

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/33614> holds various files of this Leiden University dissertation.

**Author:** Lehmann, Kathleen Corina

**Title:** Biochemistry and function of nidovirus replicase proteins

**Issue Date:** 2015-06-23

# Biochemistry and function of nidovirus replicase proteins

*Kathleen Lehmann*

ISBN: 978-94-6169-683-0

Layout and Printing: Optima Grafische Communicatie ([www.ogc.nl](http://www.ogc.nl))

The research described in this thesis was carried out at the Department of Medical Microbiology, Leiden University Medical Center, The Netherlands, and was financially supported by the European Union's Seventh Framework program (FP7/2007-2013) through the EUVIRNA project (Marie Curie Initial Training Network on (+) RNA virus replication and antiviral drug development, Grant agreement no. 264286).

Funding of the printing costs by the Department of Medical Microbiology is gratefully acknowledged.

# **Biochemistry and function of nidovirus replicase proteins**

Proefschrift

ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op dinsdag 23 juni 2015  
klokke 13:45 uur

door

**Kathleen Corina Lehmann**

geboren te Strausberg, Duitsland  
in 1985

## **PROMOTIECOMMISSIE**

**Promotores:** Prof. dr. A.E. Gorbalenya  
Prof. dr. E.J. Snijder

**Co-promotor:** Dr. C.C. Posthuma

**Overige leden:** Dr. H. van Attikum  
Prof. dr. B. Berkhout Academisch Medisch Centrum,  
Universiteit van Amsterdam  
Prof. dr. J. Ziebuhr  
Justus-Liebig-Universität Gießen, Duitsland

As we acquire more knowledge,  
things do not become more comprehensible  
but more complex and mysterious.

**Albert Schweitzer**  
**(1875-1965)**



## TABLE OF CONTENTS

<b>Chapter 1</b>	General introduction	9
<b>Chapter 2</b>	Structural basis for the regulatory function of a complex zinc-binding domain in a replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase	33
<b>Chapter 3</b>	What we know but do not understand about nidovirus helicases	69
<b>Chapter 4</b>	Arterivirus RNA-dependent RNA polymerase: vital enzymatic activity remains elusive	121
<b>Chapter 5</b>	Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses	143
<b>Chapter 6</b>	Arterivirus nsp12 versus the coronavirus nsp16 2'-O-methyltransferase: comparison of the C-terminal cleavage products of two nidovirus pp1ab polyproteins	197
<b>Chapter 7</b>	Development of an anti-coronavirus drug – a private sector opportunity?	225
<b>Chapter 8</b>	General discussion	243
	List of abbreviations	267
	Summary	271
	Samenvatting	275
	Curriculum Vitae	279
	List of publications	280





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

General introduction

# CHAPTER 1



## 1 **VIRUS DIVERSITY AND THE ORDER *NIDOVIRALES***

2  
3 Per definition viruses are inanimate organic entities that are capable to replicate them-  
4 selves. However, in contrast to living beings their DNA or RNA genomes do not encode  
5 information for the expression of proteins involved in the synthesis of the four funda-  
6 mental biological building blocks: amino acids, carbohydrates, lipids, and nucleoside  
7 triphosphates (NTPs). Because of this limitation, they are obligate intracellular parasites  
8 that strictly depend on the metabolism of a host cell. Additionally, host proteins may  
9 play essential or supporting roles during specific steps in the viral replication cycle – the  
10 most obvious being cellular receptors used for viral entry. In effect, host factors thus  
11 determine the spectrum of genetically related cellular species and cell types a virus can  
12 infect – in short the virus' host range.

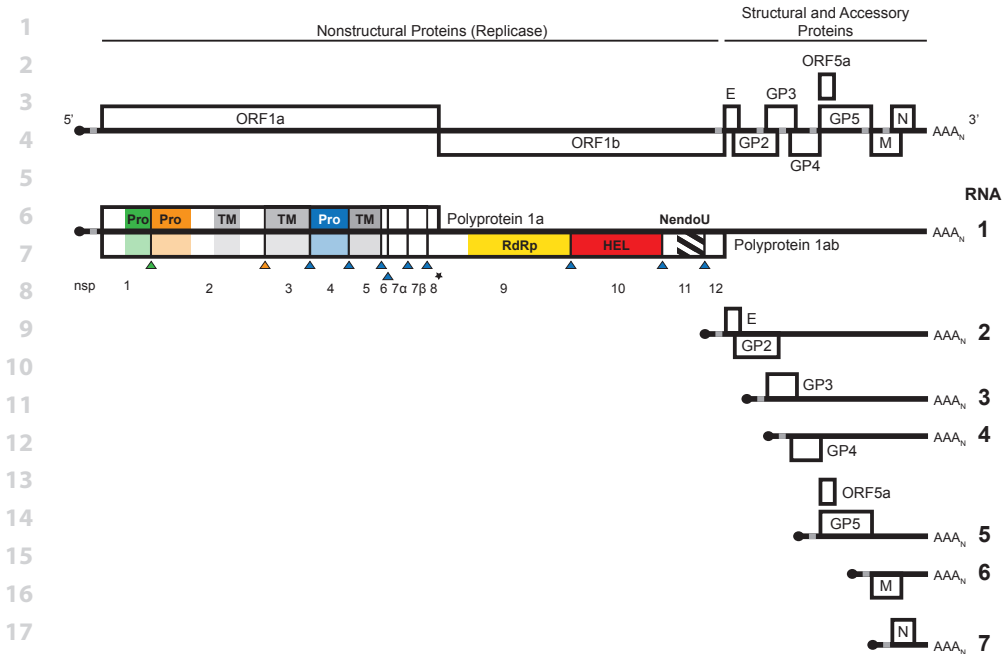
13  
14 At the moment the International Committee on Taxonomy of Viruses (ICTV) recognizes  
15 about 3000 different virus species (1). On the other hand, it has been estimated that  
16 about 15 million different cellular species (~9 million eukaryotes, ~6 million prokary-  
17 otes) live on this planet (2;3). If we assume that each of those is host to at least one virus  
18 species – likely a vast underestimation given that humans are host to 189 known viruses  
19 (4) – a lot remains to be discovered. To bring order into the known and anticipated virus  
20 diversity in terms of, for example, genome type and organization, or replication strategy,  
21 related viruses have been grouped into genera, (sub-)families, and orders (proceeding  
22 from lower to higher rank). However, due to the extreme divergence of viruses and  
23 fast evolution, the relationship between different ranks remains often obscure. In this  
24 respect, virus taxonomy stands in stark contrast to the Tree of Life that has been con-  
25 structed for organisms to reflect the course of cellular evolution.

26  
27 The viruses that are discussed in this thesis belong to the order *Nidovirales*. This name  
28 derives from the typical genome expression strategy of its members featuring a nested  
29 set of subgenomic (sg) mRNAs (in Latin, *nidus* means nest). At the moment four fami-  
30 lies with different host ranges are united in the order: *Arteriviridae* (vertebrate hosts),  
31 *Coronaviridae* (vertebrate hosts), *Mesoniviridae* (invertebrate hosts), and *Roniviridae*  
32 (invertebrate hosts) (5-8). With the exception of the *Mesoniviridae*, all families contain  
33 economically important pathogens infecting livestock, for example swine (arterivirus  
34 porcine reproductive and respiratory syndrome virus, coronaviruses porcine epidemic  
35 diarrhea virus and transmissible gastroenteritis virus), cattle (bovine coronavirus),  
36 poultry (coronavirus infectious bronchitis virus, IBV), and prawn (ronivirus yellow head  
37 virus), and hence cause severe losses to the respective industries (9-14). Additionally, es-  
38 tablished human coronaviruses may cause mild respiratory symptoms. Combined these  
39 are the second leading cause for common cold after rhinoviruses (*Picornaviridae*) (15).

1 Recent years also saw the emergence of two previously unknown and highly pathogenic  
2 zoonotic coronaviruses in the human population: severe acute respiratory syndrome  
3 coronavirus (SARS-CoV) in 2002 and Middle East respiratory syndrome coronavirus  
4 (MERS-CoV) in 2012 (16;17). In contrast to the established human coronaviruses,  
5 which are constantly circulating in the human population, these viruses were initially  
6 directly transmitted from an animal reservoir to humans. In the case of SARS-CoV it is  
7 now believed that this reservoir may be one of the numerous bat species (18;19). From  
8 these animals the virus spread to humans and caused the first pandemic of the 21st  
9 century with major outbreaks in China and Southeast Asia but also Canada (20). Despite  
10 concerns that SARS-CoV might mutate to permanently establish itself within the human  
11 population, the virus disappeared – thanks to the imposed control measures like strict  
12 quarantine protocols – from circulation in humans in 2003 after causing about 8500  
13 cases, including 812 deaths (21). The second newly-emerged coronavirus, MERS-CoV,  
14 which might be transmitted by camels (22), appears to be even more lethal with a case  
15 fatality rate of above 30%. However, thus far the case numbers have remained low, with  
16 about 1000 cases between April 2012 and November 2014 (23). Still, the threat to global  
17 public health and economy, exemplified by the SARS and MERS outbreaks, but also  
18 the combined economic damage caused by the veterinary nidoviruses call for a more  
19 thorough understanding of nidovirus biology. Ultimately these efforts might contribute  
20 to the development of countermeasures to keep future outbreaks in check.

## 23 THE NIDOVIRUS REPLICATION CYCLE

25 Nidoviruses enter a host cell by receptor-mediated endocytosis utilizing a variety of  
26 entry receptors (24). Afterwards the viral genome, which is a single RNA molecule of  
27 positive (mRNA) polarity carrying a type-1 cap structure (cap-1) (<sup>m</sup>GpppN<sub>m</sub>) and a poly-  
28 adenylate (polyA) tail at its 5' and 3' end, respectively, is released into the cytoplasm.  
29 The genome is organized into multiple open reading frames (ORFs) (Figure 1), of which  
30 ORF1a and ORF1b encode all nonstructural proteins (nsps) separated by a ribosomal  
31 frameshift site, comprising a secondary structure element called RNA pseudoknot and  
32 a uridine-rich so-called "slippery sequence". It is estimated that in equine arteritis virus  
33 (EAV) about 15-20% and in the coronaviruses mouse hepatitis virus (MHV) and IBV up to  
34 40% of the translating ribosomes perform the -1 frameshift and hence synthesize a large  
35 polyprotein called pp1ab (25-28). In the remainder of the cases the ribosome reaches  
36 a stop codon that is located just downstream of the frameshift signal. The resulting  
37 polyprotein is known as pp1a. Interestingly, all key enzymes for RNA synthesis and pro-  
38 cessing, for example the RNA-dependent RNA polymerase (RdRp), helicase, and – in the  
39 case of coronaviruses – also the proofreading exoribonuclease and capping enzymes,



**Figure 1:** Typical nidovirus genome organization illustrated using equine arteritis virus (EAV). Open reading frames (ORFs) are indicated as boxes. Cleavage sites of replicase proteins in polyproteins 1a and 1ab are marked by triangles corresponding in color to the protease responsible for cleavage. Known transmembrane and enzymatic domains are indicated. Pro, protease; TM, transmembrane domain; RdRp, RNA-dependent RNA polymerase; HEL, helicase; NendoU, endoribonuclease; E, envelope protein; GP, glycoprotein; M, membrane protein; N, nucleocapsid protein. The ribosomal frameshift site leading to expression of polyprotein 1ab is labeled with a star. Transcription-regulating sequences are indicated as gray boxes. Presumed 5' cap structures are depicted as black dots.

are encoded downstream of the frameshift while known co-factors of these enzymes, RNA-binding and membrane-anchoring proteins, as well as proteolytic enzymes are so far exclusively mapped to pp1a (29-32). The frameshift is thus an elegant way to regulate the relative abundance of these key enzymes compared to other proteins controlling genome replication and expression. Surprisingly, a second frameshift site was recently discovered in all arteriviruses except EAV. This site, located further upstream roughly in the middle of ORF1a, is able to direct a -2 as well as a -1 frameshift and thus gives rise to two additional variants of the membrane-bound nonstructural protein nsp2, one of which being predicted to be soluble. This arterivirus frameshift site with dual shift capacity, which is controlled by *trans*-activation by the upstream nsp1 $\beta$  subunit, is the only one of its kind known to date (33;34).

In addition to the nsps that are directly translated from the genome, group-specific structural proteins and – in the case of coronaviruses – accessory proteins are trans-

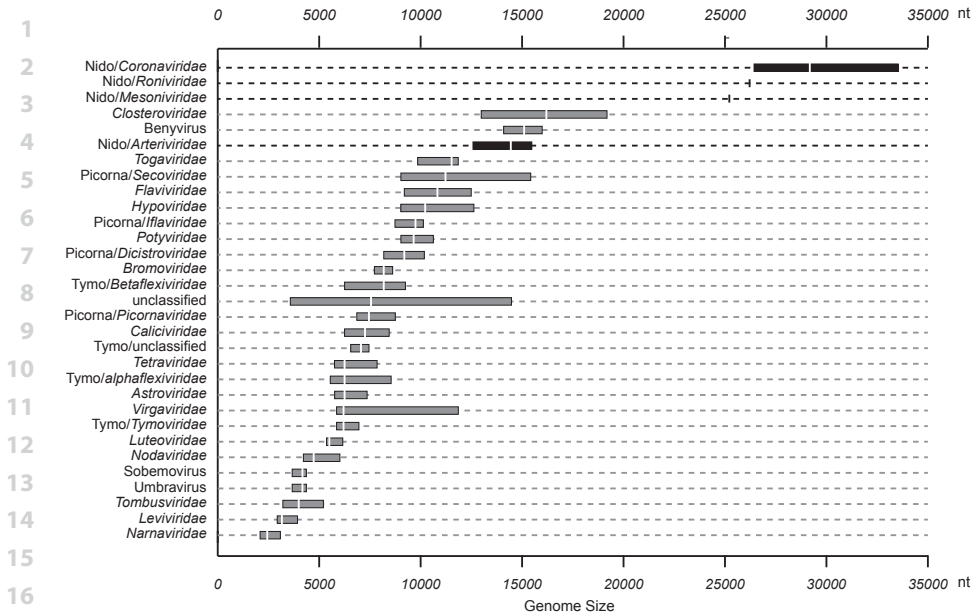
1 lated from an in part extensive set of sg mRNAs (35). These mRNAs, which carry the  
2 same 5'- and 3'-terminal sequences as the genome in most nidoviruses, are transcribed  
3 with the help a unique mechanism involving subgenome-size negative-stranded (-)  
4 templates that in part arise from discontinuous RNA synthesis (see below). As for all  
5 positive-stranded (+) RNA viruses, RNA replication (amplification of the genome) and  
6 transcription (synthesis of sg mRNAs) are thought to take place in association with an  
7 extensive network of modified membranes (36;37). For nidoviruses this membranous  
8 web takes mainly the form of interconnected double-membrane vesicles (DMVs) and  
9 convoluted membranes (CMs). It was speculated that these membrane structures may  
10 provide a scaffold for replication-transcription complex (RTC) assembly inside DMVs.  
11 Hence, it is thought that these structures support viral replication in two ways; on the  
12 one hand, by increasing local concentrations of NTPs, RNAs, and proteins required  
13 for RNA synthesis and, on the other, by shielding viral replication products, especially  
14 double-stranded replication intermediates, from detection by the host's innate immune  
15 system. After encapsidation of the viral genome, particles bud into the lumen of the  
16 smooth endoplasmic reticulum (ER) or Golgi complex. From there they are transported  
17 via the cellular secretory pathway to be released from the plasma membrane (24).

## 20 MOLECULAR DETAILS OF NIDOVIRUS REPLICATION AND TRANSCRIPTION

### 22 **Prison break: how nidoviruses with large genomes overcame size constraints**

24 The genome sizes of nidoviruses range from 12-16 kilobases (kb) for arteriviruses (from  
25 here on referred to as "small nidoviruses") to 20-34 kb for mesoni-, roni-, and coronavi-  
26 ruses ("intermediate and large nidoviruses"). With these sizes especially the latter group  
27 deviates substantially from the average size of most (+) RNA virus genomes that typically  
28 are smaller than 10 kb (Figure 2) (38). Still, even the largest RNA virus currently known,  
29 the recently discovered ball python nidovirus – a proposed member of the *Torovirinae*, a  
30 subfamily of the *Coronaviridae* – with a genome size of 33.5 kb (39), is dwarfed by some  
31 DNA viruses, whose genomes can reach sizes in the range of megabase pairs (Mbp),  
32 for example mimiviruses (~1 Mbp) and pandoraviruses (~2.5 Mbp) (40;41). Considering  
33 these tremendous size differences, two questions arise: in what way are RNA viruses so  
34 fundamentally different from DNA viruses that a genome expansion of the scale of the  
35 latter did not occur, and how did large nidoviruses, at least to some extent, overcome  
36 the size restrictions imposed on other (+) RNA viruses?

38 To answer these questions, it should be informative to explore the underlying reason for  
39 the existence of the observed size barrier in (+) RNA viruses. All of these viruses encode



**Figure 2:** Genome sizes of positive-stranded RNA viruses. Size ranges of major families or genera and unclassified viruses are indicated by black (nidoviruses) or gray boxes. The median size is marked by a white vertical bar. Nido, *Nidovirales*; Picorna, *Picornavirales*; Tymo, *Tymovirales*. Adapted from (42).

an RdRp that synthesizes copies of the viral genome during the infection of a host cell. The basic mechanism by which these enzymes fulfill this central function in the viral replication cycle can be simplified to two steps: the matching of an incoming NTP to the template and the formation of a chemical bond to extend the nascent RNA chain (43;44). The first step of this mechanism basically occurs by a trial-and-error method as polymerases lack the means to determine the identity of the nucleotide that is about to be copied or of the NTP that has entered the active site. Instead the selection and, ultimately, the incorporation of an NTP is solely based on the relative difference between its dissociation rate from the active site and the rate of phosphodiester bond formation. Because a correct Watson-Crick base pair is energetically more stable than a mismatched one or any of the alternative base pairs, the correct NTP will, on average, remain at the active site for a longer period of time than an incorrect one. If this period is long enough for bond formation to occur, the RNA chain will be extended by this one nucleotide. If not, the NTP will diffuse away, and the next NTP can be tried at random. In summary, in order to minimize the number of errors but maintain RNA synthesis, the chemical reaction rate should be much lower than the dissociation rate of incorrect NTPs but higher than the dissociation rate of correct NTPs.



1 In general, error rates of RdRps were estimated to range between  $10^{-3}$  and  $10^{-5}$  errors  
2 per nucleotide incorporated (45). That means most individual genomes of an average  
3 (+) RNA virus would differ by at least one nucleotide from each other. To emphasize this  
4 variation, the concept of a quasispecies was introduced, essentially representing a cloud  
5 of different variants of a consensus sequence that are heterogeneous in respect to their  
6 fitness (46;47). Depending on the environmental conditions, the composition of these  
7 quasispecies may differ. Interestingly, decreasing the variation within a quasispecies by  
8 increasing the replication fidelity of an RdRp was shown to strongly diminish the overall  
9 fitness of a virus population (48-52). It is therefore believed that on an evolutionary scale  
10 it is the quasispecies, rather than individual variants, that is targeted by selection. On  
11 the other hand, decreasing replication fidelity will lead to the accumulation of too many  
12 detrimental mutations, which will eventually prevent virus replication, a consequence  
13 that was termed "error catastrophe". Because of these two opposing principles, RNA  
14 viruses are thought to be optimized to exist close to the threshold of this error catastro-  
15 phe. In summary, this implies that the size of the genome is limited by the error rate of  
16 the RdRp it encodes. Interestingly, there seems to be a correlation between the size of  
17 RNA genomes and their RdRp genes (45). Whether these larger RdRps indeed operate  
18 with a lower error rate, however, remains to be seen.

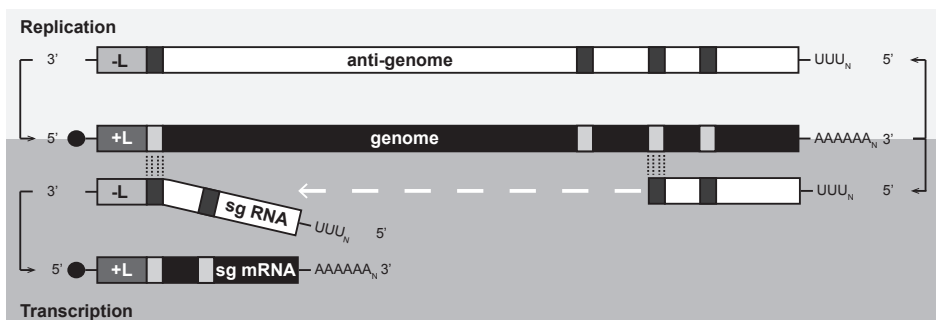
19  
20 To express the interdependence between replication fidelity, genome size, and genome  
21 complexity, the term "Eigen trap" has been coined (53). This term essentially conveys  
22 the fact that none of these three parameters can be increased without simultaneously  
23 increasing the other two. When comparing (+) RNA virus genomes, two instances where  
24 both genome complexity and size expanded by the introduction of a new enzyme have  
25 been recognized. First, an RNA helicase is encoded by all viruses with genomes larger  
26 than 7 kb (54). It was proposed that this enzyme may support the RdRp by removing  
27 double-stranded regions from the template. However, how this would directly affect  
28 fidelity, which depends, as explained above, on an interplay between NTP affinities and  
29 the chemical reaction rate, is unclear. Therefore, it cannot be excluded that there may be  
30 other reasons underlying the presence of a helicase in (+) RNA virus genomes.

31  
32 A more straightforward explanation was proposed for a second instance of genome  
33 expansion. In this case a 3'-5' exoribonuclease was acquired, which gave rise to (+) RNA  
34 genomes of more than 20 kb (55). A similar enzymatic activity, a 3'-5' exonuclease, is a  
35 vital part of typical DNA polymerases encoded by DNA viruses and cellular organisms.  
36 Contrary to general believe, the intrinsic error rates of DNA polymerases are, as a matter  
37 of fact, not significantly lower ( $10^{-4}$ - $10^{-5}$  bp<sup>-1</sup>) than that of RNA polymerases. It is the pres-  
38 ence of this associated exonuclease activity that enables the reduction of the error rate  
39 to  $10^{-5}$ - $10^{-7}$  bp<sup>-1</sup> by removing incorrectly incorporated NTPs during DNA synthesis (45).

1 By analogy, it was assumed and recently experimentally confirmed that the nidovirus  
 2 exoribonuclease confers proofreading activity to the viral RTC (32;56). Furthermore, in  
 3 reverse genetics experiments it was demonstrated that a knock-out of this proofreading  
 4 activity led to a more than 10-fold increase of the overall error rate during MHV and  
 5 SARS-CoV replication in cell culture (57;58). In contrast to large nidoviruses, arteriviruses  
 6 do not encode an exoribonuclease subunit (55). Still, their genomes are substantially  
 7 larger than those of most other (+) RNA viruses. It thus remains to be seen if another  
 8 domain acquisition event may be linked to this expansion.

### 9 10 Nidovirus discontinuous RNA synthesis

11  
 12 As already hinted at, the transcription mechanism of nidoviruses is unique in the virus  
 13 world (Figure 3). Although several non-nidovirus families utilize sg mRNAs, none gener-  
 14 ates those by a mechanism equivalent to that of most nidoviruses, involving discontinu-  
 15 ous (-) subgenome-length RNA synthesis (roni- and toroviruses do not or only in part  
 16 employ this mechanism) (59). In contrast to, for example, alphavirus sg mRNA synthesis  
 17 that is driven from an internal promoter in the full-length negative strand, nidovirus  
 18 sg mRNAs are transcribed from several co-terminal (-) subgenome-length RNAs of  
 19 different lengths. Essential protagonists in the still not well understood mechanism to  
 20 produce those templates are so-called transcription-regulating sequences (TRSs), which  
 21 are conserved AU-rich elements of a length – depending on the virus – of 5-18 nucleo-  
 22 tides located near the genome's 5' end (leader TRS) and upstream of most of the 3' ORFs  
 23 (body TRSs). During negative-strand synthesis, which always initiates at the genome's  
 24 3' end, the viral RdRp may pause at one of these sequences. Subsequently, the part of  
 25 the template between the body and leader TRS is skipped before RNA synthesis resumes  
 26 at the genome's 5' end at the so-called leader sequence. How exactly this skipping oc-  
 27 curs is still not understood, but a dissociative step during body-leader joining may be  
 28



29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39 **Figure 3:** Discontinuous negative-strand transcription model. Transcription-regulating sequences are indicated by gray (positive strands) or black (negative strands) boxes. Leader and anti-leader regions are labeled +L and -L, respectively. Presumed 5' cap structures of mRNAs are depicted as black dots. Adapted from (24).

1 involved since co-infection experiments with two different MHV strains showed that the  
2 leader sequence may derive from a different template than the rest of the transcribed  
3 product (60). Also base-pairing between the leader and anti-body TRS is important since  
4 the amount of each sg mRNA correlates with the calculated stability of the respective  
5 TRS duplex (61-64). However, it was speculated that base-pairing may not be the only  
6 factor involved. For instance, it was also shown that TRSs serve a function independent  
7 of base-pairing, potentially in secondary structure-dependent recruitment of specific  
8 proteins (59). Three viral proteins that were implicated in arterivirus transcription regu-  
9 lation are nsp1 (65;66), the nidovirus-wide conserved helicase nsp10 (67;68), and the  
10 endoribonuclease nsp11 (69). For these proteins, mutations either altered the balance  
11 between genome replication and transcription or selectively abolished sg mRNA syn-  
12 thesis altogether. Since genome replication requires the synthesis of full-length nega-  
13 tive strands, and hence a read-through through all TRSs, these results indicated that the  
14 three proteins are directly or indirectly involved in the discontinuous step. Interestingly,  
15 nsp1 as well as nsp10 contain a zinc-binding domain that may be instrumental in estab-  
16 lishing interactions with the RNA or proteins of the RTC to serve this regulatory function.

17  
18 As a consequence of discontinuous RNA synthesis, all nidovirus mRNAs, except the  
19 smallest, are structurally polycistronic. However, with a few exceptions, only the  
20 most 5'-located ORF is actually translated, meaning that the mRNAs are functionally  
21 monocistronic (70). Thus, the question arises what the advantage of this complicated  
22 transcription mechanism is compared to structurally monocistronic mRNAs expressed  
23 from multiple promoters or polycistronic mRNAs enabling internal ribosome entry or  
24 other non-canonical translation initiation mechanisms frequently employed by other  
25 viruses. Obviously, the nidovirus mechanism ensures that all RNAs carry the same 5'-  
26 and 3'-terminal sequences as the genome or anti-genome. This could be advantageous  
27 if regulatory elements are located at the ends. For example, sequences of negative  
28 strands may be required to initiate positive-strand synthesis or capping. It would also  
29 be possible that the genome ends contain translational enhancers (59). Finally, those  
30 elements could also serve to discriminate viral from host RNAs. In view of the notion  
31 that all vertebrate nidoviruses encode an endoribonuclease (55), whose substrate is still  
32 elusive but may well be a host RNA, this possibility is especially intriguing.

### 33 34 **Means to an end: nidovirus mRNA modification**

35  
36 Given the complexity of ribosomes, comprising 80 proteins and 4 rRNAs in higher eu-  
37 karyotes (71), RNA viruses cannot encode information for their components. To ensure  
38 that viral mRNAs are translated in the host cell, a variety of strategies is employed by  
39 different virus families. The most obvious is to adopt the essential modifications of cel-

1 lular mRNAs, a 5' cap structure and a 3' polyA tail. Alternatively, viral mRNAs may contain  
2 special secondary structures, called internal ribosome entry sites (IRESs) or 3' cap-  
3 independent translation enhancers (3' CITEs), that allow the non-canonical recruitment  
4 of the translational apparatus. Finally, some viruses encode proteins that may replace  
5 certain cellular translation initiation factors. Their mRNAs may thus lack some of the  
6 modifications of host mRNAs, for instance the cap-1 or polyA tail (72-74).

7  
8 Where characterized, the 3' ends of positive-stranded RNAs of a number of nidoviruses  
9 contained polyA tails (53;75-79). Furthermore, a cap-1 (<sup>m</sup>GpppN<sub>m</sub>) structure was found  
10 to be present at the 5' end of the genome and/or sg mRNAs of equine torovirus (80), the  
11 coronavirus MHV (81;82), and the arterivirus simian hemorrhagic fever virus (83). Based  
12 on common ancestry, it is thus assumed that all nidoviruses equip their mRNAs with  
13 these modifications, which would allow them to enter the cellular translation pathway.  
14 In line with this hypothesis, it was shown that MHV infection leads to phosphorylation of  
15 the cap-binding translation initiation factor eIF4E, which is required for the cellular path-  
16 way (84). This phosphorylation, which is a known regulatory mechanism in eukaryotic  
17 cells to strengthen the interaction between the cap and the protein, in turn increased  
18 the translation efficiency of viral mRNAs. Furthermore, overexpression of an inhibitor  
19 of eIF4E, 4E-BP, abolished replication of human coronavirus 229E in HeLa cells (85).  
20 Nevertheless, contradicting evidence with regard to the nature of its 5' end has been  
21 brought forward for the arterivirus lactate dehydrogenase-elevating virus, whose ge-  
22 nome appeared to be devoid of a cap and instead was monophosphorylated (86). Given  
23 this result and the fact that none of the members of the *Roniviridae* and *Mesoniviridae*  
24 was characterized so far, care should be taken in assuming that all nidovirus mRNAs  
25 carry the same 5' end modification. Such a deviation in respect to translation strategy  
26 was also observed in the *Flaviviridae*, whose members may utilize cap-dependent or  
27 -independent mechanisms (87;88). In addition to cap-dependent translation initiation,  
28 IRES elements may, at least in coronaviruses, drive expression of a second gene product  
29 from a single sg mRNA (89-91).

30  
31 In the host cell cap-1 and polyA tails are strictly generated in the nucleus during and  
32 shortly after RNA polymerase II-dependent transcription (92). Since nidoviruses and  
33 many other (+) RNA viruses replicate in the cytoplasm, they cannot benefit from this  
34 cellular machinery. Instead the polyA tail may be synthesized by (one of) the viral  
35 RdRp(s) – coronaviruses are believed to encode a main RdRp (nsp12) and an accessory  
36 RdRp (nsp8) (93-95). How exactly this is achieved was not investigated so far. However,  
37 since negative-stranded RNAs were shown to contain a short polyU stretch at their 5'  
38 end, it was speculated that iterative copying of this stretch may be involved (96). Alterna-  
39

1 tively, SARS-CoV nsp8 in complex with nsp7 was shown to possess terminal transferase,  
2 that is, non-templated extension activity, on single-stranded RNAs (95).

3  
4 In contrast to polyA-tail addition, the assembly of the cap-1 structure appears to be bet-  
5 ter understood in regard to the proteins involved, at least in large nidoviruses. In general  
6 the synthesis of the cap involves four steps and three different enzymatic activities that  
7 may be present in a single subunit with multiple domains or in multiple individual  
8 proteins (97;98). In case of the conventional capping pathway, which is employed by all  
9 eukaryotes and a number of viruses, the triphosphate end of a newly synthesized RNA  
10 is trimmed back to a diphosphate by an RNA-triphosphatase (RTPase). As this activity is  
11 mechanistically identical to the cleavage of NTPs, the NTPase domains of a viral helicase,  
12 if encoded, may execute it. In the second step, a guanylyltransferase (GTase) transfers a  
13 GMP-moiety to the RNA diphosphate end. In contrast to nucleotide bonds established  
14 by polymerases, this bond is formed via a 5'-5' linkage to generate a GpppN-RNA struc-  
15 ture. While this unusual bond cannot be cleaved by regular exo- and endoribonucleases,  
16 specially regulated cytoplasmic host decapping enzymes are employed for the removal  
17 of cap structures (99). As a consequence, capping confers protection against 5'-3' ex-  
18 oribonucleases, and hence capped RNAs exhibit much longer half-lives than uncapped  
19 ones. In order to make the second step irreversible, a methyl group is attached to the  
20 N7-position of the guanine by an N7-methyltransferase (N-MT). Although this so-called  
21 cap-0 structure is due to the specific recognition of the methyl group by eIF4E the basic  
22 requirement for translation initiation (100), a second methylation usually occurs at the 2'  
23 oxygen of the ribose of the first (cap-1) or second (cap-2) nucleotide following the cap.  
24 This second methylation step, which is catalyzed by a 2'-O-methyltransferase (O-MT)  
25 that may or may not be different from the domain utilized for N7 methylation, is con-  
26 nected to host mRNA surveillance mechanisms for self versus non-self discrimination  
27 (101;102). Next to this conventional pathway, alternative viral mechanisms have evolved  
28 that include a different order of steps leading to the same mature cap structure (98).

29  
30 It has been proposed that nidoviruses employ the canonical pathway of cap synthesis  
31 described above (98). However, this hypothesis is far from proven especially with respect  
32 to the universal conservation of this pathway in all nidoviruses. For instance, the GTase  
33 has not been identified in any of the nidoviruses, while RTPase activity was demonstrated  
34 for only two coronavirus helicases (nsp13) (103;104). Whether or not this enzyme, which  
35 belongs to the most conserved proteins of the order, actually exerts this activity in the  
36 context of capping, however, remains to be verified. Finally, two methyltransferases  
37 (MTases) residing in nsp14 (N-MT) and nsp16 (O-MT) have been experimentally identi-  
38 fied in coronaviruses (105-108). Interestingly, while other large nidoviruses – with the  
39 exception of toroviruses, which seem to lack the N-MT activity – encode homologs of

1 both MTases (55), neither of them was identified in arteriviruses. Since arteriviruses  
2 encode a unique protein (nsp12) at a genome position equivalent to that of coronavirus  
3 nsp16, the capping mechanism could be another example of biochemical variability  
4 within the diverse *Nidovirales* order.

## 5 6 7 **OUTLINE OF THIS THESIS** 8

9 The work described in this thesis addresses several poorly or uncharacterized (domains  
10 of) nsps that are likely involved in one or multiple steps during RNA replication and/or  
11 transcription of the prototypic arterivirus EAV. After the above short introduction on  
12 the nidovirus replication cycle and known molecular details of the unusual transcrip-  
13 tion and mRNA processing mechanisms, **chapter 2** presents the crystal structure of  
14 the enzymatically active EAV helicase nsp10, which was obtained and analyzed in close  
15 collaboration with Chinese colleagues. Interestingly, a strong resemblance between  
16 this viral protein and the conserved cellular helicase Upf1, in particular with respect to  
17 their N-terminal zinc-binding domains, became obvious. Since this cellular helicase is  
18 implicated in a number of eukaryotic post-transcriptional quality control mechanisms,  
19 a role for nsp10 and its nidovirus homologs in genome expansion is proposed. This and  
20 other potential functions of the nidovirus helicase in RNA replication, transcription,  
21 and translation, as well as virion biogenesis are further discussed in **chapter 3**, which  
22 presents a review of our current knowledge about nidovirus helicases. Special emphasis  
23 is placed on gaps that still remain, facts that cannot be easily reconciled with our current  
24 understanding of the nidovirus replication mechanisms, and questions that need to be  
25 addressed in future.

26  
27 **Chapters 4 and 5** focus on one of the central arterivirus replication proteins, nsp9,  
28 which harbors the RdRp domain. **Chapter 4** describes a carefully controlled study to  
29 investigate different polymerase activities that nsp9 may have, including a previously  
30 claimed primer-independent RdRp activity. Despite considerable efforts, involving ex-  
31 periments with different preparations of nsp9 and assays performed in the presence of  
32 putative polymerase co-factors, no *in vitro* activity was observed that could be clearly  
33 attributed to this protein. Moreover, circumstantial evidence suggested that the previ-  
34 ously reported activity may have been caused by a contamination of the recombinant  
35 nsp9 preparation with the T7 RNA polymerase used to drive its expression in *E. coli*.  
36 In arteriviruses, the RdRp domain is located in the C-terminal two-thirds of nsp9. In  
37 **chapter 5**, it is now described for the first time that the RdRp domain is flanked at its  
38 N-terminus by another domain that is conserved in all nidoviruses. However, unlike the  
39 situation for the RdRp domain, no homologs of this domain have been found in other

1 RNA viruses. This domain is thus proposed to be a second marker for the *Nidovirales*  
2 order, besides the N-terminal zinc-binding domain of the helicase subunit. Residues that  
3 are part of three conserved sequence motifs were without exception associated with  
4 a newly discovered nucleotidylation activity of recombinant nsp9. It is thus proposed  
5 that this activity could play a role in the modification of the 5' end of viral RNAs through  
6 either RNA ligation, protein priming of RNA synthesis, or guanylyl transfer during RNA  
7 capping. Further research is required to definitely tie nsp9 to one of these pathways.  
8 Nevertheless, alanine substitution of any of the conserved residues was either lethal to  
9 EAV and SARS-CoV or severely crippled these viruses, eventually resulting in reversion  
10 of the mutation. These results thus demonstrate the essential nature of this domain for  
11 virus replication, whatever its exact function will turn out to be.

12  
13 Two MTase activities, commonly required for capping of mRNAs, were previously identi-  
14 fied in two ORF1b-encoded coronavirus proteins, nsp14 and nsp16. While the former  
15 has no counterpart among the arterivirus nsps, the latter and the arterivirus C-terminal  
16 subunit nsp12 occupy equivalent positions in the ORF1b-encoded part of the replicase  
17 although the two proteins share no detectable sequence similarity. It is thus a long  
18 standing question, how arteriviruses may catalyze the 5' end modification of mRNAs,  
19 and we therefore performed a first characterization of the entirely uncharacterized  
20 EAV nsp12 subunit (**chapter 6**). Based on the genomic position of its coding sequence,  
21 sequence alignment, and secondary structure prediction it is hypothesized that nsp12  
22 might represent a unique arterivirus MTase, which has diverged from its homologs  
23 beyond sharing appreciated similarity. To test this hypothesis, recombinant nsp12 was  
24 expressed in and purified from *E. coli* and tested alone and in combination with poten-  
25 tial co-factors for N-MT and O-MT activity. Although positive controls represented by the  
26 SARS-CoV MTases (nsp14 and the nsp10:nsp16 complex) demonstrated the functionality  
27 of the assay, no activity was detected for EAV nsp12. Guided by the sequence alignment,  
28 an extensive set of EAV mutants was generated and characterized with respect to their  
29 plaque phenotype and progeny titer, as well as their protein expression. These reverse  
30 genetics experiments revealed a number of phenotypes ranging from wild-type-like via  
31 non-spreading to replication-incompetent, which indicated that nsp12 is essential for  
32 viral replication.

33  
34 The above chapters describing biochemical properties of selected proteins may ulti-  
35 mately contribute to the identification of drug targets to combat nidovirus infections. In  
36 **chapter 7** the prerequisites under which the marketing of such an antiviral drug would  
37 be economically viable are analyzed. This project was realized under guidance of several  
38 specialists of one of the industrial partners, Janssen Infectious Diseases, of the EUVIRNA  
39 consortium, the Marie Curie Initial Training Network to which my research project be-

1 longed. This study concludes that, at the moment, none of the circulating nidoviruses  
2 constitutes a sufficiently sized market to warrant the considerable investments required  
3 for drug development. The situation may be different if a new highly-pathogenic virus  
4 would emerge, as exemplified in 2002 by SARS-CoV or 2012 by MERS-CoV. In view of  
5 such threats, pre-pandemic drug stockpiling could be considered. However, also under  
6 those circumstances, it seems likely that the inherent financial risk would preclude an  
7 independent private initiative, even though market parameters and approval procedures  
8 appear to be favorable.

9  
10 Finally, **chapter 8** connects some of the main findings described in this thesis with  
11 previously described data. In particular, potential differences between small and large  
12 nidoviruses on the level of the molecular mechanisms of RNA synthesis initiation and  
13 mRNA capping are highlighted. To this end, alternative mechanisms are considered that  
14 would be consistent with the data on arteriviruses presented in this thesis and else-  
15 where. Furthermore, potential roles of cellular helicases in nidovirus replication and the  
16 host's immune response against nidoviruses are discussed.

17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39



## 1 REFERENCE LIST

- 2 1. Lefkowitz EJ. ICTV/MSL: International Committee on Taxonomy of Viruses / Master Species List  
3 updated. 2014.
- 4 2. Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. Proc.Natl.Acad.  
5 Sci.U.S.A 2002;99(16):10494-10499.
- 6 3. Mora C, Tittensor DP, Adl S, *et al.* How many species are there on Earth and in the ocean? PLoS.Biol.  
7 2011;9(8):e1001127.
- 8 4. Woolhouse M, Gaunt E. Ecological origins of novel human pathogens. Crit Rev.Microbiol. 2007;  
9 33(4):231-242.
- 10 5. Lauber C, Ziebuhr J, Junglen S, *et al.* *Mesoniviridae*: a proposed new family in the order *Nidovirales*  
11 formed by a single species of mosquito-borne viruses. Arch.Virol. 2012;157(8):1623-1628.
- 12 6. Cowley JA, Walker PJ, Flegel TW, *et al.* Family *Roniviridae*. In King AMQ, Adams MJ, Carstens EB *et*  
13 *al.* editors, *Virus taxonomy*. Ninth report of the international committee on taxonomy of viruses,  
14 Amsterdam, Elsevier Academic Press, 2012;829-834.
- 15 7. de Groot RJ, Baker SC, Baric R, *et al.* Family *Coronaviridae*. In King AMQ, Adams MJ, Carstens EB *et*  
16 *al.* editors, *Virus taxonomy*. Ninth report of the international committee on taxonomy of viruses,  
17 Amsterdam, Elsevier Academic Press, 2012;806-828.
- 18 8. Faaberg KS, Balasuriya UB, Brinton MA, *et al.* Family *Arteriviridae*. In King AMQ, Adams MJ, Carstens  
19 EB *et al.* editors, *Virus Taxonomy*. Ninth report of the international committee on taxonomy of  
20 viruses, Amsterdam, Elsevier Academic Press, 2012;796-805.
- 21 9. Sang Y, Rowland RR, Blecha F. Antiviral regulation in porcine monocytic cells at different activa-  
22 tion States. J.Virol. 2014;88(19):11395-11410.
- 23 10. Cowley JA, Dimmock CM, Spann KM, *et al.* Gill-associated virus of *Penaeus monodon* prawns: an  
24 invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. J.Gen.Virol.  
25 2000;81(Pt 6):1473-1484.
- 26 11. Song D, Park B. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemi-  
27 ology, diagnosis, and vaccines. Virus Genes 2012;44(2):167-175.
- 28 12. Zhang Q, Hu R, Tang X, *et al.* Occurrence and investigation of enteric viral infections in pigs with  
29 diarrhea in China. Arch.Virol. 2013;158(8):1631-1636.
- 30 13. Cavanagh D. Coronaviruses in poultry and other birds. Avian Pathol. 2005;34(6):439-448.
- 31 14. Saif LJ. Bovine respiratory coronavirus. Vet.Clin.North Am.Food Anim Pract. 2010;26(2):349-364.
- 32 15. Lai MM, Perlman S, Anderson LJ. *Coronaviridae*. In Mahan Knipe D, Howley PM, editors, *Fields'*  
33 *Virology Volume One*, Philadelphia, Lippincott, Williams & Wilkins, 2007;1305-1397.
- 34  
35  
36  
37  
38  
39

- 1 16. Coleman CM, Frieman MB. Coronaviruses: important emerging human pathogens. *J.Virol.* 2014;  
2 88(10):5209-5212.
- 3 17. Hilgenfeld R, Peiris M. From SARS to MERS: 10 years of research on highly pathogenic human  
4 coronaviruses. *Antiviral Res.* 2013;100(1):286-295.
- 5 18. Shi Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res.* 2008;  
6 133(1):74-87.
- 7 19. Ge XY, Li JL, Yang XL, *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses  
8 the ACE2 receptor. *Nature* 2013;503(7477):535-538.
- 9 20. Cherry JD, Krogstad P. SARS: the first pandemic of the 21st century. *Pediatr.Res.* 2004;56(1):1-5.
- 10 21. World Health Organization. SARS outbreak contained worldwide. [press release], 5 July 2003.  
11 Available at: <http://www.who.int/mediacentre/news/releases/2003/pr56/en> [accessed 10 No-  
12 vember 2014]
- 13 22. Drosten C, Kellam P, Memish ZA. Evidence for camel-to-human transmission of MERS coronavirus.  
14 *N.Engl.J.Med.* 2014;371(14):1359-1360.
- 15 23. World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV): Sum-  
16 mmary of current situation, literature update, and risk assessment - as of 5 February 2015. [online]  
17 Available at: [http://www.who.int/csr/disease/coronavirus\\_infections/mers-5-february-2015.pdf](http://www.who.int/csr/disease/coronavirus_infections/mers-5-february-2015.pdf)  
18 [accessed 17 April 2014]
- 19 24. Snijder EJ, Siddell SG, Gorbalenya AE. The order *Nidovirales*. In Mahy BW, ter Meulen V, editors,  
20 Topley and Wilson's Microbiology and Microbial Infections: Virology Volume, London, Hodder  
21 Arnold, 2005;390-404.
- 22 25. Den Boon JA, Snijder EJ, Chirnside ED, *et al.* Equine arteritis virus is not a togavirus but belongs to  
23 the coronaviruslike superfamily. *J.Virol.* 1991;65(6):2910-2920.
- 24 26. Firth AE, Brierley I. Non-canonical translation in RNA viruses. *J.Gen.Virol.* 2012;93(Pt 7):1385-1409.
- 25 27. Brierley I, Bournsnel ME, Binns MM, *et al.* An efficient ribosomal frame-shifting signal in the  
26 polymerase-encoding region of the coronavirus IBV. *EMBO J.* 1987;6(12):3779-3785.
- 27 28. Bredenbeek PJ, Pachuk CJ, Noten AF, *et al.* The primary structure and expression of the second  
28 open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved  
29 polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.*  
30 1990;18(7):1825-1832.
- 31 29. Sutton G, Fry E, Carter L, *et al.* The nsp9 replicase protein of SARS-coronavirus, structure and  
32 functional insights. *Structure.* 2004;12(2):341-353.
- 33 30. Bouvet M, Lugari A, Posthuma CC, *et al.* Coronavirus Nsp10, a critical co-factor for activation of  
34 multiple replicative enzymes. *J.Biol.Chem.* 2014;289(37):25783-25796.
- 35  
36  
37  
38  
39

- 1 31. Egloff MP, Ferron F, Campanacci V, *et al.* The severe acute respiratory syndrome-coronavirus  
2 replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world.  
3 Proc.Natl.Acad.Sci.U.S.A 2004;101(11):3792-3796.
- 4 32. Subissi L, Posthuma CC, Collet A, *et al.* One severe acute respiratory syndrome coronavirus pro-  
5 tein complex integrates processive RNA polymerase and exonuclease activities. Proc.Natl.Acad.  
6 Sci.U.S.A 2014;111(37):e3900
- 7 33. Li Y, Treffers EE, Naphthine S, *et al.* Transactivation of programmed ribosomal frameshifting by a  
8 viral protein. Proc.Natl.Acad.Sci.U.S.A 2014;111(21):e2172-2181.
- 9 34. Fang Y, Treffers EE, Li Y, *et al.* Efficient -2 frameshifting by mammalian ribosomes to synthesize an  
10 additional arterivirus protein. Proc.Natl.Acad.Sci.U.S.A 2012;109(43):e2920-2928.
- 11 35. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. J.Virol.  
12 2007;81(1):20-29.
- 13 36. Hagemeyer MC, Rottier PJ, de Haan CA. Biogenesis and dynamics of the coronavirus replicative  
14 structures. Viruses. 2012;4(11):3245-3269.
- 15 37. Paul D, Bartenschlager R. Architecture and biogenesis of plus-strand RNA virus replication facto-  
16 ries. World J.Virol. 2013;2(2):32-48.
- 17 38. Gorbalenya AE, Enjuanes L, Ziebuhr J, *et al.* *Nidovirales*: evolving the largest RNA virus genome.  
18 Virus Res. 2006;117(1):17-37.
- 19 39. Stenglein MD, Jacobson ER, Wozniak EJ, *et al.* Ball python nidovirus: a candidate etiologic agent  
20 for severe respiratory disease in Python regius. MBio. 2014;5(5):e01484-14.
- 21 40. Yamada T. Giant viruses in the environment: their origins and evolution. Curr.Opin.Virol. 2011;1(1):  
22 58-62.
- 23 41. Philippe N, Legendre M, Doutre G, *et al.* Pandoraviruses: amoeba viruses with genomes up to 2.5  
24 Mb reaching that of parasitic eukaryotes. Science 2013;341(6143):281-286.
- 25 42. Lauber C. On the evolution of genetic diversity in RNA virus species - Uncovering barriers to  
26 genetic divergence and gene length in picorna- and nidoviruses. 2012. [dissertation]
- 27 43. Arnold JJ, Cameron CE. Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state ki-  
28 netic analysis of ribonucleotide incorporation in the presence of Mg<sup>2+</sup>. Biochemistry 2004;43(18):  
29 5126-5137.
- 30 44. Ng KK, Arnold JJ, Cameron CE. Structure-function relationships among RNA-dependent RNA  
31 polymerases. Curr.Top.Microbiol.Immunol. 2008;320:137-156.
- 32 45. Belshaw R, Gardner A, Rambaut A, *et al.* Pacing a small cage: mutation and RNA viruses. Trends  
33 Ecol.Evol. 2008;23(4):188-193.
- 34  
35  
36  
37  
38  
39

- 1 46. Sallie R. Replicative homeostasis II: influence of polymerase fidelity on RNA virus quasispecies  
2 biology: implications for immune recognition, viral autoimmunity and other "virus receptor"  
3 diseases. *Virology*. 2005;2:70.
- 4 47. Vignuzzi M, Stone JK, Arnold JJ, *et al.* Quasispecies diversity determines pathogenesis through  
5 cooperative interactions in a viral population. *Nature* 2006;439(7074):344-348.
- 6 48. Coffey LL, Beeharry Y, Borderia AV, *et al.* Arbovirus high fidelity variant loses fitness in mosquitoes  
7 and mice. *Proc.Natl.Acad.Sci.U.S.A* 2011;108(38):16038-16043.
- 8 49. Graham RL, Becker MM, Eckerle LD, *et al.* A live, impaired-fidelity coronavirus vaccine protects in  
9 an aged, immunocompromised mouse model of lethal disease. *Nat.Med.* 2012;18(12):1820-1826.
- 10 50. Pfeiffer JK, Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selec-  
11 tive pressure in mice. *PLoS.Pathog.* 2005;1(2):e11.
- 12 51. Vignuzzi M, Wendt E, Andino R. Engineering attenuated virus vaccines by controlling replication  
13 fidelity. *Nat.Med.* 2008;14(2):154-161.
- 14 52. Weeks SA, Lee CA, Zhao Y, *et al.* A Polymerase mechanism-based strategy for viral attenuation and  
15 vaccine development. *J.Biol.Chem.* 2012;287(38):31618-31622.
- 16 53. Nga PT, Parquet MC, Lauber C, *et al.* Discovery of the first insect nidovirus, a missing evolutionary  
17 link in the emergence of the largest RNA virus genomes. *PLoS.Pathog.* 2011;7(9):e1002215.
- 18 54. Gorbalenya AE, Koonin EV. Viral proteins containing the purine NTP-binding sequence pattern.  
19 *Nucleic Acids Res.* 1989;17(21):8413-8440.
- 20 55. Lauber C, Goeman JJ, Parquet MC, *et al.* The footprint of genome architecture in the largest  
21 genome expansion in RNA viruses. *PLoS.Pathog.* 2013;9(7):e1003500.
- 22 56. Bouvet M, Imbert I, Subissi L, *et al.* RNA 3'-end mismatch excision by the severe acute respiratory  
23 syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc.Natl.*  
24 *Acad.Sci.U.S.A* 2012;109(24):9372-9377.
- 25 57. Eckerle LD, Lu X, Sperry SM, *et al.* High fidelity of murine hepatitis virus replication is decreased in  
26 nsp14 exoribonuclease mutants. *J.Virology*. 2007;81(22):12135-12144.
- 27 58. Eckerle LD, Becker MM, Halpin RA, *et al.* Infidelity of SARS-CoV Nsp14-exonuclease mutant virus  
28 replication is revealed by complete genome sequencing. *PLoS.Pathog.* 2010;6(5):e1000896.
- 29 59. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J.Gen.Virology*.  
30 2006;87(Pt 6):1403-1421.
- 31 60. Makino S, Stohman SA, Lai MM. Leader sequences of murine coronavirus mRNAs can be freely  
32 reassorted: evidence for the role of free leader RNA in transcription. *Proc.Natl.Acad.Sci.U.S.A* 1986;  
33 83(12):4204-4208.
- 34  
35  
36  
37  
38  
39

- 1 61. Pasternak AO, van den Born E, Spaan WJ, *et al.* Sequence requirements for RNA strand transfer  
2 during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J.* 2001;20(24):7220-7228.
- 3 62. Pasternak AO, van den Born E, Spaan WJ, *et al.* The stability of the duplex between sense and  
4 antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA  
5 synthesis. *J.Virol.* 2003;77(2):1175-1183.
- 6 63. van Marle G, Dobbe JC, Gultyaev AP, *et al.* Arterivirus discontinuous mRNA transcription is guided  
7 by base pairing between sense and antisense transcription-regulating sequences. *Proc.Natl.*  
8 *Acad.Sci.U.S.A* 1999;96(21):12056-12061.
- 9 64. Zuniga S, Sola I, Alonso S, *et al.* Sequence motifs involved in the regulation of discontinuous  
10 coronavirus subgenomic RNA synthesis. *J.Virol.* 2004;78(2):980-994.
- 11 65. Tijms MA, van Dinten LC, Gorbalenya AE, *et al.* A zinc finger-containing papain-like protease  
12 couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus.  
13 *Proc.Natl.Acad.Sci.U.S.A* 2001;98(4):1889-1894.
- 14 66. Tijms MA, Nedialkova DD, Zevenhoven-Dobbe JC, *et al.* Arterivirus subgenomic mRNA synthesis  
15 and virion biogenesis depend on the multifunctional nsp1 autoprotease. *J.Virol.* 2007;81(19):  
16 10496-10505.
- 17 67. van Dinten LC, van Tol H, Gorbalenya AE, *et al.* The predicted metal-binding region of the arteri-  
18 virus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion  
19 biogenesis. *J.Virol.* 2000;74(11):5213-5223.
- 20 68. van Marle G, van Dinten LC, Spaan WJ, *et al.* Characterization of an equine arteritis virus replicase  
21 mutant defective in subgenomic mRNA synthesis. *J.Virol.* 1999;73(7):5274-5281.
- 22 69. Posthuma CC, Nedialkova DD, Zevenhoven-Dobbe JC, *et al.* Site-directed mutagenesis of the  
23 Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life  
24 cycle. *J.Virol.* 2006;80(4):1653-1661.
- 25 70. Enjuanes L, Almazan F, Sola I, *et al.* Biochemical aspects of coronavirus replication and virus-host  
26 interaction. *Annu.Rev.Microbiol.* 2006;60:211-230.
- 27 71. Melnikov S, Ben-Shem A, Garreau de LN, *et al.* One core, two shells: bacterial and eukaryotic  
28 ribosomes. *Nat.Struct.Mol.Biol.* 2012;19(6):560-567.
- 29 72. Lopez-Lastra M, Ramdohr P, Letelier A, *et al.* Translation initiation of viral mRNAs. *Rev.Med.Virol.*  
30 2010;20(3):177-195.
- 31 73. Terenin IM, Andreev DE, Dmitriev SE, *et al.* A novel mechanism of eukaryotic translation initiation  
32 that is neither m7G-cap-, nor IRES-dependent. *Nucleic Acids Res.* 2013;41(3):1807-1816.
- 33 74. Nicholson BL, White KA. 3' Cap-independent translation enhancers of positive-strand RNA plant  
34 viruses. *Curr.Opin.Virol.* 2011;1(5):373-380.
- 35  
36  
37  
38  
39

- 1 75. Cowley JA, Dimmock CM, Walker PJ. Gill-associated nidovirus of *Penaeus monodon* prawns transcribes 3'-coterminally subgenomic mRNAs that do not possess 5'-leader sequences. *J.Gen.Virol.* 2002;83(Pt 4):927-935.
- 2
- 3
- 4 76. Dye C, Siddell SG. Genomic RNA sequence of feline coronavirus strain FCoV C1Je. *J.Feline.Med. Surg.* 2007;9(3):202-213.
- 5
- 6
- 7 77. Schutze H, Ulferts R, Schelle B, *et al.* Characterization of White breem virus reveals a novel genetic cluster of nidoviruses. *J.Virol.* 2006;80(23):11598-11609.
- 8
- 9 78. Wang C, Zhao Q, Liang C, *et al.* Complete genome sequence of a highly pathogenic porcine reproductive and respiratory syndrome virus variant. *J.Virol.* 2012;86(16):8906.
- 10
- 11 79. Zirkel F, Roth H, Kurth A, *et al.* Identification and characterization of genetically divergent members of the newly established family *Mesoniviridae*. *J.Virol.* 2013;87(11):6346-6358.
- 12
- 13 80. van Vliet AL, Smits SL, Rottier PJ, *et al.* Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J.* 2002;21(23):6571-6580.
- 14
- 15 81. Lai MM, Stohlman SA. Comparative analysis of RNA genomes of mouse hepatitis viruses. *J.Virol.* 1981;38(2):661-670.
- 16
- 17 82. Lai MM, Patton CD, Stohlman SA. Further characterization of mRNA's of mouse hepatitis virus: presence of common 5'-end nucleotides. *J.Virol.* 1982;41(2):557-565.
- 18
- 19 83. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion RNA. *Virology* 1986;151(1):146-150.
- 20
- 21 84. Banerjee S, Narayanan K, Mizutani T, *et al.* Murine coronavirus replication-induced p38 mitogen-activated protein kinase activation promotes interleukin-6 production and virus replication in cultured cells. *J.Virol.* 2002;76(12):5937-5948.
- 22
- 23 85. Burgui I, Yanguez E, Sonenberg N, *et al.* Influenza virus mRNA translation revisited: is the eIF4E cap-binding factor required for viral mRNA translation? *J.Virol.* 2007;81(22):12427-12438.
- 24
- 25 86. Chen Z, Faaberg KS, Plagemann PG. Determination of the 5' end of the lactate dehydrogenase-elevating virus genome by two independent approaches. *J.Gen.Virol.* 1994;75 ( Pt 4):925-930.
- 26
- 27 87. Fraser CS, Doudna JA. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nat.Rev.Microbiol.* 2007;5(1):29-38.
- 28
- 29 88. Bidet K, Garcia-Blanco MA. Flaviviral RNAs: weapons and targets in the war between virus and host. *Biochem.J.* 2014;462(2):215-230.
- 30
- 31 89. Jendrach M, Thiel V, Siddell S. Characterization of an internal ribosome entry site within mRNA 5' of murine hepatitis virus. *Arch.Virol.* 1999;144(5):921-933.
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

- 1 90. O'Connor JB, Brian DA. Downstream ribosomal entry for translation of coronavirus TGEV gene 3b.  
2 Virology 2000;269(1):172-182.
- 3 91. Thiel V, Siddell S. Translation of the MHV sM protein is mediated by the internal entry of ribosomes  
4 on mRNA 5. Adv.Exp.Med.Biol. 1995;380:311-315.
- 5 92. Neugebauer KM. On the importance of being co-transcriptional. J.Cell Sci. 2002;115(Pt 20):3865-  
6 3871.
- 7 93. Imbert I, Guillemot JC, Bourhis JM, *et al.* A second, non-canonical RNA-dependent RNA poly-  
8 merase in SARS coronavirus. EMBO J. 2006;25(20):4933-4942.
- 9 94. te Velthuis AJ, Arnold JJ, Cameron CE, *et al.* The RNA polymerase activity of SARS-coronavirus  
10 nsp12 is primer dependent. Nucleic Acids Res. 2010;38(1):203-214.
- 11 95. te Velthuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a  
12 unique multimeric RNA polymerase capable of both de novo initiation and primer extension.  
13 Nucleic Acids Res. 2012;40(4):1737-1747.
- 14 96. Hofmann MA, Brian DA. The 5' end of coronavirus minus-strand RNAs contains a short poly(U)  
15 tract. J.Virol. 1991;65(11):6331-6333.
- 16 97. Ghosh A, Lima CD. Enzymology of RNA cap synthesis. Wiley.Interdiscip.Rev.RNA. 2010;1(1):152-  
17 172.
- 18 98. Decroly E, Ferron F, Lescar J, *et al.* Conventional and unconventional mechanisms for capping viral  
19 mRNA. Nat.Rev.Microbiol. 2012;10(1):51-65.
- 20 99. Liu H, Kiledjian M. Decapping the message: a beginning or an end. Biochem.Soc.Trans. 2006;34(Pt  
21 1):35-38.
- 22 100. Rhoads RE. eIF4E: new family members, new binding partners, new roles. J.Biol.Chem. 2009;  
23 284(25):16711-16715.
- 24 101. Daffis S, Szretter KJ, Schriewer J, *et al.* 2'-O methylation of the viral mRNA cap evades host restric-  
25 tion by IFIT family members. Nature 2010;468(7322):452-456.
- 26 102. Zust R, Cervantes-Barragan L, Habjan M, *et al.* Ribose 2'-O-methylation provides a molecular  
27 signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat.  
28 Immunol. 2011;12(2):137-143.
- 29 103. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of  
30 duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. J.Virol. 2004;  
31 78(14):7833-7838.
- 32 104. Ivanov KA, Thiel V, Dobbe JC, *et al.* Multiple enzymatic activities associated with severe acute  
33 respiratory syndrome coronavirus helicase. J.Virol. 2004;78(11):5619-5632.
- 34  
35  
36  
37  
38  
39

- 1 105. Bouvet M, Debarnot C, Imbert I, *et al.* *In vitro* reconstitution of SARS-coronavirus mRNA cap  
2 methylation. *PLoS.Pathog.* 2010;6(4):e1000863.
- 3 106. Chen Y, Cai H, Pan J, *et al.* Functional screen reveals SARS coronavirus nonstructural protein nsp14  
4 as a novel cap N7 methyltransferase. *Proc.Natl.Acad.Sci.U.S.A* 2009;106(9):3484-3489.
- 5 107. Decroly E, Imbert I, Coutard B, *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding  
6 enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J.Virol.* 2008;82(16):8071-8084.
- 7  
8 108. Decroly E, Debarnot C, Ferron F, *et al.* Crystal structure and functional analysis of the SARS-  
9 coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS.Pathog.* 2011;7(5):  
10 e1002059.
- 11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

Structural basis for the regulatory  
function of a complex zinc-binding  
domain in a replicative arterivirus  
helicase resembling a nonsense-  
mediated mRNA decay helicase

## CHAPTER 2

Zengqin Deng  
Kathleen C. Lehmann  
Xiaorong Li  
Chong Feng  
Guoqiang Wang  
Qi Zhang  
Xiaoxuan Qi  
Lin Yu  
Xingliang Zhang  
Wenhai Feng  
Wei Wu  
Peng Gong  
Ye Tao  
Clara C. Posthuma  
Eric J. Snijder  
Alexander E. Gorbalenya  
and Zhongzhou Chen

*Nucleic Acids Res.* 2014; 42(5):3464-3477

**1 ABSTRACT**

2  
3 All positive-stranded RNA viruses with genomes larger than ~7 kilobases encode heli-  
4 cases, which generally are poorly characterized. The core of the nidovirus superfamily 1  
5 helicase (HEL1) is associated with a unique N-terminal zinc-binding domain (ZBD) that  
6 was previously implicated in helicase regulation, genome replication, and subgenomic  
7 mRNA synthesis. The high-resolution structure of the arterivirus helicase (nsp10), alone  
8 and in complex with a polynucleotide substrate, now provides first insights into the  
9 structural basis for nidovirus helicase function. A previously uncharacterized domain  
10 1B connects HEL1 domains 1A and 2A to a long linker of the ZBD, which further consists  
11 of a novel RING-like module and treble-clef zinc finger, together coordinating three Zn  
12 atoms. On substrate binding, major conformational changes were evident outside the  
13 HEL1 domains, notably in domain 1B. Structural characterization, mutagenesis, and bio-  
14 chemistry revealed that helicase activity depends on the extensive relay of interactions  
15 between the ZBD and HEL1 domains. The arterivirus helicase structurally resembles the  
16 cellular Upf1 helicase, suggesting that nidoviruses may also employ their helicases for  
17 post-transcriptional quality control of their large RNA genomes.

18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 1 INTRODUCTION

2  
3 Helicases and nucleic acid translocases are ATP-dependent motor proteins capable of  
4 moving along their nucleic acid substrates while either unwinding duplexed regions  
5 (helicases) or performing other functions (translocases), including protein displacement  
6 and the nucleation of larger RNA-protein complexes (1,2). These enzymes are known  
7 to be critical players in a wide variety of biological processes and are encoded by all  
8 organisms, as well as positive-stranded (+) RNA viruses with genomes larger than about  
9 7 kilobases (kb) ((3); for reviews, see (4-6)). On the basis of sequence comparisons, heli-  
10 cases/translocases have been classified into six superfamilies (SF1 to SF6) (7,8), with (+)  
11 RNA viral helicases belonging to SF1, SF2, or SF3. Based on the direction of translocation,  
12 helicases of various superfamilies have been divided into (biochemical) classes A and B,  
13 which translocate along their nucleic acid substrates in the 3'-5' or 5'-3' direction, respec-  
14 tively (7). In the case of SF1 helicases (9,10), structurally characterized cellular enzymes  
15 of class B (SF1B) are further divided into the phylogenetically compact Pif1-like (Pif1,  
16 RecD2), UvrD/Rep, and Upf1-like (Upf1, Ighmbp2) groups, with the latter being able to  
17 unwind both DNA and RNA duplexes (11).

