471-4478.
471. Osborne AR, Rapoport TA, van den Berg B (2005) Protein translocation by the Sec61/SecY channel. *Annu Rev Cell Dev Biol* 21: 529-550.
472. Roumiantzeff M, Summers DF, Maizel JV (1971) In vitro protein synthetic activity of membrane-bound poliovirus polyribosomes. *Virology* 44: 249-258.
473. Caligiuri LA, Tamm I (1969) Membranous structures associated with translation and transcription of poliovirus RNA. *Science* 166: 885-886.
474. Sawicki SG, Sawicki DL (1986) Coronavirus minus-strand RNA synthesis and effect of cycloheximide on coronavirus RNA synthesis. *J Virol* 57: 328-334.
475. Sawicki SG, Sawicki DL, Younker D, Meyer Y, Thiel V, *et al.* (2005) Functional and Genetic Analysis of Coronavirus Replicase-Transcriptase Proteins. *PLoS Pathog* 1: e39.
476. Li L, Kang H, Liu P, Makkinje N, Williamson ST, *et al.* (2008) Structural lability in stem-loop 1 drives a 5' UTR-3' UTR interaction in coronavirus replication. *J Mol Biol* 377: 790-803.
477. Pfefferle S, Schöpf J, Kögl M, Friedel CC, Müller MA, *et al.* (2011) The SARS-Coronavirus-Host Interactome: Identification of Cyclophilins as Target for Pan-Coronavirus Inhibitors. *PLoS Pathog* 7: e1002331.
478. Goldsmith CS, Tatti KM, Ksiazek TG, Rollin PE, Comer JA, *et al.* (2004) Ultrastructural characterization of SARS coronavirus. *Emerg Infect Dis* 10: 320-326.
479. van den Worm SH, Knoops K, Zevenhoven-Dobbe JC, Beugeling C, van der Meer Y, *et al.* (2011) Development and RNA-synthesizing activity of coronavirus replication structures in the absence of protein synthesis. *J Virol* 85: 5669-5673.
480. Quinkert D, Bartenschlager R, Lohmann V (2005) Quantitative analysis of the hepatitis C virus replication complex. *J Virol* 79: 13594-13605.
481. Wang T, Sawicki SG (2001) Mouse hepatitis virus minus-strand templates are unstable and turn-over during viral replication. *Adv Exp Med Biol* 494: 491-497.

## CURRICULUM VITAE

Arend Jan Wouter te Velthuis, roepnaam Aartjan, werd geboren op 19 april 1983 in het Diaconessen Ziekenhuis van Leiden. In 2001 behaalde hij zijn Gymnasium diploma aan de Bataafse Kamp te Hengelo en vervolgde hij zijn opleiding met Bachelor studies aan de Saxion Hogeschool IJsselland in Deventer (Biochemie) en de Universiteit Leiden (Biologie). Beide studies ronden hij in de zomer van 2005 af met virologische afstudeerscripties onder begeleiding van respectievelijk dr. Martine Ossevoort en Prof. dr. Rob C. Hoeben (Leids Universitair Medisch Centrum), en dr. Marjolein Kikkert en Prof. dr. Eric J. Snijder (Leids Universitair Medisch Centrum). Zijn Master studie Moleculaire Biologie volgde hij aan de Universiteit Leiden. Hij liep daarbij stages onder begeleiding van Prof. dr. Christoph P. Bagowski (Instituut Biologie Leiden), en dr. Sjoerd van der Worm en Prof. dr. Eric J. Snijder (Leids Universitair Medisch Centrum).

Na deze studie in maart 2007 *cum laude* te hebben afgesloten verhuisde hij naar de Universiteit van Oxford om daar als postgraduate research assistant in het lab van Prof. dr. Ervin Fodor aan de replicatie van influenza A virus te werken. In datzelfde jaar werd hem een Toptalent-beurs van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) toegekend en werd hij door de Sir William Dunn School of Pathology in Oxford genomineerd voor een Sir Christopher Welch Scholarship. Op basis van de NWO-beurs startte hij in oktober 2007 onder begeleiding van Prof. dr. Eric J. Snijder zijn promotie-onderzoek in de moleculaire virologie, een onderzoek dat in de zomer van 2008 werd uitgebreid met biofysisch onderzoek onder begeleiding van Prof. dr. Nynke H. Dekker aan de Technische Universiteit Delft. Sinds de afronding van zijn promotie-onderzoek werkt Aartjan weer in Oxford aan de replicatie van influenza A virus, ondersteund door een Rubicon beurs van NWO, en is hij een Kemp Postdoctoral Fellow bij Lincoln College.