18  
19 Helicase SF1 also includes a large number of (putative) helicases from a dozen (+) RNA  
20 virus families belonging to two diverse phylogenetic lineages, known as the alphavirus-  
21 like (or Sindbis virus-like) supergroup (12) and the order *Nidovirales* (13). More detailed  
22 studies on the SF1 helicases of two alphavirus-like viruses have recently been published.  
23 The helicase domain of the dendrolimus punctatus tetravirus (an insect virus from  
24 the *Alphatetraviridae* family) was found to have dsRNA-unwinding activity with 5'-3'  
25 directionality (14). The helicase domain of the plant tomato mosaic virus (ToMV; family  
26 *Virgaviridae*) was not characterized enzymatically, but its crystal structure revealed the  
27 two canonical RecA-like  $\alpha/\beta$  domains (1A and 2A) of the helicase core (15). Accessory  
28 domain insertions, an otherwise frequently observed phenomenon among cellular SF1  
29 helicases, are lacking in the ToMV helicase. The SF1 helicases of nidoviruses, one of  
30 which is the focus of this study, were characterized in some detail using bioinformatics,  
31 molecular genetics, and biochemistry (see below), but structural information was lack-  
32 ing thus far.

33  
34 Nidoviruses constitute an order of (+) RNA viruses comprised of virus groups target-  
35 ing a wide variety of mammalian, avian, and invertebrate hosts. In mammals nidovirus  
36 infection can be associated with severe respiratory disease, as in the case of porcine  
37 reproductive and respiratory syndrome (PRRS) (16), one of the leading swine diseases  
38 (caused by arteriviruses), and zoonotic coronavirus infections in humans, like severe  
39 acute respiratory syndrome (SARS) (17) and Middle East respiratory syndrome (MERS)

1 (18). The continuing outbreak of the latter disease is currently attracting worldwide  
2 attention, in particular because of its ~40% case fatality rate. Besides their pathogenic  
3 properties, nidoviruses have been studied for their extraordinary large RNA genomes;  
4 even the shortest nidovirus genome (the 12.7 kb RNA of the arterivirus equine arteritis  
5 virus, EAV) outranks almost all other mammalian (+) RNA virus genomes, whereas coro-  
6 navirus genomes (26.3-31.7 kb) are larger than those of any other RNA virus group. Their  
7 large genome size enabled nidoviruses to evolve substantial genetic complexity, which  
8 is evident from (among other properties) the acquisition of a variety of enzymatic activi-  
9 ties and accessory proteins, many of which are lacking or rare in other (+) RNA viruses  
10 (19). These proteins appear to contribute to the regulation of the complex RNA synthesis  
11 of nidoviruses, which occurs exclusively in the cytoplasm of the infected cell, and to the  
12 elaborate array of virus-host interactions needed to support efficient virus replication  
13 (13,20). For example, nidoviruses with genomes larger than 20 kb employ a proofread-  
14 ing 3'-5' RNA exonuclease that is proposed to promote the fidelity of viral RNA synthesis  
15 (19,21-27). However, it is completely unknown whether and how nidoviruses deal with  
16 translational quality control during the expression of their large multicistronic genomic  
17 RNAs, which also serve as mRNAs for the synthesis of the viral replicative enzymes.

18  
19 Compared to other (+) RNA viruses, nidovirus replicase genes encode an exceptionally  
20 large number of nonstructural proteins (nsps) (19,24,25,28). Nidovirus nsps are expressed  
21 from open reading frames (ORFs) 1a and 1b, which make up the 5'-proximal 65-75% of  
22 the genome RNA. ORF1a encodes polyprotein 1a (pp1a; size ranging from 1728 to 4550  
23 amino acids) and following a -1 ribosomal frameshift pp1a can be extended with the  
24 ORF1b-encoded polyprotein to give pp1ab (3175 to 7183 amino acids) (29) (Figure S1).  
25 Both polyproteins are subject to extensive proteolytic processing by multiple internally  
26 encoded proteinases (19,30). The nidovirus replicase backbone consists of a conserved  
27 array of domains, arranged in a nidovirus-specific order and including the ORF1b-en-  
28 coded RNA-dependent RNA polymerase (RdRp) and helicase domains, the core enzymes  
29 needed for genome RNA synthesis (replication) and subgenomic (sg) mRNA production  
30 (transcription). The latter process yields an extensive nested set of sg mRNAs, which is  
31 used to express up to a dozen structural and accessory proteins from smaller ORFs in  
32 the 3'-proximal part of the genome (31-33). In both corona- and arteriviruses, sg mRNAs  
33 contain a common leader sequence that is identical to the 5' end of the genome. Their  
34 generation from subgenome-size negative-stranded templates involves a mechanism  
35 of discontinuous negative-strand RNA synthesis (31,32).

36  
37 Previous studies identified the nsp carrying RNA helicase activity (arterivirus nsp10 and  
38 coronavirus nsp13) as one of the two most evolutionarily conserved nidovirus proteins.  
39 Biochemical studies using recombinant arterivirus and coronavirus helicases revealed

1 very similar enzymatic properties, including nucleic acid-stimulated ATPase and 5'-3' du-  
2 plex unwinding activities on both RNA and DNA substrates containing 5' single-stranded  
3 regions (34,35). A unique nidovirus helicase feature is the presence of an N-terminal  
4 (predicted) complex zinc-binding domain (ZBD) of 80-100 residues. The ZBD includes  
5 12 or 13 conserved Cys/His residues (36) and is a nidoviral genetic marker not found in  
6 any other RNA virus group (19). The ZBD is separated from the downstream helicase core  
7 domains (HEL1) by an uncharacterized domain that varies in size and sequence between  
8 arteri- and coronaviruses (37). For the arterivirus prototype EAV, the significance of the  
9 nsp10 ZBD was evaluated extensively using site-directed mutagenesis in combination  
10 with biochemical assays and reverse genetics. Amino acid substitutions in the ZBD or  
11 the adjacent "spacer" that connects it to the downstream domain can profoundly affect  
12 EAV helicase activity and RNA synthesis, with most replacements of conserved Cys or  
13 His residues yielding replication-negative virus phenotypes (36,37). Intriguingly, some  
14 mutations in the spacer region selectively inactivated transcription while not affecting  
15 replication (36,38), strongly suggesting a specific role for nsp10 in the unique mecha-  
16 nism of discontinuous sg RNA synthesis.

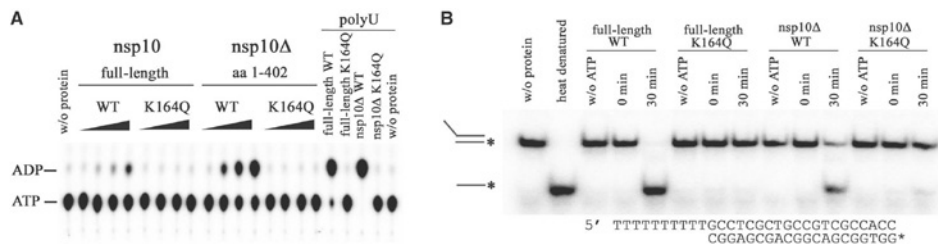
17  
18 Despite its importance as a key replicative enzyme and antiviral drug target (39), no  
19 three-dimensional structural information has been reported for any nidovirus helicase.  
20 To understand the regulatory role of the ZBD and the protein's interaction with nucleic  
21 acids, we characterized the structure of a helicase-competent derivative of EAV nsp10,  
22 alone and in complex with poly(dT). The multi-domain nsp10 includes the canonical  
23 1A and 2A core domains of a SF1 helicase, a flexible accessory domain that is sensitive  
24 to nucleic acid binding, and a complex ZBD displaying a novel structural organization.  
25 Strikingly, the protein was found to bear structural resemblance to the eukaryotic Upf1  
26 helicases, which are multi-domain proteins involved in RNA quality control, including  
27 nonsense-mediated mRNA decay (40). Thus, our study not only highlights how nidovirus  
28 helicase activity depends on the extensive relay of interactions between the ZBD and  
29 HEL1 domains but also provides a framework to propose and explore a role for the  
30 enzyme in the post-transcriptional quality control of nidovirus RNAs.

## 31 32 33 **RESULTS**

### 34 35 **C-terminally truncated EAV nsp10 retains ATPase and helicase activity**

36  
37 Full-length EAV nsp10 and a series of truncated variants were overexpressed in and puri-  
38 fied from *E. coli*. After extensive crystallization trials, diffracting crystals could only be  
39 obtained for a truncated form of nsp10 (aa 1-402) lacking the 65 C-terminal residues. For

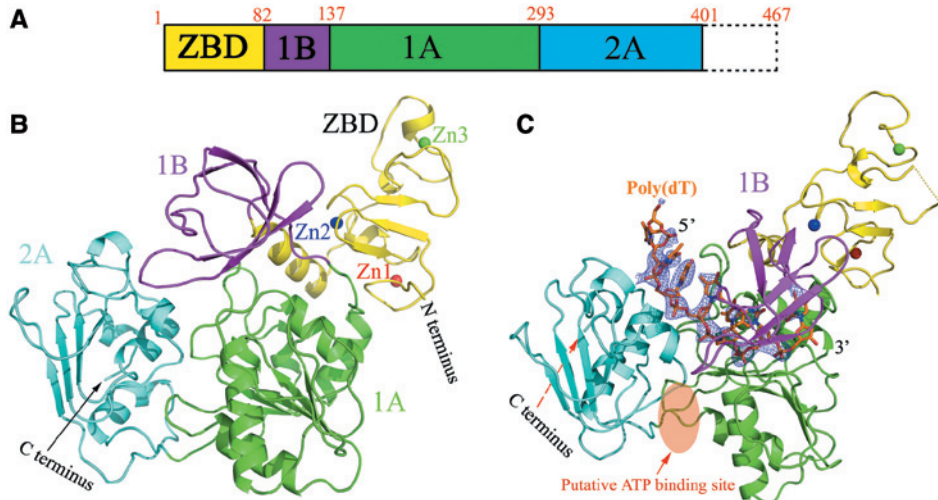
1 simplicity, we will hereafter refer to this protein as nsp10 $\Delta$ , which was used throughout  
 2 this study unless otherwise specified. To verify that nsp10 $\Delta$ , which contained all charac-  
 3 teristic SF1 helicase sequences (motifs), is enzymatically active, we performed *in vitro*  
 4 enzyme assays to compare full-length and truncated nsp10. In agreement with previ-  
 5 ously published results (35), full-length nsp10 displayed only weak ATPase activity in  
 6 the absence of nucleic acid but was strongly stimulated by the addition of poly-uridine  
 7 (polyU). In the absence of polyU nsp10 $\Delta$  showed a five-fold higher ATPase activity than  
 8 the full-length protein (Figure 1A), yet this increased ATP turnover did apparently not  
 9 translate into increased helicase activity. Unwinding of a partially double-stranded DNA  
 10 substrate by nsp10 $\Delta$  was incomplete but went to completion when using full-length  
 11 nsp10 (Figure 1B). As expected, replacement of the conserved lysine of the Walker A  
 12 motif, which is essential for ATP hydrolysis (35), with glutamine (mutant K164Q) com-  
 13 pletely abolished ATPase and consequentially also helicase activity. This confirmed that  
 14 the observed activities could be completely attributed to the recombinant EAV proteins  
 15 used rather than to potential trace amounts of contaminating bacterial enzymes.



**Figure 1.** EAV nsp10 *in vitro* enzymatic activity assays. **(A)** ATPase activity of full-length EAV nsp10, the C-terminally truncated nsp10 $\Delta$  (amino acids 1-402) and respective active site mutants carrying a Lys-164 to Gln substitution in their Walker A box were analyzed as described in Supplementary Experimental Procedures. In the absence of nucleic acid, ATPase activity was measured by incubation at 20°C for 0, 5, 15 and 30 min, respectively. ATPase activity was strongly stimulated by the presence of poly-uridine (polyU), as measured by incubation at 20°C for 5 min. **(B)** Helicase activity of full-length nsp10 and nsp10 $\Delta$ , and their respective K164Q mutants. Activity was determined with the indicated DNA substrate (the asterisk marks the position of the radioactive label). Samples were incubated for 0 or 30 min at 30°C. Control samples without protein or ATP were incubated for 30 min.

32 The observed enzymatic differences between nsp10 and nsp10 $\Delta$  may be caused by the  
 33 latter's truncation and could, in principle, be explained by one or multiple defects, like  
 34 decreased unwinding velocity and/or processivity, loss of affinity towards the substrate,  
 35 or uncoupling of ATPase from helicase activity. The results of the ATPase assay lead us  
 36 to propose that the observed reduction of duplex unwinding may be due to unproduc-  
 37 tive ATP hydrolysis, originating from the fact that the ATPase reaction is independent  
 38 of nucleic acid substrate binding. Accordingly, the input ATP in the nsp10 $\Delta$  assay may  
 39 have been depleted before complete unwinding was achieved. Regardless of which

1 interpretation is correct, the C-terminal 65 amino acids clearly are dispensable for the  
 2 helicase activity of EAV nsp10. This result is in good agreement with the fact that the  
 3 truncated protein retained all HEL1 key domains (Figure 2A) previously shown to be  
 4 evolutionary conserved and essential in both *in vitro* enzyme assays and *in vivo* studies  
 5 with virus mutants.



**Figure 2.** Overall structures of EAV nsp10 $\Delta$  and the nsp10 $\Delta$ -DNA binary complex. **(A)** Domain organization of EAV nsp10 depicting the N-terminal zinc-binding domain (ZBD; yellow), the two RecA-like domains 1A (green) and 2A (cyan) of HEL1, and an additional regulatory domain 1B (magenta). Structure of **(B)** free and **(C)** nucleic acid-bound nsp10 $\Delta$ . Also the  $F_o - F_c$  differential electron density map of the bound single-stranded part of a partially double-stranded DNA substrate at 2.5  $\sigma$  is presented. The putative ATP binding site is shown as a red oval.

## 27 The crystal structure of EAV nsp10 $\Delta$ reveals a multi-domain organization of the 28 arterivirus replicative helicase

29  
 30 Since three-dimensional structures of orthologous proteins were not available, we took  
 31 advantage of the zinc-binding properties of nsp10 and used the zinc multiple-wave-  
 32 length anomalous diffraction (MAD) method (42) to solve the EAV nsp10 $\Delta$  structure.  
 33 The presence and position of three zinc atoms were established with anomalous data  
 34 collected from the zinc absorption edge (Table 1). The final model included EAV nsp10  
 35 residues 1-401, 3 zinc ions in the N-terminal ZBD, 5 sulfate ions, and 267 water molecules.  
 36

37 Two RecA-like  $\alpha/\beta$  domains (1A and 2A) form the structure's C-terminal part (Figure 2B;  
 38 cyan and green) and constitute the helicase core (HEL1). Domain 1A contains a paral-  
 39 lel five-stranded  $\beta$ -sheet that is sandwiched by three  $\alpha$ -helices on one side and two



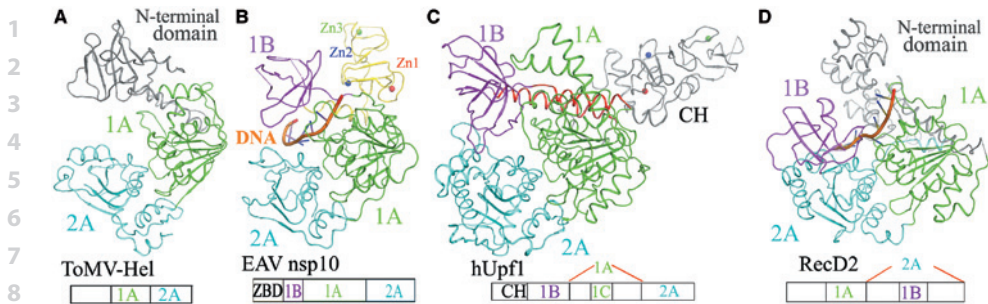
**Table 1.** Data collection and refinement statistics of nsp10 $\Delta$  and the nsp10 $\Delta$ -DNA complex\*.

Data collection	Zn-Peak	Zn-edge	Zn-remote	nsp10 $\Delta$	nsp10 $\Delta$ -DNA
Wavelength	1.2827	1.2831	1.0000	1.0000	1.0000
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	P1
<b>Cell dimensions</b>					
<i>a,b,c</i> (Å)	89.1,90.6,56.9	89.1,90.6,56.9	89.1,90.6,56.9	89.8,91.0,57.7	56.6,88.8,128.8
$\alpha,\beta,\gamma$ (°)	90,90,90	90,90,90	90,90,90	90,90,90	81.7,90.0,71.4
Resolution (Å) <sup>†</sup>	50-2.83 (2.88-2.83)	50-2.80 (2.85-2.80)	50-2.75 (2.80-2.75)	50-2.0 (2.03-2.0)	50-2.65 (2.7-2.65)
<i>R</i> <sub>merge</sub> (%)	8.7 (38.9)	9.0 (39.9)	9.3 (42.8)	7.4 (67.0)	10.0 (73.2)
<i>I</i> / $\sigma$	15.2 (2.2)	15.5 (1.9)	17 (2.8)	42.0 (2)	25.6 (3.4)
Completeness (%)	97.6 (87.2)	98.8 (84.5)	99.1 (89)	97.8 (86.1)	96.1 (95.3)
Redundancy	6.9 (5.3)	6.8 (4.3)	10.1 (6.7)	12.5 (8.9)	4.9 (4.8)
<b>Refinement</b>					
Resolution (Å)				50-2.0 (2.05-2.0)	50-2.65 (2.72-2.65)
No. of reflections				30451 (1874)	62140 (4421)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)				19.5/22.4	23.2/25.7
<b>No. of atoms</b>					
Protein				2986	11594
DNA/ion				29	567
Water				252	822
<b>B-factors</b>					
Protein				42.7	49.6
DNA/ion				44.3	62.0
Water				50.4	42.0
<b>RMS deviations</b>					
Bond lengths (Å)				0.008	0.007
Bond angles (°)				1.17	1.22
Ramachandran Plot (%) <sup>§</sup>				93.9/6.1/0/0	83.3/15.9/0.8/0

\*Three crystal experiments for each structure. <sup>†</sup> Statistics for highest resolution shell.

<sup>§</sup>Residues in most favored, additional allowed, generously allowed and disallowed regions of the Ramachandran plot.

$\alpha$ -helices on the other. Domain 2A contains a parallel four-stranded  $\beta$ -sheet with five  $\alpha$ -helices on the side facing domain 1A. Upstream of domain 1A, we identified an additional domain with a characteristic  $\beta$ -barrel fold (Figures 2A and 1B; magenta). It consists of five  $\beta$ -strands arranged as two tightly packed anti-parallel  $\beta$ -sheets and is juxtaposed to domain 1A (Figure 2B). The location of this domain in the protein sequence and its orientation relative to the HEL1 domain resemble those of domain 1B in helicases of the SF1B Upf1-like subfamily (Figures 3B and 3C), and it was therefore named accord-

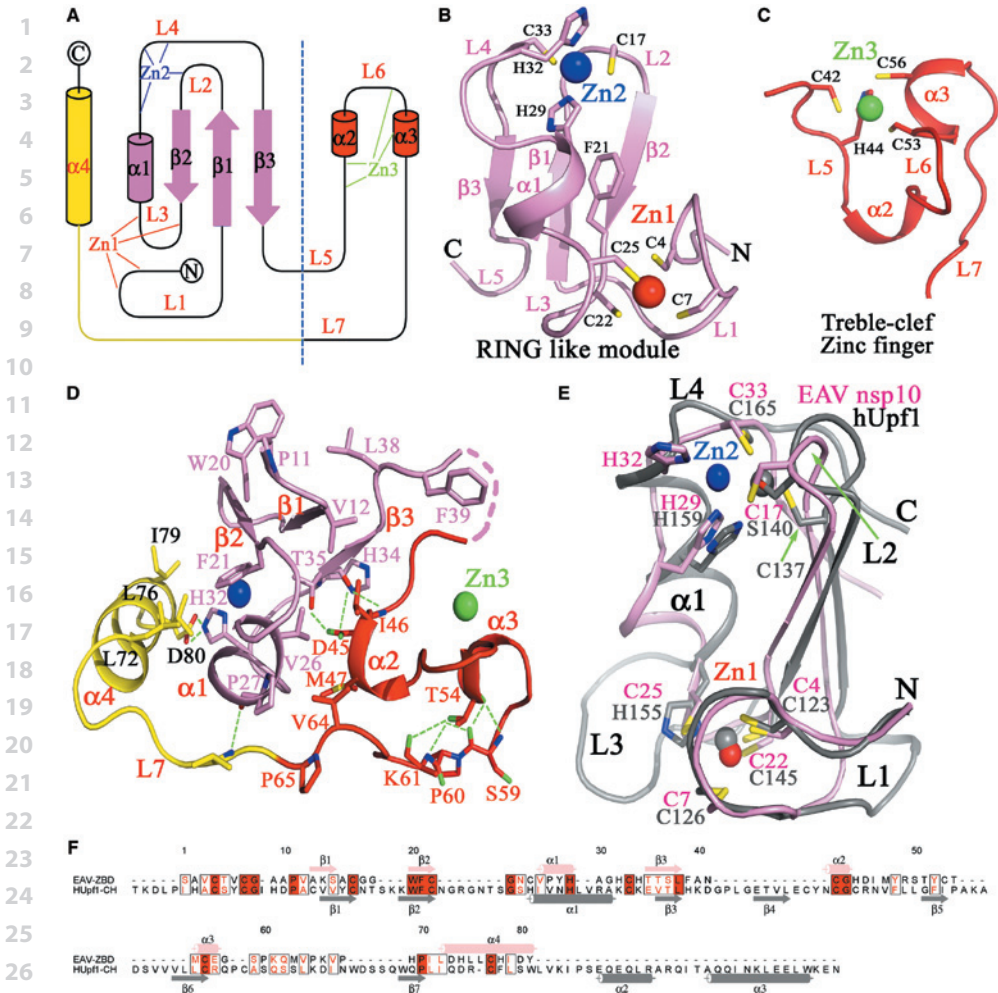


**Figure 3.** Structural comparison of EAV nsp10 $\Delta$  with selected SF1 helicases. **(A)** ToMV-HEL (pdb code: 3vkw), **(B)** EAV nsp10 $\Delta$ , **(C)** hUpf1 (pdb code: 2wjy) and **(D)** RecD2 (pdb code: 3gp8). Domain colors are the same as in Figure 2.

ingly in our nsp10 $\Delta$  structure. The domain has no counterpart in the only other solved structure of a viral SF1 helicase, that from ToMV (Figure 3A) (15), whereas its counterpart in helicases of the Pif1-like subfamily is inserted in domain 2A (Figure 3D) (48).

Our structure further revealed that the N-terminal ZBD (Figure 2; yellow) has a compact fold containing three structural zinc atoms. Based on secondary structure analysis with DIAL (49), we could partition the ZBD into three elements (Figure 4). Two adjacent and structurally different zinc fingers, an N-terminal RING-like module (residues 1 to 40, pink), and a treble-clef zinc finger (residues 41 to 65, red) constitute the main body of the ZBD. The third element is a C-terminal linker region (Linker1) that includes the long loop L7, which crosses the entire domain, and helix  $\alpha$ 4 (residues 66 to 82, yellow), which connects the two zinc fingers with domain 1B (Figure 4A). This classification is further supported by the observation that the connecting residues between the RING module and treble-clef zinc finger are disordered (Figures S2 and 4D). Only 12 out of the 13 Cys/His residues are involved in zinc binding rather than all 13 residues as proposed previously ((36); Figures 4B and 4C). Not involved is His34, which is not conserved in other arteri- and coronaviruses (Figure S3B).

The N-terminal RING-like module has a notable binuclear structure with a cross-brace topology involving six Cys and two His residues that coordinate two zinc ions (Figure 4A). A three-stranded antiparallel  $\beta$ -sheet ( $\beta$ 1- $\beta$ 3) sits in the center and packs against helix  $\alpha$ 1 following  $\beta$ 2 (Figure 4B). The first zinc ion (Zn1) is coordinated by four cysteine residues (Cys4, Cys7, Cys22, and Cys25) within a treble-clef zinc finger-like motif. Residues Cys4 and Cys7 are provided by the zinc knuckle within loop L1 whereas Cys22 is positioned at the C-terminus of  $\beta$ 2 and Cys25 comes from the N-terminus of helix  $\alpha$ 1. The second zinc ion (Zn2) is coordinated by residues Cys17, Cys33, His29, and His32, which are arranged in an  $\alpha\beta$  zinc finger-like motif. The second pair of the zinc-coordinating residues of both



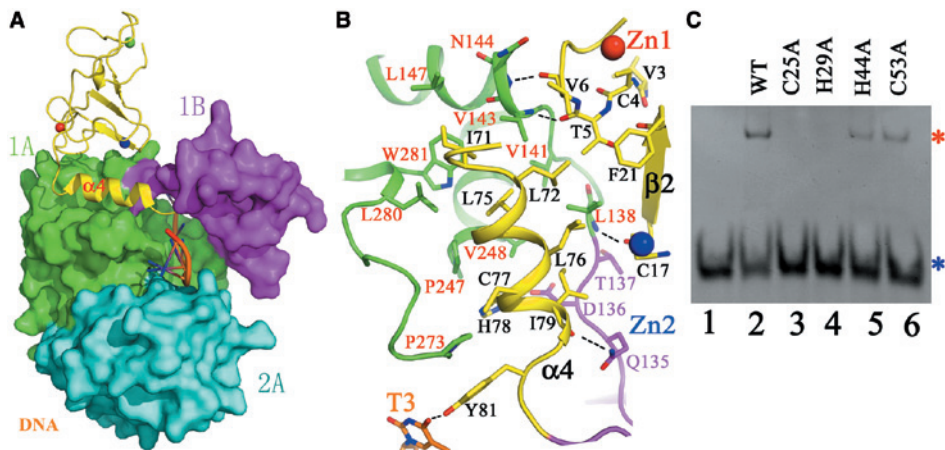
**Figure 4.** Structural characterization of the EAV nsp10 $\Delta$  ZBD. **(A)** Topology of the ZBD with its RING-like module (pink), treble-clef zinc finger (red), and Linker 1 (yellow) indicated. **(B)** Structure of the RING-like module and **(C)** treble-clef zinc finger. The residues coordinating the Zn<sup>2+</sup> ions are shown as sticks. **(D)** Interactions between the RING-like module and the treble-clef zinc finger. **(E)** Superposition of the RING-like modules of EAV nsp10 (pink) and hUpf1 (pdb code: 2wjy; gray). **(F)** Sequence alignment of ZBD with the CH domain of hUpf1.

zinc-binding motifs of the RING module may include both His and Cys residues in other arteri- and coronaviruses. Overall, the RING module of these viruses can be described by a characteristic, conserved Cys2A-CysB-CysA-[His/Cys]A-[His/Cys]3B pattern (where applicable, A and B refer to residues chelating the first and second zinc ion, respectively; brackets indicate positions at which either His or Cys can be present).

1 The C-terminal zinc finger of the ZBD adopts a treble-clef fold distinct from that of the  
 2 RING module (see above; Figure 4C). Two one-turn helices  $\alpha 2$  and  $\alpha 3$  are stabilized by a  
 3 zinc atom (Zn3) that is chelated by residues Cys42 and His44 of a Zn-knuckle within loop  
 4 L5 while Cys53 and Cys56 originate from L6 and  $\alpha 3$ , respectively. An extensive array of  
 5 hydrogen bonds is observed between the main chains of residues in loop L7 and Thr54  
 6 in  $\alpha 3$  (Figure 4D). These multiple hydrogen-bonding interactions play a major role in the  
 7 formation of a compact zinc finger. Arteri- and coronaviruses appear to tolerate replace-  
 8 ments (Cys for His or vice versa) at the second and fourth residues of this finger (36,37),  
 9 which can be described by the characteristic, conserved Cys-[His/Cys]-Cys-[His/Cys] pat-  
 10 tern. Finally, Linker1 includes only one structured element ( $\alpha 4$ ), but it plays a central role in  
 11 the interaction between the main body of the ZBD and HEL1, as detailed below.

### 13 The structural basis for the essential role of the ZBD in EAV nsp10 helicase 14 function

16 Previously, ZBD mutagenesis demonstrated the *in vitro* and *in vivo* importance of this  
 17 domain for nsp10 enzyme activities, genome replication and transcription, and arteri-  
 18 virus viability. The solved structure now provides us with a structural basis for these  
 19 observations. The ZBD packs against the HEL1 domain through extensive hydrophobic  
 20 and hydrophilic interactions (Figures 5A and 5B). Specifically, residues Leu138, Val141,  
 21 Val143, Leu147, Pro247, Val248, Leu280, and Trp281 in domain 1A together with residues  
 22 Ile71, Leu72, Leu75, Leu76, and Ile79 from  $\alpha 4$  in the ZBD create an extensive hydropho-  
 23



36 **Figure 5.** Inter-domain interactions of the ZBD and HEL1 domains 1A and 1B. (A) Overview of the spatial  
 37 orientation of the essential interaction helix  $\alpha 4$  of the ZBD. (B) Close-up view of the domain interface. Resi-  
 38 dues engaged in interactions are shown as sticks. Domain colors are the same as in Figure 2A. (C) DNA-  
 39 binding assay of EAV nsp10 $\Delta$  mutants with reduced Zn<sup>2+</sup>-binding capabilities. Position of free DNA and  
 protein-DNA complexes are indicated by blue and red asterisk, respectively.

1 bic surface. The total interface area between the ZBD and HEL1 is  $1019 \text{ \AA}^2$ , as determined  
2 by Protein Interfaces, Surfaces, and Assemblies (PISA) (50). A major part of this interface  
3 involves the  $\alpha 4$  helix, which is located in a groove formed by two helices and a loop of  
4 domain 1A while making extensive contacts to the main body of the ZBD and, to lesser  
5 extent, domain 1B (Figure 5). The interface areas between  $\alpha 4$  and domain 1A, on the  
6 one hand, and the ZBD fingers (including zinc ions) on the other hand, are  $558.1 \text{ \AA}^2$   
7 and  $402.4 \text{ \AA}^2$ , respectively. In addition, four hydrogen bonds between the ZBD and HEL1  
8 enhance the interaction (Figure 5B), and a salt bridge is observed between His78 in the  
9 ZBD and Asp136 in domain 1B (Figure 5B). The large size of these interface surfaces and  
10 the large number of interactions suggest the existence of a signaling network through  
11 which the ZBD could affect both the fold and activity of HEL1.

12  
13 The proposed signaling network can now be used to rationalize, in a structural con-  
14 text, the previously reported phenotypes of EAV ZBD mutants carrying replacements  
15 of residues not directly involved in zinc-binding. For instance, a replication-negative  
16 phenotype was described for mutant D45A (36). It is now clear that Asp45 forms two  
17 hydrogen bonds with the main and side chain of Thr35 and electrostatically interacts  
18 with the side chain of His34, which both belong to the RING-like zinc finger (Figure 4D).  
19 Replacement of Asp45 may thus greatly reduce these interactions and disrupt ZBD in-  
20 tegrity, potentially affecting the structural integrity of HEL1. Another residue, Ser59, was  
21 probed extensively by mutagenesis after the finding that a virus mutant (EAV030F) car-  
22 rying a S59P mutation replicates its genomic RNA with wild-type efficiency while being  
23 completely defective in sg mRNA synthesis (38). This transcription-negative phenotype  
24 was attributed to the severe structural constraints exerted by Pro residues on the local  
25 conformation of the proposed hinge region, since various substitutions of Ser59 alone  
26 (to Ala, Cys, Gly, His, Leu, or Thr) yielded virus mutants with a wild-type phenotype while  
27 combining the neutral S59G mutation with a P60G substitution reproduced the specific  
28 defect in sg mRNA synthesis (36). This interpretation is now further supported by the  
29 nsp10 $\Delta$  structure in which Ser59 and Pro60 are located in the hinge connecting the  
30 treble-clef zinc finger and  $\alpha 4$  of the ZBD. The main chain of Ser59 forms three hydrogen  
31 bonds with the treble-clef Thr54, which is also connected to the Pro60 side chain and  
32 Lys61 main chain (Figure 4D). Due to the unique properties of the Pro residue, the Ser59-  
33 to-Thr54 bonds are likely disrupted by the S59P mutation but are not affected by the  
34 alternative replacements tested. Consequently, also due to the main chain rigidity as-  
35 sociated with the introduction of a Pro residue, the orientation of  $\alpha 4$  relative to 1A and/  
36 or the main body of the ZBD is likely affected in mutant S59P, which carries adjacent Pro  
37 residues at positions 59 and 60. Likewise, the introduction of two Gly residues at these  
38 positions (double mutant S59G/P60G; (36)) probably gives rise to excessive flexibility of  
39 the hinge region thus compromising nsp10 function in a similar manner.

1 To further explore the role of the ZBD, we tested the effect of four mutations (C25A and  
2 H29A in the RING-like module; H44A and C53A in the treble-clef zinc finger) expected to  
3 affect the ability to bind Zn1, Zn2, or Zn3, respectively. In agreement with the proposed  
4 structural role of these zinc ions, soluble His-tagged proteins containing these muta-  
5 tions could not be obtained, and only low yields of GST-nsp10 fusion proteins carrying  
6 the same mutations could be recovered. For mutants C25A and H29A, band shift analysis  
7 revealed a complete loss of binding to a partially double-stranded DNA substrate con-  
8 taining a 5' single-stranded poly(dT) overhang (substrate 5'-DNA-T10; Figure 5C, lanes  
9 3 and 4). These results complement previous findings showing a complete loss of both  
10 ATPase and helicase activity for these mutants (37). In contrast, the level of nucleic acid  
11 binding by mutants H44A and C53A was comparable to that of the wild-type protein  
12 (Figure 5C, lanes 5 and 6), consistent with nsp10-H44A retaining a limited level of ATPase  
13 and helicase activity (37). Upon further testing, we observed that the addition of 40 mM  
14 EDTA altered the overall conformation of nsp10 $\Delta$ , as detected by changes in circular  
15 dichroism (Figure S4A), and reduced its binding to 5'-DNA-A10 (Figure S4B). In summary,  
16 these results reveal that the ZBD interacts extensively with the HEL1 domain and that its  
17 integrity is an essential determinant of nsp10 $\Delta$  properties in *in vitro* assays.

### 18 19 **Structural resemblance between EAV nsp10 $\Delta$ and mRNA decay factor Upf1**

20  
21 Next we analyzed the existence of structural similarity between EAV nsp10 $\Delta$  and other  
22 proteins by scanning a protein data bank using the DALI server (51). The structure of  
23 the nsp10 $\Delta$  HEL1 domain was found to be most similar (Z score, 20.9; root-mean-square  
24 deviation (RMSD), 3.5 Å) to the helicase core of nonsense-mediated mRNA decay factor  
25 Upf1 and its homolog Ighmbp2 (Z score, 19.9; RMSD, 3.0 Å), which both belong to the  
26 Upf1-like helicase subfamily (11). Further comparisons revealed that this resemblance  
27 extends into the respective N-terminal zinc-binding domains. The binuclear RING-like  
28 module of the nsp10 $\Delta$  ZBD was found to be most similar to RING-like module 1 in the CH-  
29 domain of Upf1 (Figure 4E). This similarity was rather limited (Z-score of 1.9 and RMSD of  
30 2.2 Å) because only six out of the eight zinc-chelating residues in the two domains could  
31 be juxtaposed (Figure 4F) and because loops L1, L3 and helix  $\alpha$ 1 in nsp10 $\Delta$  are shorter  
32 than the corresponding elements in Upf1. We did not detect significant similarity of the  
33 treble-clef zinc finger with other proteins although we note that the Upf1 CH-domain  
34 also has a zinc finger (but of a different fold) downstream of the RING1 module. Thus, the  
35 EAV nsp10 ZBD prototypes a novel and complex multi-domain zinc finger with distinct  
36 structural properties. On the other hand, EAV nsp10 and Upf1 share a similar domain  
37 organization, including structurally similar RING and helicase domains. These similarities  
38 are further enhanced by the 5'-3' directionality of duplex unwinding shared by both  
39

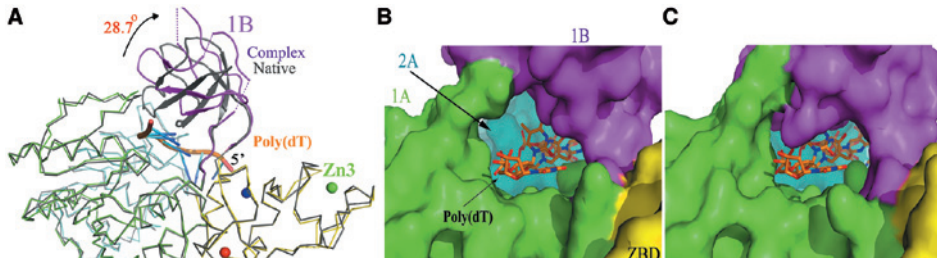
1 these helicases and likely extend to other nidovirus helicases in view of the observed  
2 sequence conservation (Figure S3B).

### 3 4 **Structure of EAV nsp10Δ in complex with a nucleic acid substrate**

5  
6 We proceeded to solve the crystal structure of nsp10Δ in complex with a nucleic acid  
7 substrate. Nidovirus RNA helicases, including EAV nsp10, were previously found to lack  
8 the ability to discriminate between RNA and DNA substrates, a property shared with  
9 only a few other helicases (34,35). This substrate promiscuity allowed us to use a partially  
10 double-stranded DNA substrate (5'-DNA-T10) containing a 5' single-stranded poly(dT)  
11 overhang for crystallographic studies. The binding of this substrate was deduced from  
12 an increase of the protein's Stokes radius in gel filtration chromatography (Figure S5).  
13 The binary complex diffracted to a resolution of 2.65 Å in space group P1 and was solved  
14 by molecular replacement (Table 1). Continuous electron density was found in the en-  
15 zyme's binding pocket (Figure 2C), which apparently corresponded to seven thymidine  
16 residues. This part presents in an extended conformation and lies in a channel formed  
17 by domains 1A, 1B, and 2A, with its 5' end in domain 2A and its 3' end in domain 1A.  
18 The remaining three unpaired thymidines and the entire double-stranded portion of  
19 the substrate could not be located. The asymmetric unit contained four nsp10Δ-DNA  
20 binary complexes with a Matthews coefficient of 2.73 Å<sup>3</sup>/Da, corresponding to a solvent  
21 content of 55%. These complexes shared a remarkably similar spatial arrangement with  
22 the RMSD of their Cα atoms being only 0.8 Å. Several connecting residues between sub-  
23 domains were missing in the structure of the complex, indicating apparent structural  
24 flexibility of these residues.

### 25 26 **Nucleic acid binding induces profound conformational changes outside the** 27 **HEL1 domain of nsp10Δ**

28  
29 The Cα atoms of domains 1A and 2A of free nsp10Δ and the nsp10Δ-DNA complex can  
30 be superimposed with an RMSD of 0.6 Å, indicating that the relative orientations of  
31 these core domains are barely affected by DNA binding (Figure 6A). However, outside  
32 these domains the effect of DNA binding was considerable, with the RMSD between  
33 the Cα atoms of the two forms of nsp10Δ increasing to 1.8 Å. Particularly large con-  
34 formational changes were observed in domain 1B, which rotates approximately 28°  
35 towards the ZBD in the nsp10Δ-DNA complex (Figure 6A). The RMSD between the Cα  
36 atoms of the two forms of domain 1B is 1.8 Å, with loop residues being affected most  
37 profoundly (Figure S6A). Both width and height of the polynucleotide substrate channel  
38 formed by domains 1A and 1B (originally about 5 and 11 Å, respectively) are increased  
39 by 2 Å upon this rotation. This reorganization makes this channel large enough to accept



**Figure 6.** Conformational changes of EAV nsp10 $\Delta$  upon nucleic acid binding. **(A)** Comparison of the free and DNA-bound states of nsp10 $\Delta$ . The arrow indicates the movement of domain 1B (cartoon) in the DNA-bound state compared to the free state. **(B)** Surface model of the channel formed by domains 1A and 1B in the DNA-bound state and **(C)** the DNA-free state. Domain colors are the same as used in Figure 2A. Note that the DNA in C was extracted from the complex structure of DNA-bound state.

single-stranded nucleic acids although it remains too narrow for a nucleic acid duplex (Figure 6B). Consequently, double-stranded nucleic acids must be unwound at the entrance of the substrate channel to let a single-stranded chain enter. Besides this large conformational change, temperature factor calculations suggest that the regions at the surface of domain 1B not directly involved in DNA binding may become flexible (Figure S2). For example, domain 1B residues Arg95, Gly125, and Ala131 become disordered after DNA binding (Figures 2C, 3B, and S2).

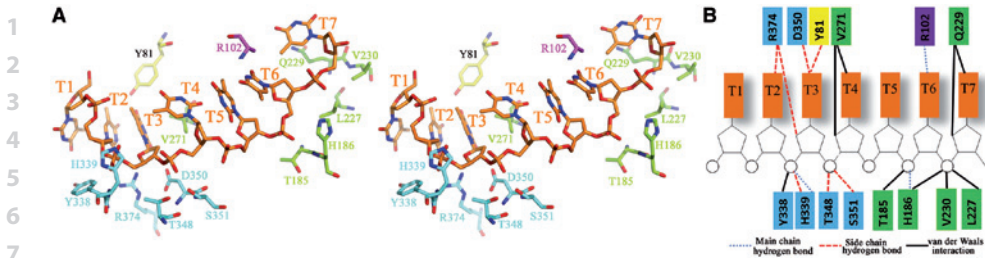
Upon DNA binding, a structural change was also observed in the treble-clef zinc finger of the ZBD, as reflected by its relatively high temperature factor (compared to that of domains 1A and 2A) in the nsp10 $\Delta$ -DNA complex as opposed to nsp10 $\Delta$  alone (Figure S2).

### Substrate recognition by EAV nsp10 $\Delta$ is sequence-independent

As outlined above, the single-stranded part of the DNA substrate is bound to a nucleic acid-binding channel formed by domains 1A, 1B, and 2A (Figure 2C). The backbone phosphates of the poly(dT) are located on top of domains 1A and 2A, with the thymine bases exposed to the solvent (Figure S7A). The majority of contacts with the bound DNA are made via the phosphodiester backbone and nonspecific protein-base interactions as depicted in Figure 7.

Consistent with this observation, the base orientation varies in the four EAV nsp10-poly(dT) complexes of the asymmetric unit while the position of the DNA backbone is rather rigid (Figures S7B and S7C). Several key residues from domains 1A and 2A contact the DNA backbone in the channel of the protein (Figures 7A and 7B). Base T1, the most 5' one, is exposed to the solvent and protrudes outwards, causing a bend in the DNA backbone between T1 and T2. The bases T2 and T3 as well as T5 and T6 stack with each





**Figure 7.** Interactions between EAV nsp10 $\Delta$  and a DNA substrate. **(A)** Stereo view of the nucleic acid-binding pocket of nsp10 $\Delta$ . Bound single-stranded DNA and the interacting residues are shown as sticks. Nucleotides are numbered (T1 to T7) in the 5'-3' direction and are shown in orange. Residues are colored according to their domain origin as indicated in Figure 2A. **(B)** Schematic representation of the contacts between nsp10 $\Delta$  residues and DNA.

other at an average distance of 3.7 Å. In contrast, base T4 is almost perpendicular to T3, with its edge exposed to protein side chains that make specific contacts. Val271 in domain 1A forms van der Waals contacts with the base and the sugar ring of T4 and thus stabilizes the DNA conformation. Moreover, the binding is stabilized by several hydrogen bonds between His186, His339, Thr348, Ser351, and the backbone of the DNA and by van der Waals contacts between Thr185, Leu227, Val230, Tyr338, and the phosphate groups of the DNA. While the interactions described above do not involve specific bases, six further interactions specific for thymine were found. For example, the backbone NH of Arg102 forms a hydrogen bond with the O4 atom of T6. The O2 and O4 atoms of base T3 form hydrogen bonds with the side chains of Asp350 and Tyr81. Also, several residues, such as Arg374 and Gln229, interact with both the base and the sugar ring. However, no interaction was observed between nsp10 $\Delta$  and position C2' of the ribose ring of the DNA substrate. This observation may explain why EAV nsp10 has the ability to unwind both DNA and RNA, in agreement with the substrate specificity observed for other helicases (52,53) possessing or lacking the ability to interact with the 2'OH moiety of the RNA backbone.

## DISCUSSION

Among (+) RNA viruses, whose RdRps generally have a high error rate, nidoviruses stand out for their large to very large genome size (13 to 32 kb). Consequently, the replication fidelity of nidoviruses, in particular coronaviruses, has been the subject of intense study. Most recently, the identification of a unique 3'-5' exoribonuclease (ExoN) activity has provided the basis for the hypothesis that a primitive proofreading mechanism operates

1 to promote the fidelity of RNA-dependent RNA synthesis in nidoviruses with >20 kb  
2 genomes (21-27).

3  
4 Despite this recent progress, the two central subunits of the nidovirus replicase, the  
5 RdRp and the unique ZBD-containing RNA helicase, have remained poorly character-  
6 ized, also due to the lack of structural information. Remarkably, our present analysis of  
7 the arterivirus helicase structure revealed a number of important similarities to Upf1  
8 helicases, eukaryotic enzymes involved in quality control of RNAs through multiple  
9 pathways, including nonsense-mediated mRNA decay (54-56). In contrast to the ExoN-  
10 driven control of replication fidelity (see above), the possibility of post-transcriptional  
11 quality control of nidovirus mRNAs has not been considered thus far. Yet, replicase  
12 ORF1ab is extremely large (from 3175 to over 7000 codons) and its correct expression  
13 by translation of the viral genome is a critical first step in the production of the enzymes  
14 directing genome replication and expression. Therefore, our study not only provides the  
15 first insights into the structural basis for nidovirus RNA helicase function but also creates  
16 a basis to propose a role for this protein in the post-transcriptional quality control of viral  
17 mRNAs. This role may be common to all nidoviruses, regardless of their genomes size,  
18 which would distinguish it from the ExoN-based proofreading mechanism that appears  
19 to be restricted to nidoviruses with a >20 kb genome. On the time scale of nidovirus  
20 evolution, the acquisition of ZBD-HEL1 may have been a critical event to facilitate the  
21 genome expansion of ancestral small-sized nidoviruses thus setting the stage for the  
22 subsequent ExoN-driven expansion towards even larger nidovirus genomes (19,57).

23

## 24 **EAV nsp10 represents a multi-domain helicase conserved in nidoviruses**

25

26 Previously, using bioinformatics, biochemistry, and molecular genetics, it was estab-  
27 lished that nsp10 of arteriviruses and its orthologs in other nidoviruses are multi-domain  
28 proteins. Of its domains the ZBD and HEL1 domains are critical for the enzyme's ATPase  
29 and helicase activities *in vitro* and for the regulation of viral replication and transcription  
30 in infected cells. Our structural and biochemical studies extended the characterization  
31 of known domains and delineated two hitherto uncharacterized domains: one (domain  
32 1B) flanked by the ZBD and HEL1, and the other (C-terminal domain) located down-  
33 stream of HEL1, with its structure remaining to be solved. Our data show that, along  
34 with the ZBD, these two non-enzymatic domains may regulate HEL1 function. Given  
35 that nsp10/nsp13 is one of only three proteins whose nidovirus-wide conservation can  
36 be detected at the sequence level (19,24,25,28), the nsp10 $\Delta$  structure should be ap-  
37 plicable to other nidovirus helicases, including those of PRRS viruses and coronaviruses.  
38 However, considerable size differences exist between arteri- and coronaviruses in the  
39 most conserved ZBD and HEL1 domains, whereas the 1B and C-terminal domains lack

1 appreciable sequence conservation. Thus, helicase structures from other small- and  
2 large-genome nidoviruses will be required to fully understand the enzyme's function.

### 3 4 **The nsp10 C-terminal domain: coupling ATPase and helicase activities?**

5  
6 While attempting to solve the EAV nsp10 structure, we were confronted with the low  
7 stability of the full-length recombinant protein expressed in *E. coli*. We solved this prob-  
8 lem by characterizing the C-terminally truncated nsp10 $\Delta$ , which lacks the 65 residues  
9 (C-terminal domain) downstream of the known HEL1 motifs. This protein was found to  
10 bind partially double-stranded DNA and display the previously reported *in vitro* ATPase  
11 and helicase activities. Since, compared to full-length nsp10, nsp10 $\Delta$  appeared to be  
12 somewhat more active as an ATPase but somewhat less active as a helicase, the C-ter-  
13 minal truncation may have affected the coupling of these two enzymatic activities. This  
14 suggests that the C-terminal domain may have evolved to (co-)regulate nsp10 helicase-  
15 mediated functions *in vivo*, implying that it must be able to communicate with the nsp10  
16 active site. This could be achieved either directly, by interacting with the nucleic acid- or  
17 ATP-binding site (the nsp10 $\Delta$  C-terminus is separated by  $\sim 22.5$  Å from the active center;  
18 Figure 2C), or indirectly, through a protein signal transduction network. Importantly,  
19 the C-terminal domain is poorly conserved among arteri- and coronaviruses in terms of  
20 both sequence and size (Figure S3A and data not shown), arguing that such a putative  
21 regulatory function could be executed in a virus- and, possibly, host-specific manner.

### 22 23 **The nsp10 structure: defining a complex ZBD**

24  
25 Our characterization of the EAV nsp10 structure verified and revised a model of the N-  
26 terminal ZBD based on prior studies (36,37,58). It resolved the uncertainty about the  
27 number of zinc ions bound (now established to be three) and the fold of this domain (a  
28 unique structure combining a RING-like module fused with a treble-clef zinc finger). Fur-  
29 thermore, it redefined the C-terminal border of the ZBD and placed it thirteen residues  
30 downstream to include a third, hitherto unrecognized, structural element (helix  $\alpha 4$ ). Pre-  
31 viously, we analyzed a variety of EAV nsp10 ZBD mutants in which putative zinc-binding  
32 residues were replaced in a manner (Cys $\rightarrow$ His or His $\rightarrow$ Cys) that could preserve zinc bind-  
33 ing (36,37). From the solved structure, it is now apparent that the replication-negative  
34 phenotypes of these virus mutants can likely be attributed to the detrimental impact  
35 of the respective mutations on ZBD integrity and, through the extensive interaction  
36 network, the HEL1 domain. It presently remains unclear why the replacement of His44  
37 by Cys in the treble-clef zinc finger was partially tolerated. On the other hand, structural  
38 superposition of the RING-like modules of nsp10 and hUpf1 (Figure 4E) reveals how the  
39 only other similarly tolerated replacement (36,37), that of the Zn1-coordinating Cys25 by

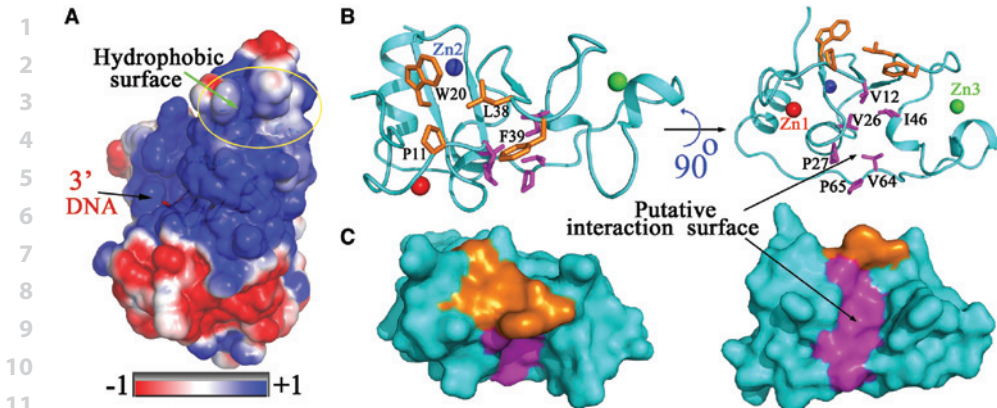
1 His (found in the equivalent position in hUpf1), could be accommodated by nsp10. The  
2 RING-like module 1 of Upf1 also shares structural similarity with RING-box domains of E3  
3 ubiquitin ligases (59), and the involvement of this module in self-ubiquitination of Upf1  
4 was indeed demonstrated (60). It would be interesting to see whether these results are  
5 relevant for nsp10 and its ZBD. Recently, arterivirus papain-like protease 2 was found to  
6 have deubiquitinase activity, which suppresses the innate immune response in infected  
7 host cells (61,62).

## 8 9 **The nsp10-nucleic acid complex: towards the dsRNA unwinding mechanism**

10  
11 To understand how nsp10 unwinds its natural dsRNA substrates, we analyzed a complex  
12 of nsp10 $\Delta$  with a partially double-stranded DNA substrate. Only seven thymidine resi-  
13 dues could be confirmed in the structure of that complex (Figure 2C). The DNA-bound  
14 nsp10 $\Delta$  structure revealed two possible RNA-binding clefts at the surface of nsp10,  
15 which are formed by domain 1B and the ZBD (named putative exit site 1), and 1A and  
16 the ZBD (putative exit site 2), respectively (Figure S8). Both have continuous positively  
17 charged surfaces, with the latter (Figure S8, right panel) being sufficiently large to bind  
18 a ssRNA longer than ten base pairs, which could be especially suited for unwinding  
19 complex secondary structures. This organization suggests that, after unwinding, one of  
20 the separated RNA chains would be guided through the narrow nucleic acid substrate  
21 tunnel formed by domains 1A, 1B, and 2A while the path of the other strand remains  
22 to be defined. No matter which cleft is actually used for RNA binding, the positively  
23 charged ZBD, and especially its RING-like module, would be involved.

24  
25 Like the protein-binding surface of the Upf1 CH domain (54), the ZBD has a putative  
26 protein interaction surface composed of two major hydrophobic zones that are almost  
27 perpendicular to each other (Figure 8). Nucleic acid binding induced a conformational  
28 change (Figure S6B) of these two zones. In addition, the temperature factor of the treble-  
29 clef zinc finger was higher and several residues are disordered in the structure of the  
30 nsp10 $\Delta$ -DNA complex (Figure S2). Together these findings imply that these two zones  
31 are readily accessible for interactions with other proteins, which may further influence  
32 nucleic acid binding.

33  
34 Substrate binding by nsp10 is accompanied by structural changes in domain 1B and  
35 the treble-clef zinc finger, which may be recognized by yet-to-be identified interaction  
36 partners modulating nsp10 function. The treble-clef zinc finger is fairly distant from the  
37 bound substrate, suggesting long-distance signal transduction within nsp10, possibly  
38 involving helix  $\alpha$ 4, which interacts with 1A, 1B, and nucleic acid and is directly connected  
39 to the treble-clef zinc finger. The flexibility of the hinge region connecting the treble-clef



**Figure 8.** Putative protein interaction surfaces of the EAV nsp10 $\Delta$  ZBD. **(A)** Overview of surface charges of nsp10 $\Delta$ . The putative protein interaction surfaces are indicated by a yellow circle. **(B)** Close-up view of zone 1 (orange) and zone 2 (magenta). Hydrophobic residues are shown as sticks. **(C)** Surface representation of the two putative interaction zones. The orientation is the same as in panel B.

finger and helix  $\alpha$ 4 is likely compromised by the previously described S59P and S59G/P60G mutations that, importantly, were found to impair viral sg mRNA synthesis but not genome replication (36). Consequently, the described inter-domain communication channel may be used by nsp10 and its partners for switching from a role in genome replication to directing viral transcription, a hypothesis that will be the subject of future studies.

### **Nidovirus helicase: a role in post-transcriptional quality control of viral mRNAs?**

The observed structural similarity between the EAV nsp10 and Upf1 helicases is most remarkable, in particular since it extends to include the multi-domain organization essential for helicase function. This organization is only found in Upf1 of all eukaryotes (59) and nidovirus helicases (19,24,25,28). For Upf1, its conservation was linked to the protein's universal role in post-transcriptional quality control of eukaryotic RNAs through multiple pathways, including nonsense-mediated mRNA decay (54-56). Upf1 interacts, commonly through its CH and 1A domains, with proteins that can modulate its function. For the nidovirus helicase subunit, the functional basis of its domain conservation remains to be firmly established although ZBD – like CH in Upf1 (63) – affects helicase activity (36,37).

If the nidovirus helicase possesses some of the properties of Upf1, this could explain the exclusive conservation of the ZBD in nidoviruses, which stand out for their large to very large single-stranded RNA genomes. For instance, by providing post-transcriptional

1 quality control of genomic RNA, i.e. detection of non-sense and/or other mutations  
2 and elimination of defective molecules, the nidovirus helicase could alleviate the con-  
3 sequences of the generally low fidelity of RNA virus genome replication. Such a role of  
4 ZBD-HEL1 may have protected an ancestral nidovirus from the mutational meltdown of  
5 its expanding genome, quite similar to the proposed fixation of the proofreading ExoN  
6 domain at a later stage of nidovirus evolution (19,24,25,28). Subsequently, the enzyme  
7 would have facilitated expansion to the genome size observed in contemporary arteri-  
8 viruses and remained a critical factor in the further ExoN-driven genome expansion to  
9 evolve middle- and large-sized nidoviruses. Thus, the proposed Upf1-like role of the  
10 nidovirus helicase can be accommodated in a meaningful evolutionary scenario incor-  
11 porating several of the structural and functional observations made in this study. The  
12 structural similarity between nsp10 and Upf1 establishes a new connection between  
13 research on viral and cellular helicases, which could be mutually insightful for under-  
14 standing the evolution and function of this group of vitally important enzymes.

## 16 MATERIALS AND METHODS

### 19 Cloning, expression, and purification of soluble EAV nsp10

21 Nsp10 of the EAV-Bucyrus isolate (NCBI Reference Sequence NC\_002532) is comprised  
22 of amino acids 2371 to 2837 of replicase pp1ab, which will throughout this study be  
23 referred to as nsp10 residues 1 to 467. The full-length nsp10 sequence or a C-terminally  
24 truncated version comprising residues 1 to 402 (nsp10 $\Delta$ ) were cloned into a modified  
25 pET28a vector with a tobacco etch virus (TEV) protease cleavage site. Mutations were  
26 generated using the QuikChange protocol and confirmed by DNA sequencing. The  
27 proteins were overexpressed at 37°C in *E. coli* strain BL21 (DE3) grown to an OD600  
28 of about 0.8 in Luria–Bertani medium in the presence of 50  $\mu$ g/ml kanamycin. Protein  
29 expression was induced with 0.2 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside for 12 h at  
30 16°C. Cell pellets were resuspended in lysis buffer (20 mM HEPES pH 7.0 for nsp10 $\Delta$  or  
31 pH 8.0 for full-length nsp10, 500 mM NaCl and 30 mM imidazole), supplemented with  
32 protease inhibitor cocktail (Roche) and disrupted by sonication. Lysates were clarified  
33 at 20,000g for 30 min, and the soluble fraction was applied to a Ni<sup>2+</sup> chelating column.  
34 After sample loading, the column was washed (20 mM HEPES, pH 7.0 or 8.0, 500 mM  
35 NaCl and 60 mM imidazole), and the protein was eluted (20 mM HEPES, pH 7.0 or 8.0,  
36 500 mM NaCl and 400 mM imidazole). Proteins intended for ATPase or helicase assays  
37 were dialyzed against storage buffer (20 mM HEPES, pH 7.0 or 8.0, 100 mM NaCl, 50%  
38 glycerol) and stored at 20°C. Truncated protein for crystallization studies was digested  
39 with 10% (w/w) TEV protease to remove the His-tag. Further purification was performed

1 by size-exclusion chromatography using a Superdex 200 column (GE Healthcare) with  
2 GF buffer (20 mM HEPES, pH 7.0, 500 mM NaCl). The peak fraction was collected and  
3 analyzed by SDS-PAGE.

#### 4 5 **Crystallization and data collection**

6  
7 Purified nsp10Δ was concentrated to 10 mg/ml and initial crystallization trials were  
8 performed at 16°C using the sitting-drop vapor-diffusion method by mixing 1 μl protein  
9 solution with 1 μl reservoir solution. The conditions were then optimized and high-  
10 quality crystals were obtained in 1.6 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M HEPES, pH 7.1, 25 mM KCl and  
11 20% ethylene glycol. To obtain crystals of the protein-DNA complex, purified protein  
12 and partially double-stranded DNA with a 5' single-stranded poly-thymidine overhang  
13 (the two partially complementary sequences were 5'-TTTTTTTTTTCAGTGCTCG-3' and  
14 5'-CGCGAGCACTGC-3') were mixed in a 1:1.5 molar ratio and incubated at 4°C overnight.  
15 The complex was further purified by size-exclusion chromatography (Superdex 200, GE  
16 Healthcare) and concentrated to 5 mg/ml. The condition for obtaining crystals was 14%  
17 PEG 3350, 0.1 M HEPES, pH 7.0, and 0.2 M calcium acetate. For data collection, crystals  
18 were cryoprotected in mother liquor containing 25% (v/v) ethylene glycol and flash  
19 cooled to 173°C.

20  
21 The MAD data for intrinsic zinc atoms were collected on beamline 1W2B at the Beijing  
22 Synchrotron Radiation Facility. The data for EAV nsp10Δ and its complex with DNA  
23 were collected at beamline NE3A at Photon Factory (KEK) and beamline BL17U1 at the  
24 Shanghai Synchrotron Radiation Facility. Data was indexed, integrated and scaled using  
25 HKL2000 (41). Data collection and processing statistics are summarized in Table 1.

#### 26 27 **Structure determination**

28  
29 The structure of nsp10Δ was determined by the MAD method. Initial phases were  
30 calculated by SOLVE, and phases were subsequently improved using RESOLVE (42).  
31 The figure of merit from the MAD phasing was 0.36, and the Z score was 15.7. Several  
32 segments of the protein could be automatically modeled into the electron-density map  
33 by RESOLVE although in part only as poly-alanine chains. Manual rebuilding was per-  
34 formed in COOT (43), and refinement was performed with REFAMC5 (44). Further rounds  
35 of refinement were done with Translation/Libration/Screw (TLS) refinement (45). The  
36 structure was refined to 2.0 Å with an Rwork of 19.5% and an Rfree of 22.4%.

37  
38 Using the structure of free nsp10Δ without domain 1B as input model, the structure  
39 of nsp10Δ in complex with DNA was successfully solved by molecular replacement.

1 The initial model was obtained by MOLREP from the CCP4 program suite (46). A good  
2 match for domains ZBD, 1A, and 2A with electron density was found. Domain 1B was  
3 manually added with the aid of 2Fo–Fc and Fo–Fc maps using COOT (43). DNA molecules  
4 were included in the final stages of refinement. Difference Fourier maps clearly showed  
5 electron densities for seven bound deoxyribonucleotides. The final model was refined to  
6 2.65 Å with an Rwork of 23.2% and an Rfree of 25.7%. All figures in this article displaying  
7 molecular structures were made using PYMOL (47).

## 8 9 10 **ACCESSION NUMBERS**

11  
12 The coordinates and structure-factor amplitudes of EAV nsp10Δ and EAV nsp10Δ–DNA  
13 complex have been deposited in the Protein Data Bank with accession codes 4N0N and  
14 4N0O, respectively.

## 15 16 17 **SUPPLEMENTARY DATA**

18  
19 Supplementary Data are available at NAR online, including Supplementary References  
20 (64–66).

## 21 22 23 **FUNDING**

24  
25 National Basic Research Program of China [973 Program, 2011CB965304 and  
26 2009CB825501]; National Natural Science Foundation of China [31370720, 31222032,  
27 90919043 and 31070664]; Chinese Universities Scientific Fund [2013QJ027]; Specialized  
28 Research Fund for the Doctoral Program of Higher Education [20100008110009]; Open  
29 Research Fund Program of the State Key Laboratory of Virology of China [2013IOV003];  
30 National Laboratory of Medical Molecular Biology (PUMC) to Zhongzhou Chen; Euro-  
31 pean Union Seventh Framework program through the EUVIRNA project [264286]; the  
32 Netherlands Organization for Scientific Research [NWO; TOP-GO 700.10.352]; the Col-  
33 laborative Agreement on Bioinformatics between Leiden University Medical Center and  
34 Moscow State University (MoBiLe); and the Leiden University Fund. Funding for open  
35 access charge: [2011CB965304] and Leiden University Medical Center.

36  
37 Conflict of interest statement. None declared.  
38  
39



**1 ACKNOWLEDGEMENT**

2

3 We thank the staff at beamline NE3A (KEK), SSRF beamline BL17U, and BSRF beamline  
4 1W2B facilities for help with crystallographic data collection, Alexander Kravchenko,  
5 Dmitry Samborskiy, and Igor Sidorov for Viralis management. AEG thanks Dr. John  
6 Ziebuhr for a decade-old discussion of the possible roles of mRNA decay regulation in  
7 nidoviruses.

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

## 1 SUPPLEMENTARY EXPERIMENTAL PROCEDURES

### 2 3 DNA substrate preparations

4  
5 All DNA oligonucleotides were purchased from Invitrogen. Radioactive labeling of the  
6 5' end of single-stranded DNA was done with T4 polynucleotide kinase (Invitrogen) and  
7 [ $\gamma$ - $^{32}$ P]ATP according to the manufacturer's protocol. DNA duplexes used for crystal-  
8 lization, helicase assays, and electrophoretic mobility shift assays were annealed in a  
9 buffer containing 10 mM Tris-HCl pH 8.0 and 100 mM NaCl. Annealing was performed  
10 by first heating the mixture at 90°C for 5 min and then slowly letting it cool to room  
11 temperature in 2 h.

### 12 13 ATPase assay

14  
15 ATPase activity of full-length nsp10 and nsp10 $\Delta$  was assayed in 5  $\mu$ l reactions containing  
16 100 nM purified recombinant protein, 1 mM ATP (including 250 nCi [ $\alpha$ - $^{32}$ P]ATP, 3000 Ci/  
17 mmol), 20 mM Tris-HCl, pH 7.5, 2 mM MgCl<sub>2</sub>, 1 mM DTT, 0.01 mg/ml BSA. Additionally  
18 5% glycerol and 10 mM NaCl were introduced by the protein storage buffer. Where indi-  
19 cated 1  $\mu$ M poly-uridine RNA with a length of 30 nucleotides was added. Samples were  
20 incubated for up to 30 min at 20°C before 1  $\mu$ l 500 mM EDTA was added to stop the  
21 reaction. A 0.25  $\mu$ l aliquot of each sample was analyzed by polyethyleneimine cellulose  
22 thin-layer chromatography with 0.45 M ammonium sulfate as mobile phase. Plates were  
23 dried and analyzed using storage phosphor screens, which were subsequently scanned  
24 on a Typhoon 9410 variable mode imager (GE Healthcare). ATP turnover was quantified  
25 using ImageQuant TL software (GE Healthcare).

### 26 27 Helicase assay

28  
29 100 nM protein and 1 nM of radioactively labeled partially double-stranded DNA (Hel1,  
30 5'-TTTTTTTTTGCCTCGCTGCCGTCGCCACC-3'; Hel3, 5'-\*GGTGGCGACGGCAGCGAGGC-3';  
31 asterisk indicates the position of the radioactive label) were incubated for 10 min at 30°C  
32 in a buffer containing 20 mM Tris-HCl, pH 7.5, 10 mM NaCl, 2 mM MgCl<sub>2</sub>, 0.01 mg/ml BSA,  
33 and 0.01% Triton X-100. Additionally 5% glycerol and 10 mM NaCl were introduced by  
34 the protein storage buffer. After the initial binding phase, unwinding was started by  
35 addition of 5 mM ATP, and incubation continued at 30°C for up to 30 min. Reactions  
36 were stopped by mixing with an equal volume of loading buffer (20 mM Tris-HCl, pH 8.0,  
37 50 mM EDTA, 60% glycerol, 0.5% SDS, 0.1% bromophenol blue). To prevent re-annealing  
38 of the unwound radioactively labeled strand, an excess of unlabeled oligonucleotide  
39 Hel3 was added to a final concentration of 5  $\mu$ M. Single- and partially double-stranded

1 DNA were separated by 12% native PAGE (acrylamide:bis-acrylamide 19:1) run in ice-  
2 cold 1×TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA). Gels were dried and analyzed  
3 using storage phosphor screens, which were subsequently scanned on a Typhoon 9410  
4 variable mode imager (GE healthcare).

5

### 6 **Electrophoretic mobility shift assays**

7

8 For EMSA, unless stated otherwise, 20  $\mu$ M DNA was incubated with 30  $\mu$ M nsp10 $\Delta$  in 10-  
9  $\mu$ l reactions containing 20 mM HEPES, pH 7.0, 100 mM NaCl at 25°C for 2 hours. Samples  
10 were then analyzed on 8% native polyacrylamide gels (acrylamide:bis-acrylamide 19:1)  
11 containing 0.5×Tris-borate buffer and visualized by staining with ethidium bromide.

12

### 13 **Circular dichroism analysis**

14

15 Circular dichroism spectra were measured at beamline 4B8 at the Beijing synchrotron ra-  
16 diation facility. Spectra of EAV nsp10 $\Delta$  were collected at 1 nm intervals ranging from 250  
17 to 190 nm in a 0.0007 cm optical path length at 16°C in 0.1 M HEPES, pH 7.0, 50 mM NaCl.  
18 Far-UV CD spectra of 3 mg/ml EAV nsp10 $\Delta$  with or without 40 mM EDTA were scanned.  
19 A pure solvent baseline was measured with the same cell and subtracted. All spectra  
20 were processed using the CD tool software package(64). The machine unit (mdeg) was  
21 converted into the per residue molar absorption unit, delta epsilon ( $\Delta\epsilon$ ) in M/cm, by  
22 normalization with respect to polypeptide concentration and path length.

23

24

25

26

27

28

29

30

31

32

33

34

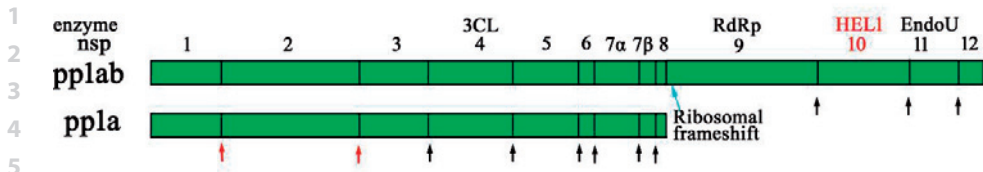
35

36

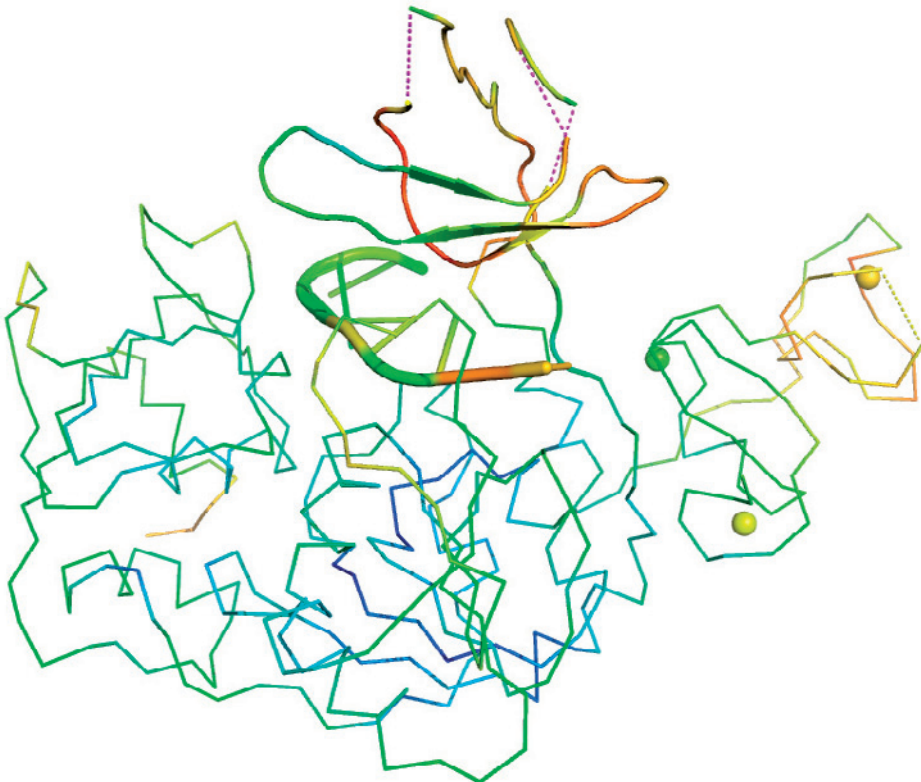
37

38

39



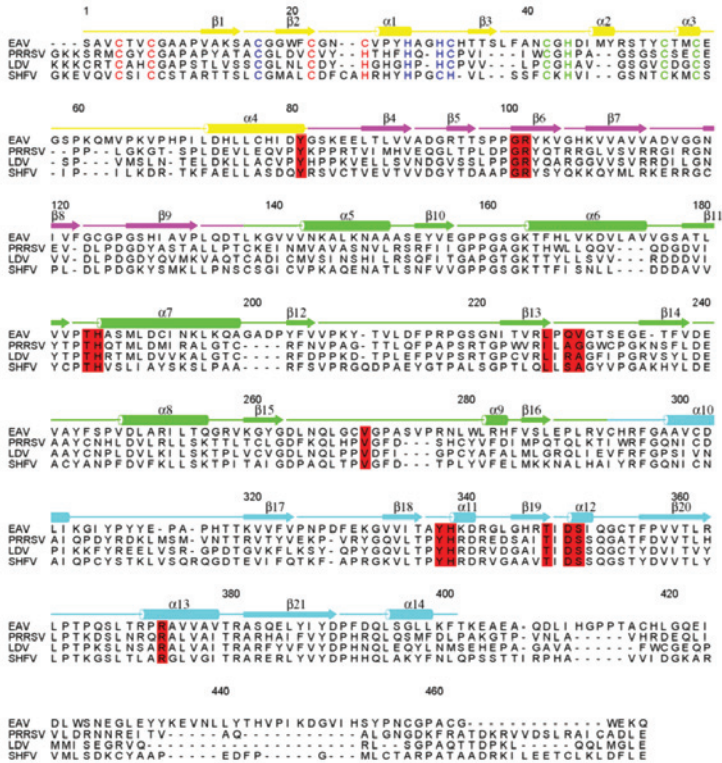
**Figure S1.** Schematic representation of the proteolytic processing of the EAV multidomain replicase polyproteins pp1a and pp1ab. The nsp1/2 and nsp2/3 cleavage sites (cleaved by the nsp1 and nsp2 autoproteases) are highlighted by red arrows, all other sites (indicated with black arrows) are cleaved by the 3CL protease in nsp4. Abbreviations: 3CL, serine/3C-like protease; RdRp, RNA-dependent RNA polymerase; HEL1, helicase core SF1; EndoU, Uridylate-specific endoribonuclease.



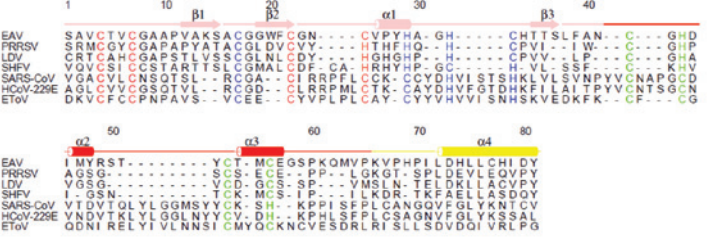
**Figure S2.** Overall B-factors of the nsp10 $\Delta$ -DNA complex, color-coded on the basis of the calculated B-factors. The colors range from blue to red corresponding to increasing fluctuations. Dashed lines mean that residues have not been resolved.

**A**

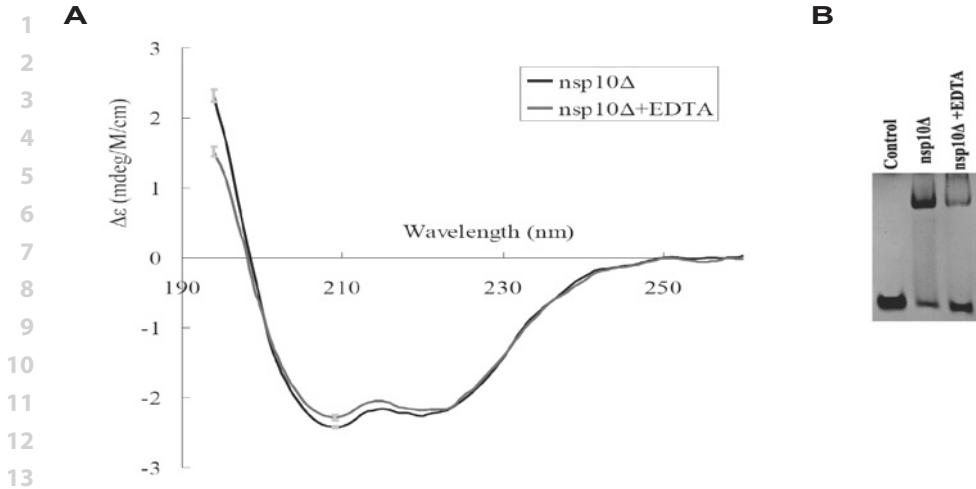
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22



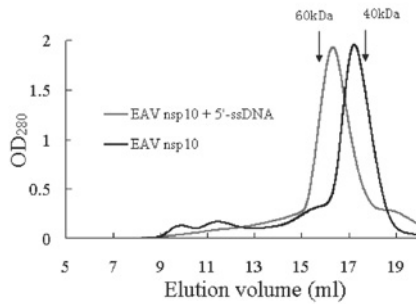
**B**



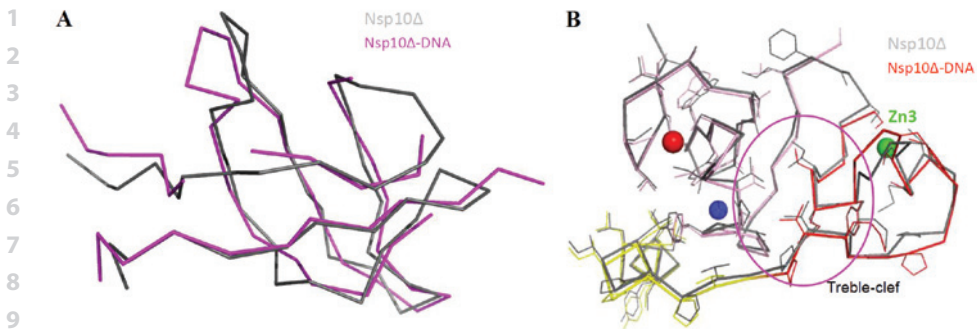
**Figure S3. (A)** Alignment of nsp10 from different arteriviruses. The zinc-binding domain (ZBD), 1B, 1A, and 2A domains are colored yellow, magenta, green, and cyan, respectively. **(B)** Alignment of the helicase-associated ZBD of selected arteri- and coronaviruses. RING-like module (pink), treble-clef zinc finger (red), and Linker1 (yellow) are indicated. Residues now known (EAV) or predicted to be involved in chelating Zn1, Zn2, and Zn3 are colored red, blue, and green, respectively. The residues involved in nucleic acid binding are highlighted with red background. The sequences were aligned using the program T-Coffee (65). The ALSRIPT program (66) was used to prepare the figure. Sequence numbering and secondary structure designations are according to the EAV nsp10 sequence and structure. The sequences used are from: EAV, equine arteritis virus (GenBank accession number X53459); LDV, lactate dehydrogenase-elevating virus (U15146); PRRSV, porcine reproductive and respiratory syndrome virus (JX317648); SHFV, simian hemorrhagic fever virus (AF180391); SARS-CoV, severe acute respiratory syndrome coronavirus (AY291315); HCoV-229E, human coronavirus 229E (X69721); ETov, equine torovirus (X56016).



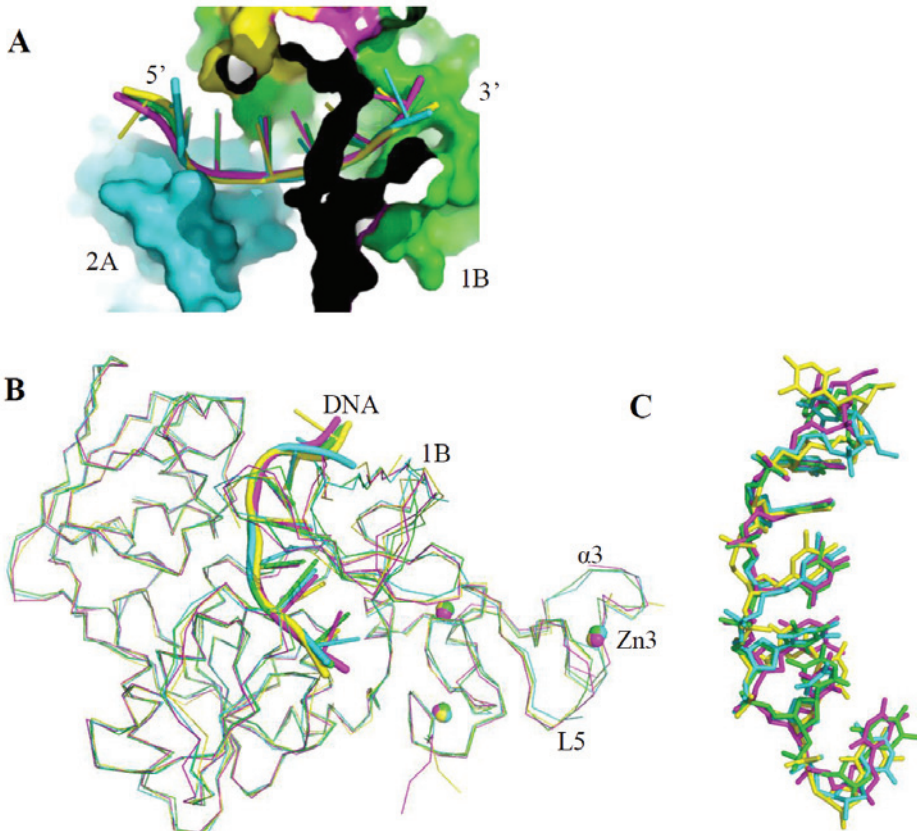
**Figure S4.** Effects of EDTA on folding and DNA-binding activity of EAV nsp10 $\Delta$ . **(A)** Superposition of far-UV CD spectra of EAV nsp10 $\Delta$  with or without EDTA. Error bars show the standard deviation. **(B)** EMSA of nsp10 $\Delta$  with 5'-DNA-A10 as substrate in absence or presence of EDTA. The control lane shows the position of free DNA.



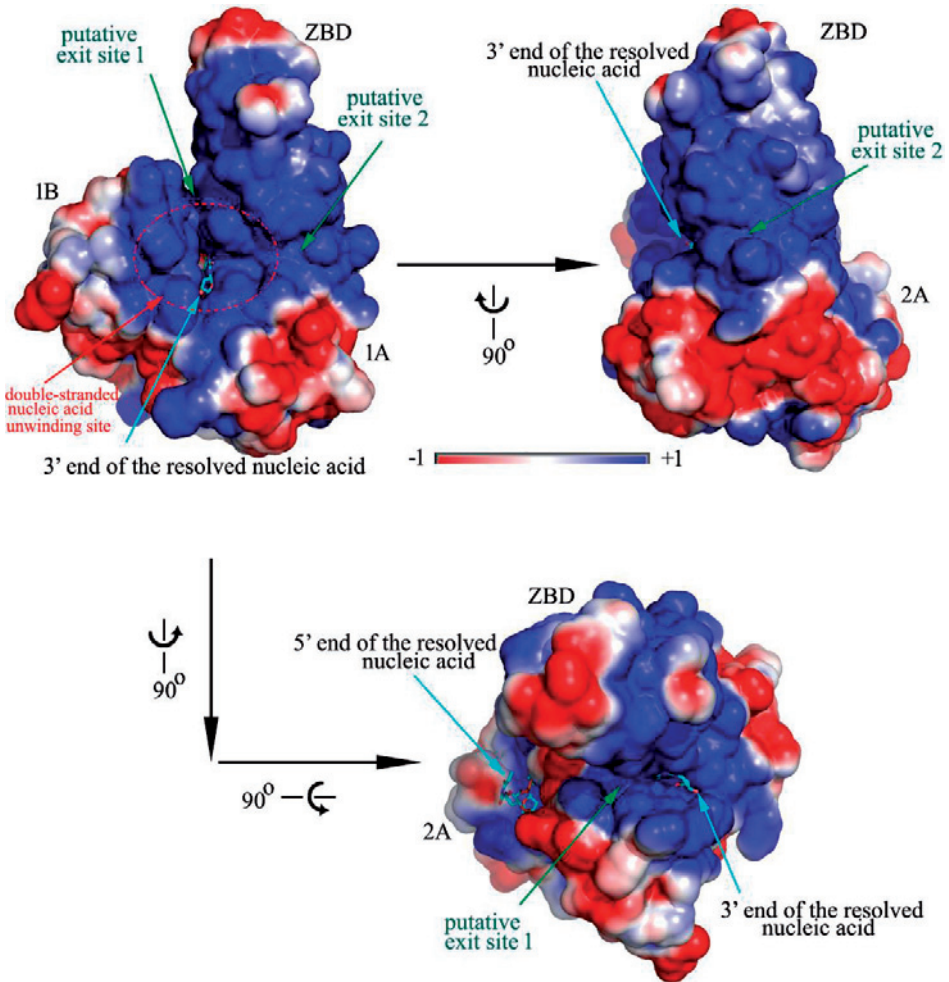
**Figure S5.** Elution profile of free and DNA-bound EAV nsp10 $\Delta$  on a Superdex 200 gel filtration column in gel filtration chromatography.



10 **Figure S6.** Overlay of (A) 1B and (B) the ZBD in presence and absence of a bound DNA substrate. The  
 11 DNA-free state is shown in gray. For the DNA-bound state, the 1B domain is shown in magenta, the RING-  
 12 like module in pink, treble-clef zinc finger in red, and Linker1 in yellow. The hydrophobic zone formed by  
 13 residues V12, V26, P27, I46, V64, and P65 is shown in magenta ellipse.



37 **Figure S7.** Overlay of the substrate channels of the four nsp10-DNA complexes of the asymmetric unit. (A)  
 38 Surface and (B) ribbon representation of the four binary complexes in the asymmetric unit. Poly(dT) is shown  
 39 as cartoon. (C) Detailed structure of the four poly(dT) molecules as they are oriented in the binding channel.



**Figure S8.** Proposed nucleic acid binding sites of EAV nsp10 $\Delta$ . Electrostatic surface representation of EAV nsp10 $\Delta$  (left), a 90°-rotated view (right) and a top view (lower). Note that the regions around the 3' end of the resolved single-stranded part of the nucleic acid substrate are rich in positive charges. The double-stranded nucleic acid unwinding site and another putative single-strand exit site are indicated by a red-dashed ellipse and dark green arrows, respectively.



## 1 REFERENCE LIST

- 2 1. Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev*  
3 *Mol Cell Biol.* 2011;12:505-516.
- 4 2. Abdelhaleem M. Helicases: an overview. *Methods Mol Biol.* 2010;587:1-12.
- 5 3. Gorbalenya AE, Koonin EV. Viral proteins containing the purine NTP-binding sequence pattern.  
6 *Nucleic Acids Res.* 1989;17(21):8413-8440.
- 7 4. Kadare G, Haenni AL. Virus-encoded RNA helicases. *J Virol.* 1997;71(4):2583-2590.
- 8 5. Steimer L, Klostermeier D. RNA helicases in infection and disease. *RNA biol.* 2012;9(6):751-771.
- 9 6. Rajagopal V, Patel SS. In Raney KD, Gotte M, Cameron CE (eds.). *Viral Genome Replication.* Springer  
10 US. 2009. pp. 429-466.
- 11 7. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid  
12 translocases. *Annu Rev Biochem.* 2007; 76:23-50.
- 13 8. Gorbalenya AE, Koonin EV. Helicases: amino acid sequence comparisons and structure-function  
14 relationships. *Curr.Opin.Struct.Biol.* 1993; 3:419-429.
- 15 9. Gorbalenya AE, Koonin EV, Donchenko AP *et al.* A novel superfamily of nucleoside triphosphate-  
16 binding motif containing proteins which are probably involved in duplex unwinding in DNA and  
17 RNA replication and recombination. *FEBS Lett.* 1988; 235(1-2):16-24.
- 18 10. Hodgman TC. A new superfamily of replicative proteins. *Nature.* 1988; 333(6168):22-23.
- 19 11. Fairman-Williams ME, Guenther UP, Jankowsky E. SF1 and SF2 helicases: family matters. *Curr.Opin.*  
20 *Struct.Biol.* 2010;20(3):313-324.
- 21 12. Strauss JH, Strauss EG. The alphaviruses: gene expression, replication, and evolution. *Microbio-*  
22 *logical reviews.* 1994; 58(3):491-562.
- 23 13. de Groot RJ, Cowley JA, Enjuanes L *et al.* Order *Nidovirales*. In King, A. M. Q., Adams, M.J., Carstens,  
24 E.B., Lefkowitz, E.J. (ed.), *Virus Taxonomy, Ninth Report of the International Committee on Tax-*  
25 *onomy of Viruses.* Elsevier Academic Press, Amsterdam. 2012. pp. 785-795.
- 26 14. Wang Q, Han Y, Qiu Y *et al.* Identification and characterization of RNA duplex unwinding and  
27 ATPase activities of an alphatetravirus superfamily 1 helicase. *Virology.* 2012; 433(2):440-448.
- 28 15. Nishikiori M, Sugiyama S, Xiang H *et al.* Crystal structure of the superfamily 1 helicase from To-  
29 mato mosaic virus. *J Virol.* 2012; 86(14):7565-7576.
- 30 16. Chand RJ, Tribble BR, Rowland RR. Pathogenesis of porcine reproductive and respiratory syndrome  
31 virus. *Curr Opin Virol.* 2012; 2(3):256-263.
- 32  
33  
34  
35  
36  
37  
38  
39

- 1 17. Peiris JS, Guan Y, Yuen KY. Severe acute respiratory syndrome. *Nat Med.* 2004; 10(12 Suppl):S88-  
2 97.
- 3 18. de Groot RJ, Baker SC, Baric RS *et al.* Middle East Respiratory Syndrome Coronavirus (MERS-CoV):  
4 Announcement of the Coronavirus Study Group. *J Virol*, 2013; 87(14):7790-7792.
- 5 19. Gorbalenya AE, Enjuanes L, Ziebuhr J *et al.* *Nidovirales*: evolving the largest RNA virus genome.  
6 *Virus Res.* 2006; 117(1):17-37.
- 7 20. Perlman S, Gallagher T, Snijder EJ. *Nidoviruses*. ASM Press, Washington, DC. 2008
- 8 21. Denison MR, Graham RL, Donaldson EF *et al.* Coronaviruses: an RNA proofreading machine regu-  
9 lates replication fidelity and diversity. *RNA Biol.* 2011; 8(2):270-279.
- 10 22. Minskaia E, Hertzog T, Gorbalenya AE *et al.* Discovery of an RNA virus 3'→5' exoribonuclease that is  
11 critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A.* 2006; 103(13):5108-5113.
- 12 23. Bouvet M, Imbert I, Subissi L *et al.* RNA 3'-end mismatch excision by the severe acute respiratory  
13 syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl*  
14 *Acad Sci U S A.* 2012; 109(24):9372-9377.
- 15 24. Snijder EJ, Bredenbeek PJ, Dobbe JC *et al.* Unique and conserved features of genome and pro-  
16 teome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol.*  
17 2003; 331(5):991-1004.
- 18 25. Nga PT, Parquet Mdel C, Lauber C *et al.* Discovery of the first insect nidovirus, a missing evolution-  
19 ary link in the emergence of the largest RNA virus genomes. *PLoS Pathog.* 2011; 7(9):e1002215.
- 20 26. Eckerle LD, Becker MM, Halpin RA *et al.* Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus  
21 Replication Is Revealed by Complete Genome Sequencing. *PLoS Pathog.* 2010; 6(5):e1000896.
- 22 27. Smith EC, Blanc H, Vignuzzi M *et al.* Coronaviruses Lacking Exoribonuclease Activity Are Suscep-  
23 tible to Lethal Mutagenesis: Evidence for Proofreading and Potential Therapeutics. *PLoS Pathog,*  
24 2013; 9(8):e1003565.
- 25 28. Ziebuhr J. The coronavirus replicase. *Curr Top Microbiol Immunol.* 2005; 287:57-94.
- 26 29. Brierley I, Digard P, Inglis SC. Characterization of an efficient coronavirus ribosomal frameshifting  
27 signal: requirement for an RNA pseudoknot. *Cell.* 1989; 57(4):537-547.
- 28 30. Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the  
29 *Nidovirales*. *J Gen Virol.* 2000; 81(Pt 4):853-879.
- 30 31. Pasternak AO, Spaan WJM, Snijder EJ. Nidovirus transcription: how to make sense...? *J Gen Virol.*  
31 2006; 87(Pt 6):1403-1421.
- 32 32. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J Virol.*  
33 2007; 81(1):20-29.
- 34  
35  
36  
37  
38  
39

- 1 33. Sola I, Mateos-Gomez PA, Almazan F *et al.* RNA-RNA and RNA-protein interactions in coronavirus  
2 replication and transcription. *RNA Biol.* 2011; 8(2):237-248.
- 3 34. Seybert A, Hegyi A, Siddell SG *et al.* The human coronavirus 229E superfamily 1 helicase has RNA  
4 and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA.* 2000; 6(7):1056-1068.
- 5 35. Seybert A, van Dinten LC, Snijder EJ *et al.* Biochemical characterization of the equine arteritis virus  
6 helicase suggests a close functional relationship between arterivirus and coronavirus helicases. *J*  
7 *Viro.* 2000; 74(20):9586-9593.
- 8 36. van Dinten LC, van Tol H, Gorbalenya AE *et al.* The predicted metal-binding region of the arterivi-  
9 rus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion  
10 biogenesis. *J Viro.* 2000; 74(11):5213-5223.
- 11 37. Seybert A, Posthuma CC, van Dinten LC *et al.* A complex zinc finger controls the enzymatic activi-  
12 ties of nidovirus helicases. *J Viro.* 2005; 79(2):696-704.
- 13 38. van Dinten LC, den Boon JA, Wassenaar AL *et al.* An infectious arterivirus cDNA clone: identifica-  
14 tion of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc Natl*  
15 *Acad Sci U S A.* 1997; 94(3):991-996.
- 16 39. Keum YS, Jeong YJ. Development of chemical inhibitors of the SARS coronavirus: viral helicase as  
17 a potential target. *Biochem Pharmacol.* 2012; 84(10):1351-1358.
- 18 40. Isken O, Maquat LE. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA  
19 function. *Genes Dev.* 2007; 21(15):1833-1856.
- 20 41. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Method*  
21 *Enzymol.* 1997; 276:307-326.
- 22 42. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol*  
23 *Crystallogr.* 1999; 55(Pt 4):849-861.
- 24 43. Emsley P, Lohkamp B, Scott WG *et al.* Features and development of Coot. *Acta Crystallogr D Biol*  
25 *Crystallogr.* 2010; 66(Pt 4):486-501.
- 26 44. Murshudov GN, Skubak P, Lebedev AA *et al.* REFMAC5 for the refinement of macromolecular  
27 crystal structures. *Acta Crystallogr D Biol Crystallogr.* 2011; 67(Pt 4):355-367.
- 28 45. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups under-  
29 going TLS motion. *Acta Crystallogr D Biol Crystallogr.* 2006; 62(Pt 4):439-450.
- 30 46. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr.*  
31 2010; 66(Pt 1):22-25.
- 32 47. Schrödinger L. The PyMOL Molecular Graphics System, Version 1.3r1. 2010.
- 33
- 34
- 35
- 36
- 37
- 38
- 39

- 1 48. Saikrishnan K, Powell B, Cook NJ *et al.* Mechanistic basis of 5'-3' translocation in SF1B helicases.  
2 Cell. 2009; 137(5):849-859.
- 3 49. Pugalenthi G, Archunan G, Sowdhamini R. DIAL: a web-based server for the automatic identifica-  
4 tion of structural domains in proteins. *Nucleic Acids Res.* 2005; 33:W130-132.
- 5 50. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.*  
6 2007; 372(3):774-797.
- 7 51. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 2010; 38:W545-  
8 549.
- 9 52. Appleby TC, Anderson R, Fedorova O *et al.* Visualizing ATP-dependent RNA translocation by the  
10 NS3 helicase from HCV. *J Mol Biol.* 2011; 405(5):1139-1153.
- 11 53. Lim SC, Bowler MW, Lai TF *et al.* The Ighmbp2 helicase structure reveals the molecular basis for  
12 disease-causing mutations in DMSA1. *Nucleic Acids Res.* 2012; 40(21):11009-11022.
- 13 54. Clerici M, Mourao A, Gutsche I *et al.* Unusual bipartite mode of interaction between the nonsense-  
14 mediated decay factors, UPF1 and UPF2. *EMBO J.* 2009; 28(15):2293-2306.
- 15 55. Lim SC, Bowler MW, Lai TF *et al.* The Ighmbp2 helicase structure reveals the molecular basis for  
16 disease-causing mutations in DMSA1. *Nucleic Acids Res.* 2012; 40(21):11009-11022.
- 17 56. Cheng Z, Muhlrads D, Lim MK *et al.* Structural and functional insights into the human Upf1 helicase  
18 core. *EMBO J.* 2007; 26(1):253-264.
- 19 57. Lauber C, Goeman JJ, Parquet Mdel C *et al.* The Footprint of Genome Architecture in the Largest  
20 Genome Expansion in RNA Viruses. *PLoS Pathog.* 2013; 9(7):e1003500.
- 21 58. Gorbalenya AE, Koonin EV, Donchenko AP *et al.* Coronavirus genome: prediction of putative func-  
22 tional domains in the non-structural polyprotein by comparative amino acid sequence analysis.  
23 *Nucleic Acids Res.* 1989; 17(12):4847-4861.
- 24 59. Kadlec J, Guilligay D, Ravelli RB *et al.* Crystal structure of the UPF2-interacting domain of  
25 nonsense-mediated mRNA decay factor UPF1. *RNA.* 2006; 12(10):1817-1824.
- 26 60. Takahashi S, Araki Y, Ohya Y *et al.* Upf1 potentially serves as a RING-related E3 ubiquitin ligase via  
27 its association with Upf3 in yeast. *RNA.* 2008; 14(9):1950-1958.
- 28 61. van Kasteren PB, Bailey-Elkin BA, James TW *et al.* Deubiquitinase function of arterivirus papain-  
29 like protease 2 suppresses the innate immune response in infected host cells. *Proc Natl Acad Sci*  
30 *U S A.* 2013; 110(9):e838-847.
- 31 62. Frias-Staheli N, Giannakopoulos NV, Kikkert M *et al.* Ovarian tumor domain-containing viral  
32 proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe.*  
33 2007; 2(6):404-416.
- 34  
35  
36  
37  
38  
39

- 1 63. Chakrabarti S, Jayachandran U, Bonneau F *et al.* Molecular mechanisms for the RNA-dependent  
2 ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell.* 2011; 41(6):693-703.
- 3 64. Lees JG, Smith BR, Wien F *et al.* CDtool-an integrated software package for circular dichroism  
4 spectroscopic data processing, analysis, and archiving. *Anal Biochem.* 2004, 332(2):285-289.
- 5 65. Barton GJ. ALSRIPT: a tool to format multiple sequence alignments. *Protein Eng.* 1993; 6(1):37-  
6 40.
- 7 66. Poirot O, O'Toole E, Notredame C. Tcoffee@igs: A web server for computing, evaluating and  
8 combining multiple sequence alignments. *Nucleic Acids Res.* 2003; 31(13):3503-3506.
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

What we know but do not understand  
about nidovirus helicases

*Virus Res. 2014; ePub ahead of print  
Reprinted with permission*

# CHAPTER 3

Kathleen C. Lehmann  
Eric J. Snijder  
Clara C. Posthuma  
and Alexander E. Gorbalenya

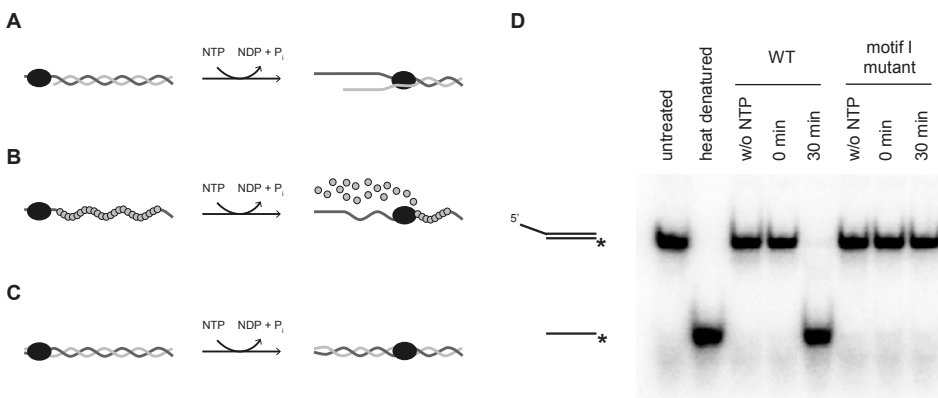
**1 ABSTRACT**

2  
3 Helicases are versatile NTP-dependent motor proteins of monophyletic origin that are  
4 found in all kingdoms of life. Their functions range from nucleic acid duplex unwinding  
5 to protein displacement and double-strand translocation. This explains their participa-  
6 tion in virtually every metabolic process that involves nucleic acids, including DNA  
7 replication, recombination and repair, transcription, translation, as well as RNA process-  
8 ing. Helicases are encoded by all plant and animal viruses with a positive-stranded RNA  
9 genome that is larger than 7 kb, indicating a link to genome size evolution in this virus  
10 class. Viral helicases belong to three out of the six currently recognized superfamilies,  
11 SF1, SF2, and SF3. Despite being omnipresent, highly conserved, and essential, only  
12 a few viral helicases, mostly from SF2, have been studied extensively. In general, their  
13 specific roles in the viral replication cycle remain poorly understood at present. The  
14 SF1 helicase protein of viruses classified in the order *Nidovirales* is encoded in replicase  
15 open reading frame 1b (ORF1b), which is translated to give rise to a large polyprotein  
16 following a ribosomal frameshift from the upstream ORF1a. Proteolytic processing of  
17 the replicase polyprotein yields a dozen or so mature proteins, one of which includes  
18 a helicase. Its hallmark is the presence of an N-terminal multi-nuclear zinc-binding  
19 domain, the nidoviral genetic marker and one of the most conserved domains across  
20 members of the order. This review summarizes biochemical, structural, and genetic data,  
21 including drug development studies, obtained using helicases originating from several  
22 mammalian nidoviruses, along with the results of the genomics characterization of a  
23 much larger number of (putative) helicases of vertebrate and invertebrate nidoviruses.  
24 In the context of our knowledge of related helicases of cellular and viral origin, it dis-  
25 cusses the implications of these results for the protein's emerging critical function(s) in  
26 nidovirus evolution, genome replication and expression, virion biogenesis, and possibly  
27 also post-transcriptional processing of viral RNAs. Using our accumulated knowledge  
28 and highlighting gaps in our data, concepts, and approaches, it concludes with a per-  
29 spective on future research aimed at elucidating the role of helicases in the nidovirus  
30 replication cycle.

31  
32  
33  
34  
35  
36  
37  
38  
39

## 1. HELICASES: CONSERVED BUT VERSATILE PLAYERS IN BIOLOGICAL PROCESSES UTILIZING NUCLEIC ACIDS IN VIRUSES AND HOSTS

Cellular life forms in all kingdoms, as well as positive-stranded (+) RNA viruses with a genome larger than 7 kb and a number of DNA viruses, encode (predicted) helicases on which they depend in various ways (1;2). Helicases are widely recognized for their capability to resolve base pairs in nucleic acid duplexes in an NTP- and metal-dependent manner (Figures 1A and D). Additionally, some helicases were shown to displace proteins present in polynucleotide-protein complexes (Figure 1B) (3) and to translocate along double-stranded DNA or RNA without unwinding it (Figure 1C) (3;4). Besides replication, their unwinding activity may be utilized in many other processes including, but not limited to, DNA repair, transcription, RNA maturation and splicing, and translation. Thus, it is not surprising that as much as 1% of all prokaryotic and eukaryotic genes were estimated to encode helicases (5).



**Figure 1.** Schematic representation of NTP-dependent helicase activities. (A) Unwinding of a nucleic acid duplex. (B) Displacement of protein bound to nucleic acid. (C) Translocation along a double-stranded nucleic acid without unwinding. (D) Typical result of a biochemical nucleic acid unwinding assay. Shown is the helicase activity of equine arteritis virus (EAV) nsp10 and a corresponding active site mutant carrying a Lys-164 to Gln substitution in motif I. The asterisk marks the position of the radioactive label. Panel D adapted from (6).

### 1.1 Helicase classification and domain organization

Helicases are classified into six evolutionary compact and interrelated superfamilies (SFs 1-6), which were established using statistically significant sequence similarity, corroborated and extended by subsequent structural analyses of selected members (3;5). Among the SF members classified are relatively few characterized helicases and numerous related proteins, whose number has steadily increased along with genom-



1 ics projects. The helicase SFs are distinguished by characteristic conserved sequences  
2 (motifs) that are often used to identify new members. Helicases of the two larger su-  
3 perfamilies, SF1 and SF2, are characterized by their monomeric or dimeric state and the  
4 presence of up to twelve conserved sequence motifs involved in NTP and nucleic acid  
5 binding and the coordination thereof (2;3;7). In contrast, SFs 3-6 require the formation  
6 of hexamers or dodecamers in order to be active and contain just three to four signature  
7 motifs. Only some of these motifs are specific to a particular superfamily, while oth-  
8 ers may be conserved in a more or less diverse group of proteins. For instance, motifs  
9 I and II (or the Walker A and B boxes, respectively (8)), are common to all helicase SFs  
10 and are also conserved in many NTPases that lack helicase activity, attesting to the fact  
11 that helicases form an evolutionary lineage within a very diverse class of NTP-utilizing  
12 enzymes. All motifs are embedded in a catalytic core. In SF1 and SF2, it is formed by two  
13 RecA-like domains, designated 1A and 2A, that emerged by duplication and extensive  
14 divergence (9-11). In contrast, members of SFs 3-6 employ a helicase core composed of  
15 a single RecA-like domain. It is this helicase core which essentially allows the conversion  
16 of chemical energy stored in NTP phosphodiester bonds to mechanical energy that  
17 fuels the directional movement of the protein along a nucleic acid strand. SF1 and SF2  
18 members typically include additional domains that may be located upstream and/or  
19 downstream of 1A and 2A, or inserted within those domains. In some helicase proteins,  
20 the size of these accessory domains exceeds that of the actual helicase core domains.  
21 Such insertions are thought to govern specific protein functions, including helicase  
22 activity *per se*, by engaging in protein-protein and protein-nucleic acid interactions or  
23 through additional enzymatic activities. This modular design contributes importantly to  
24 the versatility of this enzyme group and serves to achieve specificity (2;3;12).

## 25 26 **1.2 SF1 helicases**

27  
28 The SF1 helicases, which are the focus of this review, have been further subdivided into  
29 three distinct families using structural and biochemical considerations (Table 1) (3). Each  
30 family was named after its most prominent members of the moment: UvrD/Rep, Pif1-like,  
31 and Upf1-like. The most striking difference between these families is the direction of  
32 translocation along their nucleic acid substrate. Relative to the single strand with which  
33 they associate, members of the first family traverse their substrate in the 3'-5' direction  
34 (type A), whereas members of the other two families are moving from 5' to 3' (type B). A  
35 second distinguishing feature is the nucleic acid substrate preference, which is thought  
36 to be exclusive for DNA for the UvrD/Rep and Pif1-like families. In contrast, the Upf1-like  
37 family comprises members that may unwind either DNA or RNA, and in some cases both  
38 without a clear preference. As mentioned, the number of helicases identified through  
39 comparative genomics (so-called putative helicases) far exceeds the number of proven

1 and structurally characterized helicases. Consequently, it remains uncertain whether  
 2 the above regularities faithfully reflect a different mode of nucleic acid binding in these  
 3 families (2) or are due to a sampling bias of the characterized helicases.

4  
 5 **Table 1.** Biochemical properties of selected SF1 helicases.

Structure-based classification	Nucleotide Substrate				Nucleic Acid Substrate		Unwinding polarity
	ATP	Other NTPs <sup>a</sup>	dATP	Other dNTPs <sup>a</sup>	DNA	RNA	
<b>UvrD/Rep family</b>							
UvrD <sup>b</sup>	+	-	+	-	+	n.d.	3'-5'
Rep <sup>c</sup>	+	-	+	-	+	n.d.	3'-5'
PcrA <sup>d</sup>	+	+	+	+	+	n.d.	3'-5'
<b>Pif1-like family</b>							
Pif1 <sup>e</sup>	+	+	+	+	+	n.d.	5'-3'
RecD <sup>f</sup>	+	+	n.d.	n.d.	+	n.d.	5'-3'
Dda <sup>g</sup>	+	-	+	-	+	n.d.	5'-3'
<b>Upf1-like family</b>							
Upf1 <sup>h</sup>	+	-	+	-	+	+	5'-3'
Nidovirus helicases <sup>i</sup>	+	+	+	+	+	+	5'-3'
Alphavirus nsP2 <sup>jk</sup>	+	+	+	+	-	+	5'-3'
Hepevirus Hel <sup>kl</sup>	+	+	+	+	-	+	5'-3'
Alphatetravirus Hel <sup>km</sup>	+	+	+	+	-	+	5'-3'

23 n.d. not done; <sup>a</sup>considered negative if activity <10% of ATPase activity; <sup>b</sup>(13-15); <sup>c</sup>(16-18); <sup>d</sup>(19); <sup>e</sup>(20;21);  
 24 <sup>f</sup>(22) <sup>g</sup>(23;24); <sup>h</sup>(25;26); <sup>i</sup>(27-32); <sup>j</sup>(33-36); <sup>k</sup>based on biochemical properties and a high structural similarity of  
 25 domains 1A and 2A of tomato mosaic virus Hel to those found in Upf1 we propose to classify the related  
 26 alphavirus nsP2 and hepevirus and alphatetravirus orthologs with proven helicase activity in the Upf1-like  
 27 family; <sup>l</sup>(34); <sup>m</sup>(37).

### 28 1.3 SF1 helicases of RNA viruses

29  
 30 Viral helicases belong to one of three superfamilies, SF1 (e.g., alphavirus-like viruses, nido-  
 31 viruses, herpesviruses), SF2 (e.g., flaviviruses, herpesviruses), or SF3 (e.g., picornavirus-like  
 32 viruses), with their role in the viral replicative cycle remaining largely unknown for any  
 33 RNA virus (5;38;39). Given the apparent correlation between genome size and the pres-  
 34 ence of a helicase (1), it was speculated that these enzymes could either reduce nucleotide  
 35 misincorporation (40) or assist in unwinding of long double-stranded regions during  
 36 replication of (+) RNA viral genomes above a certain size threshold (41). Thus, expression  
 37 of a helicase protein may be particularly important for viruses of the order *Nidovirales*,  
 38 comprising the families *Arteriviridae*, *Coronaviridae*, *Mesoniviridae*, and *Roniviridae*, which  
 39 are characterized by having large to very large (+) RNA genomes (13-34 kb) (42-44).

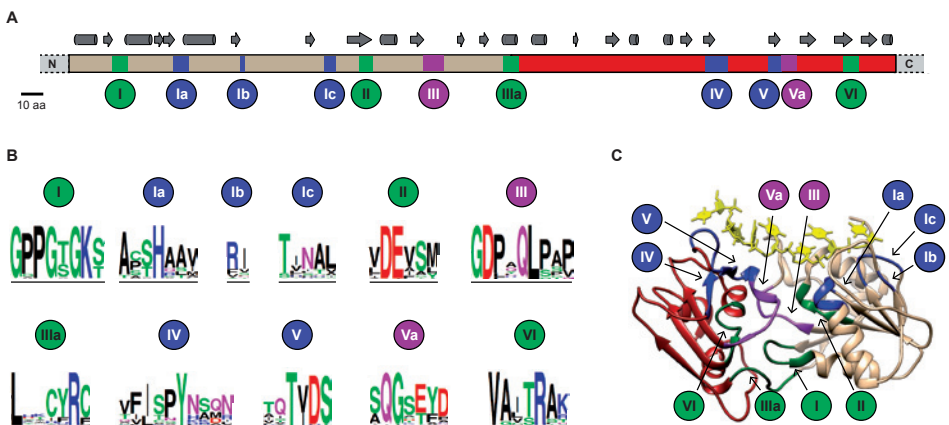
1 In this review, we will summarize the state of the art of our knowledge and recent ad-  
2 vances in our understanding of these nidovirus enzymes. In particular, we will emphasize  
3 similarities and differences between nidovirus helicases and other viral and non-viral  
4 SF1 helicases and their implications for helicase functions in the nidovirus replication  
5 cycle. We will start with a summary of our understanding of the structural organization  
6 and enzymatic activities of SF1 helicases, which is based on the analysis of few cellular  
7 enzymes. In our comparative analysis, we will frequently refer to SF1 helicases of the  
8 alphavirus-like supergroup, the only other large group of (+) RNA viruses that uniformly  
9 encodes a SF1 helicase (45). Among this supergroup, which includes a dozen animal and  
10 plant virus families with (+) RNA genomes in the size range of 7-20 kb, the helicases of  
11 the animal alphaviruses Semliki Forest virus (SFV) and chikungunya virus (CHIKV), the  
12 hepevirus hepatitis E virus (HEV), as well as the plant tobamovirus tomato mosaic virus  
13 (ToMV) have been biochemically characterized to some extent (33-35;45-47). To limit the  
14 scope of this review, comparisons are restricted to data on these proteins while the large  
15 body of literature on the genetics of helicase mutants of alphavirus-like plant viruses  
16 (e.g. see (48)) is not discussed.

## 17 18 19 **2. THE STRUCTURAL BASIS FOR HELICASE TRANSLOCATION,** 20 **DIRECTIONALITY, AND UNWINDING**

21  
22 When characterizing helicases enzymatically and structurally, some of the major ques-  
23 tions asked address the type of substrate used, the direction of movement, rate and  
24 processivity, the coupling of helicase properties to specific biological functions, and the  
25 regulation of the enzymes' activities. In this section, we will briefly summarize our un-  
26 derstanding of mechanisms of SF1 helicase unwinding and directional movement along  
27 a nucleic acid. Helicases may unwind polynucleotides that are either fully or partially  
28 double-stranded, including DNA/DNA, DNA/RNA, or RNA/RNA duplexes (Figure 1A).  
29 Since they provide the initiation sites for unwinding, single-stranded overhangs of  
30 these substrates, if present, may provide specificity to the enzyme-substrate interaction  
31 through their sequence, size, or structure or a combination thereof. SF1 and SF2 include  
32 helicases with diverse biochemical properties, including directionality and substrate  
33 preference; these properties, e.g. 5'-3' directionality of movement, may thus have evolved  
34 more than once (convergent evolution). In comparative terms, they are less conserved  
35 than the sequence motifs that define a superfamily. Consequently, understanding the  
36 underlying selection pressure that drives the conservation of these motifs may provide  
37 a separate and vital insight into helicase function.

## 2.1 Functions of conserved sequence motifs

As outlined above, despite sequence conservation and similar fold of the core domains, a fundamental difference must exist in the way in which helicase proteins of SF1A and SF1B interact with their nucleic acid substrate in order to achieve translocation, and thus unwinding, with a defined polarity. To shed light on the mechanistic basis of this difference, we first need to understand the general molecular mechanism by which directional movement is generated driven by the energy originating from NTP hydrolysis. As one would expect, the elements involved in this essential NTPase activity are generally conserved among SF1 helicases (Figures 2A and B) and (in part) beyond, reflecting considerable evolutionary constraints (2). The most notable of these motifs are motifs I and II, which harbor key residues necessary for the direct interaction with the NTP:Mg<sup>2+</sup> and NDP:Mg<sup>2+</sup> complexes. Furthermore, motif II contains a conserved glutamate thought to act as catalytic base during hydrolysis by accepting a proton from the water molecule that subsequently attacks the NTP (11). Other motifs vary in respect to sequence conservation and diversity of helicases with whom they are associated. Motifs Ia, III, IV, V, Va, and VI are found in SF1 and SF2. Of these, motifs III, IV, V, Va, and VI are most conserved, with motifs III and VI having the most distinctive signatures of each helicase superfamily.



**Figure 2.** Conservation, structure, and function of the SF1 helicase core domains 1A (tan) and 2A (red). Shown are properties of nidovirus helicases, which may deviate from other SF1 helicases. **(A)** Relative positions of motifs and secondary structure elements of a consensus nidovirus helicase. Secondary structures were predicted by Jpred (53) using a multiple sequence alignment of 31 representative nidoviruses that was assisted by tools in the Viralis software platform (54). Motifs as defined by Fairman-Williams *et al.* (2) are colored according to their predominant biochemical function. Green, NTP binding and hydrolysis; blue, nucleic acid binding; purple, coupling between NTP and nucleic acid binding sites. **(B)** Sequence conservation of helicase signature motifs of representative nidoviruses depicted using WebLogo (55). Motifs Q (LNxxQ) and Vb (xxxxVR) are absent in nidoviruses. **(C)** Position of sequence motifs in the tertiary structure of the EAV nsp10 helicase core with the modeled single-stranded nucleic acid shown in yellow (PDB accession number 4N0O)

1 Additionally, motif IIIa is found exclusively in SF1 helicases while motifs Q, Ib, Ic, and Vb  
2 may be specific to different subsets of SF1 helicases. For instance, motifs Q and Vb are  
3 apparently absent in nidovirus helicases (Figure 2).  
4

5 Additional nucleotide-binding residues reside in motifs Q, IIIa, and VI. The first two of  
6 these seem to be specifically important for the recognition of ATP. Thus, it is not surprising  
7 that the Q motif and a conserved tyrosine in motif IIIa that provides stacking interactions  
8 with the adenine base may be lacking in helicases without nucleotide specificity, such  
9 as that of ToMV (45). Moreover, a highly conserved arginine residue in motif VI, termed  
10 “arginine finger”, specifically interacts with the  $\gamma$ -phosphate of the incoming NTP. Rela-  
11 tive to motifs I, II, and Q this residue is located on the opposite side of the NTP-binding  
12 cleft formed by domains 1A and 2A (Figure 2C). This position enables it to fulfill a critical  
13 function in opening and closing of the cleft in response to NTP binding and hydrolysis,  
14 which triggers a conformational shift at the NTP binding site that is translated into a  
15 rotation of domain 2A towards 1A (49-51). Although this domain movement is a com-  
16 mon feature, the exact mechanism and kinetics of hydrolysis and product release may  
17 differ between helicases, possibly in response to specific physiological roles (52).  
18

19 To translate the rotation into movement along a nucleic acid track, the majority of the re-  
20 maining (partially) conserved motifs is primarily devoted to nucleic acid binding (motifs Ia-c,  
21 IV, IVa, V, and Vb). These motifs are located exclusively on the opposite face of the helicase  
22 core relative to the nucleotide binding site, resulting in a nucleic acid-binding channel  
23 involving both core domains on the one side and additional protein-specific domains on the  
24 other. Interestingly, while SF1A helicases are thought to establish contacts mainly via base  
25 stacking, members of SF1B seem to bind predominantly to the phosphate-sugar backbone.  
26 This would suggest that SF1A helicases are strongly inhibited by base lesions while SF1B  
27 members would not tolerate backbone modifications (50). Moreover, the crystal structure  
28 of the DNA-specific Pif1-like helicase RecD2 revealed a C3' endo conformation of the sugar  
29 backbone of a bound DNA. As this conformation appeared to be heavily stabilized by amino  
30 acid side chain interactions, it was speculated that this binding mode may be the basis for  
31 discrimination against RNA, whose 2'-OH would prevent establishing these interactions  
32 (50). In further support of this hypothesis, similar interactions are conserved in other DNA  
33 helicases like Rep, PcrA, and RecB while they cannot be found in Upf1, which can utilize both  
34 DNA and RNA (10;49;51;56). Finally, motifs III and Va contact both NTP and nucleic acid bind-  
35 ing sites to promote the coupling of the binding processes (2). Thus, it is not surprising to  
36 observe that NTP hydrolysis is inhibited if helicase translocation is prevented (57;58).  
37  
38  
39

## 2.2 General models of translocation: how the energy of NTP-hydrolysis is converted into directional movement

In an effort to explain the sequence of events that enables the conversion of the NTP hydrolysis-induced domain rotation into a directional movement, two hypotheses have been brought forward: the Brownian motor hypothesis (59) and the 'backbone stepping motor' model (18;60). The former, based on the observation that the SF2 helicase (NS3) of Hepatitis C virus (HCV) has lower nucleic acid affinity in the NTP-bound state than in the free state, explains the directional translocation on the basis of Brownian motion and a 'power stroke' caused by NTP-hydrolysis. Furthermore, it assumes an asymmetric sawtooth-shaped energy profile for helicase binding to nucleic acids over the length of the single strand. In the absence of NTP, the enzyme would bind strongly to the nucleic acid and thus be incapable of any movement, a state characterized by a local energy minimum. Conversely, once NTP binds, the enzyme's affinity for nucleic acid decreases, which triggers either the dissociation of the protein or its random movement covering a certain number of bases. During this stage, existing energy barriers may be fully or partially overcome. When the NTP is hydrolyzed and the enzyme's affinity for its nucleic acid substrate is restored, the protein may either fall into a neighboring local energy minimum or return to the original one. However, since the energy profile is assumed to be asymmetric, i.e. a shorter, steeper barrier exists on one side of the helicase, the enzyme would be more likely to overcome a local energy maximum in one direction than the other, resulting in a net forward motion. Although this model presents a simple thermodynamic perspective on protein translocation, it has remained unclear which protein or nucleic acid properties would play a role in creating the required asymmetric energy profile (59).

Two main variants have been proposed for the 'backbone stepping motor' model, which also relies on sequential NTP-dependent affinity changes: the 'active rolling' model by Wong and Lohman (60), and the 'inchworm' model, originally from Yarranton and Gefter (18) and extended by Velankar *et al.* (51). Although these models are quite similar, there is one decisive difference: the proposed number of binding sites per enzyme. The former model proposes the existence of a single nucleic acid binding site per helicase subunit, which allows binding to either single- or double-stranded nucleic acids. Thus, in order to engage simultaneously with the transition region between duplex and single strand, the so-called "fork", at least a functional enzyme dimer is required. Upon NTP hydrolysis, the individual subunits of this dimer are then assumed to "roll" over each other, resulting in a constant replacement of the leading molecule. This model thus implies a step size, defined as the distance that a helicase moves forward during a single catalytic cycle, at least equaling the size of each individual binding site. In contrast, the inchworm mecha-

1 nism postulates the existence of two binding sites within the same helicase subunit, al-  
2 lowing the protein to “slide” along a nucleic acid. Therefore, neither the oligomerization  
3 state nor the unwinding step size would be subject to any constraints.

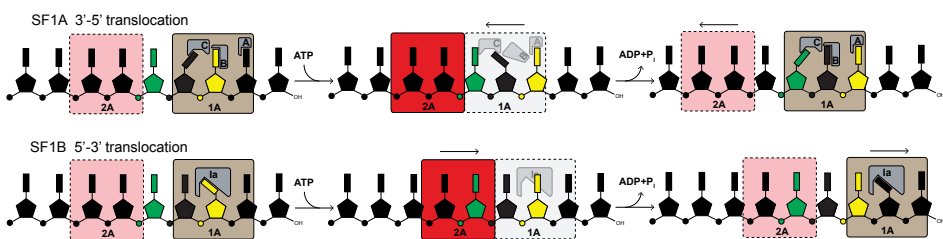
4  
5 In support of the latter model, all helicase crystal structures currently available show a  
6 monomeric protein, even in those cases where the co-crystallized nucleic acid would  
7 have allowed binding of a second subunit (51). Furthermore, biochemical data demon-  
8 strating step sizes of one base pair exist for several helicases (61;62), which is difficult  
9 to reconcile with the active rolling model unless it would include a backward motion  
10 at some point in the reaction cycle. As this would imply a deliberately low efficiency of  
11 unwinding, this is unlikely to be the case. An additional conflict with the active rolling  
12 model derives from studies with PcrA, which was not found to dimerize yet displayed a  
13 higher affinity for substrates comprised of single- and double-stranded regions than for  
14 those containing exclusively either of these states (19). This essentially argues against  
15 the strict temporal separation of single- and double-strand binding. It should be noted  
16 that the kinetic data derived from studies with Rep and UvrD that served as the basis for  
17 the rolling model are consistent with the inchworm mechanism as well.

### 18 19 **2.3 The structural basis for directionality**

20  
21 Given the large body of experimental data supporting it, the inchworm model is now  
22 widely accepted. The model (Figure 3, top panel) based on the SF1A helicase PcrA  
23 proposes two intermediate steps that are characterized by different affinities for a DNA  
24 substrate, a feature controlled by the binding and hydrolysis of NTPs. Initially, the heli-  
25 case protein binds to a single strand in the 3'-5' orientation, utilizing all binding pockets  
26 present in the 1A and 2A domains. Of special importance is a phenylalanine in motif Ia,  
27 located in domain 1A and conserved in the UvrD/Rep subfamily of SF1 (63) but replaced  
28 by different residues in nidoviruses (2nd residue in motif Ia of Figure 2B). It fulfills the  
29 role of gatekeeper for the second of three base acceptor pockets within this domain  
30 (named A to C in the 3'-5' direction) (Figure 3, top panel). Once NTP binds to the NTP  
31 binding site, inducing the closure of the binding cleft, the conformational change is  
32 relayed to this residue and causes the displacement of the base residing in pocket B  
33 into pocket A. At this point, only two of three binding sites of 1A are occupied and thus  
34 the overall affinity for DNA is reduced. Since domain 2A retains its grip on the DNA and  
35 both core domains rotate towards each other, 1A is pulled along the substrate in 3'-5'  
36 direction. Subsequently, NTP hydrolysis triggers the re-opening of pocket B and allows  
37 the next base to enter. Ultimately, binding of a base to the vacant binding pocket in 1A  
38 initiates a cascade of base movements and re-establishes the initial strong binding of  
39

1 1A. In conjunction with the re-opening of the NTP binding cleft, 2A is now pushed away  
 2 from 1A, completing the translocation step in 3'-5' direction (51).

3  
 4 In order to change the direction of translocation to fulfill alternative cellular functions,  
 5 two variations of this mechanism would be conceivable: the helicase could bind to nucleic  
 6 acids in the opposite orientation or the order of binding events could be reversed. Using  
 7 the SF1B prototype helicase RecD2 Saikrishnan *et al.* (50) demonstrated that the latter is  
 8 true for this protein. Although the overall binding site and its location resemble those of  
 9 SF1A helicases, subtle differences in what seems to be an otherwise remarkably similar  
 10 mechanism became apparent (Figure 3, bottom panel). In particular, it is noteworthy  
 11 that the conserved phenylalanine of motif Ia is not present in SF1B helicases but is, as in  
 12 SF2 helicases, often replaced by a conserved proline (64). While this replacement results  
 13 in the loss of the gatekeeping function for this residue, the pocket mentioned above  
 14 still plays a major role during translocation. Before NTP binding, it is occupied by a base,  
 15 which has to flip out of the stacked conformation with its neighboring bases in order to  
 16 bind. Together with several other contacts with the DNA backbone, this leads to a firm  
 17 interaction between domain 1A and the single strand. As soon as NTP is bound and cleft  
 18 closure induced, the conformational change of the NTP-binding cleft translates directly  
 19 into several alterations and concomitant strengthening of the interaction network of  
 20 2A. However, as a number of contacts are made with the downstream base compared  
 21 to the NTP-free state, it seems that, in contrast to the mechanism for 3'-5' translocation,  
 22 2A becomes the dominant binding domain only after the domain movement has been  
 23 completed. In parallel, the base in the motif Ia pocket moves back into a stacked con-  
 24 formation, releasing the mechanical block on 1A movement. This conformational switch  
 25 may be facilitated by relaxation of the DNA as 2A moves towards 1A. Finally, as the NTP  
 26 is hydrolyzed, the binding cleft re-opens and pushes 1A towards the 3' end, concluding  
 27 the 5'-3' translocation by one base.



36 **Figure 3.** Mechanistic models of helicase translocation with different polarities. Schematic representation  
 37 of major conformational changes upon ATP binding and hydrolysis based on crystal structures of PcrA (top)  
 38 and RecD2 (bottom). Domains with the relatively weaker binding affinity are depicted in lighter colors and  
 39 with dashed lines. Mechanistic details are explained in the accompanying section 2.3.



1 Although both models offer a compelling mechanism for the directional movement of  
2 helicases, it needs to be stressed that both are based on the analysis of just two crystal  
3 structures, one in the absence and one in the presence of NTPs, which were obtained  
4 for very different helicases representing SF1A and SF1B, respectively. Thus it is possible  
5 that, when more structures become available, additional transition states might be  
6 identified in the catalytic cycle, possibly leading to revision or elaboration of the pro-  
7 posed mechanisms. Also, these mechanisms must be correlated with the evolutionary  
8 conservation of key residues, which will require insights that may not be deducible from  
9 the helicase family classification (SF1A and SF1B) founded on structural and biochemical  
10 considerations.

#### 11 12 **2.4 Mechanisms of strand separation: do helicases play an active or passive** 13 **role?**

14  
15 As the translocation of a helicase strictly occurs along a single strand, the energy spent  
16 on this process could, in principle, also account for strand separation when the helicase  
17 reaches a double-stranded region simply by excluding the other strand from entering  
18 the binding site and prying the duplex open. In fact, the amount of energy released  
19 by hydrolysis of any NTP (~10 kcal/mol) should be more than sufficient to separate an  
20 average base pair under physiological conditions (~1.6 kcal/mol) (65). Still, the question  
21 whether the actual unwinding is actively supported by the protein or occurs passively  
22 depending on thermal fraying, that is, spontaneous opening and closing, of double  
23 strands is still an unresolved issue. In order to clarify the use of active and passive in  
24 this respect, it should be made clear that this terminology relates to the mechanism by  
25 which strand opening is achieved and not to the overall enzyme activity, which is by  
26 definition always active in the biochemical sense as it requires NTP hydrolysis.