## LIST OF PUBLICATIONS

1. **Te Velthuis A.J.**, Van de Worm S.H., Snijder E.J.: The SARS-CoV nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of *de novo* initiation and primer extension. *Nucleic Acids Res.* (2012)
2. **Te Velthuis A.J.**, Sakalis P.A., Fowler D.A., Bagowski C.P.: Genome-wide analysis of PDZ domain binding reveals inherent functional overlap within the PDZ network. *PLoS ONE.* 6:e16047 (2011)
3. **Te Velthuis A.J.**, van den Worm S.H., Sims A., Baric R., Snijder E.J., Hemert M.J.: Zn<sup>2+</sup> inhibits Coronavirus and arterivirus RNA polymerase activity *in vitro* and zinc ionophores block nidovirus replication in cell culture. *PLoS Pathog.* 6:e1001176 (2010)
4. **Te Velthuis A.J.**, Kerssemakers, J.W., Lipfert J. Dekker, N.H.: Quantitative guidelines for force calibration through spectral analysis of magnetic tweezers data. *Biophys J.* 99:1292-1302. (2010)
5. Bagowski C.P., Bruins W., **Te Velthuis A.J.**: The Nature of Protein Domain Evolution: Shaping the Interaction Network. *Curr Genomics.* 11:368-376. (2010)
6. Knoops K., Swett-Tapia C., van den Worm S.H., **Te Velthuis A.J.**, Koster A.J., Mommaas A.M., Snijder E.J., Kikkert M.: Integrity of the early secretory pathway promotes, but is not required for, severe acute respiratory syndrome coronavirus RNA synthesis and virus-induced remodeling of endoplasmic reticulum membranes. *J Virol.* 84:833-846. (2010)
7. **Te Velthuis A.J.**, Arnold J.J., Cameron C.E., van den Worm S.H., Snijder, E.J.: The RNA polymerase activity of SARS coronavirus is primer-dependent. *Nucleic Acids Res.* 38:203-214. (2010)
8. **Te Velthuis A.J.**, Bagowski C.P.: Linking fold, function and phylogeny: a comparative genomics view of protein (domain) evolution. *Curr Genomics.* 9:88-96. (2008)
9. Ott E.B., van den Akker N.M., Sakalis P.A., Gittenberger-de Groot A.C., **Te Velthuis A.J.**, Bagowski C.P.: The LIM domain only protein 7 is important for zebrafish heart development. *Dev Dyn.* 237:3940-3952. (2008)
10. **Te Velthuis A.J.**, Bagowski C.P.: PDZ and LIM domain-encoding genes: molecular interaction and their role in development. *ScientificWorldJournal.* 7:1470-1492 (2007)
11. **Te Velthuis A.J.**, Admiraal J.F., Bagowski C.P.: Evolutionary insights into origin and structure of the MAGUK family. *BMC Evol Biol.* 7:129. (2007)
12. **Te Velthuis A.J.**, Isogai T., Gerrits L., Bagowski C.P.: Convergent evolution of functional domains leads to diverse gene architectures in the PDZ/LIM family. *PLOS One.* 2:e189. (2007)
13. Ossevoort M., Zaldumbide A., **Te Velthuis A.J.**, Melchers M., Rensing M.E., Wiertz E.J., Hoeben R.C.: The Nested Open-Reading Frame in The Epstein-Barr Virus Nuclear

Antigen-1 mRNA Encodes a Protein Capable of Inhibiting Antigen Presentation In Cis. *Mol. Immunol.* 44:3588-3596. (2007)

14. Ott E.B., **Te Velthuis A.J.**, Bagowski C.P.: Comparative analysis of splice form-specific expression of LIM kinases during zebrafish development. *Gene Expr Patterns.* 7:620-629. (2007)
15. **Te Velthuis A.J.**, Ott E.B., Marquez I.J., Bagowski C.P.: Gene expression patterns of the ALP family during zebrafish development. *Gene Expr Patterns.* 7:297-305. (2007)