27  
28 In a passive unwinding model, the helicase would temporarily pause when it encounters  
29 a double-stranded region and move forward once the base pair opens, thereby prevent-  
30 ing re-annealing of the single strands (57). Although estimates on the fraying frequency  
31 are as high as  $1000\text{ s}^{-1}$ , this model would not account for helicase proteins or complexes  
32 that possess higher unwinding rates, for example RecBCD (66;67). Thus, at least some  
33 helicases seem to employ an active mechanism for duplex destabilization (12;68). Sup-  
34 port for such a mechanism was again obtained from the crystal structure of PcrA (51).  
35 In this case the optional domains 1B and 2B play a critical role by specifically binding  
36 to double-stranded DNA in the context of the NTP-bound state. Once the NTP is hydro-  
37 lyzed, their binding affinity is lost, and contacts with the DNA are predominantly limited  
38 to both core domains. Thus, NTP binding does not only start translocation, it also leads  
39 to the bending of the duplex region behind the fork. As a consequence of this distortion,

1 the first four to five base pairs at the junction start to open, which allows stabilization of  
2 the newly generated single-stranded region by the forward motion of the protein.

3  
4 In contrast to the duplex opening, the translocation mechanism described with the  
5 inchworm model implies that PcrA progresses by a single base per NTP hydrolyzed.  
6 Therefore, it is possible that translocation and unwinding are not coupled during every  
7 hydrolysis cycle in PcrA (51), a hypothesis that would also explain unwinding step sizes  
8 of more than one base pair seen with some other helicases. This theoretical concept was  
9 supported by findings for HCV NS3, which suggested that three successive hydrolysis  
10 events trigger the sudden unwinding of three base pairs in a so-called spring-load  
11 mechanism (69). Thus, it is important to distinguish between translocation step size,  
12 which may be fixed at one nucleotide per NTP, and the size of individual unwinding  
13 steps, which may involve hydrolysis of several NTP molecules.

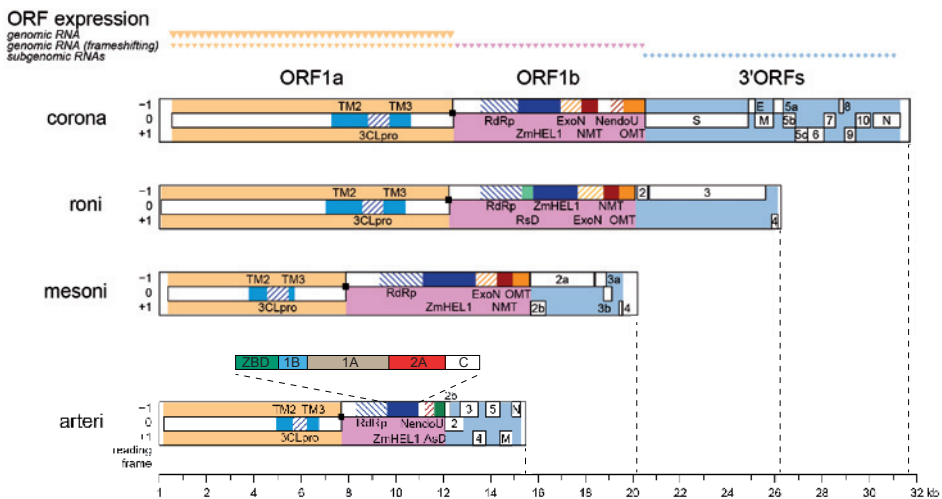
14  
15 From the study of multiple members of different helicase superfamilies, it has now  
16 become apparent that introduction of some kind of mechanical strain into the double-  
17 stranded region, to ease its opening up, may be a common part of helicase action.  
18 However, as the participation of non-conserved domains already implies, the exact  
19 mechanism of strand separation may vary between helicases. Particularly, the element  
20 responsible for the destabilization, called pin or wedge, may be part of different domains  
21 and vary in size (51;56;64;70).

### 22 23 24 **3. THE NIDOVIRUS HELICASE: A MULTI-FUNCTIONAL ENZYME CONTROLLING** 25 **KEY STEPS IN VIRAL REPLICATION**

26  
27 The presence of a helicase in nidoviruses was recognized already in the late 1980s by  
28 comparative genomics involving the first sequenced nidovirus genome, that of the avian  
29 infectious bronchitis coronavirus (71), along with known bacterial helicases, for example  
30 UvrD. These findings led to the original proposal to establish SF1, the first among the  
31 helicase superfamilies, and to classify the corresponding nidovirus protein domain as  
32 putative helicase (7;72;73). Subsequent genomic studies of other nidoviruses reported  
33 conservation of this domain and its characteristic helicase motifs, strongly supporting  
34 the helicase function in nidoviruses.

35  
36 The multidomain helicase subunit (hereafter referred to as helicase) of nidoviruses  
37 ranges in size from 50 to 70 kDa and is called nsp10 in arteriviruses and nsp13 in corona-  
38 viruses, which are the only nidovirus families that were characterized in respect to this  
39 enzyme. The protein is encoded in replicase ORF1b and proteolytically released from a

1 nonstructural polyprotein (pp1ab) that results from ORF1a/1b ribosomal frameshifting  
 2 (Figure 4). Comparative sequence analysis indicated that the nidovirus helicase subunit  
 3 consists at least of an N-terminal zinc-binding domain (ZBD; known also as zinc-binding  
 4 module, Zm, or complex zinc-finger), the helicase core domains, and an as yet undefined  
 5 C-terminal part. Despite considerable size differences, it is one of the three ubiquitous  
 6 and evolutionary most conserved proteins encoded by nidoviruses. The other two are  
 7 the chymotrypsin-like protease (3CL<sup>pro</sup> or main protease M<sup>pro</sup>), responsible for most of  
 8 the proteolytic processing of pp1a and pp1ab, and the RNA-dependent RNA polymerase  
 9 (RdRp) that is thought to synthesize the genome as well as a set of subgenomic (sg)  
 10 mRNAs (41;42). The helicase is genetically segregated with the RdRp in ORF1b, and mu-  
 11 tation of specific conserved residues can either negatively impact or block replication of  
 12 the arterivirus equine arteritis virus (EAV) (74;75), the only nidovirus for which reverse  
 13 genetics analysis of the helicase was performed so far. These genetic and genomics  
 14 observations established the helicase as an essential protein for viral replication.



29 **Figure 4.** Genomic organization and key replicase domains of four nidoviruses. The coding regions are  
 30 partitioned into ORF1a (yellow), ORF1b (purple), and the 3' ORFs (blue), which also differ in their expres-  
 31 sion mechanism or template as indicated on top. Black squares, ribosomal frameshift sites. Within ORFs  
 32 (white rectangles), colored patterns highlight domains identified in: all nidovirus replicases [transmem-  
 33 brane domain 2 (TM2), transmembrane domain 3 (TM3), 3C-like protease (3CLpro), RNA-dependent RNA  
 34 polymerase (RdRp), and Zn-binding domain fused with helicase domain (ZmHEL1)], light and dark blue;  
 35 large nidoviruses [exoribonuclease (ExoN), 2'-O-methyltransferase (OMT)], orange; certain clades [N7-  
 36 methyltransferase (NMT), endoribonuclease (NendoU)], red; ronivirus-specific domain (RsD), light green;  
 37 arterivirus-specific domain (AsD), dark green. Genomic organizations are shown for Beluga whale cor-  
 38 onavirus SW1 (corona), gill-associated virus (roni), Nam Dinh virus (mesoni), and porcine respiratory and  
 39 reproductive syndrome virus, North American genotype (arteri). Domain organization of ZmHEL1 is shown  
 for arteriviruses. Adapted from (42).

1 Biochemically, helicase activity was first demonstrated for a coronavirus and an arteri-  
2 virus (30;31) just few years before the 2003 SARS (severe acute respiratory syndrome)  
3 pandemic, which sparked renewed interest in nidoviruses in general and coronaviruses  
4 in particular. By now biochemical data are available for helicases originating from severe  
5 acute respiratory syndrome coronavirus (SARS-CoV) (27;32;76-78), human coronavirus  
6 229E (HCoV-229E) (29;31;79), porcine reproductive and respiratory syndrome virus  
7 (PRRSV) (28), and EAV (6;30;74). Furthermore, as mentioned above the latter was also  
8 extensively probed in reverse genetics studies (74;75), and the crystal structure of a  
9 truncated variant was reported recently (6). In contrast, putative helicases originating  
10 from the invertebrate families *Roniviridae* and *Mesoniviridae* have not been experimen-  
11 tally characterized calling for caution with respect to generalization of results available  
12 exclusively for vertebrate nidoviruses. For the sake of clarity, the source of the described  
13 data, that is, either from an arterivirus or a coronavirus, is stated throughout this paper  
14 unless a general conclusion applicable to all vertebrate nidoviruses is drawn.

### 16 3.1 Three different enzyme activities with contrasting substrate requirements

#### 18 3.1.1 A promiscuous NTPase sensitive to the presence of polynucleotides

20 As discussed in section 2, helicase translocation is driven by the energy released upon  
21 hydrolysis of phosphodiester bonds of NTPs. Accordingly, mutation of the conserved  
22 lysine of motif I (G/A)x(A/P)GxGK(S/T), which binds ATP, abolished all NTPase activity of  
23 the nidovirus helicases probed thus far (29-31;76). Additionally, a number of labs char-  
24 acterized the nucleotide specificity of these proteins (28;29;32;76). However, a direct  
25 comparison of the kinetic constants obtained in these studies is complicated by differ-  
26 ent experimental conditions used regarding pH, incubation temperature, and presence  
27 of homopolymeric RNAs.

29 In general, like the SF1 helicases of the alphavirus-like viruses CHIKV nsp2 (*Togaviridae*)  
30 (33) and HEV ORF1 protein (*Hepeviridae*) (34), none of the nidovirus enzymes tested  
31 seemed to be able to strongly discriminate between any of the ribo- or deoxynucleotides.  
32 EAV nsp10 was found to utilize both tested nucleotides, ATP and GTP, with comparable  
33 efficiency (30). Similarly, PRRSV nsp10 was able to hydrolyze all NTPs with a preference  
34 for purines over pyrimidines and with UTP being the least favorite substrate (28). The  
35 same order of substrate preferences was deduced in one SARS-CoV nsp13 study, where  
36 the obtained  $k_{\text{cat}}/K_m$  values ranged from  $1.9 \mu\text{M}^{-1}\text{s}^{-1}$  for ATP and GTP to  $0.4 \mu\text{M}^{-1}\text{s}^{-1}$  for UTP  
37 (76). Yet, a second study by Tanner *et al.* (32) on the same protein carrying a different  
38 affinity tag reported an almost twofold difference between ATP ( $k_{\text{cat}}/K_m$   $57.9 \text{ mM}^{-1}\text{s}^{-1}$ )  
39 and GTP ( $k_{\text{cat}}/K_m$   $27.0 \text{ mM}^{-1}\text{s}^{-1}$ ) but a similar efficiency for GTP and CTP (for a discussion

1 of the influence of affinity tags and choice of expression system on enzyme activities  
2 see below). Again UTP was the least efficiently used substrate ( $k_{\text{cat}}/K_m$   $19.1 \text{ mM}^{-1}\text{s}^{-1}$ ).  
3 However unlike other studies, the latter research was performed in the presence of a  
4 polynucleotide, polyuridine (polyU). As such homopolymeric co-factors are known to  
5 affect the NTPase rate of a number of helicases (19;25;80), this difference in experimental  
6 set-up may explain the observed variation in the relative order of substrate preference  
7 (see also below). Finally, also the HCoV-229E helicase behaved similarly in terms of NTP  
8 selectivity, with again ATP being the most and UTP the least preferred substrate ( $k_{\text{cat}}/$   
9  $K_m$   $0.9 \mu\text{M}^{-1}\text{s}^{-1}$  and  $0.3 \mu\text{M}^{-1}\text{s}^{-1}$ , respectively) (29). Furthermore, besides the differences  
10 observed in the two SARS-CoV nsp13 studies, for which a possible reason was discussed  
11 above, all groups examining coronavirus helicases reported that different dNTPs and  
12 NTPs could be utilized with the same overall relative preference. What discriminated  
13 them was an up to threefold lower efficiency of the utilization of dNTPs compared to  
14 NTPs. Given the general resemblance between the data obtained with helicases of  
15 selected coronaviruses and arteriviruses, it is tempting to speculate that this lack of  
16 specificity is conserved in these two distantly related families of vertebrate nidoviruses.  
17 It remains unknown whether this property may also be conserved in the invertebrate  
18 nidoviruses, whose helicases have not been characterized at all.

19  
20 Interestingly, the observations described above suggest that neither base-specific  
21 side chains nor the 2'-OH contribute significantly to nucleotide binding. Instead, the  
22 mere presence of a 5' triphosphate may be the sole prerequisite for promoting NTPase  
23 activity. This interpretation is further supported by a three-dimensional model of the  
24 highly conserved helicase core domains 1A and 2A of SARS-CoV nsp13, computed by  
25 Hoffmann *et al.* (81) based on structures of the PcrA, Rep, and RecB DNA helicases.  
26 After molecular dynamics simulation and energy minimization of ATP binding, six  
27 hydrogen bonds between the  $\beta$ - and  $\gamma$ -phosphates of the ATP and the side chains of  
28 (partially) conserved residues (T286 motif I, K288 motif I, R443 motif IIIa, R567 motif VI)  
29 were predicted. Moreover, an additional hydrogen bond may be established between  
30 the 3'-OH of the sugar and E540 (motif Va). Additionally, the base may be involved in a  
31 stacking interaction with H290 (extended motif I) and a cation- $\pi$  interaction with R442  
32 (motif IIIa). In contrast, no interactions that would be specific for the adenine base could  
33 be identified. Finally, a hydrogen bond was predicted to be present between the 2'-OH  
34 of the ribose and K569 (motif VI). With respect to the experimental data, the presence of  
35 this single interaction, as compared to the nine interactions that are not specific for the  
36 type of sugar, may explain the only slight reduction of enzymatic activity when using  
37 deoxynucleotide substrates.

1 In section 2, we mentioned the stimulation of the NTPase activity of eukaryotic and  
2 bacterial SF1 and SF2 helicases by single-stranded nucleic acids. This effect was also  
3 observed for the nidovirus helicases of EAV, SARS-CoV, and HCoV-229E (30-32). All stud-  
4 ies recorded the strongest stimulation by poly(U), poly(dT), and poly(dA). A remarkable  
5 finding was the unusual magnitude of the increase, 15- to 20-fold for EAV, 15- to 25-fold  
6 for SARS-CoV, and ~50-fold for HCoV-229E, which is similar to the enhancement observed  
7 for viral SF2 helicases, like HCV NS3, which could be stimulated up to 15-fold (39;82). In  
8 contrast, the nucleic acid-stimulated enhancement of NTP hydrolysis by SF1 helicases of  
9 viruses of the alphavirus-like supergroup was generally found to be minor, mostly about  
10 twofold (33;34;83;84). SARS-CoV nsp13 helicase activity was furthermore stimulated by  
11 more than 15-fold by poly(A) and poly(dC). While HCoV-229E nsp13 activity was also  
12 significantly increased by the presence of poly(C) (32-fold), neither poly(A), poly(G), nor  
13 tRNA induced a more than fivefold stimulation. Similarly, EAV nsp10 NTPase activity was  
14 only enhanced up to fourfold by tRNA and homopolymeric RNAs other than poly(U). The  
15 observed variation in the scale of stimulation of NTPase activity depending on the type  
16 of polynucleotide raises the question of its molecular basis. On the one hand, it could  
17 merely reflect a higher affinity for certain nucleic acid substrates compared to others,  
18 which would imply sequence-dependent helicase activity. Yet, as the affinity towards  
19 nucleic acids has not been evaluated for any of the nidovirus helicases so far and little is  
20 known about the exact means by which these allosteric activators may influence NTPase  
21 activity, alternative explanations should also be considered. For instance, based on crys-  
22 tal structures of HCV NS3, Frick *et al.* (85) suggested that electrostatic changes inside the  
23 NTP-binding site could be caused by a subtle rotation of domain 2A upon nucleic acid  
24 binding. From computer simulation of the ionization states of amino acid side chains,  
25 it appeared that this conformational change would lead to a switch of the protonation  
26 states of the conserved lysine in motif I and the conserved aspartate of motif II and  
27 would thus directly influence the NTPase rate. Therefore, the magnitude of nucleic acid  
28 stimulation may be indirectly governed by the relay of conformational changes from the  
29 nucleic acid binding channel in the vicinity of domains 1A and 2A to the NTP binding  
30 cleft between these domains. In this line of reasoning, different nucleic acids may differ  
31 in respect to their ability to induce conformational changes rather than their binding  
32 affinity, a property which might be also relevant for nidovirus helicases.

33  
34  
35  
36  
37  
38  
39

1 3.1.2 A DNA and RNA helicase with stringent requirements for its partially double-stranded  
2 substrate

3  
4 3.1.2.1 A helicase without nucleic acid preference

5  
6 After confirmation of their NTPase activity, the next question was whether the putative  
7 nidovirus helicases are indeed functional and capable of unwinding nucleic acid du-  
8 plexes, presumably the double-stranded RNAs that are formed during viral replication.  
9 HCoV-229E nsp13 and EAV nsp10 were the first proteins for which this question was  
10 addressed in a so-called “all-or-nothing” assay, which only records unwinding events  
11 that result in complete strand separation (30;31). In these pioneering studies, published  
12 in short succession by Seybert *et al.*, both proteins were able to unwind not only partially  
13 double-stranded RNA but also DNA substrates containing a single-stranded region at  
14 one or both of the 5' ends, irrespective of the additional presence of 3' tails. In contrast,  
15 no activity was observed with substrates containing only an unpaired region at one  
16 of the 3' ends or with blunt-ended substrates. These findings demonstrate that the  
17 nidovirus helicase recognizes and binds single-stranded RNA and DNA (see section 3.2  
18 for the structural basis for this lack of specificity) before proceeding to unwind in 5'-3'  
19 direction. Later the same polarity was also established for SARS-CoV nsp13 (32;76;77)  
20 and PRRSV nsp10 (28), confirming the classification of nidovirus helicases as members  
21 of SF1B. In comparison, the helicases of CHIKV nsp2 and HEV ORF1 protein did share the  
22 5'-3' polarity on RNA substrates but were incapable of unwinding DNA (33;34), indicat-  
23 ing that these properties could be uncoupled.

24  
25 The assays employed in the initial studies mentioned above were based on a multiple-  
26 turnover approach, which assesses and compares the ratio of single-stranded products  
27 to double-stranded substrates at a defined reaction end point. Therefore, they disregard  
28 multiple unwinding events involving the same substrate but different helicase mol-  
29 ecules and, in these particular studies, also product re-annealing (30;31). Thus, differ-  
30 ences in binding affinity, processivity, and velocity of the enzyme when comparing the  
31 two substrates may have been masked. Addressing the basis of the apparent lack of  
32 substrate specificity, the unwinding kinetics of RNA and DNA substrates with identical  
33 sequences have been examined more closely for SARS-CoV nsp13 using a single-cycle  
34 assay, which prevents re-binding of proteins that have dissociated from the nucleic acid  
35 or that are not bound when the reaction is started by ATP addition (27). Strikingly, also in  
36 this single-turnover experiment no difference was observed between the unwinding of  
37 RNA and DNA substrates. On the one hand, this lack of specificity may reflect a structural  
38 property (see section 3.2) that has evolved due to the lack of selection for discriminating  
39 between the two substrates and thus has no immediate functional implications for the

1 virus. On the other hand, while RNA unwinding may be employed during replication  
2 and/or transcription of viral RNAs, the helicase could specifically exert its DNA unwind-  
3 ing activity on host nuclear DNA as proposed for the SF2 helicase of HCV, which even  
4 showed a preference for DNA over RNA substrates in absence of protein co-factors (86).  
5 However, no nidovirus helicase was found to be traveling to the nucleus, which stands  
6 in contrast to a few other nonstructural proteins, for example nsp1 of EAV (87). Alterna-  
7 tively, the hypothesis could be modified to propose that the nidovirus helicase might  
8 target host mitochondrial DNA, a possibility that has not been discussed in the literature  
9 to the best of our knowledge. Finally, the biochemical properties described above were  
10 determined using *in vitro* assays utilizing purified recombinant helicases and may be  
11 only partially applicable to helicases within viral replication complexes, which include  
12 other proteins and co-factors.

### 13 14 3.1.2.2 The influence of sequence and size of single-stranded nucleic acids on helicase activity 15 and oligomeric state

16  
17 Next to the nature of the sugar, the sequence of the single-stranded region used to  
18 initially bind the protein may provide specificity for certain substrates. Therefore,  
19 again Seybert *et al.* (31) investigated the unwinding of DNA duplexes containing 10-  
20 nt long homopolymeric tails at one of their 5' ends. Although HCoV-229E nsp13 could  
21 utilize substrates with a dA, dC, or dT tail, it showed a marked preference for the two  
22 pyrimidines. In contrast, dG-containing duplexes were not unwound. However, this may  
23 not be due to a specific discrimination against guanine but rather to the formation of  
24 higher order structures in this particular substrate, which, as the authors speculated,  
25 may interfere with helicase activity. Later, it was also demonstrated that SARS-CoV  
26 nsp13 is able to initiate unwinding on tails with random sequences (27). Despite this  
27 apparent lack of sequence specificity in these *in vitro* assays, the function of the nido-  
28 virus helicase in infected cells may require loading at specific nucleic acid sequences or  
29 higher-order structures. Thus, it would be interesting to investigate binding affinities  
30 towards, for example, sequences located within the untranslated region of the genome,  
31 the antigenome, or any of the transcription-regulating sequences (TRSs) that direct the  
32 production of nidoviral sg RNAs (see section 3.3.2). Furthermore, it is conceivable that  
33 one (or more) of the other nonstructural proteins may interact and target the helicase to  
34 specific genome regions if it indeed would not possess any specificity itself.

35  
36 In addition to the sequence, also the length of the 5' overhang is important for binding  
37 of the protein to its nucleic acid substrate. Besides revealing the minimal spatial require-  
38 ments for helicase binding, characterization of this property may more importantly  
39 provide functional insights into the still disputed matter concerning the need for di-



1 merization of SF1 helicases, which possibly leads to cooperativity. However, it should be  
2 noted that the NTP-binding motifs of the two SF1 core domains face each other to form  
3 an intact NTP-binding site in each protein subunit. Therefore, proteins of this superfam-  
4 ily are not *per se* dependent on oligomerization. In contrast, in SF3-6 members, whose  
5 binding motifs are localized on distant sides of the protein, assembly of a functional NTP  
6 binding site requires two subunits (3). Nevertheless, it has been shown for individual  
7 SF1 members that dimerization is needed at different stages of the enzymatic cycle. For  
8 instance, the *E. coli* helicase UvrD requires dimerization to initiate and sustain unwinding  
9 but not to translocate along DNA (88-91).

10  
11 To shed light on the oligomerization requirements of nidovirus helicases, unwinding  
12 of DNA substrates with the same double-stranded region but progressively shorter  
13 tails was tested for SARS-CoV and HCoV-229E nsp13 (27;31;77). In agreement with the  
14 results summarized above, unwinding was observed in all studies if the DNA contained  
15 an overhang of ten or more nucleotides. In contrast, the results of experiments using  
16 substrates with shorter 5' overhangs and SARS-CoV nsp13 were less consistent between  
17 different groups but suggested that the minimal length required for binding may be in  
18 the range of five to seven nucleotides. However, it is important to note that all factors  
19 that may influence the extent and frequency of the fraying of the double-stranded part  
20 of the substrate, like temperature, sequence, or salt concentration, may have influenced  
21 the outcome of these experiments.

22  
23 Given the need for only short binding sequences, it is possible that multiple helicase  
24 molecules bind simultaneously to a single substrate molecule with a longer tail region,  
25 giving rise to cooperativity effects. Interestingly, the data from Lee *et al.* (77) suggests  
26 that such a scenario could be possible for nidovirus helicases. While unwinding of DNA  
27 comprising a 50-base pair duplex region and tails of 15 or less nucleotides was inefficient  
28 (<7.5% unwound) in a single-cycle all-or-nothing assay containing a 40-fold excess of  
29 protein over DNA, the ratio of single-stranded to double-stranded DNA increased step-  
30 wise to 18%, 55%, and 95% when the tail was extended to 20, 30, or 40 nt, respectively.  
31 Although this result could also be explained by an increasing affinity between a single  
32 protein and its substrate and therefore by more frequent initiation of unwinding, the  
33 absence of a direct correlation between tail length and activity makes it tempting to  
34 speculate that a second molecule entered the reaction once the tail exceeded a certain  
35 length, most likely 20 nt (77). Additionally, in the same study it was shown by cross-  
36 linking experiments with DMS (dimethyl sulfoxide) that SARS-CoV nsp13 can form  
37 dimers, trimers, and possibly also larger oligomers in solution in the absence or pres-  
38 ence of ATP and DNA. Notwithstanding a possibility of forming artificial oligomers by  
39 aggregation, this finding would, at least in principle, be consistent with an earlier yeast-

1 two-hybrid screen indicating an nsp13 self-interaction (92). Nevertheless, none of the  
2 studies has addressed the question whether helicase molecules indeed form oligomers  
3 during unwinding. Likewise, it is unknown whether oligomer formation affects biophys-  
4 ical properties of the individual monomers to reveal typical indicators of cooperativity,  
5 such as coordinated substrate binding or translocation activities (93). Accordingly, Lee  
6 *et al.* (77) postulated that the higher net product formation is plausible if at least one of  
7 the independently operating helicase molecules stayed attached to the substrate until  
8 unwinding was completed.

### 9 10 3.1.2.3 The size of double-stranded regions of the substrate affects helicase processivity

11  
12 The fact that a comparable tail length-dependent increase of unwinding was not  
13 observed in two other studies using a similar design but employing polynucleotides  
14 with slightly shorter duplex regions underlines the importance of another feature of  
15 helicases, their processivity. This feature is defined as either the average number (N)  
16 of base pairs that a helicase can unwind without dissociating from its substrate or as  
17 the probability (P) that a helicase proceeds with unwinding after each catalytic cycle.  
18 This property should not be mistaken for rate of translocation that is defined as number  
19 of base pairs unwound per second (57). In general, processivity can be influenced in  
20 two ways; first, by the protein's affinity for a nucleic acid, and second, by preventing  
21 re-annealing of freshly separated strands. The latter can be achieved by simultaneous  
22 binding of additional proteins behind the helicase irrespective of having a physical  
23 interaction with it. Additionally, this would also reduce the rate of backward movement,  
24 called slippage (94). Thus, the mere presence of a second nsp13 molecule, as opposed  
25 to true biochemical cooperativity, could also explain the data obtained by Lee *et al.* (77)  
26 with duplexes containing successively longer tails.

27  
28 To assess the processivity of the SARS-CoV helicase, DNA substrates with a 20-nt tail and  
29 increasingly longer double-stranded parts were assayed in single-cycle experiments.  
30 As expected, the percentage of unwound double strand decreased with increasing  
31 duplex length (27;77). After correction for spontaneous melting of the last eight to ten  
32 base pairs, which is frequently observed in the presence of helicases (95), Adedeji *et al.*  
33 calculated a processivity (expressed as probability) of  $0.80 \pm 0.03$ , which appears to be  
34 surprisingly low given the 32-kb genome size of SARS-CoV (27). With this value nsp13  
35 would rank far below the highly processive *E. coli* heterotrimeric complex RecBCD (pro-  
36 cessivity  $\sim 1$ ), which contains two functional SF1 helicases, but slightly above T4 phage  
37 Dda with a processivity of 0.73 (95;96) that unwinds a dsDNA genome of a size compa-  
38 rable to that of coronaviruses. However, in the context of a viral replication complex this  
39 processivity may be substantially higher due to an increase of the overall affinity of the

1 complex towards RNA and prevention of re-annealing. The kinetic step size, defined as  
2 the number of base pairs separated per enzyme cycle, hence not necessarily per NTP  
3 hydrolyzed, was estimated to be  $9.4 \pm 2.1$  bp. The catalytic rate for unwinding of DNA  
4 was determined to be 30 steps per second (27). No parameters were calculated by Lee  
5 *et al.* (77).

6  
7 Strikingly, although comparable DNA substrates were used in both studies (27;77), the  
8 unwinding kinetics differed substantially. In one case the fraction of unwound substrate  
9 plateaued after approximately 100 to 500 s (77) while in the other it took less than 1 s  
10 (27) for all duplex lengths. Conveniently, the authors of the latter study also provided the  
11 explanation for this difference when they compared three different recombinant nsp13  
12 fusion proteins containing either a GST (glutathione S-transferase), MBP (maltose-  
13 binding protein), or hexahistidine moiety at their N-termini (27;97). Besides the fusion  
14 partner, a second difference was that the GST version was expressed in eukaryotic cells  
15 with the help of a baculovirus vector while the other two were expressed in *E. coli*. Simi-  
16 lar to the his-tagged bacterially expressed protein from Lee *et al.*, they found a several  
17 hundred-fold lower ATPase rate compared to GST-nsp13 originating from baculovirus  
18 ( $0.2 \text{ s}^{-1}$  compared to  $104.1 \text{ s}^{-1}$ ) for proteins which were expressed in *E. coli*, regardless of  
19 the tag used. At the same time, however, DNA binding properties remained unaffected.  
20 When considering this difference, it is interesting to note that neither the GST nor the  
21 MBP moiety could be removed from SARS-CoV nsp13, nor the His tag from its HCoV-229E  
22 homolog (27;31;76). This may suggest that each of these foreign sequences was similarly  
23 closely associated with the N-terminal domain of the viral protein and hence inaccessi-  
24 ble for a protease. Thus, it seems likely that not the identity of the foreign sequence  
25 was the cause of the inhibition in the study by Lee *et al.* (77) but rather the expression  
26 in a bacterial host that may have resulted in misfolding. Therefore, the numbers cited  
27 above should be taken with caution and the design of any future experimental studies  
28 seeking to determine biophysical parameters of nidovirus helicases should take this  
29 effect into account.

### 30 3.1.3 RTPase: an enzymatic activity not required for helicase function

31  
32  
33 Many RNA viral helicases of SF1 and SF2 were shown to possess RNA 5'-triphosphatase  
34 (RTPase) activity, that is, the capability to specifically cleave the phosphodiester bond  
35 between the  $\beta$ - and  $\gamma$ -phosphate of the most 5' nucleotide of newly synthesized RNA  
36 (35;47;98;99). This activity was shown to be the first of four consecutive steps in the  
37 conventional RNA capping pathway (reviewed in (100)) leading to the addition of an  
38  $^m\text{GpppN}_m$ -cap to the 5' end of mRNA. It is employed for mRNA capping by all eukaryotes  
39 and, presumably, also by a number of RNA viruses with capped genomes, for example

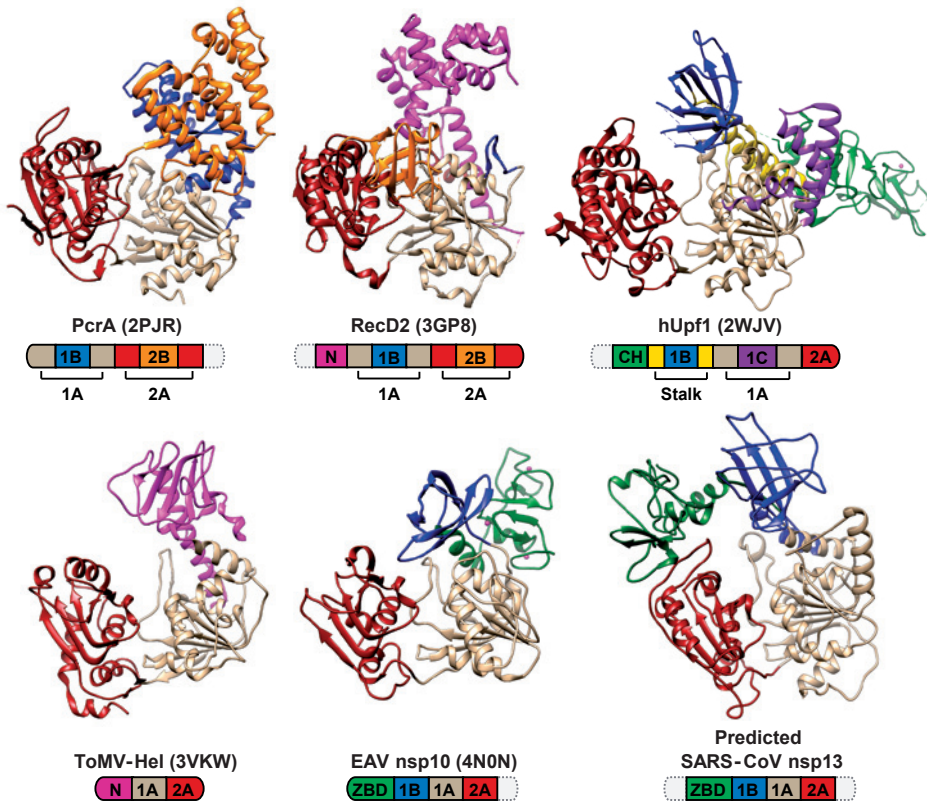
1 flaviviruses and alphaviruses utilizing their NS3 and nsP2 helicases, respectively (35;99).  
2 To test whether this activity is also shared by the helicases of nidoviruses, which also  
3 employ capped genomes (101-103), short RNA substrates were *in vitro* transcribed in the  
4 presence of [ $\gamma$ - $^{32}$ P]GTP. After incubation with SARS-CoV or HCoV-229E nsp13 (29;76) or  
5 most recently EAV nsp10 (our unpublished data), the radioactive phosphate was released  
6 while the RNA stayed otherwise intact. Conversely, when the same assay was performed  
7 with RNAs transcribed in the presence of [ $\alpha$ - $^{32}$ P]GTP, none of the helicases was able to  
8 release the radioactive label, indicating that they do not possess general phosphatase  
9 activity. As, from a mechanistic perspective, the RTPase activity is very similar to the  
10 already reported NTPase activity, the next question was whether both activities would  
11 utilize the same active site. To this end, an alanine substitution mutant of the conserved  
12 lysine located in motif I was shown to abolish both NTPase and RTPase activities for the  
13 above mentioned coronaviruses (29;76). Thus, the authors concluded that the utilization  
14 of both substrates depends on the same active site. To further support this conclusion,  
15 competition experiments between both activities were performed. In agreement with  
16 their conclusion, ATP acted as potent inhibitor for RTPase activity while AMP had almost  
17 no effect. These results imply that nidovirus helicases may also be involved in the control  
18 of the translation of their mRNAs.

### 19 20 **3.2 Structure: interplay between enzyme core, ZBD, and a new accessory** 21 **domain revealed upon polynucleotide binding**

22  
23 Recently, EAV nsp10 became the first nidovirus helicase (and only the second RNA  
24 viral SF1 helicase after that of ToMV (45)) for which a three-dimensional structure was  
25 reported (6). The structure was solved for a somewhat truncated version of the helicase,  
26 which did not include the most C-terminal 65 amino acid residues, after the full-length  
27 protein failed to form any crystals. This truncation did not involve any of the conserved  
28 helicase motifs, and neither ATPase nor helicase activity were abolished, supporting  
29 the relevance of the structure of this truncated nsp10 version. In fact, compared to  
30 the full-length wild-type protein, this engineered protein variant showed an increased  
31 ATPase activity in the absence of homopolymeric RNA. Nevertheless, its unwinding  
32 activity seemed to be moderately lower than for the full-length protein, indicative of  
33 less efficient coupling between ATPase and helicase activities. This may suggest that  
34 the C-terminal domain, which is not conserved among nidoviruses, is not only flexible  
35 but perhaps also exerts a regulatory function on the helicase core, facilitating coupling  
36 between NTPase and polynucleotide binding activities.

37  
38 Overall the EAV nsp10 structure revealed an organization comprising four domains  
39 successively encoding an N-terminal ZBD, a new domain designated 1B, and two RecA-

1 like domains (1A and 2A, together designated HEL1) containing all conserved helicase  
 2 motifs in line with prior analyses (2;72;73). In comparison to representative members of  
 3 the three SF1 helicase families, this organization is most similar to that of Upf1, which  
 4 contains three accessory domains, an N-terminal zinc-binding domain followed by 1B,  
 5 and 1C inserted into the 1A core domain (49). In contrast, UvrD and RecD2 of the UvrD/  
 6 Rep and Pif1-like families, respectively, comprise insertions, designated 1B and 2B, in  
 7 both core domains. While these insertions are the only additional domains for UvrD,  
 8 RecD2 also features an N-terminal domain that, however, does not include a zinc-finger  
 9



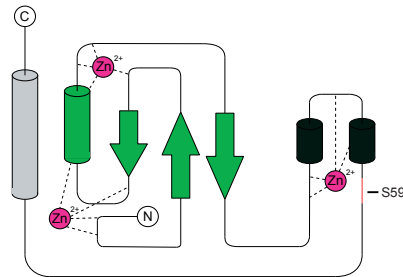
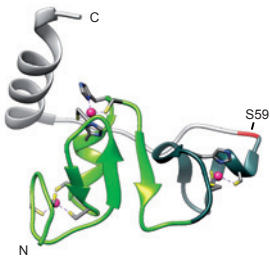
32 **Figure 5.** Three-dimensional models of prototypic prokaryotic and eukaryotic (top) and viral (bottom) SF1  
 33 helicases. Depicted are the prototypic helicases PcrA (UvrD/Rep family), RecD2 (Pif1-like family), hUpf1  
 34 (Upf1-like family), as well as the currently only available structures of viral SF1 helicases of tomato mosaik  
 35 virus (ToMV) and equine arteritis virus (EAV). Also shown is a structure prediction of severe acute respira-  
 36 tory syndrome coronavirus (SARS-CoV) nsp13 obtained with Phyre2 (104). Based on sequence and struc-  
 37 tural comparisons nidovirus helicases are classified into the Upf1-like family. Domain colors correspond to  
 38 those used for the domain organization schemes depicted below each structure. Same coloring of domains  
 39 other than 1A and 2A does not imply any evolutionary relationship. Zinc<sup>2+</sup> ions are depicted as pink spheres.  
 Dashed domains in the organization schemes represent parts that could not be modeled. Domain sizes are  
 not to scale. PDB accession numbers are listed in brackets.

1 (Figure 5) (50;105). At this point, it is important to note that an equivalent designation  
2 does not imply a divergent evolutionary relationship of accessory domains if these  
3 domains are associated with distantly related helicases.

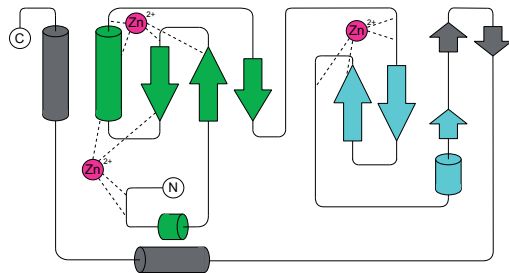
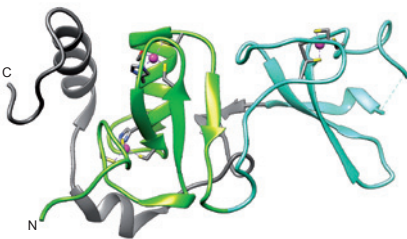
4  
5 Domain 1A (residues 138-293) of the EAV nsp10 structure folds as a parallel five-stranded  
6  $\beta$ -sheet sandwiched between three and two  $\alpha$ -helices on the sides. Conversely, domain  
7 2A (residues 294-401) is comprised of a parallel four-stranded  $\beta$ -sheet and five  $\alpha$ -helices  
8 facing domain 1A. Comparison of the domain structure with all protein structures cur-  
9 rently available yielded the closest similarity to the SF1B helicase Upf1 and its close ho-  
10 molog Ighmbp2 with Z-scores of 20.9 (RMSD 3.5 Å) and 19.9 (RMSD 3.0 Å), respectively.  
11 Likewise, the newly discovered 1B domain (residues 83-137), which contains a two- and  
12 a three-stranded antiparallel  $\beta$ -sheet arranged into a  $\beta$ -barrel fold, resembles a domain  
13 found in the Upf1-like helicase subfamily in terms of its location and orientation relative  
14 to the helicase core and was thus named accordingly. Finally, the ZBD (residues 1-82) co-  
15 ordinates three zinc ions with its twelve conserved cysteine and histidine residues and  
16 folds as two distinct zinc-binding modules, which are connected by a disordered region  
17 (Figure 6). The larger N-terminal module (residues 1-40) coordinating two zinc ions can  
18 be classified as RING-like with a binuclear structure with cross-brace topology. Within  
19 this structure a treble-clef zinc finger-like motif, involving four cysteines in the case of  
20 EAV nsp10, chelates the first metal ion while the second ion is embedded in an  $\alpha\beta$   
21 zinc finger-like motif containing two cysteines and two histidines in EAV nsp10. Based  
22 on nidovirus-wide sequence conservation the signature residues of this RING-like motif  
23 can be described with the formula  $\text{Cys}_{2A}\text{-Cys}_B\text{-Cys}_A\text{-[His/Cys]}_A\text{-[His/Cys]}_3B$  (with A and  
24 B designating the chelated zinc ions, residues in brackets indicating alternative amino  
25 acids, and numbers indicating the number of times the residue occurs in succession).  
26 The more distal zinc-binding module (residues 41-65), which is built by three cysteines  
27 and one histidine in EAV, has a treble-clef fold that is different from the one of the RING  
28 module. Its conservation pattern can be described as  $\text{Cys}\text{-[His/Cys]}\text{-Cys}\text{-[His/Cys]}$ . The  
29 residues outside of these zinc-binding modules are part of either a long loop, which  
30 enables extensive hydrogen bonding with the latter module and may thus contribute  
31 substantially to the overall rigidity of the ZBD, or an  $\alpha$ -helix connecting the zinc-binding  
32 modules with the remainder of the protein. As for the other domains, the closest similar-  
33 ity of the RING-like module was again found to the N-terminal and similarly complex  
34 zinc-binding CH-domain of Upf1 (Z-score 1.9, RMSD 2.2 Å). This similarity included struc-  
35 tural equivalents for six of the eight chelating residues. In contrast, EAV nsp10 and Upf1  
36 have structurally different zinc-binding modules downstream of the RING-like module,  
37 which suggests that the ZBD of nidovirus helicases prototypes a novel type of a complex  
38 multi-nuclear zinc-binding structure.

39

## EAV nsp10 ZBD



## hUpf1 CH



**Figure 6.** Structural comparison between the EAV nsp10 ZBD and hUpf1 CH-domain. Structure and topology of the N-terminal domains of EAV nsp10 (PDB accession number 4N0N) and hUpf1 (PDB accession number 2WJV). Both domains possess a RING-like zinc-binding module of similar fold (bright green) and a second module of different fold. Linker regions are colored in light and dark gray. Residues coordinating  $Zn^{2+}$  are shown as sticks. S59 of EAV nsp10 is shown in red.

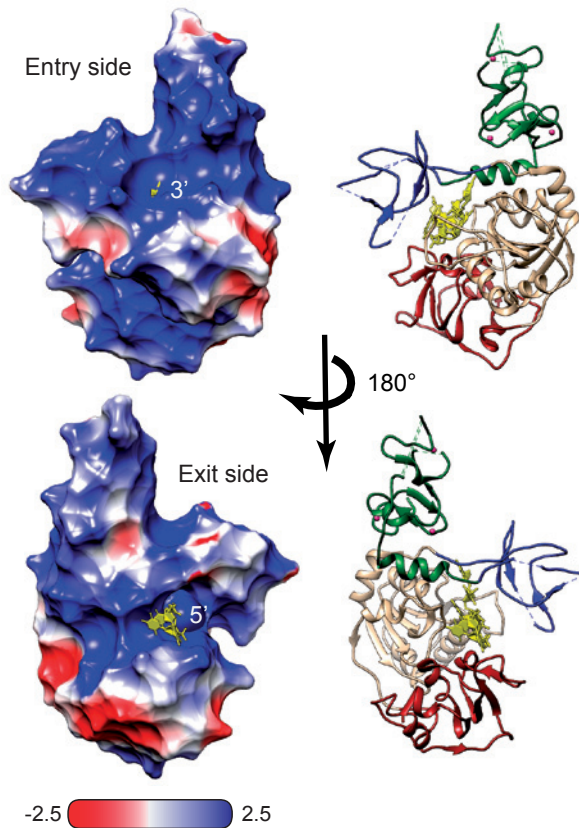
The presence and conservation of a putative complex zinc-binding domain, located between an RdRp upstream and an SF1 helicase downstream, were among the initially recognized specific features of the group of viruses, including coronaviruses and later also arteriviruses, that has now been united in the order *Nidovirales* (38;106). To date, this domain has been identified as uniquely associated with all known nidoviruses, which resulted in its recognition as a molecular marker of the order (41;107). The significance of this observation for nidoviruses is highlighted by the fact that all other conserved protein or polynucleotide domains are either conserved also in other RNA viruses, e.g., RdRp, HEL1, or 3CLpro, or not conserved in some nidoviruses, e.g., an endoribonuclease (NendoU) or an 3'-5' exoribonuclease (ExoN).

Interestingly, all mutant helicases expressed as N-terminal fusions to MBP were reported to have the same solubility as the wild-type protein while His- or GST-tagged mutants clearly were less soluble (6;74). Likewise, Seybert *et al.* (74) demonstrated that zinc ions are an essential structural co-factor required for proper folding of EAV nsp10 as well as HCoV-229E nsp13. Overall, these findings support the assignment of an important

1 role to the ZBD, which could influence the function of the helicase core by affecting  
2 its structure. The latter interpretation was further substantiated by the identification of  
3 an extensive interface area of 1019 Å<sup>2</sup> between these domains, which may be part of a  
4 signaling network (6) (see below for more on function).

5  
6 To elucidate the basis for the lack of specificity for its nucleic acid substrate but also to  
7 gain first insights into its potential unwinding mechanism, a second crystal structure of  
8 nsp10 was solved in complex with a partially double-stranded DNA but in the absence  
9 of NTPs (Figure 2C) (6). Electron density for seven of the ten thymidines of the single-  
10 stranded loading region was identified within a channel formed by domains 1A, 1B,  
11 and 2A. As biochemical data indicated, the majority of the protein DNA contacts was  
12 established either with the phosphodiester backbone or non-specifically with the base.  
13 Furthermore, none of the amino acid side chains was found to be in a position that would  
14 enable an interaction with the 2' hydroxyl group of RNA. Although we cannot exclude  
15 that binding of an RNA substrate might induce such a contact, this structural feature  
16 may explain the lack of discrimination between DNA and RNA, which all currently tested  
17 nidovirus helicases share with some cellular members of the Upf1-like family. In agree-  
18 ment with the proposed unwinding model based on RecD2 (50), the 5' and 3' ends of  
19 the substrate were located in domain 2A and 1A, respectively. Taking the demonstrated  
20 polarity of unwinding into consideration, this implies that domain 1A must be leading  
21 during translocation. The remaining three nucleotides of the single strand as well as the  
22 double-stranded part of the substrate could not be resolved, which indicates a certain  
23 degree of flexibility of the complex. Interestingly, superposition of the Ca atoms of the  
24 helicase core domains of the free and substrate-bound structures revealed that the  
25 overall conformation of these domains is not profoundly affected by DNA binding (RMSD  
26 0.6 Å). Conversely, outside domains 1A and 2A the structural change was significantly  
27 greater with an RMSD of 1.8 Å. Especially remarkable is an approximately 29° rotation of  
28 the 1B domain towards the ZBD upon nucleic acid binding. At the same time, the part  
29 of the nucleic acid substrate channel that is formed by domains 1A and 1B assumes an  
30 open conformation that is 2 Å wider than in the absence of the substrate. Nevertheless,  
31 the channel remains too narrow for entry of a duplex. This may suggest that unwind-  
32 ing is achieved by a structural element at the entrance of the substrate channel that  
33 destabilizes the duplex and makes one of the strands available for being pulled into the  
34 channel. In line with this model, the area around the entry site for the nucleic acid strand  
35 appeared to be heavily positively charged (Figure 7) and may thus be utilized to bend  
36 the duplex during active unwinding, as seen for PcrA, or to guide the displaced strand.  
37 However, as the double-stranded part of the DNA could not be modeled, neither the  
38 presence nor the identity of the putative element that facilitates unwinding of double-  
39 stranded polynucleotides were established. Intriguingly, also surface regions in domain





**Figure 7.** Surface electrostatic potential of an EAV nsp10-DNA complex. Both entry and exit side of the protein's nucleic acid binding channel are predominantly positively charged, potentially providing binding surfaces for nucleic acid. Electrostatic potential mapped onto the molecular surface of 4N0O. Ribbon diagrams are colored as in Figure 5. Red and blue colored regions denote negative and positive surface charges, respectively.

1B and the ZBD that are not part of the substrate channel itself were substantially affected by DNA binding. Thus, it seems plausible that these two domains have a direct role in the presumably substrate-dependent binding of interaction partners.

In conclusion, the crystal structure of EAV nsp10 reinforces a common notion that nidovirus helicases may have evolved N- and/or C-terminal extensions in order to utilize the central helicase core's enzymatic activities in many processes of the viral replication cycle (see section 3.3. for functional implications). Of special interest is the presence of an N-terminal ZBD representing a novel complex zinc-binding fold and having structural similarity to the CH-domain of the cellular helicase Upf1, which is known to be involved in a number of RNA quality control pathways (108). This virus-host similarity is striking since it is observed despite a pronounced divergence of the ZBD among nidoviruses.

1 Indeed, the size differences between the ZBDs of different nidovirus families are con-  
2 siderable, and just a dozen invariant residues mostly involved in zinc binding are shared  
3 by ZBDs. In line with these observations, replacement of EAV nsp10 residues 4 to 63 by  
4 the orthologous nsp10 ZBD sequence from PRRSV, another arterivirus, results in a total  
5 of 31 substitutions and six deletions and was not compatible with EAV viability (75).  
6 Thus, it may be possible that the interaction network between the ZBD and the helicase  
7 core is species-specific and may, especially in the large ZBDs of coronaviruses, impose a  
8 more complex regulation on the helicase core than is now apparent from the EAV nsp10  
9 data. Moreover, it cannot be excluded that larger ZBDs may harbor additional structural  
10 elements that may further expand their functional repertoire.

### 11 12 **3.3 Helicase function: a protein with diverse properties critically involved in** 13 **several processes**

14  
15 Our understanding of the functions of helicase proteins in the nidovirus replication  
16 cycle has been and continues to be informed by functional studies using different  
17 techniques. Some of these studies, e.g., those dealing with the characterization of the  
18 genome, enzymatic activities, and helicase structure, were already reviewed in the sec-  
19 tions above. They will be mentioned below to the extent that is sufficient to connect  
20 them to other nidovirus studies, primarily employing reverse-genetics and molecular  
21 biology techniques, and also to functional paradigms in the helicase field.

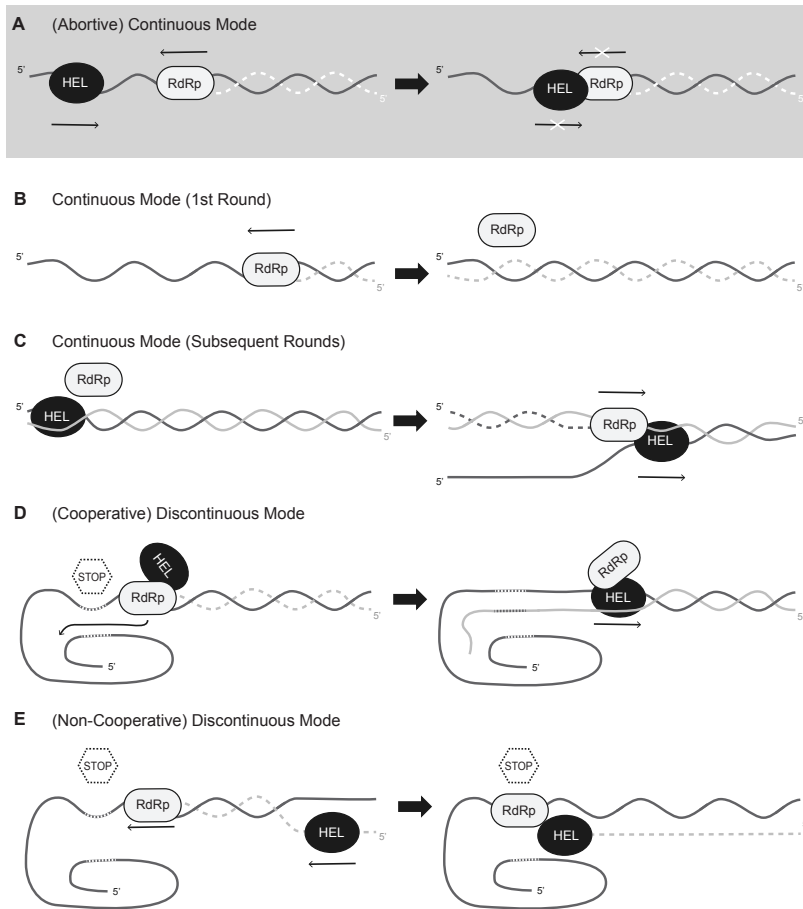
#### 22 23 *3.3.1 Helicase and genome replication*

24  
25 Due to its genetic segregation with the RdRp, helicases of single-strand RNA viruses  
26 could be considered principally replicative helicases. As such, they might function in  
27 a manner reminiscent of better studied DNA helicases that participate in the replica-  
28 tion of the double-stranded genomes of prokaryotes and eukaryotes. Obviously, this  
29 parallel is only valid if it also accounts for specifics of the replication and transcription of  
30 the single-stranded RNA genomes of nidoviruses. In this context, several helicase roles  
31 can be envisioned. First, the enzyme could support the polymerase by removing any  
32 obstacles the replication complex may encounter, for instance secondary structures, in  
33 single-stranded templates, thereby increasing polymerase processivity (39). This model  
34 assumes proximity of or even interaction between the RdRp and helicase, which was  
35 shown for a number of helicases, for example SARS-CoV nsp13 and SF2 HCV NS3, to con-  
36 siderably stimulate their activity (27;109-111). Still, since template-based nucleic acid  
37 synthesis proceeds in the 5'-3' direction, and polymerases thus translocate with the op-  
38 posite polarity along the template, only helicases with a 3'-5' polarity will be able to stay  
39 associated with a replication complex of which the RdRp and helicase move along the

1 same single strand (Figure 8A). Thus, the 5'-3' polarity of the helicases of nidoviruses and  
2 also alphaviruses (28;30-33;76) seems to be incompatible with the conventional model  
3 of the active helicase associating with a polymerase using a single-stranded template,  
4 as became clear immediately when this polarity was first described (31). Alternatively,  
5 the RdRp alone may be capable of unwinding short stretches of secondary structure in  
6 single-stranded templates during RNA synthesis (Figure 8B) while the helicase would  
7 separate the strands of dsRNA. In this model RdRp and helicase may not collide and could  
8 even cooperate as they would move in the same direction but along the complementary  
9 strands to exercise their respective activities (Figure 8C). Since molecules resembling ei-  
10 ther blunt-ended replicative forms (RFs) or 3' polyA-tailed replicative intermediates (RIs),  
11 which are expected to be formed during replication and/or transcription (see below)  
12 by different mechanisms, did not support the activity of nidovirus helicases tested so  
13 far in *in vitro* experiments (30-32;76), the model depicted in Figure 8C depends on the  
14 involvement of other co-factors. For instance and obviously, cooperation of the helicase  
15 with other proteins that bind double-stranded RNA may facilitate its interaction with  
16 RFs/RIs as a step towards initiating unwinding once thermal fraying occurs.

17

18 To formally establish interactions with other (nonstructural) proteins, two-hybrid  
19 screens of SARS-CoV proteins were conducted. These identified three possible interac-  
20 tion partners of nsp13: the accessory protein 3b encoded by an sg RNA, the RdRp nsp12,  
21 and a domain of the transmembrane nsp3, which is the largest nonstructural protein  
22 of coronaviruses and was also shown to bind to several other enzymes implicated in  
23 RNA replication, like the RdRp, the ExoN and N7-methyltransferase (N-MT) (nsp14), and  
24 the 2'-O-methyltransferase (O-MT) (nsp16) (92;112;113). Additionally, nsp13 was found  
25 to localize to presumably endoplasmic reticulum-derived membranous replication  
26 structures in infected cells (76), which is very similar to the localization of most other  
27 nonstructural proteins of related nidoviruses (114;115). These findings indicate that  
28 nsp13 is, as expected, most likely part of the membrane-bound replication complex, an  
29 assumption further supported by findings from a complementation assay. Using EAV  
30 replicons, mutants deficient in nsp10 activity could not be complemented *in trans* by  
31 either simultaneous or exclusive expression of wild-type nsp10 from an internal ribo-  
32 some entry site element inserted downstream of the replicase gene (75). The most likely  
33 explanation, besides technical reasons, is the apparently aberrant cytoplasmic localiza-  
34 tion of ectopically expressed nsp10. This suggests that nsp10 needs to be expressed in  
35 the context of the replicase polyprotein, which likely enables correct complex forma-  
36 tion, in order to fulfill its role during virus replication. On the other hand, it cannot be  
37 excluded at present that nsp10-containing cleavage intermediates may have a separate  
38 function early in virus replication (75;116). Regardless which interpretation is true, both  
39 would require the co-expression of the RdRp and helicase proteins during infection, as



**Figure 8.** Schematic representation of possible functions of the nidovirus helicase (HEL) in cooperation with the viral polymerase (RdRp) during genome replication and transcription. **(A)** Abortive replication and unwinding. Simultaneous RNA synthesis and unwinding of secondary structures of the template strand would lead to collision due to opposite polarities of HEL and RdRp with respect to a single-stranded template. **(B)** During the first round of continuous RNA synthesis on a single strand HEL may not be required. Note that this model assumes that the RdRp is able to remove secondary structures from the template strand. **(C)** Unwinding of double-stranded replication intermediates after the first round of RNA synthesis enables HEL and RdRp to traverse in the same direction along different strands: HEL along the positive strand (black), RdRp along the negative strand (gray). Note that nidovirus HEL is unable to initiate at blunt ends *in vitro* and thus presumably would require additional loading factors. **(D)** During discontinuous negative-strand RNA synthesis the RdRp may stall at a body TRS (dashed and stop symbol). Associated inactive HEL may subsequently facilitate switching to the leader TRS thus enabling addition of the anti-leader sequence without dissociation of RdRp and nascent strand. After completion of negative-strand synthesis, the RdRp becomes inactive. In order to increase HEL processivity, the RdRp stays associated with HEL that is traversing along the template strand (black) to separate the negative-stranded subgenome-length RNA. Additional proteins required for circularization were omitted from the scheme for clarity. Inspired by (121). **(E)** HEL may trail behind the synthesizing RdRp. Once RNA synthesis stalls at a body TRS, continued translocation along the nascent strand would lead to removal of this strand from the RdRp active site once HEL and RdRp collide. The 3' end of the released nascent strand, carrying the body TRS complement, may subsequently base-pair with the leader TRS. Finally, the same or a second RdRp molecule may add the anti-leader sequence. Additional proteins required for circularization and potentially TRS base-pairing were omitted from the scheme for clarity.

1 was expected from the co-segregation of the respective genetic loci. In line with these  
2 considerations, a twofold stimulation of unwinding rate and almost doubling of the ki-  
3 netic step size was detected upon addition of the cognate polymerase to the SARS-CoV  
4 helicase while the ATPase rate remained unchanged (27). Conversely, addition of the  
5 non-cognate RdRp of foot-and-mouth disease virus (3D<sup>pol</sup>, *Picornaviridae* family) had no  
6 effect on any of the nsp13 parameters, indicating that a specific interaction between  
7 RdRp and helicase is required to spur this enhanced activity.

### 8 9 3.3.2 Helicase and genome transcription

10  
11 To rationalize the above observations, it may be informative to recapitulate the main  
12 facts on the unusual transcription mechanism of nidoviruses (reviewed in (117;118)). A  
13 common feature of this virus order is the generation of subgenome-size templates for  
14 sg mRNA synthesis, which is achieved by interruption of negative-strand RNA synthesis  
15 at specific RNA sequences. These signals, termed body TRSs, are located immediately up-  
16 stream of the genes in the 3'-proximal part of the genome, whose expression depends on  
17 the production of sg mRNAs. In most but not all nidovirus groups (see below), following  
18 interruption of negative-strand synthesis, the nascent strand subsequently is translo-  
19 cated to the 5' end of the genomic template where its 3'-terminal body TRS complement  
20 can base-pair with the so-called leader TRS. Next, negative-strand synthesis is resumed  
21 to add the complement of the genomic 5' leader sequence. The subgenome-length  
22 negative strands then serve as templates for the synthesis of viral sg mRNAs, which thus  
23 carry the same 5'- and 3'-terminal sequences as the genome. Whereas all arteri-, corona-,  
24 and mesonivirus sg mRNAs appear to contain a leader that is identical to the genomic  
25 5' end, ronivirus sg mRNAs and all but the largest of the torovirus sg mRNAs lack such  
26 a common 5' sequence, suggesting that their subgenome-length negative strands are  
27 functional templates for sg mRNA synthesis immediately after their release at a body TRS  
28 (103;119;120). Thus, variations in the mechanism of subgenome-length negative-strand  
29 RNA synthesis have evolved, which differ in their ability to resume negative-strand RNA  
30 synthesis after its interruption at a body TRS (117;118).

31  
32 Although currently neither the protein complex(es) involved nor any detailed mecha-  
33 nisms for this process have been described, one of the explanations for the cooperativity  
34 between helicase and RdRp may be found in the context of this unique transcription  
35 mechanism. For example, both enzymes may be part of a complex that allows only one  
36 of the two proteins to be active at any given moment. It is conceivable that subgenome-  
37 length negative-strand RNA synthesis is started by a complex either lacking the helicase  
38 or including an enzymatically silent form (Figure 8D). Once a body TRS is reached, the  
39 RNA sequence may stall the RdRp and during this pause association of the helicase

1 may facilitate a template switch to the leader sequence. The latter may be positioned  
2 close proximity to the original template strand if we assume that the genome is circular-  
3 ized via a protein bridge as seen for several other viruses (122). Implicitly, this model  
4 proposes that the nascent strand only dissociates from the RdRp when anti-leader se-  
5 quence synthesis is completed. At this stage RdRp activity may be silenced and helicase  
6 activity may be triggered, leading to unwinding of the newly synthesized strand by the  
7 helicase's movement in the opposite direction (Figure 8D). In consequence, partially  
8 double-stranded intermediates should arise under these conditions independent from  
9 helicase activity if helicase and RdRp cannot associate during the pause. In an alterna-  
10 tive model, the nidovirus helicase may trail behind the RdRp along the newly synthe-  
11 sized strand similar to the transcription termination factor Rho (123) (Figure 8E). Upon  
12 encountering a body TRS, the RdRp could stall until the lagging helicase reaches it. At  
13 this stage, the helicase would be in the position to pull the negative strand out of the  
14 RdRp active site and thereby potentially allows polymerase dissociation. The nascent  
15 negative strand may then, possibly under guidance of additional proteins, base-pair  
16 with the leader TRS (again assuming genome circularization). Subsequently, the same or  
17 a second RdRp molecule may engage in extending the nascent negative strand with the  
18 anti-leader sequence. This model could also be modified to complete the synthesis of  
19 the negative strand using the original RdRp molecule assisted by an enzymatically inac-  
20 tive helicase molecule as depicted in Figure 8D. The models in Figures 8D and 8E could  
21 also be adapted to the utilization of sg mRNAs, rather than the genome, as template  
22 for the synthesis of complementary strands of other, smaller sg mRNAs, as described  
23 for coronaviruses (124;125). Once subgenome-length negative strands are produced,  
24 they may function as template for the respective sg mRNAs, according to the model in  
25 Figures 8B and 8C. Thus, the above considerations envision differential requirements  
26 for helicase activity and helicase interaction with the RdRp in the continuous and  
27 discontinuous modes of RNA synthesis that operate during nidovirus replication and  
28 transcription. These processes may also recruit host helicase(s) (126), indicating that the  
29 proposed models may be modified as a result of future experimental probing.

30  
31 To shed light on the validity of these hypotheses, it should be insightful to identify  
32 proteins or protein complexes that are modulated upon encountering a body TRS. Re-  
33 markably, the functional profile of a single-point mutant of EAV nsp10 showed that  
34 this protein may have properties compatible with the requirements of the template  
35 switching model (Figure 8D). This mutant, with a replacement of residue S2429 of the  
36 replicase polyprotein (subsequently referred to as S59 of nsp10) by a proline, displayed  
37 a selective reduction of negative- and positive-stranded subgenome-length RNA syn-  
38 thesis by ~100-fold while genome synthesis and polyprotein processing remained un-  
39 altered (116;127). Subsequent further probing of the same residue, which was originally

1 thought to be located at the border between the ZBD and the helicase core, by replacing  
2 it with alanine, cysteine, glycine, histidine, leucine, or threonine led to no significant  
3 difference compared to wild-type nsp10 (75). Already at that time it was speculated that  
4 the special structural properties of proline in combination with a localization within a  
5 proposed hinge-region connecting the ZBD and the helicase core might be the cause of  
6 the mutant phenotype.

7  
8 This hypothesis was recently verified and further refined on the basis of the EAV nsp10  
9 crystal structure (6). It is now evident that this region of the protein appears to be an  
10 integral part of both the ZBD and a linker to the downstream region. S59, which notably  
11 is followed by another proline (P60), is located immediately downstream of the second  
12 zinc-binding module of the ZBD (Figure 6) and contributes three main chain hydrogen  
13 bonds to the interactions with this module. These interactions are likely maintained in  
14 many mutants as long as the general backbone conformation of this loop is not sig-  
15 nificantly altered by, for instance, the introduction of a second proline (6). In agreement  
16 with this interpretation, also the introduction of consecutive glycines at positions 59  
17 and 60, which may lead to excessive flexibility, or inversion of the serine and proline  
18 resulted in phenotypes similar to that of the S59P mutant (75). Interestingly, when any  
19 of these three mutations was introduced into recombinant nsp10, the protein's activities  
20 in ATPase and helicase assays were almost indistinguishable from those of the wild-type  
21 protein while the respective mutant viruses were severely crippled (74;75). These obser-  
22 vations suggested that the ZBD may have a vital function independent of the helicase  
23 activity *per se*, for example, in interacting with partners that facilitate the regulation or  
24 utilization of helicase functions. Remarkably, nsp10 is the third nonstructural protein  
25 in arteriviruses, next to nsp1 (128) and nsp11 (129), which was directly and specifically  
26 implicated in transcription. Given this specific effect on sg RNAs it seems plausible that  
27 interactions between nsp10 and other proteins play a role in discontinuous negative-  
28 strand synthesis. Therefore, it might be possible to rescue the EAV nsp10/S59P mutant  
29 virus by supplying negative-stranded subgenome-length RNA templates separately.

### 30 31 3.3.3 Helicase and virion biogenesis

32  
33 To gain insight into the function of the ZBD, its characterization was extended to point  
34 mutants of EAV nsp10 whose conserved zinc-binding histidine and cysteine residues  
35 were individually swapped. This approach was chosen with the goal of affecting but  
36 not impairing the protein's function by retaining the metal binding capacity, which  
37 was, however, not directly measured in the study. Despite this effort, RNA synthesis in  
38 general was still abolished in most mutants and, consequently, the respective muta-  
39 tions were lethal for the virus (74;75). Where tested, this deficiency coincided with a

1 loss or severe reduction ( $\geq 80\%$ ) of ATPase activity. The only mutants that were viable,  
2 although displaying delayed replication, a small-plaque phenotype, and lower progeny  
3 titers, were C25H and H44C, which *in vitro* had an up to 40% reduced ATPase activity and  
4 therefore also diminished helicase activity. Upon a more detailed investigation of their  
5 defects, total RNA synthesis of both mutants seemed to be severely reduced compared  
6 to the parental virus. Nevertheless, while C25H had apparently lower helicase activity  
7 than H44C, its progeny titer surprisingly was 100-fold higher. Furthermore, the 5-log  
8 decrease of H44C viability was not consistent with its relatively mild decrease in overall  
9 RNA synthesis. These results may indicate that EAV nsp10 has an additional function in  
10 processes downstream of RNA synthesis.

11  
12 In line with the above findings, also a mutation that blocked the nsp10-nsp11 cleavage  
13 resulted in a lack of infectious progeny although genomic and sg RNAs were synthe-  
14 sized (130). Although this defect could also be explained by a late function of nsp11, it  
15 emphasizes the need for further investigation of the role of nsp10 in other steps than  
16 RNA synthesis. Such an additional function would not be unprecedented in the virus  
17 world as the helicase domain of the NS3 protein of several flaviviruses but not its associ-  
18 ated enzymatic activities appears to be important for particle formation, potentially by  
19 providing interaction surfaces for other proteins including the core protein (131-134).

### 20 21 3.3.4 Helicase and translation

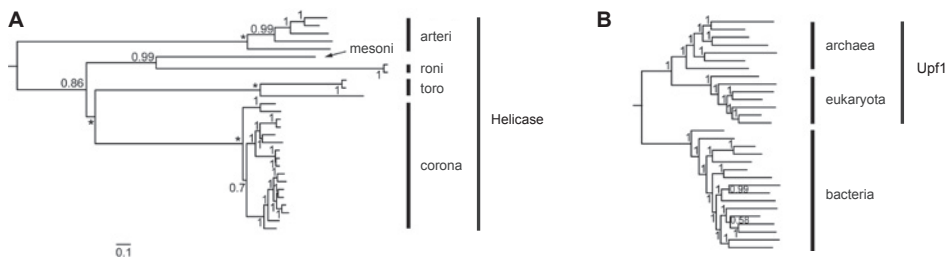
22  
23 Next to its functions in RNA synthesis, biochemical data suggested a role for the helicase  
24 in translation through providing RNA 5'-triphosphatase activity implicated in mRNA  
25 capping (29;76). This activity was demonstrated in other RNA viruses with capped RNA,  
26 including alphavirus-like viruses (35;47;98;99) and flaviviruses (135;136), which also  
27 encode other enzymatic activities involved in the production of capped RNAs. Based  
28 on this parallel, nidoviruses are expected to encode all components of the enzymatic  
29 capping machinery which has not been demonstrated yet for any nidovirus. For coro-  
30 naviruses, a guanylyltransferase has not been identified (137) while two other major  
31 nidovirus groups lack also orthologs of the coronavirus N-MT (toroviruses) or N-MT  
32 and O-MT (arteriviruses) (138). Thus, the RNA 5'-triphosphatase activity of the helicase  
33 protein may be the only enzymatic activity of the capping machinery that is conserved  
34 across nidoviruses. Clearly, further research focusing on different lineages is required  
35 to ascertain the universal link between the conservation of the RNA 5'-triphosphatase  
36 activity of the helicase and cap formation.

37  
38  
39



## 3.3.5 Helicase and post-transcriptional quality control

In addition to being involved in replication, transcription, and translation, which has been proposed for helicases of other RNA virus families as well, nidovirus helicases may also engage in a unique post-transcriptional RNA surveillance and processing pathway. This hypothesis arose from the surprising but provocative similarities to the highly conserved helicase Upf1 (6), which is universally employed by all eukaryotes (Figure 9). As described above, those similarities involved not only the widely documented 5'-3' polarity of unwinding and the lack of nucleic acid specificity but also and uniquely the recently solved domain organization and fold. Most remarkably, Upf1 as well as nidovirus helicases carry a complex bipartite multi-nuclear zinc-binding domain at their N-terminus and an unstructured domain at their C-terminus, which both (probably) exert regulatory functions on the NTPase of these proteins (6;49;139). The most conserved ZBD/CH and the helicase core domains of these two helicase lineages, nidovirus and Upf1, may be of monophyletic origin while the evolutionary relationships of the least conserved domain are understandably untraceable. Given these parallels, it was speculated that, like Upf1 (108), nidovirus helicases could be involved in processes targeting aberrant viral transcripts, including the genome, for degradation (6) in order to prevent the synthesis of potentially harmful, truncated proteins. Alternatively, nidovirus helicases may be employed to interfere with Upf1-dependent pathways of the host by directly competing for interaction partners. In theory such interference may be a means to either protect viral RNAs from being recognized by and targeted to Upf1-dependent degradation pathways or to trigger the specific degradation of antiviral host mRNAs. The latter would mimic the type of regulation of mRNA abundance that is mediated by the host's non-sense mediated decay machinery (140). Currently no data that link the



**Figure 9.** Phylogeny of nidoviruses in comparison to the Tree of Life (ToL). Bayesian phylogenies of nidoviruses (A) and ToL (B) are drawn to a common scale of 0.1 amino acid substitutions per position. Major lineages are indicated by vertical bars and names; arteri: *Arteriviridae*, mesoni: *Mesoniviridae*, roni: *Roniviridae*, toro: *Torovirinae*, corona: *Coronavirinae*. Lineages encoding the nidovirus helicase or Upf1 are indicated. Rooting was according to either (A) domain-specific outgroups or (B) as described (141). Posterior probability support values and fixed basal branch points (\*) are indicated. The nidovirus and ToL alignments include, respectively, three enzymes and 56 single-gene protein families, 604 and 3336 columns, 2.95% and 2.8% gaps. Adapted from (42).

1 nidovirus helicase with the stability of viral or cellular mRNAs are available. Similarly,  
2 RNA signals or proteins that could assist the helicase in the critical recognition of its  
3 wild-type or aberrant targets, particularly those of viral origin, remain uncharacterized.  
4 Thus, further experimental research is clearly needed to test the above hypotheses,  
5 which are not mutually exclusive.

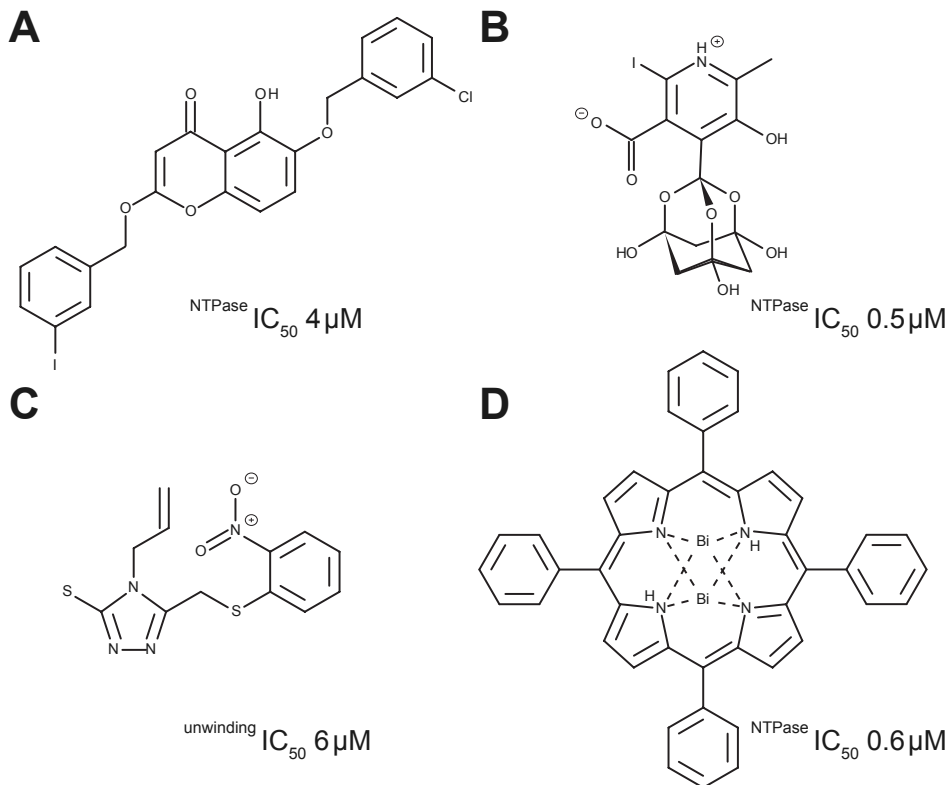
6  
7 The involvement of the helicase in non-sense mediated decay of aberrant cognate  
8 RNAs might have been used to facilitate helicase gene fixation in the genome of an  
9 ancestral nidovirus. Given the restricted genome size variation in families of (+) RNA  
10 viruses (41), such facilitation could be essential since the helicase locus is larger than  
11 1000 nucleotides and its acquisition must have increased genome size considerably. For  
12 nidoviruses, constraints on the genome size expansion were postulated to be linked  
13 to the fidelity of RNA replication and the division of labor between the three principal  
14 genome regions, ORF1a, ORF1b, and the 3'-proximal ORFs (42). These regions seem to  
15 have expanded largely in succession, with domain acquisition by ORF1b leading to the  
16 transition from small to large nidoviruses. Among the proteins acquired early on was  
17 ExoN, whose proofreading activity was likely critical for the fixation of this gene in the  
18 expanding ancestral genome. If it had not improved the fidelity of RNA synthesis, the  
19 expanded genome would have melted down due to its increased error load. The ExoN  
20 acquisition also relieved further genome expansions from fidelity constraints. Similar  
21 considerations were invoked to explain the fixation of the helicase gene in an ances-  
22 tral nidovirus genome upon its transition from a helicase-free proto-ancestor with an  
23 astrovirus-like genome organization (6). It can be argued that with the helicase being  
24 involved in post-transcriptional mRNA quality control, progeny genomes that passed  
25 this quality control would carry fewer errors, an effect similar to, although likely with  
26 smaller impact than, that of ExoN proofreading during replication. In contrast, this effect  
27 would seem unlikely if the helicase were involved in the control of viral or cellular mRNA  
28 stability in the other two ways detailed above. Importantly, although the evolutionary  
29 considerations regarding the ancestral event favor one hypothesis over the other two, it  
30 remains to be established how this likely early specialization constrained further evolu-  
31 tion of the nidovirus helicase. Consequently, all three hypotheses must be considered  
32 when studying contemporary nidoviruses.

#### 33 34 35 **4. NIDOVIRUS HELICASES AS DRUG TARGETS**

36  
37 Despite intensified drug screening efforts since the 2003 SARS pandemic, clinically ap-  
38 proved antivirals against nidoviruses are still lacking. Reflecting their importance during  
39 the viral replicative cycle, major drug targets are the chymotrypsin-like main protease

1 (nsp5 in *Coronaviridae*), the RdRp, and also the helicase. As we have seen in the previous  
2 sections, helicase activity depends on several reactions and/or interactions: NTP binding,  
3 NTP hydrolysis, NDP and phosphate release, nucleic acid binding, translocation, duplex  
4 destabilization, protein co-factor binding, and signal transduction interconnecting any  
5 of these steps. Each of these may, at least in theory, be targeted to prevent unwinding.  
6 Consequently, the diversity of candidate drug scaffolds can be expected to be extensive,  
7 comprising NTP analogs and other small molecules, nucleic acid competitors, as well as  
8 antibodies and aptamers (142). The most accessible target in the helicase subunit prob-  
9 ably is the NTP-binding site. However, given the similarity between viral and cellular  
10 NTPases and the vast number of proven and putative helicase enzymes in humans, many  
11 of the identified hits may possess significant cytotoxicity. To address this challenge, hit  
12 compounds are often first screened in unwinding assays and then counter-selected for  
13 NTPase inhibition. In this manner, molecules targeting less conserved regions and func-  
14 tions of the protein, for instance protein-protein interactions, which tend to be specific  
15 for each helicase, can be identified. Still, also this approach is hampered by the fact that  
16 helicase inhibition is often caused by intercalation into the nucleic acid substrate rather  
17 than binding to the enzyme itself (143).

18  
19 Over the past years, four drug scaffolds with inhibitory effect on the SARS-CoV helicase  
20 and low cytotoxicity have been identified. Interestingly, the available biochemical  
21 data indicate that the modes of action of these inhibitors may be very different. First,  
22 chromone derivatives (Figure 10A) and the adamantane-derived bananins (Figure 10B)  
23 have a direct effect on the NTPase activity although they are not nucleotide analogs  
24 (144;145). While chromones were not characterized in detail, bananins seemed to  
25 specifically inhibit nucleic acid-stimulated NTPase activity in a non-competitive fashion.  
26 Furthermore, these compounds did not inhibit the *E. coli* SF1 helicase DnaB, which sug-  
27 gests the presence of a non-conserved binding site on the surface of SARS-CoV nsp13.  
28 In agreement with this hypothesis, bananins exhibited good selectivity with an  $EC_{50}$   
29 of  $<10 \mu\text{M}$  and  $CC_{50}$  of  $390 \mu\text{M}$ . Whether this site may also be present in other nidovirus  
30 helicases has not been tested so far. Another non-competitive inhibitor was identified  
31 by Adedeji *et al.* (146;147). In contrast to the above compounds, it neither had an effect  
32 on the NTPase or nucleic acid-binding activities nor was it able to bind non-specifically  
33 to nucleic acids. Also this compound, designated SSYA10-001 (Figure 10C) originating  
34 from the Maybridge HitFinder chemical library, did not inhibit the NS3 proteins of the  
35 two flaviviruses HCV and Dengue virus. Moreover, it was efficacious against Middle East  
36 respiratory syndrome coronavirus (MERS CoV), murine hepatitis virus, and SARS-CoV in  
37 cell culture, with  $EC_{50}$  values ranging from  $7\text{-}25 \mu\text{M}$  while being non-toxic up to con-  
38 centrations of  $500 \mu\text{M}$ . Finally, bismuth complexes (Figure 10D) proved to be effective  
39 against SARS-CoV in cell culture ( $EC_{50}$   $6 \mu\text{M}$ ,  $CC_{50}$   $5 \text{mM}$ ) (148;149). As a mode of action it



**Figure 10.** Antiviral compounds with low cytotoxicity targeting SARS-CoV nsp13. Representative chemical structures of each inhibitor family.  $IC_{50}$  values refer to either NTPase or unwinding activity. **(A)** Chromones are inhibitors of NTPase activity. **(B)** Adamantine-derived bananines inhibit nucleic acid-stimulated NTPase activity non-competitively while they do not affect unstimulated NTPase activity. **(C)** SSYA10-001 is a non-competitive inhibitor of unwinding but has no effect on NTPase activity. **(D)** Bismuth complexes act by releasing  $Bi^{3+}$ , which compete with  $Zn^{2+}$  for binding to the ZBD.  $Bi^{3+}$  are located above and below the plane of the porphyrin rings and are additionally coordinated by a solvent molecule.

was proposed that bismuth ions can directly compete with zinc ions for their cysteine binding partners within the ZBD, thereby inhibiting NTPase and unwinding activities. It thus seems likely that these complexes may present broad-spectrum antiviral compounds against all nidoviruses.

## 5. CONCLUDING REMARKS: A LONG AND UNWINDING ROAD TO UNDERSTANDING NIDOVIRUS HELICASES

As detailed above, nidovirus helicases have been the subject of about a dozen, mostly biochemically oriented, studies involving a few mammalian viruses from the family

1 *Arteriviridae* and subfamily *Coronavirinae*. These studies established the *in vitro* require-  
2 ments for processivity and unidirectional 5'-3' movement and their dependence on  
3 partially double-stranded DNA or RNA substrates and NTPase activity. The 5'-3' direc-  
4 tionality coupled with the requirement for a single-stranded overhang to initiate duplex  
5 unwinding is particularly restrictive with respect to RNA synthesis models, which can  
6 probably accommodate the helicase only upon postulating the involvement of addi-  
7 tional and yet-to-be characterized co-factors. This and other properties of the nidovirus  
8 helicases listed above are shared with closely related helicases of viral and host origin.  
9 The overall similarity also includes the importance of a few highly conserved residues in  
10 the characteristic helicase motifs as well as the typical structural organization of the two  
11 core RecA-like domains, as revealed by the recent crystal structure of EAV nsp10.

12  
13 The latter structure was also instrumental in establishing similarities between nido-  
14 viral and Upf1 helicases, creating a novel functional dimension that may be explored  
15 in future nidovirus research to assess the potential involvement of the helicase in  
16 post-transcriptional quality control. The helicase core domains of both the arterivirus  
17 nsp10 and Upf1 enzymes are covalently linked to apparently orthologous multinuclear  
18 zinc-binding domains, whose extensive interaction with other helicase and external  
19 domains may mediate signal transduction. Known as the ZBD in nidoviruses, and distin-  
20 guished by its exclusive presence in these viruses, this domain has been characterized  
21 extensively by site-directed mutagenesis and reverse genetics. From these studies it  
22 became clear that the arterivirus helicase is required for replication, transcription, and  
23 virion biogenesis. How the enzyme controls and possibly interconnects these processes  
24 are big unknowns that may not be resolved without major advancements in several  
25 fields, including the high-resolution visualization and *in vitro* reconstitution of the virus-  
26 specific intracellular factories that mediate major processes in the nidovirus replication  
27 cycle. These lines of inquiry are at the cutting-edge of the currently pursued research  
28 efforts (see e.g. (150;151)). Their progress may be greatly stimulated by the availability  
29 of temperature-sensitive and other conditional mutants, whose parallel characteriza-  
30 tion by traditional genetics, e.g., complementation and recombination, may be equally  
31 insightful (152). Together, these studies are expected to uncover the temporal and  
32 spatial dynamics of the interactions between the helicase and its partners, which have  
33 remained totally obscure so far. Admittedly, the generation and characterization of  
34 such mutants is time-consuming and requires unique expertise. These issues must be  
35 urgently addressed if our understanding of nidovirus helicase functions is to approach  
36 the level that has already been achieved for cellular helicases involved in other complex  
37 RNA-based processes, e.g., transcription or splicing (reviewed in (153;154)).

1 To understand the specifics of helicase functions in the different phylogenetic lineages  
2 and firmly establish nidovirus-wide and lineage-specific features, their experimental  
3 characterization must be extended beyond the currently studied small number of vi-  
4 ruses. These studies may also address questions inspired by comparative genomics, like  
5 it was done with the prior verification of the helicase and ZBD assignments. For instance,  
6 genomics tells us that the helicase is expressed downstream of the RdRp in nidoviruses  
7 while it is the other way around in alpha-, flavi-, and picornavirus-like viruses. This large-  
8 scale evolutionary difference may be linked to fundamental constraints, whose nature  
9 remains as unclear as 25 years ago when this difference was first established (5;38;39).  
10 More recently, we have learned that in a helicase-based phylogeny coronaviruses cluster  
11 with invertebrate nidoviruses rather than with mammalian toroviruses with whom they  
12 share many more characteristics and form the *Coronaviridae* family (138). That study also  
13 established the mosaic conservation of ORF1b domains, including methyltransferases,  
14 in the major phylogenetic lineages of nidoviruses, which questions the universal role  
15 of the helicase's RTPase activity in forming the 5' end RNA cap of nidoviruses. Resolving  
16 these apparent conflicts, and others that could emerge from the on-going genomic  
17 characterization of nidoviruses, is a challenge for future experimental research. If met,  
18 the integration of mechanistic and evolutionary insights may help in developing effec-  
19 tive drugs targeting nidovirus helicases. It will also ensure that our understanding of  
20 the details and relative importance of helicase characteristics is informed by natural  
21 selection rather than formed by our perception.

## 22 23 24 **6. ACKNOWLEDGEMENTS**

25  
26 The authors thank Alexander Kravchenko, Dmitry Samborskiy, and Igor Sidorov for main-  
27 tenance of Viralis and its database. This work was supported by the European Union's  
28 Seventh Framework program (FP7/2007-2013) through the EUVIRNA project (European  
29 Training Network on (+) RNA virus replication and antiviral drug development, Grant  
30 agreement no. 264286), by the Netherlands Organization for Scientific Research (NWO;  
31 TOP-GO grant 700.10.352), by the Collaborative Agreement on Bioinformatics between  
32 Leiden University Medical Center and Moscow State University (MoBiLe), and by the  
33 Leiden University Fund.

34  
35  
36  
37  
38  
39

1 **REFERENCE LIST**

- 2
- 3 1. Gorbalenya AE, Koonin EV. Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.* 1989;17(21):8413-8440.
- 4
- 5 2. Fairman-Williams ME, Guenther UP, Jankowsky E. SF1 and SF2 helicases: family matters. *Curr.Opin. Struct.Biol.* 2010;20(3):313-324.
- 6
- 7 3. Singleton MR, Dillingham MS, Wigley DB. Structure and mechanism of helicases and nucleic acid translocases. *Annu.Rev.Biochem.* 2007;76:23-50.
- 8
- 9
- 10 4. Jankowsky E, Fairman ME. RNA helicases - one fold for many functions. *Curr.Opin.Struct.Biol.* 2007;17(3):316-324.
- 11
- 12 5. Gorbalenya AE, Koonin EV. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr.Opin.Struct.Biol.* 1993;3:419-429.
- 13
- 14 6. Deng Z, Lehmann KC, Li X, *et al.* Structural basis for the regulatory function of a complex zinc-binding domain in a replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase. *Nucleic Acids Res.* 2014;42(5):3464-3477.
- 15
- 16 7. Gorbalenya AE, Koonin EV, Donchenko AP, *et al.* Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* 1989;17(12):4713-4730.
- 17
- 18 8. Walker JE, Saraste M, Runswick MJ, *et al.* Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1982;1(8):945-951.
- 19
- 20 9. Kim JL, Morgenstern KA, Griffith JP, *et al.* Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure.* 1998;6(1):89-100.
- 21
- 22 10. Korolev S, Hsieh J, Gauss GH, *et al.* Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 1997;90(4):635-647.
- 23
- 24 11. Subramanya HS, Bird LE, Brannigan JA, *et al.* Crystal structure of a DExx box DNA helicase. *Nature* 1996;384(6607):379-383.
- 25
- 26 12. Singleton MR, Wigley DB. Modularity and specialization in superfamily 1 and 2 helicases. *J.Bacteriol.* 2002;184(7):1819-1826.
- 27
- 28 13. Hickson ID, Arthur HM, Bramhill D, *et al.* The *E. coli* uvrD gene product is DNA helicase II. *Mol.Gen. Genet.* 1983;190(2):265-270.
- 29
- 30 14. Matson SW. *Escherichia coli* helicase II (uvrD gene product) translocates unidirectionally in a 3' to 5' direction. *J.Biol.Chem.* 1986;261(22):10169-10175.
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

- 1 15. Matson SW, George JW. DNA helicase II of *Escherichia coli*. Characterization of the single-stranded  
2 DNA-dependent NTPase and helicase activities. *J.Biol.Chem.* 1987;262(5):2066-2076.
- 3 16. Arai N, Arai K, Kornberg A. Complexes of Rep protein with ATP and DNA as a basis for helicase  
4 action. *J.Biol.Chem.* 1981;256(10):5287-5293.
- 5 17. Wong I, Moore KJ, Bjornson KP, *et al.* ATPase activity of *Escherichia coli* Rep helicase is dramatically  
6 dependent on DNA ligation and protein oligomeric states. *Biochemistry* 1996;35(18):5726-5734.
- 7 18. Yarranton GT, Gefter ML. Enzyme-catalyzed DNA unwinding: studies on *Escherichia coli* rep pro-  
8 tein. *Proc.Natl.Acad.Sci.U.S.A* 1979;76(4):1658-1662.
- 9 19. Bird LE, Brannigan JA, Subramanya HS, *et al.* Characterisation of *Bacillus stearothermophilus* PcrA  
10 helicase: evidence against an active rolling mechanism. *Nucleic Acids Res.* 1998;26(11):2686-  
11 2693.
- 12 20. Lahaye A, Stahl H, Thines-Sempoux D, *et al.* PIF1: a DNA helicase in yeast mitochondria. *EMBO J.*  
13 1991;10(4):997-1007.
- 14 21. Lahaye A, Leterme S, Foury F. PIF1 DNA helicase from *Saccharomyces cerevisiae*. Biochemical  
15 characterization of the enzyme. *J.Biol.Chem.* 1993;268(35):26155-26161.
- 16 22. Chen HW, Ruan B, Yu M, *et al.* The RecD subunit of the RecBCD enzyme from *Escherichia coli* is a  
17 single-stranded DNA-dependent ATPase. *J.Biol.Chem.* 1997;272(15):10072-10079.
- 18 23. Hacker KJ, Alberts BM. Overexpression, purification, sequence analysis, and characterization of  
19 the T4 bacteriophage dda DNA helicase. *J.Biol.Chem.* 1992;267(29):20674-20681.
- 20 24. Jongeneel CV, Formosa T, Alberts BM. Purification and characterization of the bacteriophage T4  
21 dda protein. A DNA helicase that associates with the viral helix-destabilizing protein. *J.Biol.Chem.*  
22 1984;259(20):12925-12932.
- 23 25. Bhattacharya A, Czaplinski K, Trifillis P, *et al.* Characterization of the biochemical properties of the  
24 human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA.* 2000;6(9):  
25 1226-1235.
- 26 26. Czaplinski K, Weng Y, Hagan KW, *et al.* Purification and characterization of the Upf1 protein: a  
27 factor involved in translation and mRNA degradation. *RNA.* 1995;1(6):610-623.
- 28 27. Adedeji AO, Marchand B, te Velthuis AJ, *et al.* Mechanism of nucleic acid unwinding by SARS-CoV  
29 helicase. *PLoS.One.* 2012;7(5):e36521.
- 30 28. Bautista EM, Faaberg KS, Mickelson D, *et al.* Functional properties of the predicted helicase of  
31 porcine reproductive and respiratory syndrome virus. *Virology* 2002;298(2):258-270.
- 32 29. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of  
33 duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J.Virol.* 2004;  
34 78(14):7833-7838.
- 35  
36  
37  
38  
39



- 1 30. Seybert A, van Dinten LC, Snijder EJ, *et al.* Biochemical characterization of the equine arteritis  
2 virus helicase suggests a close functional relationship between arterivirus and coronavirus heli-  
3 cases. *J.Virol.* 2000;74(20):9586-9593.
- 4 31. Seybert A, Hegyi A, Siddell SG, *et al.* The human coronavirus 229E superfamily 1 helicase has RNA  
5 and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA.* 2000;6(7):1056-1068.
- 6 32. Tanner JA, Watt RM, Chai YB, *et al.* The severe acute respiratory syndrome (SARS) coronavirus  
7 NTPase/helicase belongs to a distinct class of 5' to 3' viral helicases. *J.Biol.Chem.* 2003;278(41):  
8 39578-39582.
- 9 33. Das PK, Merits A, Lulla A. Functional cross-talk between distant domains of chikungunya virus  
10 non-structural protein 2 is decisive for its RNA-modulating activity. *J.Biol.Chem.* 2014;289(9):  
11 5635-5653.
- 12 34. Karpe YA, Lole KS. NTPase and 5' to 3' RNA duplex-unwinding activities of the hepatitis E virus  
13 helicase domain. *J.Virol.* 2010;84(7):3595-3602.
- 14 35. Karpe YA, Aher PP, Lole KS. NTPase and 5'-RNA triphosphatase activities of Chikungunya virus  
15 nsP2 protein. *PLoS.One.* 2011;6(7):e22336.
- 16 36. Xiang H, Ishibashi K, Nishikiori M, *et al.* Expression, purification, and functional characterization of  
17 a stable helicase domain from a tomato mosaic virus replication protein. *Protein Expr.Purif.* 2012;  
18 81(1):89-95.
- 19 37. Wang Q, Han Y, Qiu Y, *et al.* Identification and characterization of RNA duplex unwinding and  
20 ATPase activities of an alphatetravirus superfamily 1 helicase. *Virology* 2012;433(2):440-448.
- 21 38. Gorbalenya AE, Koonin EV, Donchenko AP, *et al.* Coronavirus genome: prediction of putative func-  
22 tional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* 1989;17(12):4847-4861.
- 23 39. Kadare G, Haenni AL. Virus-encoded RNA helicases. *J.Virol.* 1997;71(4):2583-2590.
- 24 40. Koonin EV. Similarities in RNA helicases. *Nature* 1991;352(6333):290.
- 25 41. Gorbalenya AE, Enjuanes L, Ziebuhr J, *et al.* *Nidovirales*: evolving the largest RNA virus genome.  
26 *Virus Res.* 2006;117(1):17-37.
- 27 42. Lauber C, Goeman JJ, Parquet MC, *et al.* The footprint of genome architecture in the largest  
28 genome expansion in RNA viruses. *PLoS.Pathog.* 2013;9(7):e1003500.
- 29 43. de Groot RJ, Cowley JA, Enjuanes L, *et al.* Order *Nidovirales*. In King AMQ, Adams MJ, Carstens EB *et al.*  
30 *et al.* editors, *Virus taxonomy*. Ninth report of the international committee on taxonomy of viruses,  
31 Amsterdam, Elsevier Academic Press, 2012.
- 32 44. Stenglein MD, Jacobson ER, Wozniak EJ, *et al.* Ball python nidovirus: a candidate etiologic agent  
33 for severe respiratory disease in Python regius. *MBio.* 2014;5(5):e01484-14.
- 34
- 35
- 36
- 37
- 38
- 39

- 1 45. Nishikiori M, Sugiyama S, Xiang H, *et al.* Crystal structure of the superfamily 1 helicase from  
2 Tomato mosaic virus. *J.Virol.* 2012;86(14):7565-7576.
- 3 46. Gomez de Cedron M, Ehsani N, Mikkola ML, *et al.* RNA helicase activity of Semliki Forest virus  
4 replicase protein NSP2. *FEBS Lett.* 1999;448(1):19-22.
- 5 47. Karpe YA, Lole KS. RNA 5'-triphosphatase activity of the hepatitis E virus helicase domain. *J.Virol.*  
6 2010;84(18):9637-9641.
- 7 48. Wang X, Lee WM, Watanabe T, *et al.* Brome mosaic virus 1a nucleoside triphosphatase/helicase  
8 domain plays crucial roles in recruiting RNA replication templates. *J.Virol.* 2005;79(21):13747-  
9 13758.
- 10 49. Cheng Z, Muhlrud D, Lim MK, *et al.* Structural and functional insights into the human Upf1 heli-  
11 case core. *EMBO J.* 2007;26(1):253-264.
- 12 50. Saikrishnan K, Powell B, Cook NJ, *et al.* Mechanistic basis of 5'-3' translocation in SF1B helicases.  
13 *Cell* 2009;137(5):849-859.
- 14 51. Velankar SS, Soutanas P, Dillingham MS, *et al.* Crystal structures of complexes of PcrA DNA heli-  
15 case with a DNA substrate indicate an inchworm mechanism. *Cell* 1999;97(1):75-84.
- 16 52. Toseland CP, Martinez-Senac MM, Slatter AF, *et al.* The ATPase cycle of PcrA helicase and its cou-  
17 pling to translocation on DNA. *J.Mol.Biol.* 2009;392(4):1020-1032.
- 18 53. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*  
19 2008;36(Web Server issue):W197-201.
- 20 54. Gorbalenya AE, Lieutaud P, Harris MR, *et al.* Practical application of bioinformatics by the multidis-  
21 ciplinary VIZIER consortium. *Antiviral Res.* 2010;87(2):95-110.
- 22 55. Crooks GE, Hon G, Chandonia JM, *et al.* WebLogo: a sequence logo generator. *Genome Res.* 2004;  
23 14(6):1188-1190.
- 24 56. Singleton MR, Dillingham MS, Gaudier M, *et al.* Crystal structure of RecBCD enzyme reveals a  
25 machine for processing DNA breaks. *Nature* 2004;432(7014):187-193.
- 26 57. Lohman TM, Bjornson KP. Mechanisms of helicase-catalyzed DNA unwinding. *Annu.Rev.Biochem.*  
27 1996;65:169-214.
- 28 58. Maine IP, Kodadek T. Inhibition of the DNA unwinding and ATP hydrolysis activities of the bacte-  
29 riophage T4 DDA helicase by a sequence specific DNA-protein complex. *Biochem.Biophys.Res.*  
30 *Commun.* 1994;198(3):1070-1077.
- 31 59. Levin MK, Gurjar M, Patel SS. A Brownian motor mechanism of translocation and strand separa-  
32 tion by hepatitis C virus helicase. *Nat.Struct.Mol.Biol.* 2005;12(5):429-435.
- 33  
34  
35  
36  
37  
38  
39

- 1 60. Wong I, Lohman TM. Allosteric effects of nucleotide cofactors on *Escherichia coli* Rep helicase-  
2 DNA binding. *Science* 1992;256(5055):350-355.
- 3 61. Dillingham MS, Wigley DB, Webb MR. Demonstration of unidirectional single-stranded DNA  
4 translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry*  
5 2000;39(1):205-212.
- 6 62. Tomko EJ, Fischer CJ, Niedziela-Majka A, *et al.* A nonuniform stepping mechanism for *E. coli* UvrD  
7 monomer translocation along single-stranded DNA. *Mol.Cell* 2007;26(3):335-347.
- 8 63. Dillingham MS, Soultanas P, Wiley P, *et al.* Defining the roles of individual residues in the single-  
9 stranded DNA binding site of PcrA helicase. *Proc.Natl.Acad.Sci.U.S.A* 2001;98(15):8381-8387.
- 10 64. Saikrishnan K, Griffiths SP, Cook N, *et al.* DNA binding to RecD: role of the 1B domain in SF1B  
11 helicase activity. *EMBO J.* 2008;27(16):2222-2229.
- 12 65. von Hippel PH, Delagoutte E. A general model for nucleic acid helicases and their "coupling"  
13 within macromolecular machines. *Cell* 2001;104(2):177-190.
- 14 66. Chen YZ, Zhuang W, Prohofsky EW. Energy flow considerations and thermal fluctuational opening  
15 of DNA base pairs at a replicating fork: unwinding consistent with observed replication rates.  
16 *J.Biomol.Struct.Dyn.* 1992;10(2):415-427.
- 17 67. Liu B, Baskin RJ, Kowalczykowski SC. DNA unwinding heterogeneity by RecBCD results from static  
18 molecules able to equilibrate. *Nature* 2013;500(7463):482-485.
- 19 68. Pyle AM. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu.Rev.Biophys.*  
20 2008;37:317-336.
- 21 69. Myong S, Bruno MM, Pyle AM, *et al.* Spring-loaded mechanism of DNA unwinding by hepatitis C  
22 virus NS3 helicase. *Science* 2007;317(5837):513-516.
- 23 70. Singleton MR, Scaife S, Wigley DB. Structural analysis of DNA replication fork reversal by RecG.  
24 *Cell* 2001;107(1):79-89.
- 25 71. Boursnell ME, Brown TD, Foulds IJ, *et al.* Completion of the sequence of the genome of the coro-  
26 navirus avian infectious bronchitis virus. *J.Gen.Virol.* 1987;68 ( Pt 1):57-77.
- 27 72. Gorbalenya AE, Koonin EV, Donchenko AP, *et al.* A novel superfamily of nucleoside triphosphate-  
28 binding motif containing proteins which are probably involved in duplex unwinding in DNA and  
29 RNA replication and recombination. *FEBS Lett.* 1988;235(1-2):16-24.
- 30 73. Hodgman TC. A new superfamily of replicative proteins. *Nature* 1988;333(6168):22-23.
- 31 74. Seybert A, Posthuma CC, van Dinten LC, *et al.* A complex zinc finger controls the enzymatic activi-  
32 ties of nidovirus helicases. *J.Virol.* 2005;79(2):696-704.
- 33  
34  
35  
36  
37  
38  
39

- 1 75. van Dinten LC, van Tol H, Gorbalenya AE, *et al.* The predicted metal-binding region of the arteri-  
2 virus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion  
3 biogenesis. *J.Virol.* 2000;74(11):5213-5223.
- 4 76. Ivanov KA, Thiel V, Dobbe JC, *et al.* Multiple enzymatic activities associated with severe acute  
5 respiratory syndrome coronavirus helicase. *J.Virol.* 2004;78(11):5619-5632.
- 6 77. Lee NR, Kwon HM, Park K, *et al.* Cooperative translocation enhances the unwinding of duplex DNA  
7 by SARS coronavirus helicase nsP13. *Nucleic Acids Res.* 2010;38(21):7626-7636.
- 8 78. Thiel V, Ivanov KA, Putics A, *et al.* Mechanisms and enzymes involved in SARS coronavirus genome  
9 expression. *J.Gen.Virol.* 2003;84(Pt 9):2305-2315.
- 10 79. Seybert A, Ziebuhr J. Guanosine triphosphatase activity of the human coronavirus helicase. *Adv.*  
11 *Exp.Med.Biol.* 2001;494:255-260.
- 12 80. Preugschat F, Averett DR, Clarke BE, *et al.* A steady-state and pre-steady-state kinetic analysis of  
13 the NTPase activity associated with the hepatitis C virus NS3 helicase domain. *J.Biol.Chem.* 1996;  
14 271(40):24449-24457.
- 15 81. Hoffmann M, Eitner K, von Grotthuss M, *et al.* Three dimensional model of severe acute respiratory  
16 syndrome coronavirus helicase ATPase catalytic domain and molecular design of severe acute  
17 respiratory syndrome coronavirus helicase inhibitors. *J.Comput.Aided Mol.Des* 2006;20(5):305-  
18 319.
- 19 82. Suzich JA, Tamura JK, Palmer-Hill F, *et al.* Hepatitis C virus NS3 protein polynucleotide-stimulated  
20 nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes.  
21 *J.Virol.* 1993;67(10):6152-6158.
- 22 83. Gros C, Wengler G. Identification of an RNA-stimulated NTPase in the predicted helicase sequence  
23 of the Rubella virus nonstructural polyprotein. *Virology* 1996;217(1):367-372.
- 24 84. Kadare G, David C, Haenni AL. ATPase, GTPase, and RNA binding activities associated with the  
25 206-kilodalton protein of turnip yellow mosaic virus. *J.Virol.* 1996;70(11):8169-8174.
- 26 85. Frick DN, Rypma RS, Lam AM, *et al.* Electrostatic analysis of the hepatitis C virus NS3 helicase  
27 reveals both active and allosteric site locations. *Nucleic Acids Res.* 2004;32(18):5519-5528.
- 28 86. Pang PS, Jankowsky E, Planet PJ, *et al.* The hepatitis C viral NS3 protein is a processive DNA heli-  
29 case with cofactor enhanced RNA unwinding. *EMBO J.* 2002;21(5):1168-1176.
- 30 87. Tijms MA, van der Meer Y, Snijder EJ. Nuclear localization of non-structural protein 1 and nucleo-  
31 capsid protein of equine arteritis virus. *J.Gen.Virol.* 2002;83(Pt 4):795-800.
- 32 88. Ali JA, Maluf NK, Lohman TM. An oligomeric form of *E. coli* UvrD is required for optimal helicase  
33 activity. *J.Mol.Biol.* 1999;293(4):815-834.
- 34  
35  
36  
37  
38  
39

- 1 89. Fischer CJ, Maluf NK, Lohman TM. Mechanism of ATP-dependent translocation of *E.coli* UvrD  
2 monomers along single-stranded DNA. *J.Mol.Biol.* 2004;344(5):1287-1309.
- 3 90. Maluf NK, Fischer CJ, Lohman TM. A Dimer of *Escherichia coli* UvrD is the active form of the heli-  
4 case *in vitro*. *J.Mol.Biol.* 2003;325(5):913-935.
- 5 91. Maluf NK, Lohman TM. Self-association equilibria of *Escherichia coli* UvrD helicase studied by  
6 analytical ultracentrifugation. *J.Mol.Biol.* 2003;325(5):889-912.
- 7 92. von Brunn A, Teepe C, Simpson JC, *et al.* Analysis of intraviral protein-protein interactions of the  
8 SARS coronavirus ORFeome. *PLoS.One.* 2007;2(5):e459.
- 9 93. Levin MK, Wang YH, Patel SS. The functional interaction of the hepatitis C virus helicase molecules  
10 is responsible for unwinding processivity. *J.Biol.Chem.* 2004;279(25):26005-26012.
- 11 94. Manosas M, Xi XG, Bensimon D, *et al.* Active and passive mechanisms of helicases. *Nucleic Acids*  
12 *Res.* 2010;38(16):5518-5526.
- 13 95. Eoff RL, Raney KD. Intermediates revealed in the kinetic mechanism for DNA unwinding by a  
14 monomeric helicase. *Nat.Struct.Mol.Biol.* 2006;13(3):242-249.
- 15 96. Lucius AL, Vindigni A, Gregorian R, *et al.* DNA unwinding step-size of *E. coli* RecBCD helicase  
16 determined from single turnover chemical quenched-flow kinetic studies. *J.Mol.Biol.* 2002;324(3):  
17 409-428.
- 18 97. Adedeji AO, Singh K, Sarafianos SG. Structural and biochemical basis for the difference in the  
19 helicase activity of two different constructs of SARS-CoV helicase. *Cell Mol.Biol.(Noisy.-le-grand)*  
20 2012;58(1):114-121.
- 21 98. Balistreri G, Caldentey J, Kaariainen L, *et al.* Enzymatic defects of the nsP2 proteins of Semliki  
22 Forest virus temperature-sensitive mutants. *J.Virol.* 2007;81(6):2849-2860.
- 23 99. Benarroch D, Selisko B, Locatelli GA, *et al.* The RNA helicase, nucleotide 5'-triphosphatase, and  
24 RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg<sup>2+</sup>-dependent and require a  
25 functional Walker B motif in the helicase catalytic core. *Virology* 2004;328(2):208-218.
- 26 100. Decroly E, Ferron F, Lescar J, *et al.* Conventional and unconventional mechanisms for capping viral  
27 mRNA. *Nat.Rev.Microbiol.* 2012;10(1):51-65.
- 28 101. Lai MM, Patton CD, Stohlman SA. Replication of mouse hepatitis virus: negative-stranded RNA  
29 and replicative form RNA are of genome length. *J.Virol.* 1982;44(2):487-492.
- 30 102. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion  
31 RNA. *Virology* 1986;151(1):146-150.
- 32 103. van Vliet AL, Smits SL, Rottier PJ, *et al.* Discontinuous and non-discontinuous subgenomic RNA  
33 transcription in a nidovirus. *EMBO J.* 2002;21(23):6571-6580.
- 34  
35  
36  
37  
38  
39

- 1 104. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre  
2 server. *Nat.Protoc.* 2009;4(3):363-371.
- 3 105. Lee JY, Yang W. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke.  
4 *Cell* 2006;127(7):1349-1360.
- 5 106. Den Boon JA, Snijder EJ, Chirnside ED, *et al.* Equine arteritis virus is not a togavirus but belongs to  
6 the coronaviruslike superfamily. *J.Virol.* 1991;65(6):2910-2920.
- 7 107. Lauber C, Ziebuhr J, Junglen S, *et al.* *Mesoniviridae*: a proposed new family in the order *Nidovirales*  
8 formed by a single species of mosquito-borne viruses. *Arch.Virol.* 2012;157(8):1623-1628.
- 9 108. Imamachi N, Tani H, Akimitsu N. Up-frameshift protein 1 (UPF1): multitasking entertainer in RNA  
10 decay. *Drug Discov.Ther.* 2012;6(2):55-61.
- 11 109. Cha TA, Alberts BM. The bacteriophage T4 DNA replication fork. Only DNA helicase is required for  
12 leading strand DNA synthesis by the DNA polymerase holoenzyme. *J.Biol.Chem.* 1989;264(21):  
13 12220-12225.
- 14 110. Sladewski TE, Hetrick KM, Foster PL. *Escherichia coli* Rep DNA helicase and error-prone DNA  
15 polymerase IV interact physically and functionally. *Mol.Microbiol.* 2011;80(2):524-541.
- 16 111. Zhang C, Cai Z, Kim YC, *et al.* Stimulation of hepatitis C virus (HCV) nonstructural protein 3 (NS3)  
17 helicase activity by the NS3 protease domain and by HCV RNA-dependent RNA polymerase.  
18 *J.Virol.* 2005;79(14):8687-8697.
- 19 112. Imbert I, Snijder EJ, Dimitrova M, *et al.* The SARS-Coronavirus PLnc domain of nsp3 as a replica-  
20 tion/transcription scaffolding protein. *Virus Res.* 2008;133(2):136-148.
- 21 113. Pan J, Peng X, Gao Y, *et al.* Genome-wide analysis of protein-protein interactions and involvement  
22 of viral proteins in SARS-CoV replication. *PLoS.One.* 2008;3(10):e3299.
- 23 114. Li Y, Tas A, Snijder EJ, *et al.* Identification of porcine reproductive and respiratory syndrome virus  
24 ORF1a-encoded non-structural proteins in virus-infected cells. *J.Gen.Virol.* 2012;93(Pt 4):829-839.
- 25 115. van der Meer Y, Snijder EJ, Dobbe JC, *et al.* Localization of mouse hepatitis virus nonstructural  
26 proteins and RNA synthesis indicates a role for late endosomes in viral replication. *J.Virol.* 1999;  
27 73(9):7641-7657.
- 28 116. van Dinten LC, Wassenaar AL, Gorbalenya AE, *et al.* Processing of the equine arteritis virus repli-  
29 case ORF1b protein: identification of cleavage products containing the putative viral polymerase  
30 and helicase domains. *J.Virol.* 1996;70(10):6625-6633.
- 31 117. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J.Gen.Virol.*  
32 2006;87(Pt 6):1403-1421.
- 33 118. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J.Virol.*  
34 2007;81(1):20-29.
- 35  
36  
37  
38  
39

- 1 119. Cowley JA, Dimmock CM, Walker PJ. Gill-associated nidovirus of *Penaeus monodon* prawns transcribes 3'-coterminal subgenomic mRNAs that do not possess 5'-leader sequences. *J.Gen.Virol.* 2002;83(Pt 4):927-935.
- 2
- 3
- 4 120. Zirkel F, Roth H, Kurth A, *et al.* Identification and characterization of genetically divergent members of the newly established family *Mesoniviridae*. *J.Virol.* 2013;87(11):6346-6358.
- 5
- 6 121. Enjuanes L, Almazan F, Sola I, *et al.* Biochemical aspects of coronavirus replication and virus-host interaction. *Annu.Rev.Microbiol.* 2006;60:211-230.
- 7
- 8
- 9 122. Sola I, Mateos-Gomez PA, Almazan F, *et al.* RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA.Biol.* 2011;8(2):237-248.
- 10
- 11 123. Ciampi MS. Rho-dependent terminators and transcription termination. *Microbiology* 2006;152(Pt 9):2515-2528.
- 12
- 13 124. Wu HY, Brian DA. Subgenomic messenger RNA amplification in coronaviruses. *Proc.Natl.Acad. Sci.U.S.A* 2010;107(27):12257-12262.
- 14
- 15 125. Sawicki D, Wang T, Sawicki S. The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. *J.Gen.Virol.* 2001;82(Pt 2):385-396.
- 16
- 17 126. Wu CH, Chen PJ, Yeh SH. Nucleocapsid Phosphorylation and RNA Helicase DDX1 Recruitment Enables Coronavirus Transition from Discontinuous to Continuous Transcription. *Cell Host.Microbe* 2014;16(4):462-472.
- 18
- 19
- 20 127. van Marle G, van Dinten LC, Spaan WJ, *et al.* Characterization of an equine arteritis virus replicase mutant defective in subgenomic mRNA synthesis. *J.Virol.* 1999;73(7):5274-5281.
- 21
- 22
- 23 128. Nedialkova DD, Gorbalenya AE, Snijder EJ. Arterivirus Nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs. *PLoS.Pathog.* 2010; 6(2):e1000772.
- 24
- 25
- 26 129. Posthuma CC, Nedialkova DD, Zevenhoven-Dobbe JC, *et al.* Site-directed mutagenesis of the Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life cycle. *J.Virol.* 2006;80(4):1653-1661.
- 27
- 28
- 29 130. van Dinten LC, Rensen S, Gorbalenya AE, *et al.* Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication. *J.Virol.* 1999;73(3):2027-2037.
- 30
- 31
- 32 131. Jones DM, Atoom AM, Zhang X, *et al.* A genetic interaction between the core and NS3 proteins of hepatitis C virus is essential for production of infectious virus. *J.Virol.* 2011;85(23):12351-12361.
- 33
- 34
- 35 132. Kummerer BM, Rice CM. Mutations in the yellow fever virus nonstructural protein NS2A selectively block production of infectious particles. *J.Virol.* 2002;76(10):4773-4784.
- 36
- 37
- 38
- 39

- 1 133. Liu WJ, Sedlak PL, Kondratieva N, *et al.* Complementation analysis of the flavivirus Kunjin NS3 and  
2 NS5 proteins defines the minimal regions essential for formation of a replication complex and  
3 shows a requirement of NS3 in cis for virus assembly. *J.Virol.* 2002;76(21):10766-10775.
- 4 134. Ma Y, Yates J, Liang Y, *et al.* NS3 helicase domains involved in infectious intracellular hepatitis C  
5 virus particle assembly. *J.Virol.* 2008;82(15):7624-7639.
- 6 135. Bartelma G, Padmanabhan R. Expression, purification, and characterization of the RNA 5'-triphos-  
7 phatase activity of dengue virus type 2 nonstructural protein 3. *Virology* 2002;299(1):122-132.
- 8 136. Wengler G, Wengler G. The carboxy-terminal part of the NS 3 protein of the West Nile flavivirus  
9 can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated  
10 NTPase. *Virology* 1991;184(2):707-715.
- 11 137. Subissi L, Imbert I, Ferron F, *et al.* SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replica-  
12 tive enzymes as antiviral targets. *Antiviral Res.* 2014;101:122-130.
- 13 138. Nga PT, Parquet MC, Lauber C, *et al.* Discovery of the first insect nidovirus, a missing evolutionary  
14 link in the emergence of the largest RNA virus genomes. *PLoS.Pathog.* 2011;7(9):e1002215.
- 15 139. Fiorini F, Boudvillain M, Le HH. Tight intramolecular regulation of the human Upf1 helicase by its  
16 N- and C-terminal domains. *Nucleic Acids Res.* 2013;41(4):2404-2415.
- 17 140. Belew AT, Meskauskas A, Musalgaonkar S, *et al.* Ribosomal frameshifting in the CCR5 mRNA is  
18 regulated by miRNAs and the NMD pathway. *Nature* 2014;512(7514):265-269.
- 19 141. Boussau B, Blanquart S, Necsulea A, *et al.* Parallel adaptations to high temperatures in the Ar-  
20 chaeon eon. *Nature* 2008;456(7224):942-945.
- 21 142. Kwong AD, Rao BG, Jeang KT. Viral and cellular RNA helicases as antiviral targets. *Nat.Rev.Drug*  
22 *Discov.* 2005;4(10):845-853.
- 23 143. Shadrick WR, Ndjomou J, Kolli R, *et al.* Discovering new medicines targeting helicases: challenges  
24 and recent progress. *J.Biomol.Screen.* 2013;18(7):761-781.
- 25 144. Kim MK, Yu MS, Park HR, *et al.* 2,6-Bis-arylmethoxy-5-hydroxychromones with antiviral activity  
26 against both hepatitis C virus (HCV) and SARS-associated coronavirus (SCV). *Eur.J.Med.Chem.*  
27 2011;46(11):5698-5704.
- 28 145. Tanner JA, Zheng BJ, Zhou J, *et al.* The adamantane-derived bananins are potent inhibitors of the  
29 helicase activities and replication of SARS coronavirus. *Chem.Biol.* 2005;12(3):303-311.
- 30 146. Adedeji AO, Singh K, Calcatera NE, *et al.* Severe acute respiratory syndrome coronavirus replica-  
31 tion inhibitor that interferes with the nucleic acid unwinding of the viral helicase. *Antimicrob.*  
32 *Agents Chemother.* 2012;56(9):4718-4728.
- 33 147. Adedeji AO, Singh K, Kassim A, *et al.* Evaluation of SSYA10-001 as a Replication Inhibitor of SARS,  
34 MHV and MERS Coronaviruses. *Antimicrob.Agents Chemother.* 2014;58(8):4894-4898.
- 35  
36  
37  
38  
39



- 1 148. Yang N, Tanner JA, Wang Z, *et al.* Inhibition of SARS coronavirus helicase by bismuth complexes.  
2 Chem.Commun.(Camb.) 2007;4413-4415.
- 3 149. Yang N, Tanner JA, Zheng BJ, *et al.* Bismuth complexes inhibit the SARS coronavirus. *Angew.*  
4 *Chem.Int.Ed Engl.* 2007;46(34):6464-6468.
- 5 150. Knoops K, Kikkert M, Worm SH, *et al.* SARS-coronavirus replication is supported by a reticulove-  
6 sicular network of modified endoplasmic reticulum. *PLoS.Biol.* 2008;6(9):e226.
- 7  
8 151. Subissi L, Posthuma CC, Collet A, *et al.* One severe acute respiratory syndrome coronavirus pro-  
9 tein complex integrates processive RNA polymerase and exonuclease activities. *Proc.Natl.Acad.*  
10 *Sci.U.S.A* 2014;111(37):e3900
- 11 152. Sawicki SG, Sawicki DL, Younker D, *et al.* Functional and genetic analysis of coronavirus replicase-  
12 transcriptase proteins. *PLoS.Pathog.* 2005;1(4):e39.
- 13 153. Cordin O, Beggs JD. RNA helicases in splicing. *RNA.Biol.* 2013;10(1):83-95.
- 14  
15 154. Eisen A, Lucchesi JC. Unraveling the role of helicases in transcription. *Bioessays* 1998;20(8):634-  
16 641.
- 17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

Arterivirus RNA-dependent RNA  
polymerase: vital enzymatic activity  
remains elusive

# CHAPTER 4

Kathleen C. Lehmann  
Alexander E. Gorbalenya  
Eric J. Snijder  
and Clara C. Posthuma

**1 ABSTRACT**

2  
3 Polynucleotide polymerases are the central enzymes involved in nucleic acid-based func-  
4 tions of all organisms and viruses. Reflecting this importance, a detailed understanding  
5 of their activities is crucial for deciphering biologically important processes like genome  
6 replication, transcription, and repair. All plus-stranded RNA viruses encode a conserved  
7 RNA-dependent RNA polymerase (RdRp), which was extensively characterized only in  
8 viruses of few families. In the order *Nidovirales*, which includes viruses with (very) large  
9 genomes, the RdRp is expressed in association with other replicative enzymes as part  
10 of the polyprotein encoded in open reading frame 1b (ORF1b). Based on sequence  
11 conservation, it was mapped to the C-terminal domain of nonstructural protein (nsp) 9  
12 in arteriviruses and nsp12 in coronaviruses, the two families of mammalian nidoviruses.  
13 Potent primer-dependent RdRp activity was demonstrated for the severe acute respira-  
14 tory syndrome coronavirus enzyme. In contrast, the only study focusing on nsp9 of the  
15 arterivirus equine arteritis virus (EAV) reported *de novo* polymerase activity on certain  
16 homopolymeric RNA templates in biochemical assays. However, this activity was not  
17 maintained when  $Mn^{2+}$  ions, which are known to relieve the sequence dependency of  
18 polymerases, were omitted or when biologically more relevant templates representing  
19 viral sequences were supplied. Due to these observations, we sought to revisit the  
20 biochemical properties of this polymerase. We describe here the results of a carefully  
21 controlled study involving several preparations of purified recombinant EAV nsp9 that  
22 included the wild-type and a set of active site mutants, which were tested for *de novo*  
23 and primer-dependent polymerase and terminal transferase activities. However, we were  
24 unable to reproduce the published EAV nsp9 activity as the RdRp domain of nsp9 was  
25 found not to be associated with any of the activities observed in these assays. Also we  
26 noticed a striking resemblance between the product profiles of one of the tested prepa-  
27 rations of nsp9 and that of T7 phage RNA polymerase. Our results hence emphasize the  
28 need to employ diverse controls when utilizing highly sensitive biochemical assays.

29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 1 INTRODUCTION

2  
3 Polymerases, which catalyze the templated synthesis of polynucleotides in the 5'-3'  
4 direction, are enzymes encoded by all organisms and RNA viruses, as well as some DNA  
5 viruses. Reflecting the principal differentiation into DNA- and RNA-based processes and  
6 functions, those enzymes can be grouped into four classes each possessing a distinct  
7 combination of specificities for their substrate (NTPs or dNTPs) and template (RNA or  
8 DNA) under physiological conditions. Despite these fundamental differences regarding  
9 the requirements for their substrates, many polymerases of the four classes, including  
10 all characterized RNA-dependent RNA polymerases (RdRps), employ the same catalytic  
11 mechanism and a similar three-dimensional fold resembling the shape of a right hand  
12 with finger, thumb, and palm domains (1;2). At the sequence level, these polymerases  
13 share two sequence motifs, motifs A and C, found in the most conserved palm domain  
14 (3). Few conserved residues, primarily aspartates, located in these motifs are implicated  
15 in (d)NTP binding and/or catalysis (4;5), and consequently their replacement should  
16 abolish or at least severely decrease nucleic acid synthesis (2).

17  
18 Based on their requirements for initiation of nucleic acid synthesis, two types of poly-  
19 merases are recognized: primer-dependent and *de novo*-initiating enzymes (2;6). The  
20 latter, to our knowledge exclusively RNA polymerases (DNA- or RNA-dependent), are  
21 capable of positioning two NTPs, typically two purines, in a manner that allows the  
22 formation of a starting dinucleotide. In contrast, primer-dependent polymerases are un-  
23 able to accommodate the required stable association between the first (d)NTP and the  
24 template. As a result the formation of the first dinucleotide is an energetically extremely  
25 unfavorable event in these proteins. To overcome this problem, short RNA primers must  
26 be produced and placed on the template. For this purpose, organisms and viruses have  
27 evolved different initiation mechanisms that are all assisted by additional proteins or  
28 domains. They may involve the synthesis of short RNA fragments (by e.g., eukaryotic  
29 DNA primase (7)), the formation of covalent RNA-protein complexes (e.g., picornavirus  
30 VPg-RNA complexes (8)), or the utilization of tRNAs (by lentivirus tRNA-binding domains  
31 (9)) or 5' fragments of cellular mRNAs (generated by influenza virus, bunyavirus, and  
32 arenavirus endoribonuclease and cap-binding domains (10-12)).

33  
34 For genome replication many viruses rely on a polymerase that is encoded within  
35 their genome. In viruses of the order *Nidovirales* (comprising the families *Arteriviridae*,  
36 *Coronaviridae*, *Mesoniviridae*, and *Roniviridae*), which are characterized by their large to  
37 exceptionally large single-stranded RNA genomes (13;14), a canonical RdRp possessing  
38 common motifs of other polymerases with right-hand structure is expressed from ORF1b  
39 as part of the pp1ab replicase polyprotein (15-17). After proteolytic cleavage, a protein

1 subunit (nonstructural protein (nsp) 9 in *Arteriviridae*, nsp12 in *Coronaviridae*) harbor-  
2 ing conserved motifs of an RdRp in its C-terminal two-thirds is released (14;15;18;19).  
3 Eventually, this cleavage product becomes a key subunit of the membrane-associated  
4 multi-subunit replication-transcription complex (RTC) that mediates the synthesis of  
5 diverse viral RNAs (20-22). This complex has been characterized *in situ* and through  
6 reconstitution of its activities *in vitro*. In one of these studies coronaviruses, prototyped  
7 by severe acute respiratory syndrome coronavirus (SARS-CoV), were proposed to ex-  
8 press a second, non-canonical RNA polymerase subunit: the ORF1a-encoded nsp8 (23).  
9 In agreement with early studies describing nsp8 as an obligatory *de novo* polymerase  
10 capable of synthesizing products of less than six nucleotides (23) and nsp12 as strictly  
11 primer-dependent (24), it was speculated that the two proteins may work sequentially  
12 on the same template, with nsp8 providing the primers required by the nsp12 “main  
13 RdRp”. Subsequently, also recombinant feline coronavirus nsp8 and human coronavirus  
14 229E nsp7-10 (an nsp8-containing precursor) were reported to be able to synthesize  
15 RNA oligonucleotides with a length of up to six nucleotides. Upon addition of the cog-  
16 nate nsp7, the activity of feline coronavirus nsp8 was further enhanced, generating RNA  
17 products of up to 67 nucleotides (25).

18  
19 However, recent studies question this clear division of labor. First, it was shown that  
20 recombinant nsp8 expressed without any artificial terminal residues also possesses  
21 primer-dependent activity. Furthermore, in complex with its co-factor nsp7, this activity  
22 was estimated to be only 2.5-fold lower in terms of NTP incorporation per active site  
23 than that of nsp12 (26). Additionally, one study also reported *de novo* activity for nsp12  
24 (27). Finally, in the most recent study, SARS-CoV nsp12 showed non-processive primer  
25 extension activity in an *in vitro* assay, which was substantially enhanced by the addition  
26 of nsp7 and nsp8. The same combination of three proteins was also required for *de novo*  
27 initiation of RNA synthesis. A complex of just nsp7 and nsp8, on the other hand, did not  
28 show any activity in this study. Hence it was concluded that the nsp7-nsp8 complex  
29 serves as an activator and processivity factor, rather than primase, for the nsp12 RdRp  
30 (16). The background of the reported differences and apparent contradictions with  
31 respect to the properties of SARS-CoV nsp8 (in complex with nsp7) and nsp12 remain  
32 unknown, but technical differences are likely to play a role, especially concerning the  
33 expression constructs, protein purification, and templates used.

34  
35 Besides SARS-CoV nsp12, RdRp activity was characterized for only one other nidovirus  
36 “main RdRp”, the arterivirus equine arteritis virus (EAV) nsp9 (28). In that study, *de novo*  
37 RdRp activity was reported on poly-uridine (pU) and poly-cytidine (pC) single-stranded  
38 RNAs while no primer extension or terminal transferase activity, that is, the untemplated  
39 elongation of RNA strands, was detected. Thus, it was concluded that EAV nsp9 activity

1 is restricted to *de novo* initiation. However, the applicability of the observed activity to  
2 virus replication remained uncertain since activity on templates containing appropriate  
3 virus-specific sequences could not be detected, and the *in vitro* activity required the pres-  
4 ence of  $Mn^{2+}$ , which is known to relieve template requirements for other polymerases  
5 (29). One possible explanation for the lack of initiation on virus-specific templates could  
6 be that additional co-factors, e.g. higher-order RNA structures or proteins, are needed  
7 for genuine *de novo* initiation *in vivo*. Therefore, the aim of this study was to character-  
8 ize the RNA polymerase activity of EAV nsp9 in more detail. We report the results of a  
9 carefully controlled study involving several preparations of purified recombinant EAV  
10 nsp9 that included the wild-type protein and a set of active-site mutants, which were  
11 tested for *de novo* and primer-dependent polymerase and terminal transferase activi-  
12 ties. However, we were unable to reproduce the published EAV nsp9 activity as the RdRp  
13 domain of nsp9 was found not to be associated with any of the activities observed in  
14 these assays. Also we noticed a striking resemblance between the product profiles of  
15 one of the tested preparations of nsp9 and that of T7 phage RNA polymerase. Our results  
16 hence emphasize the need to employ diverse controls when utilizing highly sensitive  
17 biochemical assays.

## 18 19 20 **RESULTS AND DISCUSSION**

### 21 22 **Expression and purification of EAV nsp9 using two vectors**

23  
24 Previously, the purification and *de novo* polymerase activity of recombinant EAV nsp9  
25 were described (28). In that study the viral protein (subsequently designated as nsp9/  
26 pDEST) was cloned into a pDEST vector including a C-terminal hexahistidine tag and  
27 expressed in *E. coli* BL21 (DE3). As typical for bacterially expressed proteins, an unknown  
28 fraction of nsp9 may contain an N-terminal formylmethionine due to saturation of the  
29 endogenous protein processing pathway by nsp9 overexpression. Such an N-terminal  
30 extension would modify the authentic N-terminus of nsp9, which is expected to be a  
31 glycine residue following the proteolytic release of nsp9 from the pp1ab polyprotein by  
32 nsp4-mediated cleavage of the Glu1677 ↓ Gly1678 site (30). Previously, it was reported  
33 for SARS-CoV nsp8 and nsp12 that artificial tags at the N-terminus may influence RdRp  
34 activity and stability, respectively (24;26).

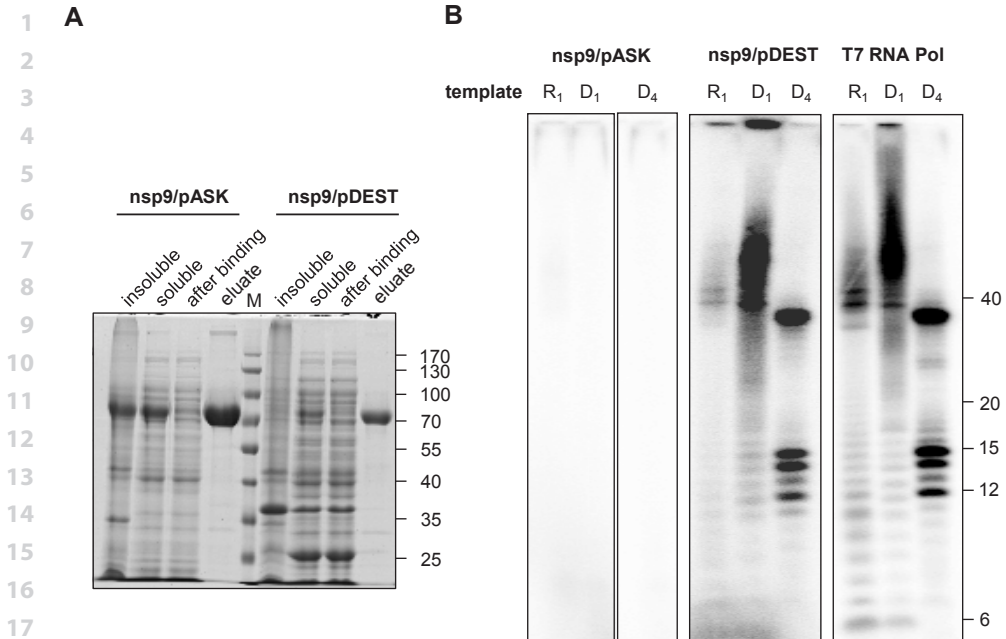
35  
36 To circumvent this potential problem, we decided to express EAV nsp9 as part of a ubiqui-  
37 tin fusion protein by using a so-called pASK vector (31), the resulting protein is hereafter  
38 referred to as nsp9/pASK. In combination with co-expression of the ubiquitin-specific  
39 protease UBP1, which will remove the N-terminal ubiquitin fusion partner *in bacterio*,

1 this enabled us to obtain the natural glycine N-terminus of nsp9 when expressed in  
2 the *E. coli* BL21 derived strain C2523/pCG1. An additional advantage of the pASK vector  
3 was that its backbone allowed us to drive expression via the endogenous pool of *E. coli*  
4 RNA polymerase after induction with anhydrotetracycline. In contrast, nsp9/pDEST was  
5 expressed from a T7 promoter after over-expression of the T7 phage RNA polymerase.  
6 Although this expression system is well characterized and has proven suitable for a  
7 wide range of proteins, the potential presence of this phage RNA polymerase in the  
8 ultimate nsp9 preparations could be of concern. Indeed, since the demonstrated activ-  
9 ity of recombinant EAV nsp9 was shown to be low (28), even trace quantities of this  
10 potent phage polymerase might cause a significant background activity complicating  
11 the interpretation of the obtained results.

12  
13 Both variants of recombinant nsp9 were expressed in their respective *E. coli* strains un-  
14 der identical growth conditions. They were subsequently batch purified in a single step  
15 using metal ion chromatography with  $\text{Co}^{2+}$  targeting the C-terminal hexahistidine tag  
16 of both polypeptides. As Figure 1A shows, both proteins could be obtained with similar  
17 purity, but nsp9/pASK was expressed in higher quantities than nsp9/pDEST. Attempts to  
18 further purify both proteins by gel filtration did not result in a significant improvement  
19 as judged by silver staining of SDS-PAGE gels (not shown).

### 21 **T7 RNA polymerase contamination may account for *de novo* activity observed** 22 **with EAV nsp9 preparations**

23  
24 nsp9/pDEST and nsp9/pASK preparations were tested side-by-side in a *de novo* polymerase  
25 assay in the presence of radioactive ATP using similar reaction conditions as described before  
26 (28). The only noteworthy difference from the published protocol was the length of the pU  
27 template, which was 30 nucleotides in our experiments compared to an undefined mixture  
28 containing RNAs of up to 300 nucleotides in the study of Beerens *et al.*. To our surprise neither  
29 of the preparations showed any activity on this template even when the ATP concentration  
30 was 15-fold increased to 1.5 mM with the goal to favor polymerase initiation (not shown).  
31 Next we tested the RdRp activity using a template whose 3'-terminal dinucleotide matched  
32 the CC dinucleotide that is present immediately upstream of the poly(A) tail at the 3' end of  
33 the EAV genome. Indeed, as previously shown for homopolymeric pC templates, nsp9/pDEST  
34 exhibited some activity with this RNA template, while nsp9/pASK remained essentially inac-  
35 tive (Figure 1B, middle and left panel, respectively, lanes R<sub>1</sub>). As noted earlier, the former and  
36 latter preparations differed in two respects: the presence of an artificial N-terminal residue  
37 in nsp9/pDEST and the induction of T7 RNA polymerase production to achieve expression of  
38 nsp9/pDEST. Only this expression of an additional polymerase can reasonably be linked to  
39 the (gain of) activity in the nsp9/pDEST preparation.



**Figure 1.** Expression, purification, and *de novo* polymerase activity of two recombinant EAV nsp9-His preparations. **(A)** Coomassie brilliant blue-stained SDS-PAGE gel of samples taken during metal ion chromatography using  $\text{Co}^{2+}$ . Insoluble and soluble: respective fractions after cell lysis; after binding: unbound protein after removal of  $\text{Co}^{2+}$  resin; eluate: elution fraction after purification. The molecular weight of nsp9-His is 78 kDa. Size markers are depicted on the right in kDa. **(B)** *de novo* polymerase assay using nsp9 expressed from pASK (final protein concentration 2  $\mu\text{M}$ ) or pDest (final protein concentration 0.6  $\mu\text{M}$ ) vectors, or using commercial T7 RNA polymerase (0.05 U per sample). R and D indicate the use of RNA and DNA templates, respectively. Identical numbers indicate templates with equivalent sequences. Template sequences are listed in Table 2. Product lengths (nt) are indicated on the right. Note that products longer than template length, 30 nt for R<sub>1</sub> and D<sub>1</sub>, result from terminal transferase activity acting on either the template or the newly synthesized strand. D<sub>4</sub> template length 45 nt.

To test this hypothesis, we extended our analysis to include also a highly diluted sample (0.01 U/ $\mu\text{l}$  final concentration) of a commercially available T7 RNA polymerase. Since this enzyme is DNA-dependent, we included two single-stranded DNA templates: a DNA variant (D<sub>1</sub>) of the RNA template used and a DNA template containing the negative-stranded T7 promoter sequence (D<sub>4</sub>). The rationale for the second template was to provide a specific recognition signal in the template for the enzyme and thus increase the chances to observe its activity. Although it has been reported that this DNA-dependent enzyme strictly requires its cognate promoter in a double-stranded form (32), we reasoned that providing DNA with the same polarity as the template that is transcribed under physiological conditions, may at least support some residual activity. Surprisingly, T7 RNA polymerase was active on all of these templates under the employed conditions (Figure 1B, right panel) with the expected preference for DNA templates.



1 Interestingly, the product pattern from the T7 promoter-containing template was  
2 markedly different from the one expected. As already mentioned, *de novo* initiation  
3 on any given template can be forced by increasing the concentration of the required  
4 NTPs. Likewise, decreasing the concentration of one of the NTPs will force a polymerase  
5 to pause and eventually dissociate from the template (or incorporate a non-matching  
6 nucleotide) once it encounters the complementary base. In this manner synthesis by  
7 enzymes with low processivity can be shifted from the production of evenly distributed  
8 but low-intensity products towards a few predominant, high-intensity bands. Thus, lim-  
9 iting the concentration of one nucleotide, in this case CTP, may increase the probability  
10 of detecting polymerase activity if the signal-to-noise ratio is a concern. However, this  
11 was not evident with nsp9/pDEST (Figure 1B, right panel). While the lack of these promi-  
12 nent bands in lanes R<sub>1</sub> and D<sub>1</sub> may be explained by misincorporation of nucleotides,  
13 favored by the high ratio between correct and incorrect NTPs as well as the presence of  
14 error-inducing Mn<sup>2+</sup>, the preference for synthesizing the products of a length of 12, 14,  
15 15, and approximately 38 nucleotides seen in lane D<sub>4</sub> is difficult to reconcile with the  
16 template's sequence. Instead it would be expected that, if at all, synthesis would termi-  
17 nate at positions preceding a G residue in the template (nucleotides 5 (in which case  
18 the product would not be visible), 8, 10, 12, etc.) as incorporation of CTP is unfavorable  
19 under the conditions applied. A possible explanation for the observed product pattern  
20 could be internal initiation on this template lacking a strong promoter sequence. Thus, it  
21 is tempting to speculate that this particular template interacts with T7 RNA polymerase  
22 in a distinctive manner that may not be shared by other polymerases.

23  
24 With that said, it remains to be noted that the nsp9/pDEST preparation showed the same  
25 overall pattern, including the preference for DNA templates, as the commercial T7 RNA  
26 polymerase. In line with this notion, an nsp9/pASK preparation gained *de novo* activity  
27 once it was expressed in BL21 (DE3) under addition of IPTG (not shown). Hence, this  
28 circumstantial evidence suggests that contaminating T7 RNA polymerase, rather than  
29 EAV nsp9 itself, is responsible for the *de novo* polymerase activity observed here.

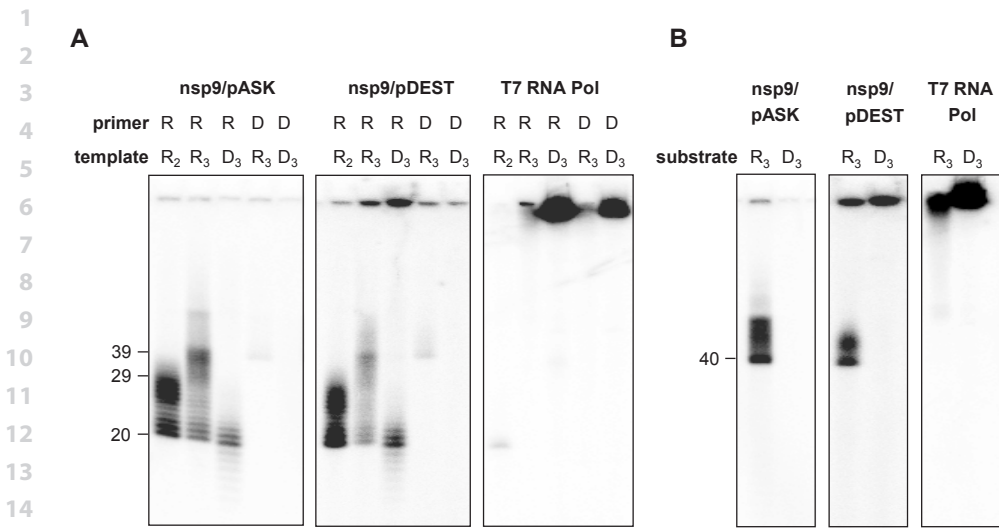
30  
31 Whether or not this contamination was also present in the nsp9 preparations described  
32 in Beerens *et al.* (28), and later on also by te Velhuis *et al.* (33), cannot be established  
33 with certainty as the experiments presented here and those published previously devi-  
34 ated in some aspects. Particularly the previously described purification protocol could  
35 not be reproduced in our experiments due to technical difficulties with the described  
36 purification buffer, which in our hands induced protein precipitation during purification.  
37 Furthermore, as mentioned above, we could not observe an additional purifying effect  
38 of a second chromatography step. Nevertheless, we investigated whether inclusion of  
39 a gel filtration step with a low flow rate (0.3 ml/min) would remove the suspected trace

1 contamination with T7 RNA polymerase (molecular weight 99 kDa) from a preparation of  
2 nsp9/pDEST (molecular weight 78 kDa). We found that this was not the case (not shown).

3  
4 In conclusion, our results revealed that the radioactive polymerase assay used in this  
5 and previous studies is sensitive enough to detect trace activities of contaminating  
6 T7 RNA polymerase and also enables this polymerase to act on templates lacking the  
7 established T7 promoter requirements. Still, the fact that we did not detect any RdRp  
8 activity for nsp9/pDEST, and therefore also not for its suspected contaminant, on a pU  
9 template may be used to argue for the detection of genuine nsp9 activity in the previous  
10 studies. In this context it is noteworthy that the two coronavirus RdRps were addressed  
11 in six independent studies (16;23-27), none of which succeeded in exactly reproducing  
12 results of any other. This may indicate that nidovirus RdRps are highly delicate proteins  
13 responding to minute changes during purification or in their reaction environment.

#### 14 15 **EAV nsp9/pASK preparations possess primer-dependent polymerase and** 16 **terminal transferase activity**

17  
18 Besides *de novo* activity, we decided to test whether EAV nsp9 may possess primer-depend-  
19 ent polymerase activity like its larger coronavirus homolog nsp12 (16;24). To detect this  
20 activity, we used a similar assay as the one described above but this time providing partially  
21 double-stranded templates. We found that both nsp9 preparations were enzymatically ac-  
22 tive on these templates and showed the highest extension activity if the template and  
23 primer were RNAs (Figure 2A, left and middle panel). This differential reaction towards the  
24 type of substrate showed that the measured activity was a direct response to the added  
25 nucleic acids, hence not to a co-purified *E. coli*-derived RNA or DNA template. Furthermore,  
26 as the presence of a DNA template significantly decreased processivity, it also demonstrated  
27 that the responsible polymerase was RNA dependent. Interestingly, while the use of a DNA  
28 primer in combination with an RNA template precluded any extension (no products in  
29 the size range between 20 and 39 nucleotides), a product corresponding to a length of 40  
30 nucleotides was detected. This suggested that the polymerase possesses terminal transfer-  
31 ase activity but only on RNA substrates. To investigate this further, we also compared the  
32 elongation of single-stranded RNA and DNA substrates in an assay otherwise identical to  
33 the one used for measuring primer-dependent polymerase activity (Figure 2B). As expected,  
34 both nsp9 preparations showed a clear selectivity in favor of RNA, again emphasizing their  
35 dependence on this substrate type. In this context it is also noteworthy that neither the  
36 primer extension nor the terminal transferase assay included  $Mn^{2+}$  ions, which can favor ac-  
37 tivity on sub-optimal templates (29). Together with the demonstrated DNA specificity of T7  
38 RNA polymerase (Figures 2A and B, right panels) this supports the reliability of these assays  
39 with respect to the reproduction of physiologically relevant substrate preferences.



15 **Figure 2.** Polymerase assays using recombinant EAV nsp9-His expressed from pASK (final protein concentra-  
 16 tion 1  $\mu$ M) or pDEST (final protein concentration 0.3  $\mu$ M) vectors or using commercial T7 RNA polymerase  
 17 (0.025 U per sample). R and D indicate RNA and DNA strands, respectively. Identical numbers indicate nu-  
 18 cleic acids with equivalent sequences. Nucleic acid sequences are listed in Table 2. Product lengths (nt)  
 19 longer than template length (29 nt for R<sub>2</sub>, 39 nt for R<sub>3</sub> and D<sub>3</sub>) must have resulted from terminal transferase  
 20 activity acting on either the template or the newly synthesized strand. **(B)** Results from terminal transferase  
 21 assay. The signal at the very top of the gel likely represents products of >200nt that cannot be resolved in  
 22 the high-percentage acrylamide gel used here. Note that products resulting from end-labeling with ATP  
 23 may be further extended by a back-priming mechanism.

24  
 25 Finally, to conclude the characterization of the polymerase, its nucleotide preference  
 26 was examined. To this end, a primed RNA template (Figure 3A) was first elongated in  
 27 the presence of a low concentration of radioactive ATP, resulting in frequent abortion  
 28 of transcription after incorporation of the first nucleotide. Subsequently, either dATP  
 29 or ATP was supplied in a concentration that should allow restarting and completion of  
 30 the reaction (Figure 3B). As expected, addition of ATP enabled the synthesis of almost  
 31 fully extended products while dATP did not support any extension beyond one or two  
 32 nucleotides (Figure 3C). In agreement with the lack of DNA primer extension and the  
 33 known inability of the prototype viral RdRp of poliovirus to further extend deoxynucleo-  
 34 tide chains (4;34), we thus conclude that the observed activity originated from an RNA-  
 35 dependent RNA polymerase.

36  
37  
38  
39



1 sion to wild-type virus later in the experiment (by 48 h p.t.; Table 1). In all cases a single  
 2 nucleotide point mutation was sufficient to restore the codon for the wild-type residue.  
 3 Nevertheless, this finding is somewhat unexpected given the universal conservation of  
 4 all four aspartates in positive-stranded RNA viruses. To our knowledge replication, even  
 5 though severely decreased and undetectable until reversion had occurred, of a single  
 6 mutant of the enzyme's active site has not been reported for any other RNA virus thus far.

7  
8 **Table 1.** Summary of reverse genetics data of EAV nsp9 mutants.

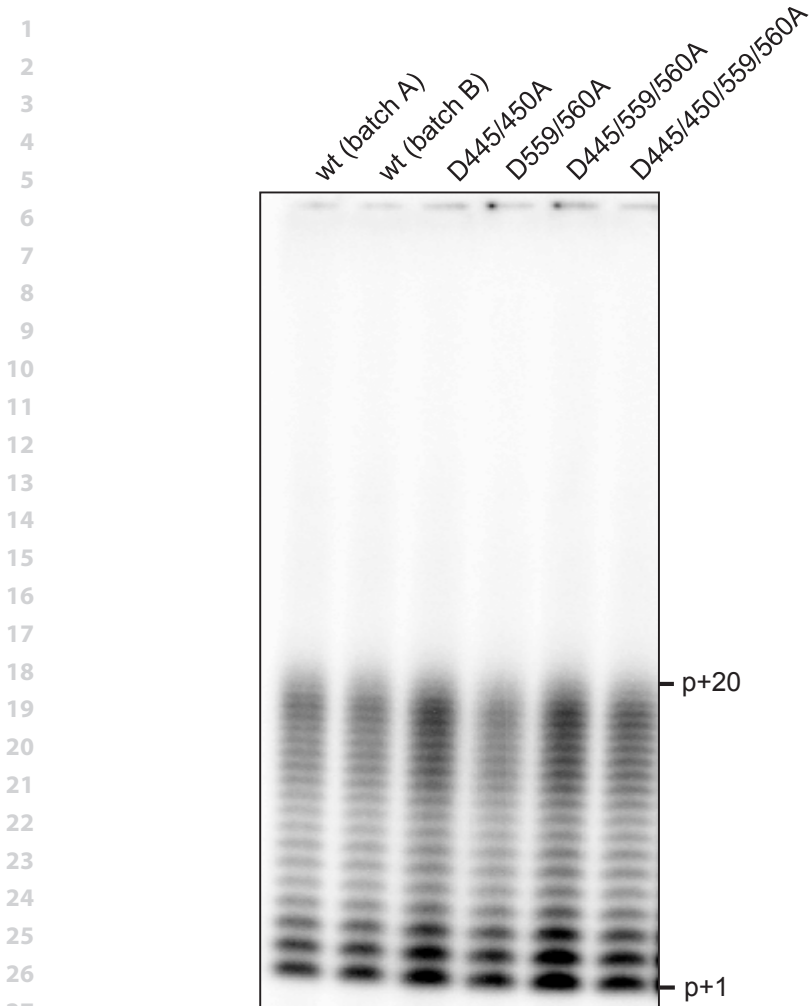
	IFA				plaque phenotype (48 h p.t.)	titer (48 h p.t.) (PFU/ml)	nsp9 sequence of P1*
	16 h p.t.	24 h p.t.	48 h p.t.	72 h p.t.			
wt	++	+++	cells dead	cells dead	large	$2 \cdot 10^7$	n.d.
D445A	-	-	+	+++	large	$6 \cdot 10^5$	wt
D450A	-	-	++	+++	large	$2 \cdot 10^5$	wt
D445/450A	-	-	-	-	-	-	n.d.
D560A	-	-	++	+++	large	$6 \cdot 10^7$	wt
D559/560A	-	-	-	-	-	-	n.d.

17 IFA's were done with antibodies directed against nsp3 and N proteins; -, negative; +, few, separated positive;  
 18 ++, clustered positive; +++, all positive; p.t., post transfection; n.d., not done; \*P1 was generated by infec-  
 19 tion of fresh BHK-21 cells with supernatant harvested at 72 h p.t.

### 22 **Observed primer extension and terminal transferase activities are not** 23 **correlated with EAV nsp9**

24  
 25 Following the results described above, we transferred the same mutations into the nsp9/  
 26 pASK expression construct to obtain negative controls for the biochemical RdRp assays  
 27 described in the previous paragraphs. However, none of the proteins with double, triple,  
 28 and quadruple aspartate-to-alanine substitutions tested showed a decreased primer  
 29 extension activity compared to two independently purified batches of wild-type recom-  
 30 binant nsp9/pASK (Figure 4). Likewise, D445A and D560A mutant proteins maintained  
 31 terminal transferase activity (not shown). Thus, the observed activities either derived  
 32 from a second active site within nsp9, which was not targeted by mutagenesis, or may  
 33 have originated from a different (contaminating) protein altogether. Both these expla-  
 34 nations are quite extraordinary since none of the described RdRps is known to have a  
 35 second active site and no RdRp activity from *E. coli* has been reported to the best of our  
 36 knowledge.

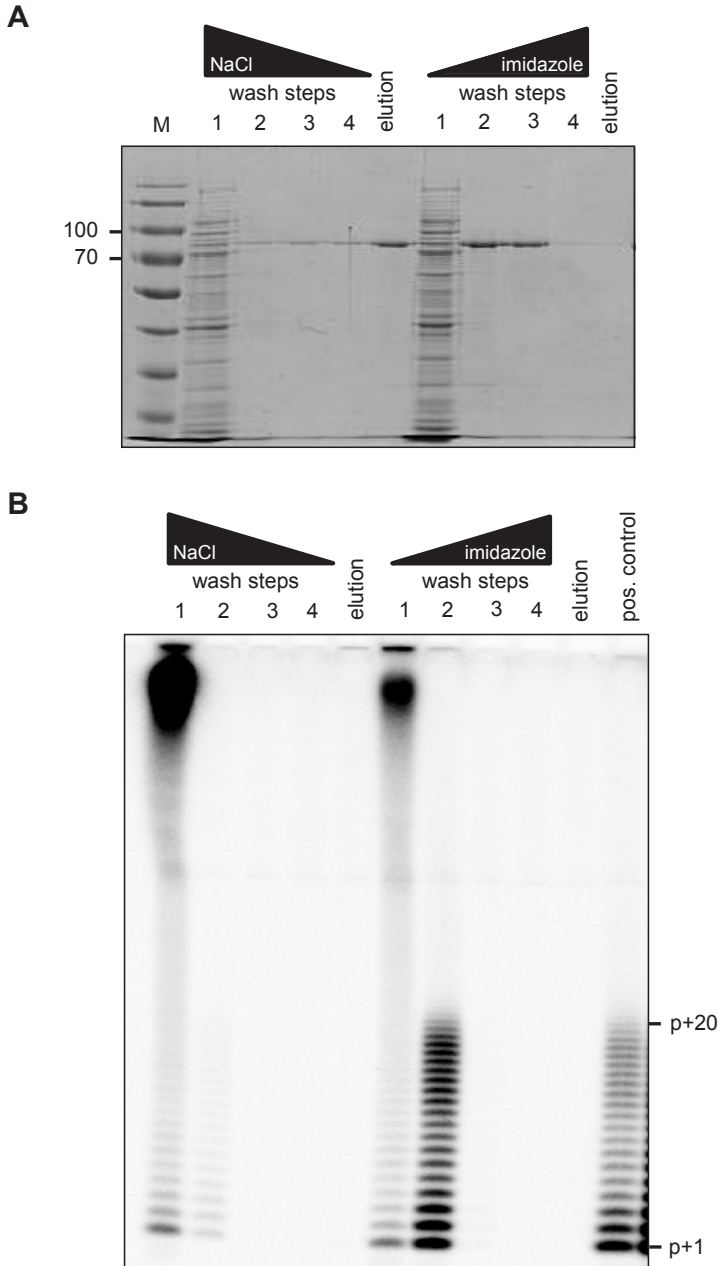
37  
 38 To discriminate between these possibilities, we asked whether it was possible to  
 39 separate nsp9-containing fractions from biochemically active ones during purification



**Figure 4.** Primer extension assay on primer/template R/R<sub>3</sub> using wild-type EAV nsp9-His expressed from a pASK vector and mutants in which essential aspartate residues of the RdRp domain were replaced with alanine (D445 and D450 of motif A, D559 and D560 of motif C). The sizes of primer extension products are indicated on the right.

of the quadruple mutant of nsp9/pASK. To this end, the wash steps of the previously established purification protocol were modified in either of two ways; first, a decreasing salt gradient was introduced to weaken (disrupt) hydrophobic interactions between a contaminant and nsp9, and second, an increasing imidazole gradient was employed in order to eliminate any contaminant from the Co<sup>2+</sup>-resin. As shown in Figure 5, the NaCl elution fraction and wash steps 2 and 3 of the imidazole gradient contained almost identical amounts of nsp9-His, as judged by SDS-PAGE, while two of these three fractions were inactive in the polymerase assay. This partial correlation between the presence of

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39



**Figure 5.** Correlation between EAV nsp9-containing fractions and primer extension activity. **(A)** Coomassie brilliant blue-stained SDS-PAGE gel of samples taken during the purification of nsp9/pASK by  $\text{Co}^{2+}$  affinity chromatography using wash buffers with either a decreasing NaCl concentration or an increasing imidazole concentration. Size markers are depicted on the left in kDa. **(B)** The samples shown in A were examined for primer extension activity on primer/template R/R<sub>3</sub>. The sizes of primer extension products are indicated on the right.

1 recombinant nsp9-His and primer extension activity could be due to either the presence  
2 of two forms of nsp9, enzymatically active and defective, or the presence of a second  
3 enzyme responsible for the activity. Resolving the remaining uncertainty is challeng-  
4 ing since bacteria are, to our knowledge, not known to encode RNA-dependent RNA  
5 polymerases, and the nature and origin of the possible heterogeneity of nsp9 remained  
6 elusive.

7

## 8 **Concluding remarks**

9

10 In a final effort to activate the polymerase activity of recombinant EAV nsp9, we included  
11 several potential protein co-factors in our primer extension assays. For these experi-  
12 ments we chose the poorly characterized arterivirus subunits encoded immediately up-  
13 stream of the ORF1a/1b ribosomal frameshift site, at the genomic position equivalent to  
14 those of the proven coronavirus nsp12-RdRp co-factors nsp7 and nsp8 (16). Although  
15 the proteins of the distantly related corona- and arterivirus families share little similarity,  
16 they might have diverged beyond recognition while retaining similar functions. These  
17 poorly characterized arterivirus proteins include nsp6, a 22-amino acid peptide in EAV,  
18 nsp7 $\alpha$  (123 amino acids in EAV), nsp7 $\beta$  (102 amino acids in EAV), and nsp8 (50 amino  
19 acids in EAV), the subunit located immediately upstream of the frameshift site that  
20 corresponds to the N-terminus of nsp9. Additionally, as these four subunits are known  
21 to be contained in, in part, long lasting cleavage intermediates (nsp6-7 $\alpha$ , nsp6-7, nsp6-  
22 7-8, nsp7, nsp7-8) (35;36) also those were tested. Finally, the EAV helicase nsp10 was  
23 included since its SARS-CoV homolog (nsp13) was shown to interact with its cognate  
24 RdRp nsp12 (37;38). Unfortunately, neither of these subunits had any positive impact  
25 on the polymerase activity of recombinant EAV nsp9/pASK (not shown) or showed  
26 any evidence of interaction with nsp9 in native gel and cross-linking experiments (not  
27 shown). However, we should note that we did not probe this possibility exhaustively  
28 using different experimental conditions to facilitate complex formation or maybe even  
29 co-expression of multiple partners. Hence, there is certainly room to explore the co-  
30 factor hypothesis in more detail.

31

32 To conclude, in this study we could neither confirm the previously reported *de novo* poly-  
33 merase activity nor detect any other RNA polymerase activity originating from purified  
34 recombinant EAV nsp9-His, indicating that the characterization of the arterivirus RdRp  
35 presents a formidable challenge. While the reason(s) underlying the differences to ear-  
36 lier studies remains to be elucidated, the outcome of the present study emphasizes the  
37 need for selecting proper controls especially when utilizing highly sensitive biochemical  
38 assays for characterizing enzymes with low activity. Furthermore, it demonstrates that  
39 *in vitro* assays may reveal activities that are not biologically relevant under physiological



1 conditions and/or in the presence of interaction partners that may alter substrate prefer-  
2 ences by modifying an enzyme's conformation. Being aware of this pitfall probably is  
3 one of the most fundamental prerequisites for the deduction of biological roles from  
4 biochemical assays.

## 5 6 7 **MATERIAL AND METHODS**

### 8 9 **Protein expression and purification**

10  
11 C-terminally His-tagged fusion proteins of wild-type and mutant EAV nsp9 were ex-  
12 pressed under the control of a tetracycline promoter from a pASK vector in the *E. coli*  
13 BL21 derivative C2523/pCG1 as described (26). As a reference, a previously used pDEST  
14 construct of nsp9-His<sub>6</sub> was expressed in *E. coli* BL21 (DE3) cells after IPTG induction under  
15 otherwise identical conditions. Proteins were purified by metal affinity chromatography  
16 using Co<sup>2+</sup> (Talon beads) as described (26) using a buffer containing 20 mM HEPES,  
17 pH 7.5, 500 mM NaCl, 10% glycerol (v/v), 10 mM imidazole, and 5 mM β-mercaptoethanol  
18 unless it is explicitly stated otherwise. Where indicated, a second purification step using  
19 a Superdex 200 10/300 GL gel filtration column with 20 mM HEPES, pH 7.5, 300 mM NaCl,  
20 1 mM DTT was performed at 4°C using a flow rate of 0.3 ml/min.

### 21 22 **Polymerase assays**

23  
24 Three different types of polymerase assays were performed: *de novo*, primer exten-  
25 sion, and terminal transferase assays. For *de novo* assays samples contained 10 mM  
26 Tris, pH 8.0, 5 mM KCl, 25 mM NaCl (including 20 mM from the protein storage buffer),  
27 6 mM MgCl<sub>2</sub>, 1.5 mM MnCl<sub>2</sub>, 1.5 mM DTT, 12.5% glycerol (including 10% from the protein  
28 storage buffer), 0.005% Triton X-100, 1.5 U RiboLock RNase inhibitor (Thermo Scientific),  
29 0.5 μM single-stranded nucleic acid template, 1.5 mM ATP, if required 0.7 mM GTP and  
30 0.7 mM UTP, 0.17 μM [α-<sup>32</sup>P]CTP (Perkin Elmer, 3000 Ci/mmol), and 2 μM nsp9/pASK or  
31 0.6 μM nsp9/pDEST or 0.05 U T7 RNA polymerase from a commercial source (Life Tech-  
32 nologies). Primer extension and terminal transferase assays were performed in 20 mM  
33 Tris, pH 8.0, 10 mM KCl, 20 mM NaCl (including 10 mM from the protein storage buffer),  
34 6 mM MgCl<sub>2</sub>, 1 mM DTT, 10% glycerol (including 5% from the protein storage buffer),  
35 0.01% Triton X-100, 0.5 U RiboLock RNase inhibitor, 1 μM partially double-stranded  
36 (primer extension) or single-stranded (terminal transferase) nucleic acid, 50 μM ATP,  
37 0.17 μM [α-<sup>32</sup>P]ATP (Perkin Elmer, 3000 Ci/mmol), and 1 μM nsp9/pASK or 0.3 μM nsp9/  
38 pDEST or 0.025 U T7 RNA polymerase (Life Technologies). Sequences of used nucleic  
39 acids are listed in Table 2.

**Table 2.** Sequences of nucleic acids used for polymerase assays.

Primers	
R	GCUAUGUGAGAUUAAGUUA
D	GCTATGTGAGATTAAGTTA
templates/substrates	
R <sub>1</sub>	UUUUUUUUUUGCCUCGUCGCCGCCACC
R <sub>2</sub> <sup>*</sup>	UUUUUUUUUU <u>UAACUUAUCUCACAUAGC</u>
R <sub>3</sub> <sup>*</sup>	UUUUUUUUUUUUUUUUUUUU <u>UAACUUAUCUCACAUAGC</u>
D <sub>1</sub>	TTTTTTTTTGCCTCGCTGCCGTCGCCACC
D <sub>3</sub> <sup>*</sup>	TTTTTTTTTTTTTTTTTTTTT <u>AACTTAATCTCACATAGC</u>
D <sub>4</sub>	GCTATGTGAGATTAAGTTATCTGAGCCCTATAGTGAGTCGTATTA

\*Sequences complementary to both primers are underlined

Nucleic acids were annealed with complementary primers by heating to 95°C for 2 min, then keeping them at 52°C for 30 min, and finally letting them cool to room temperature in 30 min.

In all three assays, samples were incubated for 1 h at 30°C before the reaction was stopped by addition of an equal volume of formamide gel loading buffer (95% formamide, 18 mM EDTA, 0.025% SDS, xylene cyanol, bromophenol blue) and 2 min denaturing at 95°C. Products were separated by gel electrophoresis in 20% polyacrylamide gels (19:1) containing 7 M urea. Gels were run in 0.5x TBE and subsequently exposed to phosphorimager screens overnight. Screens were scanned on a Typhoon variable mode scanner (GE Healthcare), and band intensities were analyzed with ImageQuant TL software (GE Healthcare).

## Reverse genetics of EAV

Alanine-encoding mutations of codons specifying conserved nsp9 residues were generated using the QuikChange protocol and were introduced into full-length cDNA clone pEAV211 (39) using appropriate shuttle vectors and restriction enzymes. The presence of the mutations was confirmed by sequencing. pEAV211 plasmid DNA was *in vitro* transcribed and full-length RNA was transfected into BHK-21 cells as described previously (40). Transfected cells were monitored until 72 h post transfection (p.t.) by immunofluorescence microscopy using antibodies directed against the nsp3 and N proteins as described (41). To monitor the production of viral progeny, supernatants were harvested at 48 h p.t. and plaque assays were performed as described (40). To verify the presence of the introduced mutations or reversions in viable mutants, fresh BHK-21 cells were

1 infected with supernatants harvested at 72 h p.t., RNA was isolated with TriPure after  
2 18 h, and the nsp9-coding region was amplified by RT-PCR and sequenced.

3

4

## 5 **ACKNOWLEDGEMENTS**

6

7 This work was supported by the European Union Seventh Framework program through  
8 the EUVIRNA project (European Training Network on (+) RNA virus replication and Anti-  
9 viral Drug Development, grant agreement no. 264286). AEG acknowledges support from  
10 the Leiden University Fund and MoBiLe Program. The authors thank Jessika Zevenhoven-  
11 Dobbe and Aartjan te Velthuis for technical assistance and helpful discussions.

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

## 1 REFERENCE LIST

- 2
- 3 1. Iyer LM, Koonin EV, Aravind L. Evolutionary connection between the catalytic subunits of DNA-
- 4 dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of
- 5 RNA polymerases. *BMC.Struct.Biol.* 2003;31.
- 6 2. Ng KK, Arnold JJ, Cameron CE. Structure-function relationships among RNA-dependent RNA poly-
- 7 merases. *Curr.Top.Microbiol.Immunol.* 2008;320:137-156.
- 8 3. Delarue M, Poch O, Tordo N, *et al.* An attempt to unify the structure of polymerases. *Protein Eng* 1990;
- 9 3(6):461-467.
- 10 4. Arnold JJ, Cameron CE. Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic
- 11 analysis of ribonucleotide incorporation in the presence of  $Mg^{2+}$ . *Biochemistry* 2004;43(18):5126-
- 12 5137.
- 13 5. Arnold JJ, Vignuzzi M, Stone JK, *et al.* Remote site control of an active site fidelity checkpoint in a viral
- 14 RNA-dependent RNA polymerase. *J.Biol.Chem.* 2005;280(27):25706-25716.
- 15 6. van Dijk AA, Makeyev EV, Bamford DH. Initiation of viral RNA-dependent RNA polymerization. *J.Gen.*
- 16 *Viro.* 2004;85(Pt 5):1077-1093.
- 17 7. Pellegrini L. The Pol alpha-primase complex. *Subcell.Biochem.* 2012;62:157-169.
- 18 8. Steil BP, Barton DJ. Cis-active RNA elements (CREs) and picornavirus RNA replication. *Virus Res.* 2009;
- 19 139(2):240-252.
- 20 9. Kleiman L. tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB.Life* 2002;53(2):107-
- 21 114.
- 22 10. Boivin S, Cusack S, Ruigrok RW, *et al.* Influenza A virus polymerase: structural insights into replication
- 23 and host adaptation mechanisms. *J.Biol.Chem.* 2010;285(37):28411-28417.
- 24 11. Morin B, Coutard B, Lelke M, *et al.* The N-terminal domain of the arenavirus L protein is an RNA endo-
- 25 nuclease essential in mRNA transcription. *PLoS.Pathog.* 2010;6(9):e1001038.
- 26 12. Reguera J, Weber F, Cusack S. Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-
- 27 like endonuclease domain, essential for viral cap-dependent transcription. *PLoS.Pathog.* 2010;6(9):
- 28 e1001101.
- 29 13. de Groot RJ, Baker SC, Baric R, *et al.* Family *Coronaviridae*. In King AMQ, Adams MJ, Carstens EB *et al.*
- 30 editors, *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*, Amster-
- 31 dam, Elsevier Academic Press, 2012;806-828.
- 32 14. Lauber C, Ziebuhr J, Junglen S, *et al.* *Mesoniviridae*: a proposed new family in the order *Nidovirales*
- 33 formed by a single species of mosquito-borne viruses. *Arch.Viro.* 2012;157(8):1623-1628.
- 34
- 35
- 36
- 37
- 38
- 39

- 1 15. Gorbalenya AE, Koonin EV, Donchenko AP, *et al.* Coronavirus genome: prediction of putative functional  
2 domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic*  
3 *Acids Res.* 1989;17(12):4847-4861.
- 4 16. Subissi L, Posthuma CC, Collet A, *et al.* One severe acute respiratory syndrome coronavirus protein  
5 complex integrates processive RNA polymerase and exonuclease activities. *Proc.Natl.Acad.Sci.U.S.A*  
6 2014;111(37):e3900
- 7 17. Den Boon JA, Snijder EJ, Chirnside ED, *et al.* Equine arteritis virus is not a togavirus but belongs to the  
8 coronaviruslike superfamily. *J.Virol.* 1991;65(6):2910-2920.
- 9 18. van Dinten LC, Wassenaar AL, Gorbalenya AE, *et al.* Processing of the equine arteritis virus replicase  
10 ORF1b protein: identification of cleavage products containing the putative viral polymerase and  
11 helicase domains. *J.Virol.* 1996;70(10):6625-6633.
- 12 19. Ziebuhr J, Siddell SG. Processing of the human coronavirus 229E replicase polyproteins by the virus-  
13 encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to  
14 pp1a and pp1ab. *J.Virol.* 1999;73(1):177-185.
- 15 20. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J.Virol.* 2007;  
16 81(1):20-29.
- 17 21. van Hemert MJ, van den Worm SH, Knoops K, *et al.* SARS-coronavirus replication/transcription com-  
18 plexes are membrane-protected and need a host factor for activity *in vitro*. *PLoS.Pathog.* 2008;4(5):  
19 e1000054.
- 20 22. van Hemert MJ, de Wilde AH, Gorbalenya AE, *et al.* The *in vitro* RNA synthesizing activity of the isolated  
21 arterivirus replication/transcription complex is dependent on a host factor. *J.Biol.Chem.* 2008;283(24):  
22 16525-16536.
- 23 23. Imbert I, Guillemot JC, Bourhis JM, *et al.* A second, non-canonical RNA-dependent RNA polymerase in  
24 SARS coronavirus. *EMBO J.* 2006;25(20):4933-4942.
- 25 24. te Velthuis AJ, Arnold JJ, Cameron CE, *et al.* The RNA polymerase activity of SARS-coronavirus nsp12 is  
26 primer dependent. *Nucleic Acids Res.* 2010;38(1):203-214.
- 27 25. Xiao Y, Ma Q, Restle T, *et al.* Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer  
28 that exhibits primer-independent RNA polymerase activity. *J.Virol.* 2012;86(8):4444-4454.
- 29 26. te Velthuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8 complex is a unique  
30 multimeric RNA polymerase capable of both *de novo* initiation and primer extension. *Nucleic Acids*  
31 *Res.* 2012;40(4):1737-1747.
- 32 27. Ahn DG, Choi JK, Taylor DR, *et al.* Biochemical characterization of a recombinant SARS coronavirus  
33 nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch.Virol.* 2012;  
34 157(11):2095-2104.
- 35  
36  
37  
38  
39

- 1 28. Beerens N, Selisko B, Ricagno S, *et al.* De novo initiation of RNA synthesis by the arterivirus RNA-  
2 dependent RNA polymerase. *J.Virol.* 2007;81(16):8384-8395.
- 3 29. Arnold JJ, Ghosh SK, Cameron CE. Poliovirus RNA-dependent RNA polymerase (3D(pol)). Divalent  
4 cation modulation of primer, template, and nucleotide selection. *J.Biol.Chem.* 1999;274(52):37060-  
5 37069.
- 6 30. Snijder EJ, Wassenaar AL, van Dinten LC, *et al.* The arterivirus nsp4 protease is the prototype of a novel  
7 group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J.Biol.Chem.* 1996;271(9):4864-  
8 4871.
- 9 31. Gohara DW, Ha CS, Kumar S, *et al.* Production of "authentic" poliovirus RNA-dependent RNA poly-  
10 merase (3D(pol)) by ubiquitin-protease-mediated cleavage in *Escherichia coli*. *Protein Expr.Purif.* 1999;  
11 17(1):128-138.
- 12 32. Rong M, He B, McAllister WT, *et al.* Promoter specificity determinants of T7 RNA polymerase. *Proc.Natl.*  
13 *Acad.Sci.U.S.A* 1998;95(2):515-519.
- 14 33. te Velthuis AJ, van den Worm SH, Sims AC, *et al.* Zn(2+) inhibits coronavirus and arterivirus RNA poly-  
15 merase activity *in vitro* and zinc ionophores block the replication of these viruses in cell culture. *PLoS.*  
16 *Pathog.* 2010;6(11):e1001176.
- 17 34. Gohara DW, Crotty S, Arnold JJ, *et al.* Poliovirus RNA-dependent RNA polymerase (3Dpol): structural,  
18 biochemical, and biological analysis of conserved structural motifs A and B. *J.Biol.Chem.* 2000;275(33):  
19 25523-25532.
- 20 35. van Aken D, Zevenhoven-Dobbe J, Gorbalenya AE, *et al.* Proteolytic maturation of replicase polypro-  
21 tein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication and includes  
22 internal cleavage of nsp7. *J.Gen.Virol.* 2006;87(Pt 12):3473-3482.
- 23 36. Wassenaar AL, Spaan WJ, Gorbalenya AE, *et al.* Alternative proteolytic processing of the arterivirus  
24 replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease.  
25 *J.Virol.* 1997;71(12):9313-9322.
- 26 37. Adedeji AO, Marchand B, te Velthuis AJ, *et al.* Mechanism of nucleic acid unwinding by SARS-CoV  
27 helicase. *PLoS.One.* 2012;7(5):e36521.
- 28 38. von Brunn A, Teepe C, Simpson JC, *et al.* Analysis of intraviral protein-protein interactions of the SARS  
29 coronavirus ORF1a. *PLoS.One.* 2007;2(5):e459.
- 30 39. van den Born E, Gulyaev AP, Snijder EJ. Secondary structure and function of the 5'-proximal region of  
31 the equine arteritis virus RNA genome. *RNA.* 2004;10(3):424-437.
- 32 40. Nedialkova DD, Gorbalenya AE, Snijder EJ. Arterivirus Nsp1 modulates the accumulation of minus-  
33 strand templates to control the relative abundance of viral mRNAs. *PLoS.Pathog.* 2010;6(2):e1000772.
- 34 41. van der Meer Y, van TH, Locker JK, *et al.* ORF1a-encoded replicase subunits are involved in the mem-  
35 brane association of the arterivirus replication complex. *J.Virol.* 1998;72(8):6689-6698.
- 36  
37  
38  
39



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

Discovery of an essential  
nucleotidylating activity associated  
with a newly delineated conserved  
domain in the RNA polymerase-  
containing protein of all nidoviruses

## CHAPTER 5

C. Lehmann  
Anastasia Gulyaeva  
Jessika C. Zevenhoven-Dobbe  
George M. C. Janssen  
Mark Ruben  
Hermen S. Overkleeft  
Peter A. van Veelen  
Dmitriy V. Samborskiy  
Alexander A. Kravchenko  
Andrey M. Leontovich  
Igor A. Sidorov  
Eric J. Snijder  
Clara C. Posthuma  
and Alexander E. Gorbalenya

*Submitted*



**1 ABSTRACT**

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

RNA viruses encode an RNA-dependent RNA polymerase (RdRp) that catalyzes the synthesis of their RNA(s). In the case of positive-stranded RNA viruses belonging to the order *Nidovirales*, the RdRp resides in a replicase subunit that is unusually large. Bioinformatics analysis of this nidoviral nonstructural protein has now revealed a signature domain (genetic marker) that is N-terminally adjacent to the RdRp and has no apparent homologs elsewhere. Based on its conservation profile, this domain is proposed to have nucleotidyltransferase activity. Using recombinant nonstructural protein 9 of the arterivirus equine arteritis virus (EAV), we have demonstrated the manganese-dependent covalent binding of guanosine and uridine phosphates to a basic residue in the newly identified domain, most likely an invariant lysine residue. Substitution of this lysine with alanine severely diminished binding. Furthermore, this mutation crippled EAV and prevented the replication of severe acute respiratory syndrome coronavirus (SARS-CoV) in cell culture, indicating that this domain, named **nidovirus RdRp-associated nucleotidyltransferase (NiRAN)**, is essential for nidoviruses. Potential functions supported by NiRAN include nucleic acid ligation, mRNA capping, and protein-primed RNA synthesis.

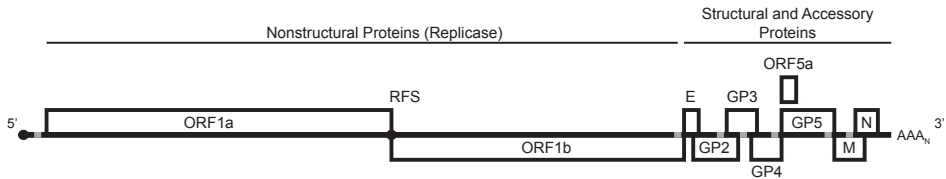
## 1 INTRODUCTION

2  
3 Positive-stranded (+) RNA viruses of the order *Nidovirales* can infect either vertebrate  
4 (families *Arteriviridae* and *Coronaviridae*) or invertebrate hosts (*Mesoniviridae* and  
5 *Roniviridae*) (1;2). Examples of nidoviruses with high economical and societal impact are  
6 the arterivirus porcine reproductive and respiratory syndrome virus (PRRSV) (3) and the  
7 zoonotic coronaviruses (CoVs) causing severe acute respiratory syndrome (SARS) and  
8 Middle East respiratory syndrome (MERS) in humans (4;5). Besides the need to control  
9 these life-threatening diseases, studies of nidoviruses are motivated by the quest to  
10 understand the molecular biology and evolution of the largest RNA genomes known  
11 to-date. Although nidoviruses constitute a monophyletic group, their genome size dif-  
12 ferences are striking, with genomes ranging from 13-16 kb for arteriviruses to 25-34 kb  
13 for roniviruses and coronaviruses. Some major transitions must therefore have occurred  
14 during their evolution, which have been postulated to be reflected in the intermediate  
15 genome size (20-21 kb) of the mesoniviruses. Genome expansion may have proceeded  
16 in a highly ordered but lineage-specific manner that was constrained or promoted by  
17 genome organization, host, and mutation, and was likely facilitated by the acquisition of  
18 enzymes providing quality control mechanisms for newly synthesized RNAs (6).

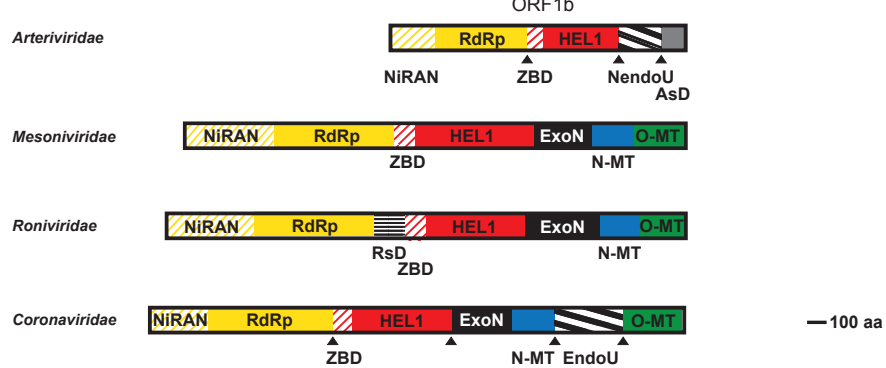
19  
20 Nidoviruses are characterized by their distinct polycistronic genome organization,  
21 the conservation of key replicative enzymes, and a common genome expression and  
22 replication strategy (2). Their distinctive transcription mechanism, which provided the  
23 basis for the name nidoviruses, involves the synthesis of subgenome-length negative-  
24 stranded RNAs that serve as templates for the production of a set of subgenomic (sg)  
25 mRNAs, which are 3' co-terminal with the viral genome and may vary considerably in  
26 number between nidoviruses (7). In most but not all nidoviruses, sg mRNAs and the  
27 genome also share a common 5' leader sequence. It derives from a unique mechanism  
28 of discontinuous negative-strand RNA synthesis that is used to equip the subgenome-  
29 length negative-stranded RNAs with the complement of the genomic leader sequence  
30 (Figure 1A). The synthesis of sg mRNAs (transcription) and genome RNA (replication)  
31 is performed by a poorly characterized replication-transcription complex (RTC) that is  
32 comprised of multiple protein subunits and is associated with virus-induced cytoplasmic  
33 membrane structures (reviewed in (8)). The viral subunits of this complex are encoded  
34 in two large open reading frames (ORFs), ORF1a and ORF1b, that are translated from  
35 the nidoviral genome. Translation starts from a single initiation codon at the 5' end of  
36 ORF1a and proceeds to either the ORF1a or the ORF1b termination codon. In the latter  
37 case, which applies to an estimated 20-40% of the ribosomes, a programmed ribosomal  
38 frameshift occurs in the short ORF1a/ORF1b overlap region. The two polyproteins (pp)  
39 resulting from nidovirus genome translation, pp1a and pp1ab, are auto-catalytically

1 processed by multiple internal proteases, one of which (the 3C-like (3CL<sup>pro</sup>) or main (M<sup>pro</sup>)  
 2 protease) is responsible for the large majority of cleavages. Downstream of ORF1b, nido-  
 3 virus genomes contain multiple smaller ORFs, known as the 3' ORFs, which are expressed  
 4 from the sg mRNAs described above. The ORF1a-ORF1b-3' ORFs array is flanked by 5'- and  
 5 3'-terminal untranslated regions, which account for 5-9% of the nidoviral genome size (6).

## A



## B



24 **Figure 1.** Genome organization and ORF1b-encoded enzymes and domains of nidoviruses. (A) Genome  
 25 organization of *Equine arteritis virus* (EAV) including replicase open reading frames (ORFs) 1a and 1b, and 3'  
 26 ORFs encoding structural proteins. Genomes of other nidoviruses employ similar organizations while they  
 27 may vary in respect to size of different regions and number of 3' ORFs. RFS, ribosomal frameshift site. (B)  
 28 ORF1b size and domain comparison between the four nidovirus families shown for EAV (*Arteriviridae*), *Nam*  
 29 *Dinh virus* (*Mesoniviridae*), *Gill-associated virus* (*Roniviridae*), and *Severe acute respiratory syndrome coro-*  
 30 *navirus* (*Coronaviridae*). NiRAN, nidovirus RdRp-associated nucleotidyltransferase; RdRp, RNA-dependent  
 31 RNA polymerase; ZBD, zinc-binding domain; Hel1, helicase superfamily 1 core domain; NendoU, nidovirus  
 32 uridylyte-specific endoribonuclease; ExoN, exoribonuclease; N-MT, N7-methyltransferase; O-MT, 2'-O-  
 33 methyltransferase; AsD, arterivirus-specific domain; RsD, ronivirus-specific domain. Depicted is a simplified  
 34 domain organization since most enzymes are multidomain proteins. Note that viruses of the *Coronaviridae*  
 35 family that do not belong to the subfamily of *Coronavirinae* encode a truncated version of N-MT. Triangles,  
 36 established cleavage sites by 3CL<sup>pro</sup> in two virus families; ORF1b-encoded proteins of other viruses may be  
 37 proteolytically processed in a similar way.

1 During evolution, most conserved proteins of nidoviruses have accepted substitutions  
2 at a higher frequency per residue than those of organisms of the Tree of Life. In line with  
3 the principal function of each region, genome conservation increases from 3' ORFs to  
4 ORF1a to ORF1b (6). Accordingly, the 3' ORF region encodes virion proteins and, option-  
5 ally, accessory proteins that are predominantly group- or family-specific and mediate  
6 virus-host interactions. ORF1a encodes a variable number of proteins that include co-  
7 factors of the RNA-dependent RNA polymerase (RdRp) and 2'-O-methyltransferase, three  
8 hydrophobic proteins mediating the association of the RTC with membranes, and the  
9 viral proteases (7;9;10). The latter include the 3CL<sup>pro</sup>, which is the only ORF1a-encoded  
10 enzyme conserved in all nidoviruses. In contrast, ORF1b is highly conserved and encodes  
11 different RNA-processing enzymes that critically control viral RNA synthesis (Figure 1B).  
12 These invariantly include the RdRp and a superfamily 1 helicase domain (HEL1), which  
13 is fused with a multinuclear zinc-binding domain (ZBD). Both enzymes are expressed  
14 as part of two different cleavage products residing next to each other in pp1ab (7). The  
15 RdRp is believed to mediate the synthesis of all viral RNA molecules, while over the  
16 years the unwinding activity of the helicase was implicated in the control of replication,  
17 transcription, translation, virion biogenesis, and, most recently, post-transcriptional  
18 RNA quality control (reviewed in (11)). Among the lineage-specific proteins encoded in  
19 ORF1b are four enzymes. A 3'-5' exoribonuclease (ExoN, in *Coronaviridae*, *Mesoniviridae*,  
20 and *Roniviridae*) and an N7-methyltransferase (N-MT, in the *Coronavirinae* subfamily,  
21 *Mesoniviridae*, and *Roniviridae*) constitute adjacent domains in the same pp1b cleavage  
22 product. They were implicated in RNA proofreading (12-14) and in 5' end cap formation  
23 (15;16), respectively. Downstream of this subunit, nidoviruses encode an uridylate-  
24 specific endoribonuclease of unknown function (NendoU, in *Arteriviridae* and *Corona-*  
25 *viridae*) (17;18) and/or a 2'-O-methyltransferase (O-MT, in *Coronaviridae*, *Mesoniviridae*,  
26 and *Roniviridae*), which was implicated in 5' end cap modification and immune evasion  
27 (15;19-21). All six ORF1b-encoded enzymes have distantly related viral and/or cellular  
28 homologs. Additionally, *Roniviridae* and *Arteriviridae* encode family-specific domains of  
29 unknown origin and function, RsD and AsD, respectively. RsD is located between the  
30 subunits containing the RdRp and ZBD-HEL1 domains (22), respectively, while AsD is the  
31 most C-terminal subunit of the arteriviral pp1ab (23).

32

33 The protein subunit containing the RdRp domain is known as nonstructural protein  
34 (nsp) 9 in the *Arteriviridae* and nsp12 in the *Coronaviridae* (7). Its major ORF1b-encoded  
35 part (~95% of its full size in all nidoviruses excluding mammalian toroviruses) varies  
36 in size from ~700 to ~900 amino acid residues and is N-terminally extended by a por-  
37 tion encoded in ORF1a, which can be as few as five residues long. The borders of the  
38 corresponding RdRp-containing proteins of the *Mesoniviridae* and *Roniviridae* have not  
39 been computationally or experimentally identified, but based on our bioinformatics

1 analyses ((2;22) and also see below) these proteins are unlikely to be smaller than those  
2 of arteriviruses. The RdRp-containing replicase subunit of nidoviruses thus seems to be  
3 larger than the characterized RdRps of other RNA viruses, which commonly comprise  
4 less than 500 amino acid residues (24;25).

5  
6 RdRps are known to adopt variations of an  $\alpha/\beta$  fold that is often described as a cupped  
7 right hand, with the palm domain being most conserved and accommodating structural  
8 elements of the active site while the less conserved fingers and thumb play an assisting  
9 role (reviewed in (26;27)). Since the fingers vary in size between known RdRps, nidovi-  
10 ruses – of all or some lineages – might have evolved unusually large fingers that could  
11 account for most of the observed size difference. Alternatively, another domain, either  
12 upstream or inside of the RdRp domain, might have been acquired.

13  
14 Prior bioinformatics analyses mapped conserved sequences (motifs), which are known  
15 to be predominantly associated with the palm domain, to the C-terminal one-third of the  
16 nidovirus RdRp-containing protein. Accordingly, the C-terminal two-thirds of SARS-CoV  
17 nsp12 were sufficient to generate three-dimensional RdRp models using as a template  
18 the RdRp structures of either rabbit hemorrhagic disease virus or a combination of those  
19 of hepatitis C virus, poliovirus, rabbit hemorrhagic disease virus, reovirus, phage  $\Phi 6$ , and  
20 human immunodeficiency virus1 (28;29).

21  
22 With one notable exception (N-MT) (16), all ORF1b-encoded enzymes were initially  
23 identified by comparative genomic analysis involving viral and cellular proteins (23;30).  
24 These assignments were fully corroborated by the subsequent biochemical character-  
25 ization of these enzymes (17;18;21;31-36). Furthermore, the (in)tolerance to replace-  
26 ment of active site residues as tested in reverse genetics studies of coronaviruses and  
27 arteriviruses in general correlated well with the observed enzyme conservation at the  
28 scale of nidovirus diversity. Accordingly, the replacement of conserved residues of the  
29 nidovirus-wide conserved RdRp, ZBD, and HEL1 were lethal for the viruses tested (37-39)  
30 while viruses were crippled upon inactivation of ExoN, NendoU, or O-MT enzymes (40-  
31 42), which are conserved in only some of the nidovirus families (22). This correlation is  
32 noteworthy since it coherently links the results of the experimental characterization of  
33 a few nidoviruses in cell culture systems to evolutionary patterns that were shaped by  
34 natural selection in many hosts over an extremely large time frame. The fact that this  
35 correlation is evident for nidoviruses overall, rather than for separate families, indicates  
36 that nidovirus-wide comparative genomics provides sensible models to the functional  
37 characterization of the most conserved replicative proteins in experimental settings *in*  
38 *vitro* and *in vivo*.

1 In the present study, we aimed to elucidate the domain organization, origin, and func-  
2 tion of the RdRp-containing proteins of nidoviruses by integrating bioinformatics,  
3 biochemistry, and reverse genetics in a manner that was validated in many prior studies.  
4 Our extensive bioinformatics analysis revealed a novel domain, encoded upstream of  
5 the RdRp domain but within the same (predicted) polyprotein cleavage product, which  
6 is conserved in all nidoviruses and has no apparent viral or cellular homologs, making it  
7 a second genetic marker for the order *Nidovirales*. Based on a conservation pattern in-  
8 volving lysine, arginine, glutamate, and aspartate residues, this domain was proposed to  
9 have nucleotidyltransferase activity. Subsequently, using recombinant nsp9 of the prototypic  
10 arterivirus equine arteritis virus (EAV), the covalent binding of guanosine and uridine  
11 phosphates was demonstrated, which was found to be extremely sensitive to replace-  
12 ment of conserved residues. The replication of both EAV and SARS-CoV was found to be  
13 severely affected by substitution of these conserved residues. Amongst those was also an  
14 invariant lysine residue that presumably binds the nucleoside phosphate. Accordingly,  
15 the domain was named **nidovirus RdRp-associated nucleotidyltransferase (NiRAN)**. We  
16 discuss the potential functions in nidovirus replication in which this essential NiRAN  
17 activity may be involved, which include RNA ligation, protein-primed RNA synthesis, and  
18 the guanylyltransferase function that is necessary for mRNA capping.

## 19 20 21 **RESULTS**

### 22 23 **Delineation of a novel, unique domain that is conserved immediately upstream** 24 **of the RdRp in polyproteins of all nidoviruses**

25  
26 To shed light on the cause of the large size of nidoviral RdRp-containing proteins, we  
27 have conducted several bioinformatics analyses of their sequences (see Materials and  
28 Methods for technical details). We have produced family-wide multiple sequence align-  
29 ments (MSAs) of nsp12 of coronaviruses, nsp9 of arteriviruses, and their counterparts of  
30 mesoniviruses and roniviruses, whose borders have been tentatively mapped through  
31 limited similarity with known 3CL<sup>pro</sup> cleavage sites of these viruses (43;44) (Figure S1).  
32 For simplicity, we will refer to the proteins of mesoni- and roniviruses as nsp12t, with “t”  
33 standing for tentative. The final subsets include 35, 10, 6, and 2 sequences representing  
34 all established and putative taxa of corona-, arteri-, mesoni-, and roniviruses, respec-  
35 tively. To scan different databases, MSAs were split into the N-terminal and C-terminal  
36 parts, which were converted into Hidden Markov Model (HMM) profiles to conduct  
37 profile-sequence (HMMER 3.1) and profile-profile (HH suite 2.0.15) comparisons and  
38 into position-specific scoring matrix (PSSM) profiles for profile-tertiary structure (Gen-  
39 THREADER 8.9) comparisons.

1 In comparisons with the Protein Data Bank (PDB) ([www.rcsb.org](http://www.rcsb.org), (45)) using Gen-  
2 THREADER, RdRps of different viruses dominated the hit list for the best sampled nidovi-  
3 ruses, corona- and arteriviruses, and they were consistently present among the top hits  
4 for the two other families (Table S2). Typically the similarity between a nidovirus query  
5 and a target encompassed the entire target and was limited to the C-terminal part of  
6 the query, with the N-terminal ~250 and 350 amino acid residues remaining unmatched  
7 in arteriviruses and other nidoviruses, respectively (Figures 2A and S2). Likewise, the C-  
8 terminal part of nsp9/nsp12/nsp12t matched the RdRp profiles of different virus families  
9 in PFAM (46) and an in-house database although this analysis was complicated by the  
10 presence of nidovirus sequences in the top-hit PFAM profile (see below). Based on these  
11 results we concluded that nsp9, nsp12, and nsp12t contain N-terminal domains that are  
12 not part of canonical RdRps.

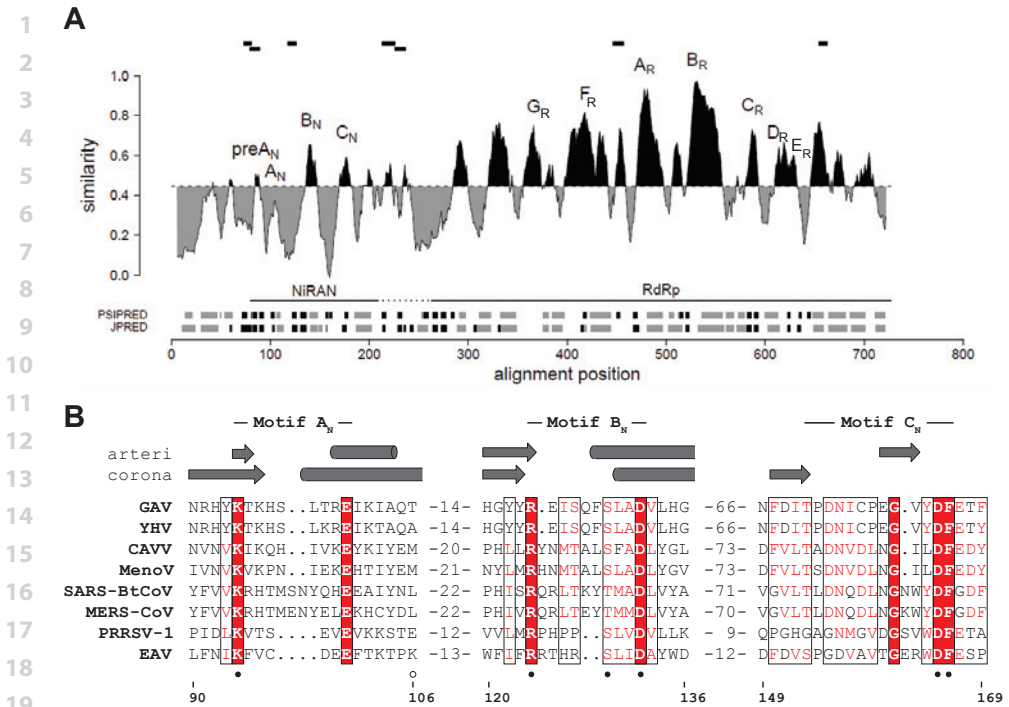
13

14 Inspection of the intra-family sequence conservation for MSAs of nsp9, nsp12, and  
15 nsp12t using a two-dimensional plot (Figure S2) revealed the association of character-  
16 istic RdRp motifs with some of the most prominent conservation peaks, located in the  
17 C-terminal half of nsp9 and nsp12. For nsp12t (Figure S2), similar conclusions could be  
18 drawn although the conservation profiles of these viruses, especially roniviruses, were  
19 of lesser resolution due to the overall higher similarity that was the result of the limited  
20 virus sampling and divergence. Importantly, also the N-terminal half of nsp9 and nsp12  
21 included a few above-average conservation peaks although the overall conservation  
22 was evidently highest around the established RdRp motifs (Figures 2A and S2). We  
23 concluded from this analysis that the N-terminal parts of at least nsp9 and nsp12 share  
24 characteristic conserved motifs (the domain is hereafter referred to as NiRAN, see below).

25

26 To investigate the relation of the NiRAN domains of the four different families, the  
27 HAlign program from the HH-suite software package was used to conduct pair-wise  
28 profile-profile comparisons, which were visualized in dot-plot format (Figure S3). This  
29 analysis revealed strong support (~98% confidence and  $E=7.7e-09-1.7e-08$ ) for the  
30 similarity between NiRANs of coronavirus nsp12 and mesonivirus nsp12t, and moder-  
31 ate support (~21-30% confidence and  $E=0.00091-0.00051$ ) for the similarity between  
32 the respective domains of mesoni- and roniviruses. Based on this observation, we have  
33 aligned the NiRAN domain of coronavirus nsp12 and mesonivirus nsp12t using the  
34 profile mode of ClustalX, with the MSA being slightly adjusted taking into account the  
35 HHsearch-mediated results. This MSA of two families was superior compared to each  
36 of the two family-specific MSAs with respect to its similarity to the MSA of roniviruses  
37 (~54-75% confidence and  $E=0.00049-0.00011$ ). Consequently, the ronivirus MSA was  
38 added to the MSA of corona- and mesoniviruses to generate an MSA of the NiRAN of

39



**Figure 2:** Delineation of the NiRAN domain in RdRp-containing proteins of nidoviruses. **(A)** Sequence variation, domain organization, and secondary structure of the RdRp-containing protein of arteriviruses, and location of peptides identified by mass spectrometry after FSBG-labeling of arterivirus nsp9. Shown is the similarity density plot obtained for the multiple sequence alignment (MSA) of proteins including NiRAN and RdRp domains of arteriviruses. To highlight the regional deviation of conservation from that of the MSA average, areas above and below the mean similarity are shaded in black and gray, respectively. Uncertainty in respect to the domain boundary between NiRAN and RdRp is indicated by a dashed horizontal line. Sequence motifs of NiRAN and RdRp are labeled. Below the similarity density plot, predicted secondary structure elements are presented in gray for  $\alpha$ -helices, black for  $\beta$ -strands. Relative positions of peptides identified by mass spectrometry after FSBG-labeling of arterivirus nsp9 are shown at the top. **(B)** MSA of the three conserved NiRAN motifs of eight representative nidoviruses and their predicted secondary structures. Absolutely conserved residues are highlighted in red boxes. Partially conserved residues are indicated in red font. Secondary structure predictions were made with JPred (91) based on arterivirus (arteri) or coronavirus (corona) MSAs. Residues mutated in recombinant equine arteritis virus (EAV, *Arteriviridae*) nonstructural protein (nsp) 9 are indicated by filled (conserved) and empty (control) circles. Amino acid numbers refer to EAV nsp9. GAV, gill-associated virus (*Roniviridae*); YHV, yellow head virus (*Roniviridae*); CAVV, Cavally virus (*Mesoniviridae*); MenoV, Meno virus (*Mesoniviridae*); SARS-BtCoV, bat severe acute respiratory syndrome coronavirus (*Coronaviridae*); MERS-CoV, Middle East respiratory syndrome coronavirus (*Coronaviridae*); PRRSV-1, porcine reproductive and respiratory syndrome virus EU-type (*Arteriviridae*).

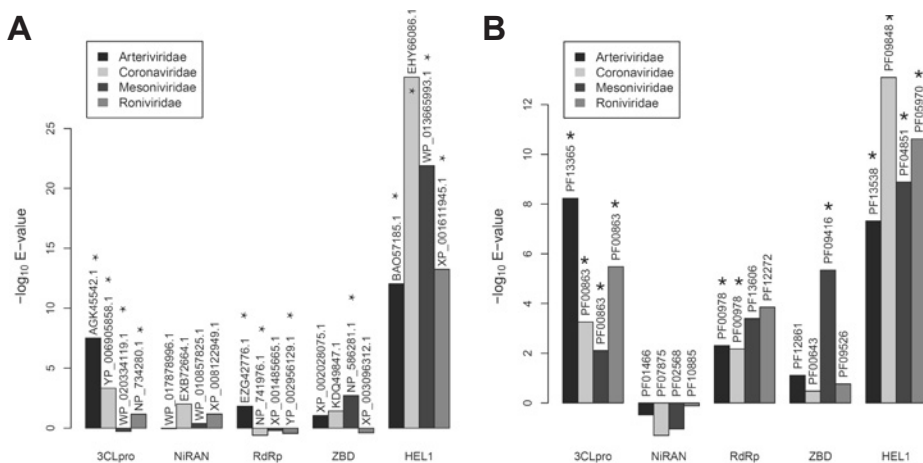


1 the three families, which are hereafter called ExoN-encoding nidoviruses, with reference  
2 to the feature that distinguishes them as a group compared to arteriviruses (Figure 1B).

3  
4 In contrast to the above observations, the support for any similarity between the NiRAN  
5 MSAs of arteriviruses and ExoN-encoding nidoviruses in our HHalign-based analysis was  
6 considered as weak, particularly with respect to confidence ( $E=0.03-0.04$  and  $\sim 1\%$  con-  
7 fidence, when comparing the MSA of arteriviruses versus ExoN-encoding nidoviruses).  
8 This experience prompted us to compare conserved motifs and predicted secondary  
9 structures of the domains of these families (Figures S1 and S2). Ten residues were found  
10 to be invariant in the conserved NiRAN of the ExoN-encoding nidoviruses. They map to  
11 three motifs designated  $A_N$  (with a  $K-x[6-9]-E$  pattern in ExoN-encoding nidoviruses),  $B_N$   
12 ( $R-x[8-9]-D$ ) and  $C_N$  ( $T-x-DN-x4-G-x[2,4]-DF$ ), respectively (Figure 2A), with motifs  $B_N$  and  
13  $C_N$  representing the most prominent conservation peaks of this domain in coronaviruses  
14 (Figure S2). Remarkably, similar conserved motifs are present in the NiRAN of arterivirus-  
15 es (Figure 2A), where  $B_N$  and  $C_N$  again occupy the two most prominent peaks (Figure S2).  
16 The three motifs are similarly positioned relative to the ORF1a/ORF1b frameshift signal  
17 in all nidoviruses, and they were aligned in the HHalign-based analysis discussed above.  
18 Specifically, all four invariant residues of motifs  $A_N$  and  $B_N$  of ExoN-encoding nidoviruses  
19 are also conserved in arteriviruses although with slightly smaller distances separating  
20 the two residues of each pair (Figure S1). In the most highly conserved motif  $C_N$ , the  
21 aspartate-phenylalanine dipeptide and likely glycine (the only deviating arginine at this  
22 position in the lactate dehydrogenase-elevating virus isolate U15146 may result from  
23 a sequencing error) are absolutely conserved among all nidoviruses while the other  
24 invariant residues of ExoN-encoding nidoviruses may be replaced by similar residues in  
25 arteriviruses. Additionally, there is a good agreement between the predicted secondary  
26 structure for the domains of arteriviruses and ExoN-encoding nidoviruses, particularly in  
27 the area encompassing the sequence motifs as well as regions immediately upstream of  
28 motif  $A_N$  (named preA motif) and downstream of motif  $C_N$  (Figure S1). In ExoN-encoding  
29 nidoviruses, motifs  $B_N$  and  $C_N$  are separated by a variable region of 40-60 amino acid  
30 residues that does not include absolutely conserved residues, while in arteriviruses mo-  
31 tifs  $B_N$  and  $C_N$  are adjacent. Also, we noted that the C-terminal border of the N-terminal  
32 conserved domain was close to that identified in the GenTHREADER analysis discussed  
33 above (Figure S2). Based on these observations, we concluded that nsp9, nsp12, and  
34 nsp12t contain the NiRAN domain, which is conserved in all nidoviruses.

35  
36 To gain insight into the origin and function of this domain, MSA-based profiles of this  
37 domain and its individual motifs of different nidovirus families and the entire order  
38 were compared with the PFAM, GenBank, Viral DB, and PDB databases. As a control,  
39 we used the HMM profiles of four other domains that are conserved in all nidoviruses,

1 3CL<sup>pro</sup>, RdRp, ZBD, and HEL1. None of the database scans involving the NiRAN retrieved  
 2 a non-nidovirus hit whose E value was better than 0.065 for HMMER and 1.3 for the  
 3 HHsearch program from HH-suite (Figure 3), and none of these hits had sequences  
 4 similar to the motifs of the NiRAN. In contrast, statistically significant hits with virus  
 5 and/or host proteins were identified for the nidoviral control proteins either in both or  
 6 one of the scans; at least some of these hits were true positives in the functional and/  
 7 or structural dimension as well. Likewise, in scans of the PDB using GenTHREADER, all  
 8 top hits for the NiRAN of the four virus families had low support ( $p=0.014$  or worse)  
 9 with no match of the conserved motifs. In contrast, top hits for four RdRp queries were  
 10 supported with P values of 0.0003 or better and targeted RdRps of other viruses, at least  
 11 for arteri- and coronavirus queries (Table S2). Based on these results and those involving  
 12 the comparison of arteriviruses and ExoN-containing nidoviruses, we concluded that  
 13 the NiRAN domain could have diverged from its homologs in other organisms beyond  
 14 the level of sequence similarity that can be recognized with the available HMM- and  
 15 PSSM-based tools.



**Figure 3:** Comparison of nidovirus-wide conserved domains with sequence databases. Shown are histograms depicting E values of the best non-nidovirus hits obtained during HMMER-mediated profile-sequence (A) and HHsearch-mediated profile-profile (B) searches of the GenBank and PFAM A databases, respectively, using MSA profiles of five nidovirus-wide conserved domains encoded by four nidovirus families. The identity of the non-nidovirus top-hit in the respective databases is specified. Stars indicate hits whose homologous relationship with the respective query is also supported by the functional and/or structural annotation of the respective targets.

## 1 **EAV nsp9 has Mn<sup>2+</sup>-dependent nucleotidylation activity with UTP/GTP** 2 **preference**

3

4 Since we could not identify any homologs of the NiRAN domain whose prior charac-  
5 terization would facilitate the formulation of a hypothesis about its function, we have  
6 reviewed the available information about nidovirus genome organization and the  
7 analyses described above. The data were most compatible with the hypothesis that this  
8 domain is an RNA processing enzyme, in view of i) the abundance of RNA processing  
9 enzymes in the ORF1b-encoded polyprotein (Figure 1B), ii) the predicted  $\alpha/\beta$  structural  
10 organization (Figure S1), and iii) the profile of invariant residues, composed of aspartate,  
11 glutamate, lysine, arginine, and phenylalanine (and possibly glycine) (Figure 2B), the  
12 first four of which are among the most frequently employed catalytic residues (47). We  
13 hypothesized that, because the domain is uniquely conserved in nidoviruses, its activity  
14 might work in concert with that of another, similarly unique RNA processing enzyme.  
15 At the time of this consideration, the NendoU endoribonuclease of nidoviruses was  
16 believed to be such an enzyme (17) (assessment revised in 2011, (22)). Consequently, we  
17 reasoned that a ligase function would be a natural counterpart for the endoribonucle-  
18 ase, as observed in many biological processes, and would fit in the functional coopera-  
19 tion framework outlined in our analysis of the SARS-CoV proteome (30). This hypothesis  
20 was also compatible with the lack of detectable similarity between the NiRAN and the  
21 highly diverse nucleotidyltransferase superfamily, to which nucleic acid ligases belong,  
22 as this superfamily is known to include groups that differ even in the most conserved  
23 sequence motifs, especially in proteins of viral origin (48;49). Based on mechanistic  
24 insights obtained with other ligases, it was expected that the conserved lysine is the  
25 principal catalytic residue of the NiRAN domain.

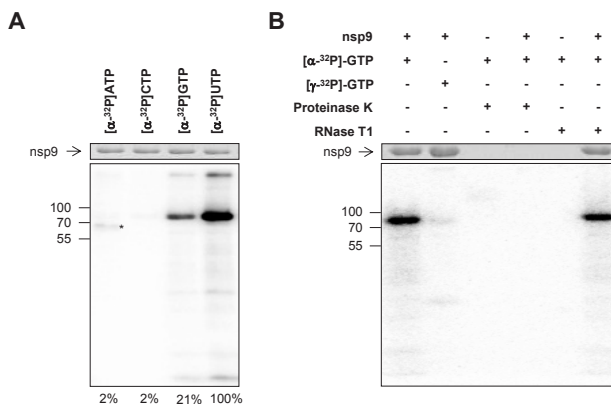
26

27 To detect this putative NTP-dependent RNA ligase activity, we took advantage of the  
28 universal ligase mechanism, which can be separated into three steps (50). First, an NTP  
29 molecule, typically ATP, is bound to the enzyme's binding pocket, and a covalent bond  
30 is established between the nucleotide's  $\alpha$ -phosphate and the side chain of either lysine  
31 or histidine, while pyrophosphate is released. Since this protein-NMP is a true, temporar-  
32 ily stable intermediate, it can be readily detected by biochemical methods. In contrast,  
33 demonstration of the following two steps, NMP transfer to the 5' phosphate of an RNA  
34 substrate and subsequent ligation of a second RNA molecule under release of the NMP,  
35 depends on the availability of target RNA sequences whose identification is often not as  
36 straightforward. Thus, we first assessed our hypothesis by testing the covalent binding of  
37 a nucleotide, known as nucleotidylation.

38

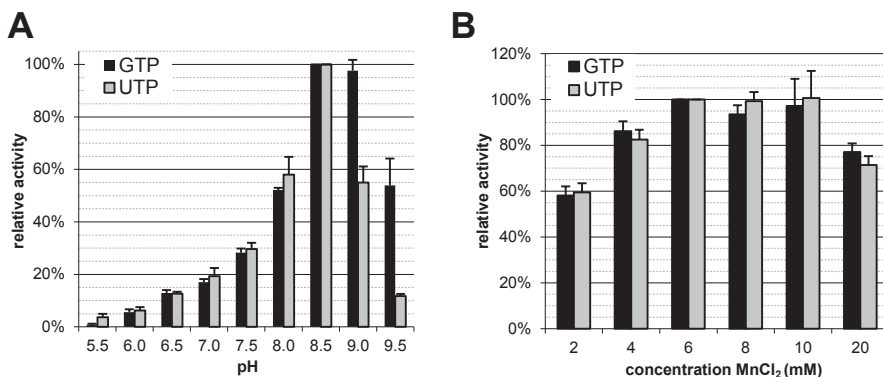
39

1 To this end, recombinant EAV nsp9 was purified and incubated with each of the four NTPs,  
 2 which were  $^{32}\text{P}$ -labeled at the  $\alpha$ -position, and run on denaturing SDS-PAGE gels to discrimi-  
 3 nate between covalent and affinity-based nucleotide binding. As can be seen in Figure 4A,  
 4 we could indeed detect a radioactively labeled product with a mobility comparable to  
 5 that of nsp9 in the presence of GTP and UTP. To verify that this labeled band corresponded  
 6 to a protein and did not result from 3' end labeling of co-purified *E. coli* RNA or polyG  
 7 synthesis by the RNA polymerase residing in the C-terminal domain of nsp9, guanylylation  
 8 was followed by the addition of either proteinase K or RNase T1, which cleaves single-  
 9 stranded RNA after G residues. As expected, only protease treatment removed the band  
 10 while incubation with RNase T1 had no effect on the product (Figure 4B). The same result  
 11 was obtained after uridylylation using RNase A, which cleaves after pyrimidines in single-  
 12 stranded RNA (data not shown). Furthermore, as the use of GTP labeled in the  $\gamma$ -position  
 13 did not result in a radioactive product, we conclude that this phosphate is, in agreement  
 14 with the general nucleotidylation mechanism, released during the reaction (Figure 4B).  
 15 Since these results were compatible with the bioinformatics results described above and  
 16 were corroborated further in experiments described below, the N-terminal domain was  
 17 named **nidovirus RdRp-associated nucleotidyltransferase (NiRAN)**.



25  
26  
27  
28  
29  
30  
31  
32  
33 **Figure 4.** EAV nsp9 has nucleotidylation activity. Purified recombinant EAV nsp9 (78 kDa) was incubated  
 34 with the indicated [ $^{32}\text{P}$ ]NTP in the presence of  $\text{MnCl}_2$ . Reaction products were visualized after denaturing  
 35 SDS-PAGE by Coomassie brilliant blue staining (top panels) and phosphor imaging (bottom panels). Posi-  
 36 tions of molecular weight markers are depicted on the left in kDa. **(A)** Uridylylation and guanylylation activ-  
 37 ity as revealed by covalent binding of the respective radioactive nucleotide to nsp9. Note that the protein  
 38 indicated with an asterisk likely is an *E. coli*-derived impurity reacting with ATP. Relative band intensities  
 39 are shown at the bottom. **(B)** Guanylylation was distinguished from RNA polymerization by incubating the  
 products generated during the nucleotidylation assay with proteinase K (1 mg/ml) or with RNase T1 (0.5 U),  
 which cleaves single-stranded RNA after G residues, for 30 min at 37°C.

1 Unexpectedly, nsp9 showed a marked substrate specificity for UTP, which resulted in the  
 2 accumulation of 5 times more enzyme-nucleotide complex than observed with GTP. In  
 3 contrast, no covalent binding was observed with ATP or CTP as substrates (Figure 4A).  
 4 The observed substrate preferences are remarkable for two reasons. First, since both  
 5 UTP and GTP are present in significantly lower concentrations under physiological con-  
 6 ditions than ATP (51) and are in general not used as primary energy source, it suggests  
 7 that the identity of the base, rather than the energy stored within the phosphodiester  
 8 bonds, may be critical for a subsequent step in the reaction pathway. Obviously, this im-  
 9 plies that the involvement of these transitory covalent complexes in reaction pathways  
 10 other than RNA ligation must be considered. Second, the selective utilization of only one  
 11 pyrimidine and one purine substrate raised questions about the nature and number of  
 12 active sites involved, for instance, whether both nucleotides bind to separate binding  
 13 sites or utilize different catalytic residues within the same binding site. Unfortunately,  
 14 there are no crystal structures for any of the nidovirus nsp9/nsp12/nsp12t subunits  
 15 available to date, which might have been used to resolve this matter in docking studies.  
 16  
 17 Therefore, to address this question indirectly we compared the pH dependence of both  
 18 activities as a signal for structural differences in the immediate environment of the  
 19 catalytic residue. Interestingly, while the relative activities below pH 8.5 were identical  
 20 with both substrates, the relative guanylylation activity was exceedingly higher than  
 21 uridylylation at a pH above 8.5 (Figure 5A). To test whether a difference in the metal ion  
 22 requirement could be the cause for the observed dependence, we determined the opti-  
 23  
 24



25  
 26 **Figure 5.** EAV nsp9 guanylylation has a slightly broader or shifted pH optimum compared to uridylylation  
 27 while the metal ion requirement is identical. **(A)** The pH optimum in the range from 5.5 to 9.5 was deter-  
 28 mined using the buffers listed in Material and Methods. **(B)** Assessment of the optimal MnCl<sub>2</sub> concentration  
 29 for nucleotidylation. Error bars represent the standard deviation of the mean based on three independent  
 30 experiments.  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39

mal manganese concentration for nucleotidylation with both substrates. As is apparent from Figure 5B, both activities share the same broad optimum between 6 and 10 mM  $\text{MnCl}_2$ . This result made it unlikely that manganese oxidation and a concomitant decrease of available  $\text{Mn}^{2+}$  ions, as we observed at a pH above 9.0, would selectively favor the utilization of one of the two substrates. The observed difference between guanylylation and uridylylation with regard to its pH optimum may thus be genuine. For instance, this slightly broadened or – more likely – shifted pH optimum of guanylylation may be the result of a GTP-induced spatial reorientation of amino acid side chains in the vicinity of the catalytic residue and a concomitant alteration of its  $\text{pK}_a$ . Alternatively, it may also be explained by the two substrates using different binding sites. These possibilities were partially addressed in the experiments described in the subsequent sections.

### **FSBG labeling of nsp9 suggests the presence of a nucleotide binding site in the NiRAN domain**

To verify that the newly discovered nucleotidylation activity is associated with the NiRAN domain, we first sought to establish the presence of the expected nucleotide binding site. To this end, we replaced the substrate in the nucleotidylation assay with the reactive guanosine analog 5'-(4-fluorosulfonylbenzoyl)guanosine (FSBG) (Figure S4A) (52). Depending on the exact shape of the nucleotide binding pocket this compound may be suitable for binding and reacting with any nucleophile within the pocket, leaving behind a stable sulfonylbenzoyl tag that can be readily detected by mass spectrometry. In this way, residues that are lining the binding site can be identified. However, because the points of attack of FSBG (sulfonyl group sulfur) and GTP ( $\alpha$ -phosphorus) are spatially separated ( $\sim 4\text{\AA}$ , Figures S4A and B), these residues are not necessarily of biological relevance to nucleotidylation but rather mark the environment of the nucleotidylation.

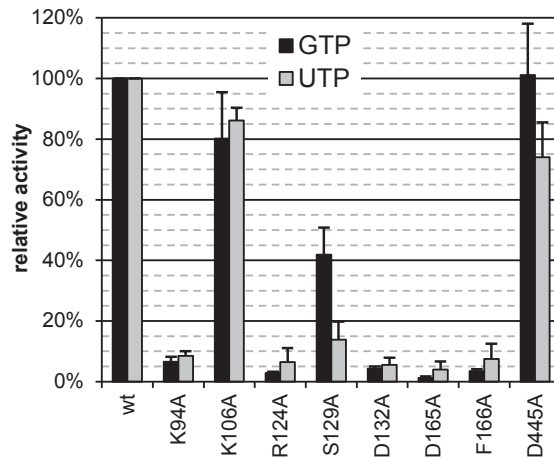
After analysis of the nucleotidylation reaction mixture by mass spectrometry, seven modified peptides representing five distinct nsp9 regions could be assigned: three in (the vicinity of) the NiRAN domain and two in the RdRp domain (Figures 2A and S5C). In agreement with previously published results (52), only lysine and tyrosine residues were found to be modified, as these are thought to provide the chemically most stable bonds. Selectivity of the modification was evident in the fact that only seven lysine and tyrosine residues served as nucleophile for the reaction. Furthermore, all these peptides were identified in independent experiments using FSBG concentrations ranging from 25  $\mu\text{M}$  to 2 mM. Within this range a concentration of 100  $\mu\text{M}$  was sufficient to detect all seven peptides. Together this strongly suggests that the reaction with FSBG only occurred after binding to a specific site(s) and did not originate from random collisions. Furthermore, the two modified residues in the EAV RdRp are located in either a predicted

1  $\alpha$ -helix or loop not far upstream and downstream of the  $A_R$  and  $E_R$  motifs, respectively,  
2 which are involved in NTP binding in other better characterized RdRps. The five modified  
3 residues in the EAV NiRAN domain are poorly conserved in related arteriviruses and are  
4 located in the vicinity of one of the three major motifs in either a predicted loop region  
5 (1 residue) or a  $\beta$ -strand (4 residues). These findings are compatible with the expected  
6 properties of the FSBG modification that may label any nucleophile within a 4 Å distance  
7 from the NTP-binding site(s). We therefore conclude that the peptides identified in this  
8 experiment reflect the presence of a nucleotide binding site(s) within the RdRp required  
9 for RNA synthesis and a second binding site that is located in the NiRAN domain, which  
10 could serve for nucleotidylation.

### 11 12 **Conserved residues of the NiRAN domain but not of the RdRp domain are** 13 **required for nucleotidylation activity**

14  
15 In a next step, the importance of conserved NiRAN residues for the guanylylation and  
16 uridylylation activities was examined by characterization of alanine substitution mutants  
17 of several residues, including five invariant residues, in recombinant EAV nsp9. Notably,  
18 none of these mutations significantly reduced expression or stability (data not shown),  
19 indicating that they are most likely compatible with the protein's structure. Subsequent  
20 characterization demonstrated that all conserved NiRAN residues that were probed are  
21 important for nucleotidylation activity, as their replacement with alanine led, with the  
22 exception of S129A, to a drop to below 10% of wild-type protein activity. In contrast, ala-  
23 nine substitution of a non-conserved N-terminal residue (K106A) as well as a conserved  
24 residue in the RdRp domain (D445A of motif  $A_R$ ), which is known to be essential for the  
25 polymerase activity in other RNA viruses (27), had only a mild effect, preserving at least  
26 75% of the activity (Figure 6). Thus, we concluded that the identified sequence motifs in  
27 the EAV nsp9 NiRAN domain are functionally connected to the nucleotidylation activity.  
28 In addition, as the level of remaining activity (again with exception of the S129A mutant)  
29 did not depend on the substrate used, both guanylylation and uridylylation are likely  
30 catalyzed by the same active site.

31  
32 In contrast to these results, the mutation at position S129, the only targeted residue  
33 that is fully conserved in arteriviruses but may be replaced by threonine in other nido-  
34 viruses, exhibited a slightly different effect on guanylylation and uridylylation. Mutant  
35 S129A displayed an intermediate activity when using GTP but was almost as deficient  
36 as mutants of the nidovirus-wide conserved residues when UTP was used as substrate  
37 (Figure 6). This finding may indicate that S129 is specifically involved in the hydrogen  
38 bond network between protein and UTP. Alternatively, as the covalent binding of the  
39 nucleotide occurs via a nucleophilic attack on the  $\alpha$ -phosphate, this serine may in prin-



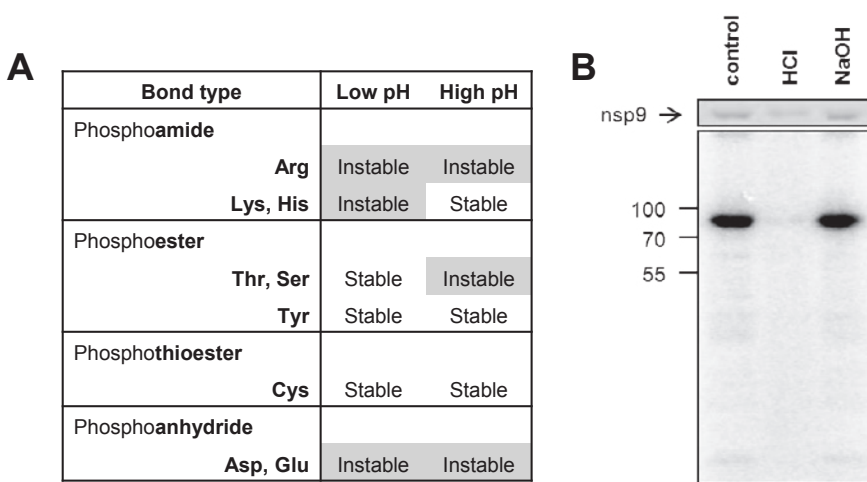
**Figure 6.** Alanine substitution of conserved NiRAN residues dramatically decreased the nucleotidylation activity of nsp9. In contrast, mutation of the non-conserved K106 in the NiRAN domain or the conserved D445 in the RdRp domain had only a mild effect on activity. Error bars represent the standard deviation of the mean based on three independent experiments.

could be suitable to play this role. Although to our knowledge nucleic acid ligases typically employ lysine and rarely histidine as catalytic residues (50;53), we cannot exclude that uridylylation occurs via this S129 while guanylylation utilizes another amino acid.

### Nucleotidylation occurs via the formation of a phosphoamide bond

In order to identify which type of amino acid is the catalytic residue involved in nucleotidylation, the chemical stability of the bond formed between enzyme and nucleotide was probed. To this end, the nucleotidylation product was subjected to either a higher or a lower pH for 4 min, while the protein was heat denatured. The loss of the radioactive label under acidic or alkaline conditions is an indicator for the type of bond that is formed (Figure 7A) (54). As evident from Figure 7B, the bond between guanosine phosphate and nsp9 was acid-labile but stable under alkaline conditions, which was indicative of a phosphoamide bond originating either from a lysine or histidine. This result was also confirmed for uridylylation (data not shown), excluding a direct role for S129 in the attachment of the uridine phosphate. Since there is no conserved histidine present in the NiRAN domain, K94 is the most likely candidate within this domain to fulfill the role of catalytic residue.





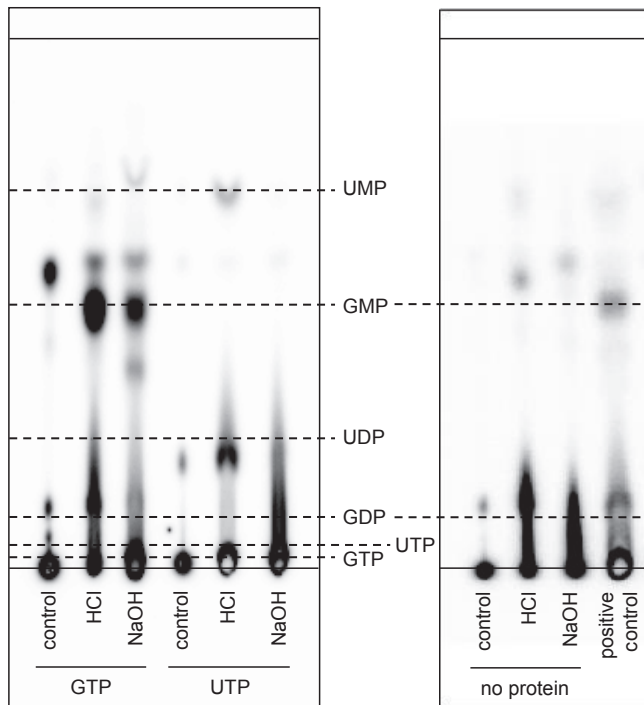
15 **Figure 7.** A phosphoamide bond is formed between nsp9 and the guanosine phosphate. **(A)** Chemical stability of different phosphoamino acid bonds. Adapted from (54). **(B)** The protein was labeled with [ $\alpha$ - $^{32}$ P] GTP and subsequently incubated at pH 8.5 (control) or under acidic or alkaline conditions. Reaction products were visualized after denaturing SDS-PAGE by Coomassie brilliant blue staining (top panel) and phosphor imaging (bottom panel). Size markers are depicted on the left in kDa.

16  
17  
18  
19  
20 **Guanosine and uridine phosphates may be attached via different phosphate groups**

21  
22  
23 So far we have demonstrated that guanylation and uridylation are essentially equally  
24 sensitive to replacement of N1RAN residues, share the same metal ion requirements, and  
25 that both rely on the formation of a phosphoamide bond. We therefore concluded that  
26 there is only one active site responsible for nucleotidylation, which allows utilization of  
27 both substrates. Interestingly, if this is true, discrimination of GTP and UTP against ATP  
28 and CTP would be solely based on the presence of an oxygen at C6 of GTP and C4 of  
29 UTP. However, given the pronounced size difference between UTP and GTP, the position  
30 of both substrates within the binding site is unlikely to be equivalent. In principle, two  
31 binding scenarios are possible. First, ribose and phosphates of both nucleotides could  
32 occupy the same position within the binding site, for example by forming hydrogen  
33 bonds via the ribose's 2' and 3' hydroxyl groups and charge interactions between the  
34 protein and the phosphates. Yet, due to the size difference of the bases (pyrimidine  
35 vs. purine), any additional interactions between protein and bases would involve dif-  
36 ferent hydrogen bond networks, potentially involving water molecules in the case of  
37 the smaller UTP. Alternatively, due to stacking interactions between an aromatic residue  
38 of the protein and the bases, uracil and the pyrimidine ring of guanine might occupy  
39 equivalent positions. As this would inevitably lead to the relative misplacement of the

1 ribose and phosphates of UTP compared to GTP, the catalytic residue may compensate  
 2 for the size difference by re-adjusting and attacking the  $\beta$ - instead of the  $\alpha$ -phosphate  
 3 of UTP.  
 4

5 To explore this possibility, nsp9 was nucleotidylated as before and non-bound label was  
 6 removed by extensive washing until no residual radioactivity was detected in the wash  
 7 buffer. The nucleotide-protein bond was subsequently broken by lowering of the pH  
 8 and the released nucleotide was analyzed by thin layer chromatography. While nsp9  
 9 incubated with GTP clearly released significantly more of the expected GMP in an acidic  
 10 environment than under alkaline conditions, the results after uridylylation were not  
 11 as conclusive. Although also in this case the monophosphate was released after HCl  
 12  
 13



34 **Figure 8.** GMP is released from labeled EAV nsp9 under acidic conditions. (A) nsp9 was labeled with  $[\alpha\text{-}^{32}\text{P}]$   
 35 GTP or  $[\alpha\text{-}^{32}\text{P}]$ UTP and was incubated at pH 8.5 (control) or under acidic or alkaline conditions after removal  
 36 of non-incorporated nucleotides. Resulting products were separated with PEI-cellulose TLC. Solid lines rep-  
 37 resent the position where samples have been spotted (bottom) and the running front (top). Dashed lines  
 38 represent the respective mobilities of the indicated nucleotides. (B)  $[\alpha\text{-}^{32}\text{P}]$ GTP was incubated under the  
 39 same conditions as in A but omitting nsp9. An nsp9-containing sample treated with HCl served as positive  
 control.

1 treatment, the intensity did not match that of GMP and a second product was present  
 2 in higher quantities (Figure 8A). This may indicate that UMP is either further hydrolyzed  
 3 under these conditions or that in fact a UMP-protein adduct is only the minor product  
 4 during uridylylation. Therefore, it remains unclear whether the binding of UTP indeed  
 5 forces an attack of the  $\beta$ -phosphate. To exclude that the observed GMP release is caused  
 6 by the treatment with HCl, control samples lacking nsp9 were also investigated. As  
 7 expected this did not result in a product with equivalent mobility to GMP (Figure 8B).

## 9 NiRAN nucleotidylation is essential for EAV and SARS-CoV replication in cell 10 culture

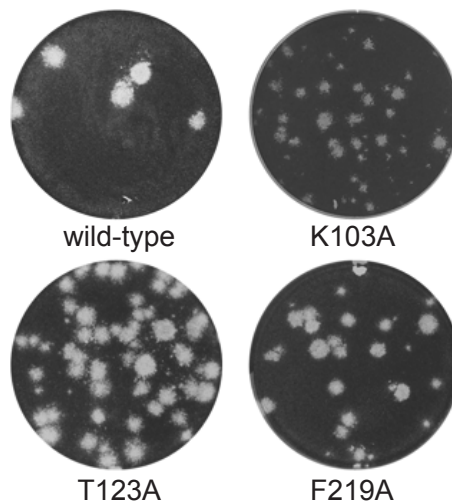
12 To establish the importance of the NiRAN domain for nidoviral replication, reverse ge-  
 13 netics was used to engineer both EAV and SARS-CoV mutants in which conserved NiRAN  
 14 residues were substituted with alanine. Following transfection of *in vitro*-transcribed full-  
 15

17 **Table 1:** Reverse genetics analysis of EAV nsp9 and SARS-CoV nsp12 mutants.

	motif	mutant	mutation	virus titers (PFU/ml at 16-18 h p.t.)	nsp9/nsp12 sequence of P1 virus <sup>a</sup>
EAV	-	wt	-	$1 \cdot 10^7, 2 \cdot 10^8$	n.d.
	A <sub>N</sub>	K94A	AAA <u>GCA</u>	<20, <20	Reversion
	Non-conserved	K106A	AAA <u>GCA</u>	$3 \cdot 10^5, 2 \cdot 10^6$	GCA
	B <sub>N</sub>	R124A	CGU <u>GCU</u>	<20, <20	Reversion
	B <sub>N</sub>	S129A	UCG <u>GCG</u>	$1 \cdot 10^4, 5 \cdot 10^3$	Reversion
	B <sub>N</sub>	D132A	GAU <u>GCU</u>	$3 \cdot 10^4, 6 \cdot 10^3$	Reversion
	C <sub>N</sub>	D165A	GAU <u>GCU</u>	$3 \cdot 10^3, 1 \cdot 10^4$	Reversion
	C <sub>N</sub>	F166A	UUU <u>GCU</u>	<20, <20	n.a.
	A <sub>R</sub>	D445A	GAC <u>GCC</u>	<20, $1 \cdot 10^4$	Reversion
SARS-CoV	-	wt	-	$4 \cdot 10^6, 3 \cdot 10^5$	n.d.
	A <sub>N</sub>	K73A	AAG <u>GCC</u>	<20, <20	n.a.
	Non-conserved	K103A	AAG <u>GCA</u>	<20, <20	GCA
	B <sub>N</sub>	R116A	CGU <u>GCU</u>	<20, <20	n.a.
	B <sub>N</sub>	T123A	ACA <u>GCU</u>	$1 \cdot 10^5, 4 \cdot 10^5$	GCU
	B <sub>N</sub>	D126A	GAU <u>GCG</u>	<20, <20	n.a.
	C <sub>N</sub>	D218A	GAU <u>GCU</u>	<20, <20	n.a.
	C <sub>N</sub>	F219A	UUC <u>GCG</u>	$2 \cdot 10^4, 8 \cdot 10^2$	GCG
	A <sub>R</sub>	D618A	GAU <u>GCG</u>	<20, <20	n.a.

37 <sup>a</sup>Virus-containing supernatants were collected at 72 h p.t. and subsequently used for re-infection of fresh  
 38 BHK-21 (EAV) or Vero-E6 (SARS-CoV) cells. Total RNA was isolated after appearance of CPE, and nsp9/nsp12  
 39 coding regions were sequenced. All results were confirmed in a second independent experiment. n.d., not  
 done; n.a., not applicable (non-viable phenotype).

1 length RNA into permissive cells, viral protein expression and progeny production were  
2 monitored (Table 1). As expected for such conserved residues, most alanine substitutions  
3 were either lethal for the virus or resulted in a severely crippled virus that reverted, thus  
4 confirming the essential role of the nucleotidylase activity during the viral replication  
5 cycle. Similarly, also replacement of a conserved aspartate in motif A of the downstream  
6 RdRp domain, which is known to be required for the activity of polymerases in other  
7 (+) RNA viruses (27), was tolerated in neither EAV nor SARS-CoV. Notable exceptions to  
8 this general pattern, in addition to the replacements of non-conserved lysine residues  
9 included as controls, were the T123A and F219A mutations in SARS-CoV nsp12. These  
10 mutations were stably maintained although they produced a mixed plaque phenotype  
11 comprising wild-type-sized and smaller plaques, with F219A also demonstrating a  
12 markedly lower progeny titer (at least 2 logs) than the wild-type control (Figure 9). The  
13 reason for this differential behavior of these two SARS-CoV mutants in comparison to  
14 those of EAV is unclear at the moment.



30 **Figure 9:** Plaque phenotypes of viable SARS-CoV NiRAN mutants. Virus-containing supernatants obtained  
31 72 h post transfection were used to infect BHK-21 cells. After 72 h cells were fixed with 4% formaldehyde  
32 and stained with crystal violet.

## 1 DISCUSSION

### 3 NiRAN is the first enzymatic genetic marker of the order *Nidovirales*

5 The NiRAN domain described in this study is the fourth ORF1b-encoded enzyme  
6 involved in RNA-dependent processes identified in arteriviruses and the seventh in  
7 coronaviruses. Its existence was not predicted by prior nidovirus research, which attests  
8 to our poor understanding of the molecular machinery that governs nidovirus replica-  
9 tion. As in most prior studies of nidoviral replicative proteins, this identification was  
10 initiated by comparative genomics analysis, whose results made it clear why this par-  
11 ticular enzyme, now called the NiRAN domain, was not identified earlier. Unlike all other  
12 nidovirus enzymes, NiRAN was found to have no appreciable sequence similarity with  
13 proteins outside the order *Nidovirales*. The analysis suggested the extreme divergence  
14 of nidovirus NiRAN domains from their prototypes, since even the similarity between  
15 the arteriviral NiRAN and that of other nidoviruses was found to be marginal. Five out  
16 of the seven amino acid residues that are evolutionary invariant in the NiRAN domain  
17 belong to the most frequently occurring residues in proteins, which likely complicated  
18 the recognition of NiRAN conservation by even the most powerful HMM-based tools.

20 Besides technical challenges in the identification of NiRAN, this domain also stands out  
21 for its properties that are indicative of an unknown but critical role in nidovirus replica-  
22 tion (see below). NiRAN is the only ORF1b-encoded domain that is located upstream of  
23 the RdRp and resides within the same nonstructural protein. This implies that NiRAN may  
24 influence the folding of the downstream RdRp domain. It would be reasonable to expect  
25 that these domains cross-talk to couple the reactions and processes they catalyze. Thus,  
26 NiRAN is a prime candidate to be a regulator and/or co-factor of the RdRp, a property  
27 that should be taken into account in future experiments aiming at the characterization  
28 of the RdRp or reconstitution of RTC activity *in vitro*.

30 The exclusive conservation of NiRAN in nidoviruses makes it a genetic marker of this  
31 order, only the second after the previously identified ZBD and the first with enzymatic  
32 activity. It may not be a coincidence that each of these markers is associated with a key  
33 enzyme in (+) RNA virus replication, RdRp and HEL1, respectively. The modulating role of  
34 the ZBD for HEL1 and its involvement in all major processes of the nidovirus replicative  
35 cycle have been documented (reviewed in (11)). Similar studies could be performed to  
36 probe the function(s) of NiRAN. This emerging parallel between NiRAN-RdRp and ZBD-  
37 HEL1 highlights the fruitful cooperation between nidovirus-wide comparative genomics  
38 and experimental studies during the functional characterization of these proteins.

## 1 Possible functions of conserved NiRAN residues

2  
3 We here demonstrated that NiRAN is essential for EAV and SARS-CoV replication in cell  
4 culture by testing mutants in which conserved residues had been replaced. The mutated  
5 viruses were either crippled (and in most cases reverted to wt) or dead, depending on  
6 the targeted residue and the virus studied. Importantly the magnitude of the observed  
7 effect paralleled that caused by the replacement of an RdRp active site residue of the  
8 respective virus, which can be expected to put the greatest possible constraints on viral  
9 replication with the RdRp being the central enzyme involved in this process. This similar-  
10 ity between the two enzymes is most notable because of the much higher divergence of  
11 the NiRAN sequence compared to the RdRp. These results also show that the significance  
12 of NiRAN for virus replication must be different from that of NendoU, the only other  
13 ORF1b-encoded enzyme that has been probed extensively by mutagenesis in reverse  
14 genetics in both corona- and arteriviruses (17;41;55). Two of those studies revealed that  
15 EAV and mouse hepatitis virus (MHV) NendoU mutants with replacements in the active  
16 site were stable and in the latter case even displayed similar plaque phenotypes as the  
17 wild-type virus while being only slightly delayed in growth (41;55).

18  
19 In our biochemical assays we detected a second enzymatic activity that is associated  
20 with the nidovirus RdRp subunit (31;33;56). This new activity, which was categorized as  
21 nucleotidylation, is associated with the N-terminal domain of EAV nsp9, as demonstrated  
22 by mass spectrometry analysis (Figures 2A and S4) and the importance of conserved  
23 NiRAN residues for this activity (Figure 6). Nucleotidylation was most pronounced with  
24 UTP as substrate but was also observed with GTP (Figure 4A). Despite their size differ-  
25 ence, both substrates appeared to be utilized by the same NiRAN binding site since  
26 uridylylation as well as guanylylation depended on the same conserved residues. To  
27 our knowledge such dual specificity has never been reported for a protein of an RNA  
28 virus and (likely) a host. Our results strongly suggested the nucleotidylated residue to  
29 be either a lysine or a histidine (Figure 7). Since NiRAN lacks a conserved histidine, K94  
30 in EAV nsp9 is the most likely target for nucleotidylation. Alternatively, reminiscent of  
31 the protein kinase mechanism, the conserved NiRAN residues might merely constitute a  
32 nucleotide binding site that presents the nucleotide to a catalytic residue located in the  
33 C-terminal RdRp domain.

34  
35 Next to K94 and/or R124, which may mediate NTP binding via interactions with the  
36 negatively charged phosphates, a third conserved residue which may contribute to NTP  
37 binding is F166 in EAV. Since phenylalanine would most likely interact with the nucleo-  
38 tide substrate by base stacking, its contribution in terms of binding energy would be one  
39 order of magnitude lower than that of electrostatic interactions of lysine/arginine with

1 the phosphates (57). Based solely on these considerations, F166 could be expected to  
2 be of “lesser” importance than the basic residues. However, this was apparently not the  
3 case since the replacement of the aromatic residue with alanine was lethal for EAV while  
4 substitution of either of the basic residues led to a low level of replication that eventu-  
5 ally facilitated reversion (Table 1). When analyzing these results, a consideration must be  
6 made about the feasibility of reversion for different engineered substitutions, which all  
7 require two nucleotide point mutations to revert back to wild-type. As simultaneous re-  
8 version of both nucleotides during a single round of replication should be an extremely  
9 rare event, the dead phenotype of the F166A mutant may hint at a lower tolerance of  
10 single-nucleotide partial revertants (F166V or F166S) in comparison to those originating  
11 from K94A (K94T or K94E) and R124A (R124P or R124G). Alternatively, the observed dead  
12 F166A phenotype may be explained by a vital interaction between NiRAN and RdRp  
13 or other proteins involving F166. In contrast to EAV, the homologous residue in SARS-  
14 CoV nsp12, F219, appeared to be less essential since its replacement merely reduced  
15 progeny titers and altered the plaque phenotype, while the nucleotide changes were  
16 maintained. At present, the exact reason for this difference between EAV and SARS-CoV  
17 is unclear, but it suggests that the role and/or regulation of this conserved phenylalanine  
18 may have evolved in these distantly related nidoviruses, whose NiRAN domains are of  
19 strikingly different sizes; such evolution has parallels in other enzymes (58).

20  
21 Since neither binding of phosphates nor base stacking would enable the enzyme to  
22 discriminate between the four bases, it is likely that some of the conserved residues are  
23 involved in the formation of a hydrogen bond network that is specific for GTP or UTP. We  
24 already speculated on the participation of nsp9 S129 in such a network, as substitution  
25 of this serine was the only mutation that had a differential effect on guanylylation and  
26 uridylylation (Figure 6). Finally, in agreement with observations for other nucleotidylate-  
27 forming enzymes (59-61), also nsp9 nucleotidylation is metal-dependent (Figure 4B),  
28 potentially due to an important role for metal ions in coordination of the triphosphate  
29 or charge neutralization of the pyrophosphate leaving group. We thus propose that at  
30 least one of the three acidic conserved residues (E100, D132, and D165 in EAV nsp9) is  
31 directly involved in the binding of the essential manganese ion(s).

### 32 **Possible roles of nucleotidylation in the context of viral replication**

33  
34  
35 The identification of the nucleotidylation activity raises the question which role it may  
36 play in the nidovirus replicative cycle. Given that the roles of other replicative enzymes of  
37 nidoviruses are far from firmly established, considerable challenges may be expected in  
38 the characterization of the NiRAN domain, starting from the identification of the ultimate  
39 target of the nucleotidylation. In this respect, it is relevant that many cellular enzymes

1 employ covalent binding of NMPs to catalyze different reactions, which are dominated  
2 by those that generate essential metabolites in an energy-dependent manner. These  
3 host metabolites are utilized by RNA viruses, whose relatively small genomes can thus  
4 be used to encode NMP-binding enzymes for other, virus-specific purposes. Therefore,  
5 in the discussion that follows we will consider the pros and cons of the involvement of  
6 NiRAN's nucleotidyltransferase activity in three previously described functions that are not  
7 involved in metabolism: nucleic acid ligation, mRNA capping, and protein-primed RNA  
8 synthesis.

### 9 10 *Ligase function*

11  
12 We initially considered NiRAN to be a non-canonical ATP-dependent RNA ligase. It was  
13 reasoned that in the context of nidovirus replication such an activity would be the  
14 functional complement of the NendoU endoribonuclease (6). Moreover, at that time  
15 both enzymes were considered to have been conserved across all taxa during evolu-  
16 tion of the nidovirus lineage. Prompted by nidovirus comparative genomics, it recently  
17 became clear that NendoU is conserved only in nidoviruses infecting vertebrate hosts.  
18 Consequently, our original hypothesis would not explain why this putative ligase would  
19 be conserved in roni- and mesoniviruses, which do not encode the endoribonuclease.  
20 Another complication regarding that original hypothesis has emerged from the present  
21 study, which identified NiRAN as being UTP/GTP-specific. Although the hydrolysis of all  
22 NTPs results in the release of the same amount of energy, ATP-dependent RNA ligases,  
23 which dominate the ligase family, are – as their name already suggests – restricted in  
24 their substrate use. It would therefore be surprising, if nidoviruses encoded a ligase that  
25 strongly discriminates against ATP. To our knowledge the GTP-specific tRNA-splicing  
26 ligase RtcB is the only currently known example of a protein involved in nucleic acid  
27 strand joining exhibiting this kind of substrate specificity (53). Also no substrates which  
28 would require a ligase function were identified in the nidovirus replication, which how-  
29 ever remains poorly characterized.

### 30 31 *5' end cap guanylyltransferase function*

32  
33 Besides RNA ligases, there is another group of enzymes, known as guanylyltransferases  
34 (GTases), that employ a very similar mechanism of nucleotidyltransferase and may be relevant  
35 to nidovirus replication. Unlike ligases, the covalent binding of GMP by GTases does not  
36 occur for energetic reasons. Rather, the bound GMP is used to permanently modify the 5'  
37 end of RNA in a process called RNA capping (reviewed in (62)). Intriguingly, three of the  
38 four enzyme activities required for this pathway have been identified in coronaviruses  
39 (35;63), with the missing activity being the GTase. Furthermore, recent characterization



1 of EAV nsp10 in our lab (unpublished) showed that it resembles its coronavirus homolog  
2 in terms of possessing RNA-triphosphatase activity, which is required prior to GTase  
3 activity in the conventional capping pathway. In line with these findings, experimental  
4 evidence supporting the presence of a cap structure on genomic RNA was reported  
5 for three very distantly related species of the *Nidovirales* order, namely for MHV (64),  
6 *Equine torovirus* (EToV) (65) (both *Coronaviridae*), and *Simian hemorrhagic fever virus*  
7 (SHFV) (*Arteriviridae*) (66). Thus, the NiRAN domain could be a candidate for catalyzing  
8 the important GTase reaction in the nidovirus capping pathway. Like ligases, canonical  
9 cellular GTases share the characteristic Kx(D/N)G motif including the principal catalytic  
10 lysine, which has no match in NiRAN. Although this deviation is notable, it is not unprec-  
11 edented in established viral GTases. For instance, upstream of its RdRp domain, flavivirus  
12 NS5 contains the GTase domain, which neither has homology to any other GTase nor  
13 contains the canonical Kx(D/N)G motif (67). Likewise, the GTase activity of alphavirus  
14 nsP1 and related proteins is associated with a unique domain (60;68). Thus, NiRAN being  
15 a cap-synthesizing GTase could be reconciled with our current knowledge about GTase  
16 structural and sequence diversity.

17

18 The same cannot be said about NiRAN's substrate preference for UTP over GTP, which  
19 has not been reported for GTases mediating cap formation. To reconcile this property  
20 with the considered functional model, we would therefore have to assume that either  
21 NiRAN has another substrate or that uridylylation is an *in vitro* artifact due to the absence  
22 of essential interaction partners of NiRAN. For instance, it would be conceivable that  
23 the association with other proteins modulates the binding site allowing discrimination  
24 against UTP.

25

### 26 *Protein-priming function*

27

28 If UTP binding by NiRAN faithfully reflects a genuine property of the enzyme, a plausible  
29 explanation for the nucleotidylation activity of nsp9 may be its involvement in protein-  
30 primed RNA synthesis. This mechanism is used by many viruses including a large group of  
31 picornavirus-like viruses, which notably have evolutionary affinity to nidoviruses (69;70).  
32 In these viruses a nucleotide is covalently attached to a protein commonly known as  
33 VPg (viral protein genome-linked), which may then be extended to a dinucleotide. This  
34 dinucleotide is subsequently base-paired to the 3' end of the viral RNA where it serves  
35 as the primer for synthesis of the complementary RNA strand (71). Interestingly, the  
36 first nucleotide of the EAV genome is a G while the 3' end is equipped with a poly(A)  
37 tail. Thus, the dual specificity of nsp9 for GTP and UTP would be compatible with the  
38 different requirements for the initiation of (+) and (-) strand synthesis of genomic and  
39 subgenomic mRNAs.

1 However, there are also observations that distinguish nidoviruses from viruses that  
2 use a VPg. First, to our knowledge, all currently described nucleotide-VPg bonds are  
3 realized via the hydroxyl group of either a tyrosine or a serine/threonine (72-76) while  
4 NiRAN is most likely to use the invariant lysine residue (Figure 7). Second, at least for  
5 coronaviruses, the VPg-based mechanism would compete with the already proposed  
6 primase-based mechanism (77) for the initiation of RNA synthesis. The latter mechanism  
7 is yet to be fully established since it assigns primase activity to a protein complex that  
8 may merely be a processivity co-factor for the nsp12 RdRp according to a recent study  
9 (78). Finally, as mentioned before, nidovirus mRNAs were concluded to be capped at  
10 their 5' end, a modification that is not observed in known VPg-utilizing viruses. To use  
11 both capping and VPg, it would thus be necessary for nidoviruses to actively or passively  
12 remove the attached protein in order to allow mRNA capping to commence. Such a  
13 reaction sequence would also imply a variation of the capping pathway as the RNA 5'  
14 end would not be di- or triphosphorylated after removal of the VPg, a requirement for  
15 entering any of the known viral capping pathways (62).

16  
17 In view of the considerations outlined for each of the three possible scenarios employ-  
18 ing nucleotidylation activity, it is evident that presently none of these can be fully  
19 reconciled with the evolutionary, structural, and functional characteristics of NiRAN  
20 described in this study. This may reflect yet-to-be revealed specifics of the nidovirus RTC  
21 and its unparalleled complexity. On the other hand, the unique NiRAN is now part of  
22 this complexity and its properties must be taken into account in future experiments  
23 involving RdRp-encoding and other replicative proteins, as well as in theoretical models  
24 describing the molecular biology of nidoviruses.

## 25 26 27 **MATERIAL AND METHODS**

### 28 29 **Virus genomes**

30  
31 Genomes of nidoviruses were retrieved from GenBank (79) and RefSeq (80) using  
32 Homology-Annotation hYbrid retrieval of GENetic Sequences (HAYGENS) tool [http://](http://veb.lumc.nl/HAYGENS)  
33 [veb.lumc.nl/HAYGENS](http://veb.lumc.nl/HAYGENS). Genomes of all viruses were used to produce sequence align-  
34 ments (see below), which were purged to retain only subsets of viruses representing  
35 the known diversity of each nidovirus family for downstream bioinformatics analyses.  
36 For the *Arteriviridae* and *Coronaviridae* families, one representative was drawn randomly  
37 from each evolutionary compact cluster corresponding to known and tentative species  
38 that were defined with the help of DEmARC1.3 (81). Twenty nine viruses of the family  
39 *Mesoniviridae* were clustered into six groups, whose intra- and inter-group evolution-

1 ary distance was below and above 0.075, respectively. One representative was chosen  
2 randomly from each of the six groups. For the *Roniviridae* family, two viruses, each  
3 prototyping a species, were used. To retrieve information about genomes, the SNAD  
4 program (82) was used.

## 6 **Multiple sequence alignments**

8 MSAs of five nidovirus-wide conserved protein domains: 3C-like protease (3CL<sup>Pro</sup>),  
9 RNA-dependent RNA polymerase (RdRp), RdRp-associated nucleotidyltransferase (Ni-  
10 RAN), superfamily 1 helicase (HEL1) and zinc-binding domain fused with HEL1 (ZBD)  
11 were obtained for four nidovirus families using the Viralis platform (83) and assisted by  
12 HMMER 3.1 (84), Muscle 3.8.31 (85), and ClustalW 2.0.12 (86) programs. Family-specific  
13 MSAs of the NiRAN domain were combined in a step-wise manner using the HH-suite  
14 2.0.15 software (87;88) and the profile mode of ClustalW with subsequent manual  
15 refinement to produce MSAs that included two, three, and four families, respectively,  
16 namely: *Coronavirinae*, *Torovirinae*, and *Mesoniviridae* (named CoToMe), *Coronaviridae*,  
17 *Mesoniviridae*, and *Roniviridae* (CoToMeRo), *Coronaviridae*, *Mesoniviridae*, *Roniviridae*,  
18 and *Arteriviridae* (CoToMeRoAr). To reveal all local similarities between two MSAs, their  
19 profiles were compared in a dot-plot fashion using a routine in HH-suite 2.0.15, whose  
20 results were visualized. Distribution of similarity density in MSAs was plotted using R  
21 package Bio3D (89) under the conservation assessment method “similarity”, substitution  
22 matrix Blosum62 (90) and a sliding window of 11 MSA columns. Peaks of similarity were  
23 attributed to the known RdRp motifs G, F, A, B, C, D, E (69), or named and assigned to the  
24 newly recognized motifs of NiRAN, preA, A, B, and C. To facilitate distinguishing between  
25 the RdRp and NiRAN motifs, suffix R and N were added to motif labels of the RdRp and  
26 NiRAN domain, respectively. Based on family-specific MSAs of NiRAN and RdRp, the  
27 secondary structure of these domains was predicted using software Jpred 3 (91) and  
28 PSIPRED (92). In both cases, the sequence with the least gaps was selected from the  
29 sequences forming the MSA. The prediction was made only for columns of the MSA in  
30 which the selected sequence does not contain gaps. The MSAs were converted into the  
31 final figure using ESPript (93).

## 33 **Homology detection**

35 The obtained MSAs were converted into HMM profiles or PSSMs and used as queries  
36 to search for homologs in three different types of databases composed of: individual  
37 sequences (nr database, including GenBank CDS translations, RefSeq proteins, Swis-  
38 sProt, PIR and PRF (94)), profiles (PFAM A (46)), and protein 3D structures (PDB (45)).  
39 For GenBank scanning, HMMER 3.1 software (84) was used under E value significance

1 threshold -10. To search for homologs among profiles, HH-suite 2.0.15 software (87;88)  
2 was used. To search for homologs among protein 3D structures pGenTHREADER 8.9  
3 software (95-97) was used.

## 4 5 **Protein Expression and Purification**

6  
7 Nucleotides 5256 to 7333 of the EAV Bucyrus strain were cloned into a pASK3 (IBA)  
8 vector essentially as described (38) to yield a construct that expresses nsp9 that is N-  
9 terminally fused to ubiquitin and tagged with hexahistidine at its C-terminus. Mutations  
10 were introduced according to the QuikChange protocol and verified by sequencing.  
11 Plasmids were transformed into *E. coli* C2523/pCG1, which constitutively express the  
12 Ubp1 protease to remove the ubiquitin tag during expression and thereby generate the  
13 native nsp9 N-terminus. Cells were cultured in Luria Broth in the presence of ampicillin  
14 (100 µg/ml) and chloramphenicol (34 µg/ml) at 37°C until an OD<sub>600</sub> >0.7. At this point  
15 protein expression was induced by the addition of anhydrotetracycline to a final con-  
16 centration of 200 ng/ml, and incubation was continued at 20°C overnight. Cell pellets  
17 were harvested by centrifugation and stored at -20°C until further use.

18  
19 Proteins were batch purified by immobilized metal ion affinity chromatography using  
20 Co<sup>2+</sup> Talon beads. In short, cell pellets were resuspended in lysis buffer (20 mM HEPES,  
21 pH 7.5, 10% glycerol (v/v), 10 mM imidazole, 5 mM β-mercaptoethanol) supplemented  
22 with 500 mM NaCl. Lysis was achieved by a 30-min incubation with 0.1 mg/ml lysozyme  
23 and five subsequent cycles of 10-s sonication to shear genomic DNA. Cellular debris  
24 was removed by centrifugation at 20,000g for 20 min. The cleared supernatant was  
25 recovered, and equilibrated Talon-beads were added. After 1 h of binding under agita-  
26 tion, beads were washed four times for 15 min with a 25-times bigger volume of lysis  
27 buffer containing first 500 mM, than 250 mM, and finally twice 100 mM NaCl. In the  
28 end, proteins were eluted twice with lysis buffer containing 100 mM NaCl and 150 mM  
29 imidazole. Both fractions were pooled and dialyzed twice for 6 h or longer against an  
30 at least 100-fold bigger volume of 20 mM HEPES, pH 7.5, 50% glycerol (v/v), 100 mM  
31 NaCl, 2 mM DTT. All steps of the purification were performed at 4°C or on ice. All mutant  
32 proteins were expressed and purified in parallel with the wild-type protein used as refer-  
33 ence in nucleotidylation assays. Protein concentrations were measured by absorbance  
34 at 280 nm using a calculated extinction coefficient of 93,170 M<sup>-1</sup>cm<sup>-1</sup> and a molecular  
35 mass of 77,885 Da for wild-type nsp9. Typical protein yields were 5 mg/l culture and  
36 nucleotidylation activity was observed for at least 4 months if stored at -20°C at a  
37 concentration below 15 µM. Finally, the absence of the N-terminal ubiquitin tag was  
38 confirmed by mass spectrometry.

## 1 Nucleotidylation Assay

2

3 Nucleotidylation assays were performed in a total volume of 10  $\mu$ l containing, unless  
4 specified otherwise, 50 mM Tris, pH 8.5, 6 mM  $MnCl_2$ , 5 mM DTT, up to 2.5  $\mu$ M nsp9,  
5 and 0.17  $\mu$ M [ $\alpha$ - $^{32}P$ ]NTP (Perkin Elmer, 3000 Ci/mmol). Furthermore, 12.5% glycerol  
6 (v/v), 25 mM NaCl, 5 mM HEPES, pH 7.5, and 0.5 mM DTT were carried over from the  
7 protein storage buffer. In preliminary experiments magnesium (1-20 mM) did not sup-  
8 port nucleotidylation activity and was consequently not pursued further. Samples were  
9 incubated for 30 min at 30°C. Reactions were stopped by addition of 5  $\mu$ l gel loading  
10 buffer (62.5 mM Tris, pH 6.8, 100 mM DTT, 2.5% SDS, 10% glycerol, 0.005% bromophe-  
11 nol blue) and denaturing of the proteins by heating at 95°C for 5 min. 12% SDS-PAGE  
12 gels were run, stained with Coomassie G-250, and destained overnight. After drying,  
13 phosphorimager screens were exposed to gels for 5 h and scanned on a Typhoon vari-  
14 able mode scanner (GE healthcare), after which band intensities were analyzed with  
15 ImageQuant TL software (GE healthcare). The buffers used to find the pH optimum of  
16 the nucleotidylation reaction were MES (pH 5.5 – 6.5), MOPS (pH 7.0), Tris (pH 7.5 – 8.5),  
17 and CHES (pH 9.0 – 9.5) (20 mM).

18

19 To assess the chemical nature of the nucleotide-protein bond, the pH was temporarily  
20 shifted after product formation. To this end, 1  $\mu$ l HCl or NaOH (both 1 M) was added  
21 before incubation at 95°C for 4 min. Afterwards the original pH was restored by addi-  
22 tion of the complementary base or acid, and samples were separated and analyzed as  
23 described.

24

## 25 FSBG Labeling and Mass Spectrometry

26

27 Reaction mixtures were the same as described for the nucleotidylation assay with  
28 two modifications. Radioactive nucleotides were replaced by the reactive GTP analog  
29 5'-(4-fluorosulfonylbenzoyl)guanosine (FSBG) (52) (up to 2 mM) (see supplementary  
30 Materials and Methods for the synthesis protocol), and samples were incubated for 1 h  
31 at 30°C to increase the ratio between labeled and unlabeled protein. Subsequently, the  
32 protein (20  $\mu$ g) was reduced by addition of 5 mM DTT and denatured in 1% SDS for  
33 10 min at 70°C. Next, the samples were alkylated by addition of 15 mM iodoacetamide  
34 and incubation for 20 min at RT. Next, the protein was applied to a centrifugal filter (Mil-  
35 lipore Microcon, MWCO 30 kDa) and washed three times with  $NH_4HCO_3$  (25 mM) before  
36 a protease digestion was performed with 2  $\mu$ g trypsin in 100  $\mu$ l  $NH_4HCO_3$  overnight at  
37 RT. Recovered peptides were treated with 50 mM NaOH for 25 min, desalted using Oasis  
38 spin columns (Waters), and finally analyzed by on-line nano-liquid chromatography tan-  
39 dem mass spectrometry on an LTQ-FT Ultra (Thermo, Bremen, Germany). Tandem mass

1 spectra were searched against the Uniprot database, using mascot version 2.2.04, with a  
2 precursor accuracy of 2 ppm, and product ion accuracy of 0.5 Da. Carbamidomethyl was  
3 set as a fixed modification, and oxidation, N-acetylation (protein N-terminus), and FSBG  
4 were set as variable modifications.

## 5 6 **Label Release**

7  
8 For analysis of the released nucleotides, 350 pmol of nsp9 were nucleotidylated with  
9 [ $\alpha$ -<sup>32</sup>P]NTPs as described above for 1 h at 30°C. After the reaction free NTPs were re-  
10 moved by buffer exchange and extensive washing with the help of a centrifugal filter  
11 (Millipore ultrafree-0.5, MWCO 10 kDa). Protein was precipitated with a 5-times greater  
12 volume of acetone overnight at -20°C. The resulting pellet was resuspended in 20 mM  
13 Tris, pH 8.5, 100 mM NaCl. Equal amounts of the solutions were incubated at 95°C for  
14 4 min after addition of HCl or NaOH (1 M). Samples were adjusted to their original pH  
15 and spotted onto polyethylenimine cellulose thin layer chromatography plates, which  
16 were developed in 80% acetic acid (1 M), 20% ethanol (v/v), 0.5 M LiCl. Plates were dried  
17 and phosphorimaging was performed as described above. Non-radioactive nucleotide  
18 standards were run on each plate and visualized by UV-shadowing to allow the identifi-  
19 cation of the radioactive products.

## 20 21 **Reverse Genetics of EAV**

22  
23 Alanine-coding mutations for conserved and control residues were introduced into  
24 full-length cDNA clone pEAV211 (98) using appropriate shuttle vectors and restriction  
25 enzymes. The presence of the mutations was confirmed by sequencing. pEAV plasmid  
26 DNA was *in vitro* transcribed with the mMessage-mMachine T7 kit (Ambion), and the  
27 synthesized RNA was transfected into BHK-21 cells after LiCl precipitation as described  
28 previously (99). Virus replication was monitored by immunofluorescence microscopy  
29 until 72 h post transfection (p.t.) using antibodies directed against nsp3 and N protein  
30 as described (100) and by plaque assays (99) using transfected cell culture supernatants,  
31 to monitor the production of viral progeny.

32  
33 Sequence analysis of the nsp9-coding region was performed to either verify the pres-  
34 ence of the introduced mutations or to monitor the presence of (second site) reversions.  
35 For this purpose, fresh BHK-21 cells were infected with virus-containing cell culture  
36 supernatants and total RNA was extracted with Tripure Isolation Reagent (Roche Ap-  
37 plied Science) after appearance of cytopathic effect (CPE) (typically at 18 h post infec-  
38 tion (p.i.)). EAV-specific primers were used to reverse transcribe RNA and PCR amplify  
39 the nsp9-coding region (nt 5256-7333). RT-PCR fragments of the EAV genome were

1 sequenced after gel purification and sequences compared to those of the respective  
2 RNA used for transfection.

3

#### 4 **Reverse Genetics of SARS-CoV**

5

6 Mutations in the SARS-CoV nsp12-coding region were engineered in prSCV, a pBelo-  
7 Bac11 derivative containing a full-length cDNA copy of the SARS-CoV Frankfurt-1 se-  
8 quence (101) by using “en passant recombineering” as described in Tischer *et al.* (102).

9 The (mutated) BAC DNA was linearized with *NotI*, extracted with phenol-chloroform,  
10 and transcribed with T7 RNA Polymerase (mMessage-mMachine T7 kit; Ambion) using  
11 an input of 2 µg of BAC DNA per 20-µL reaction. Viral RNA transcripts were precipitated  
12 with LiCl according to the manufacturer’s protocol. Subsequently, 6 µg of RNA were  
13 electroporated into  $5 \times 10^6$  BHK-Tet-SARS-N cells, which expressed the SARS-CoV N  
14 protein following 4 h induction with 2 µM doxycycline as described previously (78).  
15 Electroporated BHK-Tet-SARS-N cells were seeded in a 1:1 ratio with Vero-E6 cells. Viral  
16 protein expression and the production of viral progeny was followed until 72 h p.t. by  
17 immunofluorescence microscopy using antibodies directed against nsp4 and N protein  
18 and by plaque assays of cell culture supernatants, respectively (both methods were  
19 described previously in Subissi *et al.* (78)). All work with live SARS-CoV was performed in-  
20 side biosafety cabinets in a biosafety level 3 facility at Leiden University Medical Center.

21

22 For sequence analysis of viral progeny, fresh Vero-E6 cells were infected with harvests  
23 from viable mutants taken at 72 h p.t., and SARS-CoV RNA was isolated 18 h p.i. using  
24 TriPure Isolation Reagent (Roche Applied Science) as described in the manufacturer’s  
25 instructions. Random hexamers were used to prime the RT reaction, which was followed  
26 by amplification of the nsp12-coding region (nt 13398-16166) by using SARS-CoV-spe-  
27 cific primers. RT-PCR products were sequenced to verify the presence of the introduced  
28 mutations.

29

30

#### 31 **FUNDING**

32

33 This work was supported by the European Union Seventh Framework program through  
34 the EUVIRNA project (European Training Network on (+) RNA virus replication and An-  
35 tiviral Drug Development, grant agreement no. 264286) and the SILVER project (grant  
36 agreement no. 260644); the Netherlands Organization for Scientific Research (NWO)  
37 through TOP-GO grant 700.10.352; the Leiden University Fund; and through the Col-  
38 laborative Agreement in Bioinformatics between Leiden University Medical Center and  
39 Moscow State University (MoBiLe).

**1 ACKNOWLEDGEMENTS**

2

3 The authors thank Bruno Canard, Etienne Decroly, Isabelle Imbert, Barbara Selisko,  
4 Lorenzo Subissi, and Aartjan te Velthuis for helpful discussions; Chris Lauber and Erik  
5 Hoogendoorn for help with the DEmARC-based analysis, and Daniel Cupac and Linda  
6 Boomaars for technical assistance.

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39



**1 SUPPLEMENTARY DATA****3 Supplementary Material and Methods****5 Synthesis of 5'-(4-fluorosulfonylbenzoyl)guanosine (FSBG)**

7 Guanosine monohydrate (875 mg, 2.90 mmol) was co-evaporated twice with anhydrous  
8 DMF and subsequently dissolved in DMPU with gentle warming. The clear solution was  
9 cooled in an ice bath, and 4-(fluorosulfonyl)benzoyl chloride (812 mg, 3.65 mmol) was  
10 added. After 15 minutes the mixture was warmed to room temperature and stirred  
11 for another 4 hours. Petroleum ether 40/60 (50 ml) was added and a white precipitate  
12 formed. The organic layer was decanted and the residue triturated twice with a 1/1  
13 mixture of ethyl acetate/diethyl ether (2 x 50 ml). The residue was re-crystallized from  
14 MeOH/water and further purified by C18-RP-HPLC (Phenomenex Gemini C18, pore size  
15 110Å, particle size 5 µm, 150 x 21.2 mm, gradient 20 – 50% Acetonitrile in 0.1 % aque-  
16 ous TFA, 20 ml/min) to yield the title compound as a white solid (232 mg, yield 17%)  
17 (Supplementary Figure 5).

**Table S1:** GenBank accession number, name, and acronym of each virus genome used for the bioinformatics analyses.

Accession number	Virus name	Acronym	Species
AF227196	Gill-associated virus	GAV	<i>Gill-associated virus</i>
EU487200	Yellow head virus	YHV	to be established
HM746600	Cavally virus	CAVV	<i>Alphamesonivirus 1</i>
NC_023986	Casuarina virus	CASV	to be established
AB753015.2	Dak Nong virus	DKNV	to be established
JQ957872	Hana virus	HanaV	to be established
JQ957874	Nse virus	NseV	to be established
JQ957873	Meno virus	MenoV	to be established
DQ412042	Bat SARS coronavirus Rf1	SARS-Rf1-BtCoV	<i>Severe acute respiratory syndrome-related coronavirus</i>
JN874560	Rabbit coronavirus HKU14	RbCoV_HKU14	<i>Betacoronavirus 1</i>
AF201929	Murine hepatitis virus strain 2	MHV-2	<i>Murine coronavirus</i>
AY884001	Human coronavirus HKU1	HCoV_HKU1	<i>Human coronavirus HKU1</i>
KC545383	Betacoronavirus Erinaceus/VMC/DEU/2012	EriCoV	to be established
DQ648794	Bat coronavirus (BtCoV/133/2005)	BtCoV/133/2005	<i>Tylonycteris bat coronavirus HKU4</i>
EF065509	Bat coronavirus HKU5-1	BtCoV_HKU5	<i>Pipistrellus bat coronavirus HKU5</i>
JX869059.2	MERS coronavirus EMC/2012	HCoV-EMC/2012	to be established
HM211101	Bat coronavirus HKU9-10-2	BtCoV_HKU9	<i>Rousettus bat coronavirus HKU9</i>
KF430219	Bat coronavirus CDPHE15/USA/2006	BtCoV_CDPHE15	to be established
AY567487	Human coronavirus NL63	HCoV-NL63	<i>Human coronavirus NL63</i>
EU420139	Miniopterus bat coronavirus HKU8	BtCoV_HKU8	<i>Miniopterus bat coronavirus HKU8</i>
EF203064	Rhinolophus bat coronavirus HKU2	BtCoV_HKU2	<i>Rhinolophus bat coronavirus HKU2</i>
EU420138	Bat coronavirus 1A	BtCoV_1A	<i>Miniopterus bat coronavirus 1</i>
JQ410000	Alpaca respiratory coronavirus	ACoV	<i>Human coronavirus 229E</i>
DQ648858	Bat coronavirus (BtCoV/512/2005)	BtCoV/512/2005	<i>Scotophilus bat coronavirus 512</i>
KC140102	Porcine epidemic diarrhea virus	PEDV	<i>Porcine epidemic diarrhea virus</i>
JQ989271	Rousettus bat coronavirus HKU10	BtCoV_HKU10	to be established
HM245925	Mink coronavirus strain WD1127	MCoV	to be established
FJ938060	Feline coronavirus UU2	FCoV_UU2	<i>Alphacoronavirus 1</i>
KC008600	Infectious bronchitis virus	IBV	<i>Avian coronavirus</i>
KF793824	Bottlenose dolphin coronavirus HKU22	BdCoV_HKU22	<i>Beluga whale coronavirus SW1</i>
JQ065045	Sparrow coronavirus HKU17	SpCoV_HKU17	to be established
FJ376622	Munia coronavirus HKU13-3514	MuCoV_HKU13	<i>Munia coronavirus HKU13</i>

**Table S1:** GenBank accession number, name, and acronym of each virus genome used for the bioinformatics analyses. (continued)

Accession number	Virus name	Acronym	Species
JQ065049	Common-moorhen coronavirus HKU21	CMCoV_HKU21	to be established
FJ376619.2	Bulbul coronavirus HKU11-934	BuCoV_HKU11	<i>Bulbul coronavirus HKU11</i>
FJ376621	Thrush coronavirus HKU12-600	ThCoV_HKU12	<i>Thrush coronavirus HKU12</i>
JQ065044	White-eye coronavirus HKU16	WECoV_HKU16	to be established
JQ065047	Night-heron coronavirus HKU19	NHCoV_HKU19	to be established
JQ065048	Wigeon coronavirus HKU20	WiCoV_HKU20	to be established
NC_022787	Porcine torovirus	PToV_SH1	<i>Porcine torovirus</i>
AY427798	Breda virus	BRV-1	<i>Bovine torovirus</i>
DQ898157	White bream virus	WBV	<i>White bream virus</i>
GU002364.2	Fathead minnow nidovirus	FHMNV	to be established
NC_024709	Ball python nidovirus	BPNV	to be established
JN116253	Possum nidovirus	WPDV	to be established
AF180391	Simian hemorrhagic fever virus	SHFV-LVR	<i>Simian hemorrhagic fever virus</i>
JX473847	Simian hemorrhagic fever virus	SHFV-krtg1	to be established
JX473848	Simian hemorrhagic fever virus	SHFV-krtg2	to be established
HQ845737	Simian hemorrhagic fever virus	SHFV-krc1	to be established
JX138233	Porcine reproductive and respiratory syndrome virus	PRRSV-2	<i>Porcine reproductive and respiratory syndrome virus</i>
GU737264.2	Porcine reproductive and respiratory syndrome virus	PRRSV-1	<i>Porcine reproductive and respiratory syndrome virus</i>
L13298	Lactate dehydrogenase-elevating virus	LDV-C	<i>Lactate dehydrogenase-elevating virus</i>
U15146	Lactate dehydrogenase-elevating virus	LDV-P	<i>Lactate dehydrogenase-elevating virus</i>
DQ846750	Equine arteritis virus	EAV-VBS	<i>Equine arteritis virus</i>

**Table S2:** GenTHREADER comparisons of nidovirus nsp9/nsp12(t) with known RdRps

query: Arteriviridae NiRAN (alignment of nsp9, columns 1-223, first sequence JN116253)						
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)	
3t3l (chain A)	1 (top hit)	31,382	0.034	LOW, NA, GUESS, NA	29-171	
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>		
1-121	121	Homo sapiens		Mitochondrial friedreich ataxia protein		
query: Coronaviridae NiRAN (alignment of nsp12, columns 1-310, first sequence DQ412042)						
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)	
1e8y (chain A)	1 (top hit)	35,365	0.014	NA, LOW, GUESS, NA	14-310	
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>		
517-790	841	Homo sapiens		Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform		
query: Mesoniviridae NiRAN (alignment of nsp12t, columns 1-238, first sequence HM746600)						
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)	
3s44 (chain A)	1 (top hit)	32,68	0.025	GUESS, NA, LOW, NA	1-238	
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>		
22-276	388	Pasteurella multocida		Alpha-2,3,2,6-sialyltransferase/sialidase		

**Table S2:** GenTHREADER comparisons of nidovirus nsp9/nsp12(t) with known RdRps (continued)

target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)
1usu (chain A)	1 (top hit)	27,856	0,078	GUESS, NA, NA, LOW	1-211
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>	
4-246	246	Saccharomyces cerevisiae		ATP-dependent molecular chaperone HSP82	
<b>query: Arteriviridae RdRp (alignment of nsp9, columns 224-727, first sequence JN116253)</b>					
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)
2ckw (chain A)	1 (top hit)	73,964	2,00E-06	CERT, CERT, NA, MEDIUM	263-727
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>	
1-486	487	Sapporo virus			
<b>query: Coronaviridae RdRp (alignment of nsp12, columns 311-1012, first sequence DQ412042)</b>					
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)
3uqs (chain A)	1 (top hit)	73,91	2,00E-06	CERT, CERT, LOW, MEDIUM	483-965
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>	
1-478	478	Murine norovirus 1			

**Table S2:** GenTHREADER comparisons of nidovirus nsp9/nsp12(t) with known RdRps (continued)

target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)
4m5d (chain A)	1 (top hit)	62,877	2,00E-05	NA, MEDIUM, CERT, NA	239-1054
3bso (chain A)	71 (top +ssRNA virus hit)	46,01	0,001	CERT, CERT, MEDIUM, NA	536-987
Coordinates on target (aa residues)	Target length	Target species	Target description		
30-881	881	Saccharomyces cerevisiae S288c	U3 small nucleolar RNA-associated protein 22		
1-479	479	Norwalk virus	RdRp		
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)
4ooj (chain A)	1 (top hit)	51,624	0,0003	GUESS, LOW, HIGH, HIGH	253-841
3n6m (chain A)	12 (top +ssRNA virus hit)	44,591	0,002	CERT, CERT, NA, MEDIUM	486-975
Coordinates on target (aa residues)	Target length	Target species	Target description		
83-604	604	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	SidC, interaptin		
1-462	462	Enterovirus A71	RdRp		

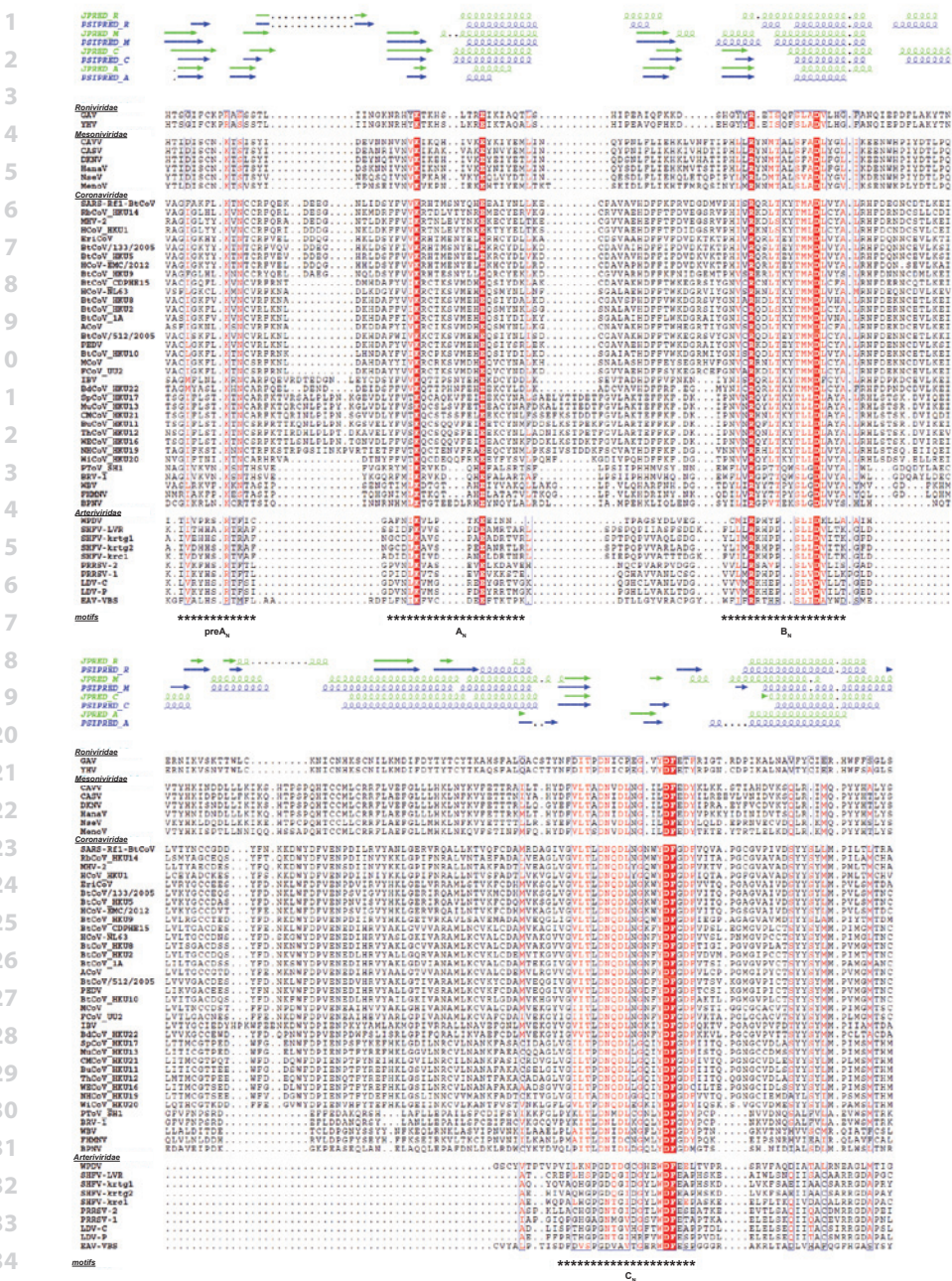
**Table S2:** GenTHREADER comparisons of nidovirus nsp9/nsp12(t) with known RdRps (continued)

query: Arteriviridae NiRAN+RdRp (alignment of nsp9, columns 1-727, first sequence JN116253)						
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)	
2ckw (chain A)	1 (top hit)	73,669	2,00E-06	CERT, CERT, NA, MEDIUM	263-727	
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>		
1-486	487	Sapporo virus		RdRp		
query: Coronaviridae NiRAN+RdRp (alignment of nsp12, columns 1-1012, first sequence DQ412042)						
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)	
3uqs (chain A)	1 (top hit)	73,16	2,00E-06	CERT, CERT, LOW, MEDIUM	483-965	
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>		
1-478	478	Murine norovirus 1		RdRp		
query: Mesoniviridae NiRAN+RdRp (alignment of nsp12, columns 1-1103, first sequence HM746600)						
target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)	
4cei (chain B)	1 (top hit)	63,048	2,00E-05	NA, NA, CERT, LOW	71-1097	
3bso (chain A)	59 (top +ssRNA virus hit)	47,569	0,0008	CERT, CERT, HIGH, NA	536-987	
<b>Coordinates on target (aa residues)</b>	<b>Target length</b>	<b>Target species</b>		<b>Target description</b>		
1-992	992	Bacillus subtilis subsp. subtilis str. 168		ATP-dependent helicase/deoxyribonuclease subunit B		
1-479	479	Norwalk virus		RdRp		

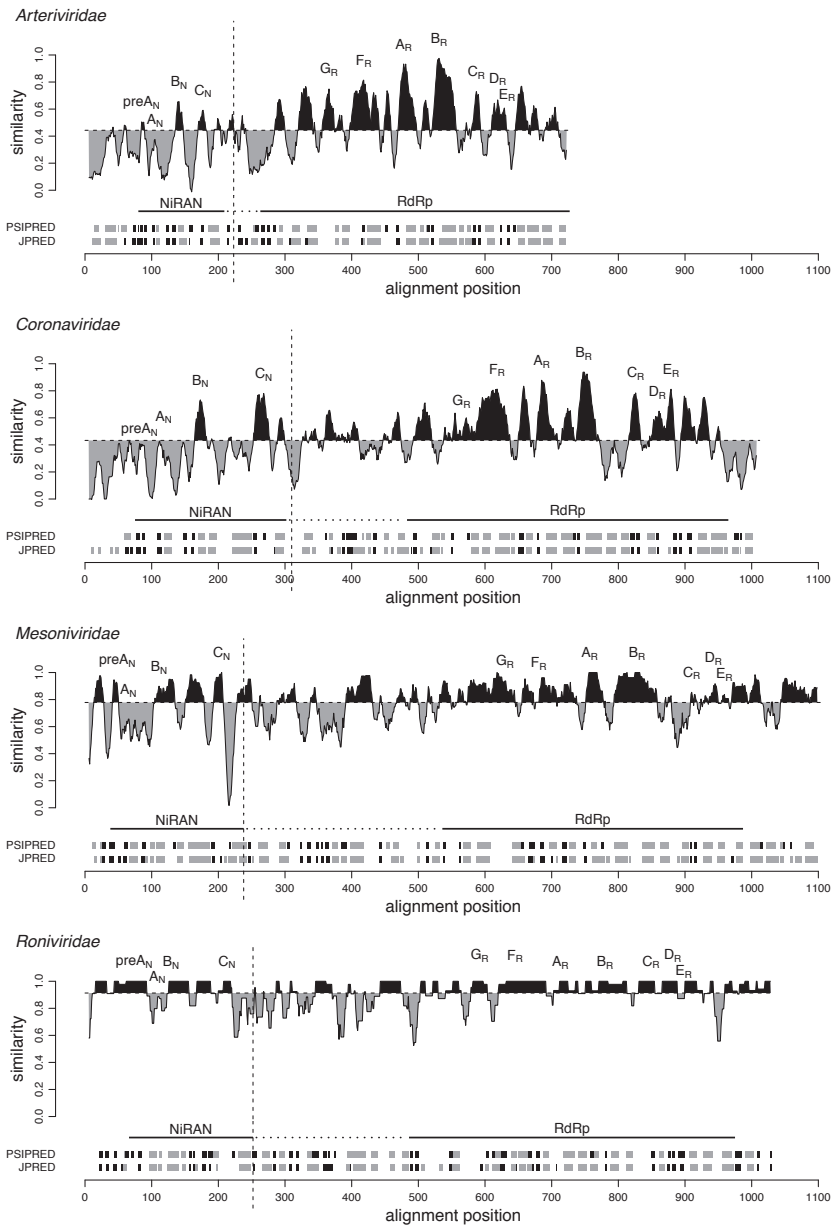
**Table S2:** GenTHREADER comparisons of nidovirus nsp9/nsp12(t) with known RdRps (continued)  
 query: **Roniviridae NiRAN+RdRp (alignment of nsp12t, columns 1-1033, first sequence AF227196)**

target PDB ID	Hit #	Score	P-value	Confidence assigned to the hit (query: Ar, Co, Me, Ro)	Coordinates on query (columns of alignment)
3izx (chain A)	1 (top hit)	60,044	4,00E-05	GUESS, GUESS, CERT, CERT	1-1033
3n6m (chain A)	14 (top +ssRNA virus hit)	43,099	0,002	CERT, CERT, NA, MEDIUM	486-975
Coordinates on target (aa residues)	Target length	Target species	Target description		
100-992	1057	Bombyx mori cypovirus 1	Structural protein VP3		
1-462	462	Enterovirus A71	RdRp		





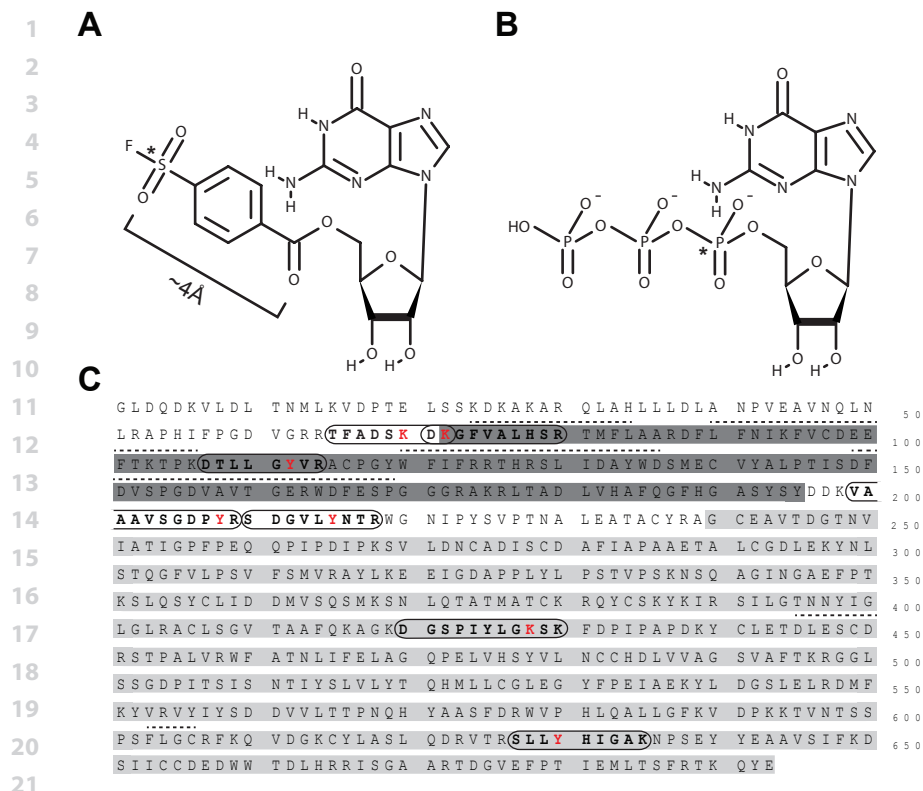
**Figure S1:** Core part of the nidovirus-wide NiRAN MSA encompassing conserved motifs. Virus names and accession numbers are listed in Table S1. Fully and partially conserved residues are depicted in red boxes or red font, respectively. Sequence motifs are indicated by stars. Secondary structure predictions are shown on the top of the MSA. The name of each prediction indicates what software (Jpred 3 (91) or PSIPRED (91)) and which family-specific NiRAN MSA (R, *Roniviridae*; M, *Mesoniviridae*; C, *Coronaviridae*; A, *Arteriviridae*) was used to produce it. The plot was generated with ESPrpt (93).



34 **Figure S2:** Sequence variation, domain organization, and secondary structure of NiRAN-RdRp-containing  
 35 proteins of nidovirus families. For each family, the similarity density plot obtained for the MSA of proteins  
 36 including the NiRAN and RdRp domains is shown. To highlight the regional deviation of conservation from  
 37 that of the MSA average, areas above and below the mean similarity are shaded in black and gray, respec-  
 38 tively. Sequence motifs of NiRAN and RdRp are labelled. Uncertainty in respect to the domain boundary  
 39 between NiRAN and RdRp is indicated by dashed horizontal lines. Domain boundaries used for all bioinfor-  
 matics analyses are indicated by dashed vertical lines. Below each similarity density plot predicted second-  
 ary structure elements are presented in gray for  $\alpha$ -helices and black for  $\beta$ -strands.

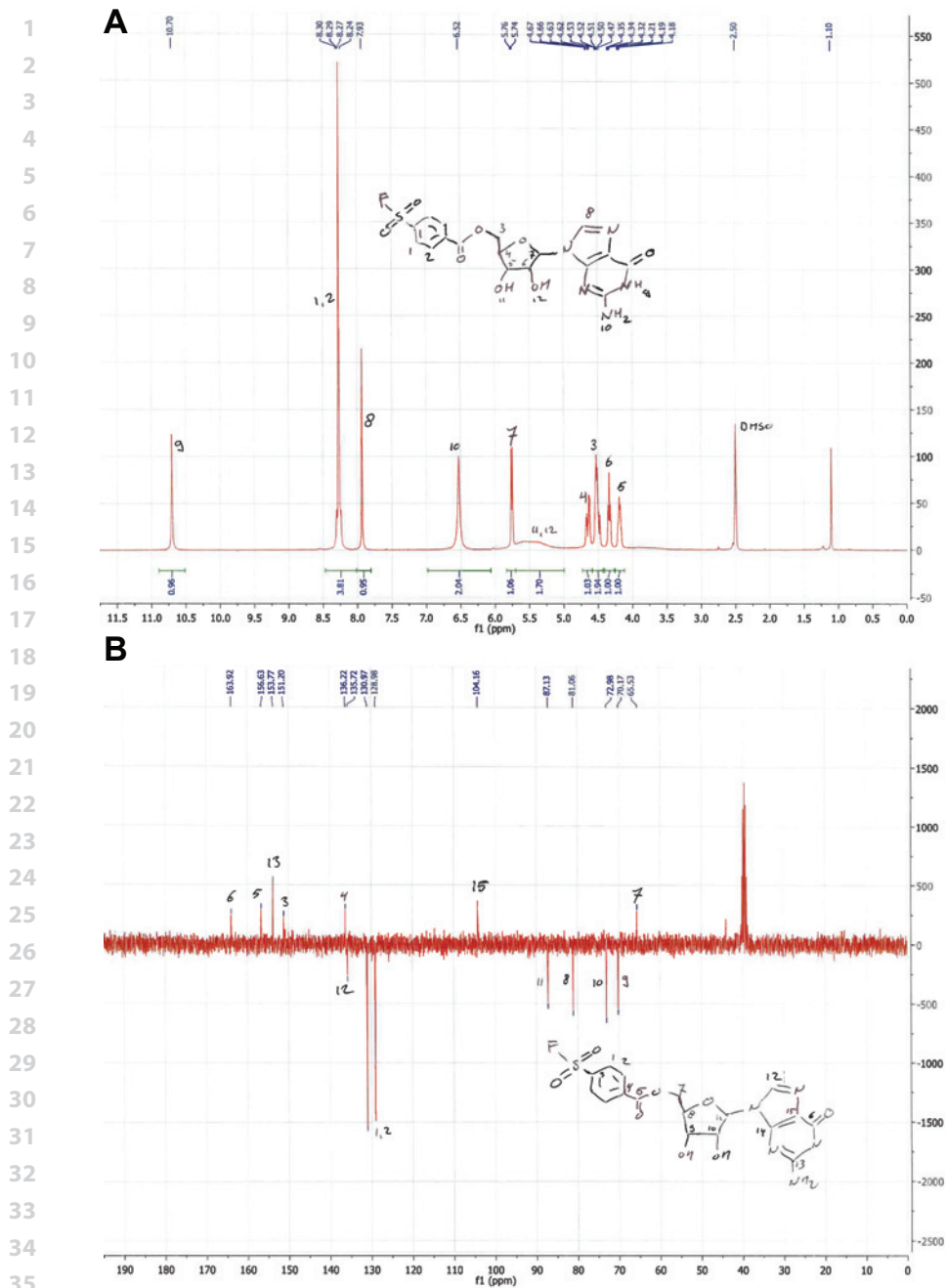


**Figure S3:** Pairwise MSA-based HMM-HMM comparison of NiRANs of different origins. Each MSA of NiRAN was converted to an HMM profile, all possible pairs of obtained HMMs were aligned with the help of HH-suite 2.0.15 software (87,88). Information about each HMM-HMM comparison is presented in a pseudo-symmetrical matrix whose row (left) and column (top) label specifies the group of viruses used as query and target, respectively. Below each dot-plot the probability of the target being homologous to the query and the E value of all aligned pairs of match states are shown in black and green, respectively.



22 **Figure S4:** (A) FSBG and (B) GTP structures indicating the spatial separation of the points of attack in FSBG  
23 and GTP. Asterisks mark the positions of the nucleophilic attack. (C) Mass spectrometry analysis of FSBG-  
24 linked EAV nsp9 identified seven unique, modified peptides (outlined) located either in vicinity of the Ni-  
25 RAN (dark gray background) or within the C-terminal RdRp domain (light gray background). Residues carry-  
26 ing the sulfonylbenzoyl modification are colored in red. Sequence or structural motifs are indicated by  
27 dashed lines above the sequence in the order preA<sub>N</sub>, A<sub>N</sub>, B<sub>N</sub>, C<sub>N</sub>, A<sub>R</sub>, and E<sub>R</sub>. See also Figure 2A.

28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39



**Figure S5:** NMR analysis of 5'-(4-fluorosulfonylbenzoyl)guanosine. **(A)**  $^1\text{H}$  NMR (300 MHz,  $\text{DMSO-d}_6$ )  $\delta$  10.70 (s, 1H), 8.38 – 8.12 (m, 4H), 7.93 (s, 1H), 6.52 (broad s, 2H), 5.75 (d,  $J = 4.8$  Hz, 1H), 5.75 (broad s, 2H), 4.65 (dd,  $J = 11.9, 3.6$  Hz, 1H), 4.59 – 4.42 (m, 2H), 4.34 (t,  $J = 5.1$  Hz, 1H), 4.25 – 4.12 (m, 1H). **(B)**  $^{13}\text{C}$  NMR (75 MHz,  $\text{DMSO-d}_6$ )  $\delta$  163.92, 156.63, 153.77, 151.20, 136.22, 135.72, 130.97, 128.98, 104.16, 87.13, 81.06, 72.98, 70.17, 65.53. Corresponding peaks and atoms are indicated by numbers.

## 1 REFERENCE LIST

- 2
- 3 1. de Groot RJ, Baker SC, Baric R, *et al.* Family *Coronaviridae*. In King AMQ, Adams MJ, Carstens EB *et al.* editors, *Virus Taxonomy*, Ninth Report of the International Committee on Taxonomy of Viruses, Oxford, Elsevier, 2012;806-828.
- 4
- 5
- 6 2. Lauber C, Ziebuhr J, Junglen S, *et al.* *Mesoniviridae*: a proposed new family in the order Nidovirales formed by a single species of mosquito-borne viruses. *Arch.Virol.* 2012;157(8):1623-1628.
- 7
- 8 3. Neumann EJ, Kliebenstein JB, Johnson CD, *et al.* Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. *J.Am.Vet.Med. Assoc.* 2005;227(3):385-392.
- 9
- 10 4. Coleman CM, Frieman MB. Coronaviruses: important emerging human pathogens. *J.Virol.* 2014; 88(10):5209-5212.
- 11
- 12 5. Hilgenfeld R, Peiris M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res.* 2013;100(1):286-295.
- 13
- 14 6. Lauber C, Goeman JJ, Parquet MC, *et al.* The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS.Pathog.* 2013;9(7):e1003500.
- 15
- 16 7. Snijder EJ, Siddell SG, Gorbalenya AE. The order *Nidovirales*. In Mahy BW, ter Meulen V, editors, *Topley and Wilson's Microbiology and Microbial Infections: Virology Volume*, London, Hodder Arnold, 2005;390-404.
- 17
- 18 8. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J.Gen.Virol.* 2006;87(Pt 6):1403-1421.
- 19
- 20 9. Gorbalenya AE, Enjuanes L, Ziebuhr J, *et al.* *Nidovirales*: evolving the largest RNA virus genome. *Virus Res.* 2006;117(1):17-37.
- 21
- 22 10. Subissi L, Imbert I, Ferron F, *et al.* SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. *Antiviral Res.* 2014;101:122-130.
- 23
- 24 11. Lehmann KC, Snijder EJ, Posthuma CC, *et al.* What we know but do not understand about nidovirus helicases. *Virus Res.* 2014;(in press)
- 25
- 26 12. Denison MR, Graham RL, Donaldson EF, *et al.* Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA.Biol.* 2011;8(2):270-279.
- 27
- 28 13. Bouvet M, Imbert I, Subissi L, *et al.* RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc.Natl.Acad.Sci.U.S.A* 2012;109(24):9372-9377.
- 29
- 30 14. Minskaia E, Hertzig T, Gorbalenya AE, *et al.* Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc.Natl.Acad.Sci.U.S.A* 2006;103(13):5108-5113.
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39

- 1 15. Bouvet M, Debarnot C, Imbert I, *et al.* *In vitro* reconstitution of SARS-coronavirus mRNA cap  
2 methylation. *PLoS.Pathog.* 2010;6(4):e1000863.
- 3 16. Chen Y, Cai H, Pan J, *et al.* Functional screen reveals SARS coronavirus nonstructural protein nsp14  
4 as a novel cap N7 methyltransferase. *Proc.Natl.Acad.Sci.U.S.A* 2009;106(9):3484-3489.
- 5 17. Ivanov KA, Hertzog T, Rozanov M, *et al.* Major genetic marker of nidoviruses encodes a replicative  
6 endoribonuclease. *Proc.Natl.Acad.Sci.U.S.A* 2004;101(34):12694-12699.
- 7 18. Nedialkova DD, Ulferts R, van den Born E, *et al.* Biochemical characterization of arterivirus non-  
8 structural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease.  
9 *J.Virol.* 2009;83(11):5671-5682.
- 10 19. Chen Y, Su C, Ke M, *et al.* Biochemical and structural insights into the mechanisms of SARS coro-  
11 navirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS.Pathog.* 2011;7(10):  
12 e1002294.
- 13 20. Daffis S, Szretter KJ, Schriewer J, *et al.* 2'-O methylation of the viral mRNA cap evades host restric-  
14 tion by IFIT family members. *Nature* 2010;468(7322):452-456.
- 15 21. Decroly E, Imbert I, Coutard B, *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding  
16 enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J.Virol.* 2008;82(16):8071-8084.
- 17 22. Nga PT, Parquet MC, Lauber C, *et al.* Discovery of the first insect nidovirus, a missing evolutionary  
18 link in the emergence of the largest RNA virus genomes. *PLoS.Pathog.* 2011;7(9):e1002215.
- 19 23. Gorbalenya AE. Big nidovirus genome. When count and order of domains matter. *Adv.Exp.Med.*  
20 *Biol.* 2001;49:41-17.
- 21 24. Gibrat JF, Mariadassou M, Boudinot P, *et al.* Analyses of the radiation of birnaviruses from diverse  
22 host phyla and of their evolutionary affinities with other double-stranded RNA and positive  
23 strand RNA viruses using robust structure-based multiple sequence alignments and advanced  
24 phylogenetic methods. *BMC.Evol.Biol.* 2013;13:154.
- 25 25. Cerny J, Cerna BB, Valdes JJ, *et al.* Evolution of tertiary structure of viral RNA dependent polymer-  
26 ases. *PLoS.One.* 2014;9(5):e96070.
- 27 26. Kadare G, Haenni AL. Virus-encoded RNA helicases. *J.Virol.* 1997;71(4):2583-2590.
- 28 27. Ng KK, Arnold JJ, Cameron CE. Structure-function relationships among RNA-dependent RNA  
29 polymerases. *Curr.Top.Microbiol.Immunol.* 2008;320:137-156.
- 30 28. Azzi A, Lin SX. Human SARS-coronavirus RNA-dependent RNA polymerase: activity determinants  
31 and nucleoside analogue inhibitors. *Proteins* 2004;57(1):12-14.
- 32 29. Xu X, Liu Y, Weiss S, *et al.* Molecular model of SARS coronavirus polymerase: implications for  
33 biochemical functions and drug design. *Nucleic Acids Res.* 2003;31(24):7117-7130.
- 34  
35  
36  
37  
38  
39

- 1 30. Snijder EJ, Bredenbeek PJ, Dobbe JC, *et al.* Unique and conserved features of genome and pro-  
2 teome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J.Mol.Biol.*  
3 2003;331(5):991-1004.
- 4 31. Ahn DG, Choi JK, Taylor DR, *et al.* Biochemical characterization of a recombinant SARS coronavirus  
5 nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch.Virol.* 2012;  
6 157(11):2095-2104.
- 7 32. Bautista EM, Faaberg KS, Mickelson D, *et al.* Functional properties of the predicted helicase of  
8 porcine reproductive and respiratory syndrome virus. *Virology* 2002;298(2):258-270.
- 9 33. Beerens N, Selisko B, Ricagno S, *et al.* De novo initiation of RNA synthesis by the arterivirus RNA-  
10 dependent RNA polymerase. *J.Virol.* 2007;81(16):8384-8395.
- 11 34. Ivanov KA, Thiel V, Dobbe JC, *et al.* Multiple enzymatic activities associated with severe acute  
12 respiratory syndrome coronavirus helicase. *J.Virol.* 2004;78(11):5619-5632.
- 13 35. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of  
14 duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J.Virol.* 2004;  
15 78(14):7833-7838.
- 16 36. Seybert A, van Dinten LC, Snijder EJ, *et al.* Biochemical characterization of the equine arteritis  
17 virus helicase suggests a close functional relationship between arterivirus and coronavirus heli-  
18 cases. *J.Virol.* 2000;74(20):9586-9593.
- 19 37. Seybert A, Posthuma CC, van Dinten LC, *et al.* A complex zinc finger controls the enzymatic activi-  
20 ties of nidovirus helicases. *J.Virol.* 2005;79(2):696-704.
- 21 38. te Velthuis AJ, van den Worm SH, Sims AC, *et al.* Zn(2+) inhibits coronavirus and arterivirus RNA  
22 polymerase activity *in vitro* and zinc ionophores block the replication of these viruses in cell  
23 culture. *PLoS.Pathog.* 2010;6(11):e1001176.
- 24 39. van Dinten LC, van Tol H, Gorbalenya AE, *et al.* The predicted metal-binding region of the arteri-  
25 virus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion  
26 biogenesis. *J.Virol.* 2000;74(11):5213-5223.
- 27 40. Eckerle LD, Lu X, Sperry SM, *et al.* High fidelity of murine hepatitis virus replication is decreased in  
28 nsp14 exoribonuclease mutants. *J.Virol.* 2007;81(22):12135-12144.
- 29 41. Posthuma CC, Nedialkova DD, Zevenhoven-Dobbe JC, *et al.* Site-directed mutagenesis of the  
30 Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life  
31 cycle. *J.Virol.* 2006;80(4):1653-1661.
- 32 42. Zust R, Cervantes-Barragan L, Habjan M, *et al.* Ribose 2'-O-methylation provides a molecular  
33 signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat.*  
34 *Immunol.* 2011;12(2):137-143.
- 35  
36  
37  
38  
39



- 1 43. Ziebuhr J, Bayer S, Cowley JA, *et al.* The 3C-like proteinase of an invertebrate nidovirus links  
2 coronavirus and potyvirus homologs. *J.Virol.* 2003;77(2):1415-1426.
- 3 44. Blanck S, Stinn A, Tsiklauri L, *et al.* Characterization of an alphamesonivirus 3C-like protease  
4 defines a special group of nidovirus main proteases. *J.Virol.* 2014;88(23):13747-13758.
- 5 45. Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-  
6 242.
- 7 46. Finn RD, Bateman A, Clements J, *et al.* Pfam: the protein families database. *Nucleic Acids Res.*  
8 2014;42(Database issue):D222-230.
- 9 47. Bartlett GJ, Porter CT, Borkakoti N, *et al.* Analysis of catalytic residues in enzyme active sites. *J.Mol.*  
10 *Biol.* 2002;324(1):105-121.
- 11 48. Henderson BR, Saeedi BJ, Campagnola G, *et al.* Analysis of RNA binding by the dengue virus NS5  
12 RNA capping enzyme. *PLoS.One.* 2011;6(10):e25795.
- 13 49. Shuman S, Schwer B. RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl  
14 transferases. *Mol.Microbiol.* 1995;17(3):405-410.
- 15 50. Shuman S, Lima CD. The polynucleotide ligase and RNA capping enzyme superfamily of covalent  
16 nucleotidyltransferases. *Curr.Opin.Struct.Biol.* 2004;14(6):757-764.
- 17 51. Traut TW. Physiological concentrations of purines and pyrimidines. *Mol.Cell Biochem.* 1994;  
18 140(1):1-22.
- 19 52. Hanouille X, Van DJ, Staes A, *et al.* A new functional, chemical proteomics technology to identify  
20 purine nucleotide binding sites in complex proteomes. *J.Proteome.Res.* 2006;5(12):3438-3445.
- 21 53. Chakravarty AK, Subbotin R, Chait BT, *et al.* RNA ligase RtcB splices 3'-phosphate and 5'-OH ends  
22 via covalent RtcB-(histidinyl)-GMP and polynucleotide-(3')pp(5')G intermediates. *Proc.Natl.Acad.*  
23 *Sci.U.S.A* 2012;109(16):6072-6077.
- 24 54. Duclos B, Marcandier S, Cozzone AJ. Chemical properties and separation of phosphoamino acids  
25 by thin-layer chromatography and/or electrophoresis. *Methods Enzymol.* 1991;201:10-21.
- 26 55. Kang H, Bhardwaj K, Li Y, *et al.* Biochemical and genetic analyses of murine hepatitis virus Nsp15  
27 endoribonuclease. *J.Virol.* 2007;81(24):13587-13597.
- 28 56. te Velhuis AJ, Arnold JJ, Cameron CE, *et al.* The RNA polymerase activity of SARS-coronavirus  
29 nsp12 is primer dependent. *Nucleic Acids Res.* 2010;38(1):203-214.
- 30 57. Kumar NV, Govil G. Theoretical studies on protein-nucleic acid interactions. III. Stacking of  
31 aromatic amino acids with bases and base pairs of nucleic acids. *Biopolymers* 1984;23(10):2009-  
32 2024.
- 33  
34  
35  
36  
37  
38  
39

- 1 58. Bartlett GJ, Borkakoti N, Thornton JM. Catalysing new reactions during evolution: economy of  
2 residues and mechanism. *J.Mol.Biol.* 2003;331(4):829-860.
- 3 59. Schmelz S, Naismith JH. Adenylate-forming enzymes. *Curr.Opin.Struct.Biol.* 2009;19(6):666-671.
- 4 60. Ahola T, Laakkonen P, Vihinen H, *et al.* Critical residues of Semliki Forest virus RNA capping enzyme  
5 involved in methyltransferase and guanylyltransferase-like activities. *J.Virol.* 1997;71(1):392-397.
- 6 61. Nandakumar J, Shuman S, Lima CD. RNA ligase structures reveal the basis for RNA specificity and  
7 conformational changes that drive ligation forward. *Cell* 2006;127(1):71-84.
- 8 62. Decroly E, Ferron F, Lescar J, *et al.* Conventional and unconventional mechanisms for capping viral  
9 mRNA. *Nat.Rev.Microbiol.* 2012;10(1):51-65.
- 10 63. Bouvet M, Debarnot C, Imbert I, *et al.* *In vitro* reconstitution of SARS-coronavirus mRNA cap  
11 methylation. *PLoS.Pathog.* 2010;6(4):e1000863.
- 12 64. Lai MM, Patton CD, Stohman SA. Further characterization of mRNA's of mouse hepatitis virus:  
13 presence of common 5'-end nucleotides. *J.Virol.* 1982;41(2):557-565.
- 14 65. van Vliet AL, Smits SL, Rottier PJ, *et al.* Discontinuous and non-discontinuous subgenomic RNA  
15 transcription in a nidovirus. *EMBO J.* 2002;21(23):6571-6580.
- 16 66. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion  
17 RNA. *Virology* 1986;151(1):146-150.
- 18 67. Issur M, Geiss BJ, Bougie I, *et al.* The flavivirus NS5 protein is a true RNA guanylyltransferase that  
19 catalyzes a two-step reaction to form the RNA cap structure. *RNA.* 2009;15(12):2340-2350.
- 20 68. Ahola T, Ahlquist P. Putative RNA capping activities encoded by brome mosaic virus: methylation  
21 and covalent binding of guanylate by replicase protein 1a. *J.Virol.* 1999;73(12):10061-10069.
- 22 69. Gorbalenya AE, Pringle FM, Zeddam JL, *et al.* The palm subdomain-based active site is internally  
23 permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J.Mol.Biol.* 2002;324(1):  
24 47-62.
- 25 70. Gorbalenya AE, Koonin EV, Donchenko AP, *et al.* Coronavirus genome: prediction of putative func-  
26 tional domains in the non-structural polyprotein by comparative amino acid sequence analysis.  
27 *Nucleic Acids Res.* 1989;17(12):4847-4861.
- 28 71. Paul AV, Rieder E, Kim DW, *et al.* Identification of an RNA hairpin in poliovirus RNA that serves as  
29 the primary template in the *in vitro* uridylylation of VPg. *J.Virol.* 2000;74(22):10359-10370.
- 30 72. Ambros V, Baltimore D. Protein is linked to the 5' end of poliovirus RNA by a phosphodiester  
31 linkage to tyrosine. *J.Biol.Chem.* 1978;253(15):5263-5266.
- 32 73. Pan J, Lin L, Tao YJ. Self-guanylylation of birnavirus VP1 does not require an intact polymerase  
33 activity site. *Virology* 2009;395(1):87-96.
- 34  
35  
36  
37  
38  
39

- 1 74. Mitra T, Sosnovtsev SV, Green KY. Mutagenesis of tyrosine 24 in the VPg protein is lethal for feline  
2 calicivirus. *J.Virol.* 2004;78(9):4931-4935.
- 3 75. Jiang J, Laliberte JF. The genome-linked protein VPg of plant viruses-a protein with many part-  
4 ners. *Curr.Opin.Virol.* 2011;1(5):347-354.
- 5 76. Zeddarn JL, Gordon KH, Lauber C, *et al.* Euprosterna elaeasa virus genome sequence and evolu-  
6 tion of the Tetraviridae family: emergence of bipartite genomes and conservation of the VPg  
7 signal with the dsRNA Birnaviridae family. *Virology* 2010;397(1):145-154.
- 8 77. Imbert I, Guillemot JC, Bourhis JM, *et al.* A second, non-canonical RNA-dependent RNA poly-  
9 merase in SARS coronavirus. *EMBO J.* 2006;25(20):4933-4942.
- 10 78. Subissi L, Posthuma CC, Collet A, *et al.* One severe acute respiratory syndrome coronavirus pro-  
11 tein complex integrates processive RNA polymerase and exonuclease activities. *Proc.Natl.Acad.*  
12 *Sci.U.S.A* 2014;111(37):e3900
- 13 79. Benson DA, Cavanaugh M, Clark K, *et al.* GenBank. *Nucleic Acids Res.* 2013;41(Database issue):  
14 D36-42.
- 15 80. Pruitt KD, Brown GR, Hiatt SM, *et al.* RefSeq: an update on mammalian reference sequences.  
16 *Nucleic Acids Res.* 2014;42(Database issue):D756-763.
- 17 81. Lauber C, Gorbalenya AE. Partitioning the genetic diversity of a virus family: approach and evalu-  
18 ation through a case study of picornaviruses. *J.Virol.* 2012;86(7):3890-3904.
- 19 82. Sidorov IA, Reshetov DA, Gorbalenya AE. SNAD: Sequence Name Annotation-based Designer.  
20 *BMC.Bioinformatics.* 2009;10:251.
- 21 83. Gorbalenya AE, Lieutaud P, Harris MR, *et al.* Practical application of bioinformatics by the multidis-  
22 ciplinary VIZIER consortium. *Antiviral Res.* 2010;87(2):95-110.
- 23 84. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching.  
24 *Nucleic Acids Res.* 2011;39(Web Server issue):W29-37.
- 25 85. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput.  
26 *Nucleic Acids Res.* 2004;32(5):1792-1797.
- 27 86. Larkin MA, Blackshields G, Brown NP, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics.*  
28 2007;23(21):2947-2948.
- 29 87. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;21(7):  
30 951-960.
- 31 88. Remmert M, Biegert A, Hauser A, *et al.* HHblits: lightning-fast iterative protein sequence searching  
32 by HMM-HMM alignment. *Nat.Methods* 2012;9(2):173-175.
- 33  
34  
35  
36  
37  
38  
39

- 1 89. Grant BJ, Rodrigues AP, ElSawy KM, *et al.* Bio3d: an R package for the comparative analysis of  
2 protein structures. *Bioinformatics*. 2006;22(21):2695-2696.
- 3 90. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc.Natl.Acad.*  
4 *Sci.U.S.A* 1992;89(22):10915-10919.
- 5 91. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*  
6 2008;36(Web Server issue):W197-201.
- 7 92. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices.  
8 *J.Mol.Biol.* 1999;292(2):195-202.
- 9 93. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server.  
10 *Nucleic Acids Res.* 2014;42(Web Server issue):W320-324.
- 11 94. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*  
12 2015;43(Database issue):D6-17.
- 13 95. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic  
14 sequences. *J.Mol.Biol.* 1999;287(4):797-815.
- 15 96. McGuffin LJ, Jones DT. Improvement of the GenTHREADER method for genomic fold recognition.  
16 *Bioinformatics*. 2003;19(7):874-881.
- 17 97. Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: new methods for  
18 improved protein fold recognition and superfamily discrimination. *Bioinformatics*. 2009;25(14):  
19 1761-1767.
- 20 98. van den Born E, Gulyaev AP, Snijder EJ. Secondary structure and function of the 5'-proximal  
21 region of the equine arteritis virus RNA genome. *RNA*. 2004;10(3):424-437.
- 22 99. Nedialkova DD, Gorbalenya AE, Snijder EJ. Arterivirus Nsp1 modulates the accumulation of  
23 minus-strand templates to control the relative abundance of viral mRNAs. *PLoS.Pathog.* 2010;  
24 6(2):e1000772.
- 25 100. van der Meer Y, van TH, Locker JK, *et al.* ORF1a-encoded replicase subunits are involved in the  
26 membrane association of the arterivirus replication complex. *J.Virol.* 1998;72(8):6689-6698.
- 27 101. Pfefferle S, Krahling V, Ditt V, *et al.* Reverse genetic characterization of the natural genomic  
28 deletion in SARS-Coronavirus strain Frankfurt-1 open reading frame 7b reveals an attenuating  
29 function of the 7b protein in-vitro and in-vivo. *Virol.J.* 2009;6:131.
- 30 102. Tischer BK, Smith GA, Osterrieder N. En passant mutagenesis: a two step markerless red recombina-  
31 tion system. *Methods Mol.Biol.* 2010;634:421-430.
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

Arterivirus nsp12 versus  
the coronavirus nsp16  
2'-O-methyltransferase: comparison  
of the C-terminal cleavage products of  
two nidovirus pp1ab polyproteins

# CHAPTER 6

Kathleen C. Lehmann  
Lisa Hooghiemstra  
Anastasia Gulyaeva  
Dmitry Samborskiy  
Jessika C. Zevenhoven-Dobbe  
Eric J. Snijder  
Alexander E. Gorbalenya  
and Clara C. Posthuma

*Submitted*

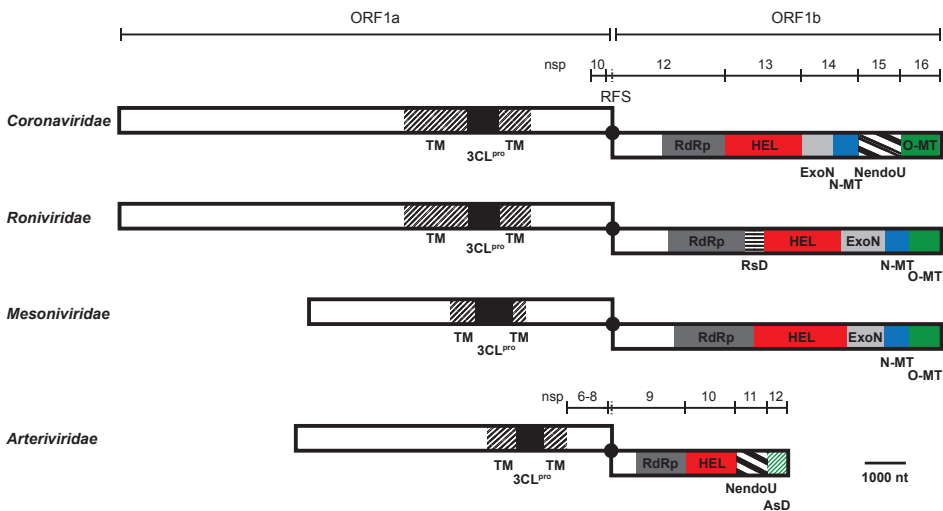
**1 ABSTRACT**

2  
3 The 3'-terminal domain of the most conserved open reading frame 1b (ORF1b) in three  
4 of the four families of the order *Nidovirales* (except the *Arteriviridae*) encodes a (putative)  
5 2'-O-methyltransferase (O-MT), known as nonstructural protein (nsp) 16 in coronaviruses  
6 and implicated in methylation of the 5' cap structure of nidoviral mRNAs. Like coronavi-  
7 rus transcripts, arterivirus mRNAs are assumed to possess a 5' cap although no candidate  
8 methyltransferases (MTases) were identified thus far. To address this knowledge gap, we  
9 analyzed the uncharacterized nsp12 of arteriviruses, which occupies the ORF1b position  
10 equivalent to that of coronavirus nsp16. In our in-depth bioinformatics analysis of nsp12,  
11 the protein was confirmed to be family-specific while having diverged much farther  
12 than other nidovirus ORF1b-encoded proteins, including those of the *Coronaviridae*.  
13 Only one invariant and several partially conserved, predominantly aromatic residues  
14 were identified in nsp12, which may adopt a structure with alternating  $\alpha$ -helices and  
15  $\beta$ -strands, an organization also found in known MTases. However, no statistically signifi-  
16 cant similarity was found between nsp12 and the two-fold larger coronavirus nsp16, nor  
17 could we detect MTase activity in biochemical assays using recombinant equine arteritis  
18 virus nsp12. Our further analysis established that this subunit is essential for replica-  
19 tion of this prototypic arterivirus. Using reverse genetics, we assessed the impact of 25  
20 substitutions at 14 positions, yielding virus phenotypes ranging from wild-type-like to  
21 nonviable. Notably, replacement of the invariant phenylalanine 109 with tyrosine was  
22 lethal. We conclude that nsp12 plays an essential role during EAV replication, possibly by  
23 acting as a co-factor for another enzyme.

24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 1 INTRODUCTION

2  
3 Arteriviruses (family *Arteriviridae*) are positive-stranded RNA viruses with genome sizes  
4 ranging from 13 to 16 kilobases. The family currently comprises a single genus that  
5 includes four species: *Equine arteritis virus* (EAV), *Simian hemorrhagic fever virus* (SHFV),  
6 *Lactate dehydrogenase-elevating virus* (LDV), and *Porcine reproductive and respiratory*  
7 *syndrome virus* (PRRSV) (1;2). Among those, the latter is the economically most relevant  
8 species causing annual losses to the American swine industry alone of about \$800  
9 million (3). Additionally, several recently identified arteriviruses remain to be formally  
10 classified, but are likely to prototype multiple novel species or even higher order taxa  
11 (4-7). Arterivirus genomes are polycistronic and contain 10 to 15 (known) open reading  
12 frames (ORFs). The 5'-proximal ORFs 1a and 1b are expressed as polyproteins (pps) 1a  
13 and 1ab that are autoproteolytically processed into the nonstructural proteins (nsps)  
14 required for genome replication and transcription (Figure 1) (8). The remaining ORFs  
15 mostly encode structural proteins that are expressed from a set of subgenomic (sg)  
16



31 **Figure 1:** Organization of key replicase domains encoded by nidovirus open reading frames (ORFs) 1a  
32 and 1b. Proteolytic cleavage products described in the text for the *Corona*- and *Arteriviridae* are indicated.  
33 Matching colors/patterns indicate domain conservation between families. Domains (putatively) involved  
34 in capping (HEL, N-MT, O-MT, AsD) are depicted in bright colors. nsp, nonstructural protein; TM, transmem-  
35 brane domain; 3CL<sup>pro</sup>, 3C-like protease; black dot and RFS, ribosomal frameshift site; RdRp, RNA-dependent  
36 RNA polymerase; HEL, helicase/RNA triphosphatase; ExoN, exoribonuclease; N-MT, N7-methyltransferase;  
37 NendoU, endoribonuclease; O-MT, 2'-O-methyltransferase; RsD, Ronivirus-specific domain; AsD, arterivirus-  
38 specific domain (nsp12). Genomic organizations are shown for Beluga whale coronavirus SW1 (*Corona*-  
39 *viridae*), gill-associated virus (*Roniviridae*), Nam Dinh virus (*Mesoniviridae*), and porcine respiratory  
and reproductive syndrome virus, North American genotype (*Arteriviridae*). Depicted is a simplified domain or-  
ganization since most enzymes are multidomain proteins. Note that viruses of the *Coronaviridae* family that  
do not belong to the subfamily of *Coronavirinae* encode a truncated version of N-MT. Adapted from (61).



1 mRNAs (9). Based on overall similarities in terms of genome expression and organization  
2 as well as synteny and homology of key replicase domains, arteriviruses were united in  
3 the order *Nidovirales* with the families *Mesoniviridae*, *Roniviridae*, and *Coronaviridae*, the  
4 latter including two distantly related subfamilies, *Coronavirinae* and *Torovirinae* (10;11).  
5 In the nidovirus tree, the arteriviruses form a basal lineage next to the one that combines  
6 the three other families, which have substantially larger genomes (12).

7  
8 ORF1b is the most conserved part of the nidovirus genome, and all ORF1b-encoded  
9 proteins characterized thus far are enzymes conserved in two or more nidovirus families.  
10 The RNA-dependent RNA polymerase and a zinc-binding domain (ZBD) fused with a  
11 superfamily 1 helicase (HEL1) are conserved in all nidoviruses. In contrast, six other do-  
12 mains are lineage specific. Four of these are conserved in two or three nidovirus families  
13 only: exoribonuclease (ExoN), N7-methyltransferase (N-MT), nidovirus uridylate-specific  
14 endoribonuclease (NendoU), and 2'-O-methyltransferase (O-MT). Two other domains  
15 are yet uncharacterized and unique to either roniviruses (RsD, ronivirus-specific domain)  
16 or arteriviruses (AsD, arterivirus-specific domain). Since five of the six lineage-specific  
17 domains occupy a unique position in the genome, the pattern of their conservation  
18 could be explained by loss or acquisition of a single domain during nidovirus evolu-  
19 tion (12). The exception is AsD, which resides in the most C-terminal subunit of the  
20 arterivirus ORF1b polyprotein (nsp12), the position occupied by the O-MT protein in  
21 all other nidoviruses (nsp16 in coronaviruses, Figure 1). If these positionally equivalent  
22 proteins are unrelated, as reported 14 years ago based on the analysis of only a few  
23 genome sequences and prior to the identification of the O-MT (13), their emergence  
24 would require the consideration of complex evolutionary hypotheses. Thus, the relation  
25 of AsD with the O-MT and other proteins must be re-evaluated while taking advantage  
26 of the increased availability of sequences and improved techniques.

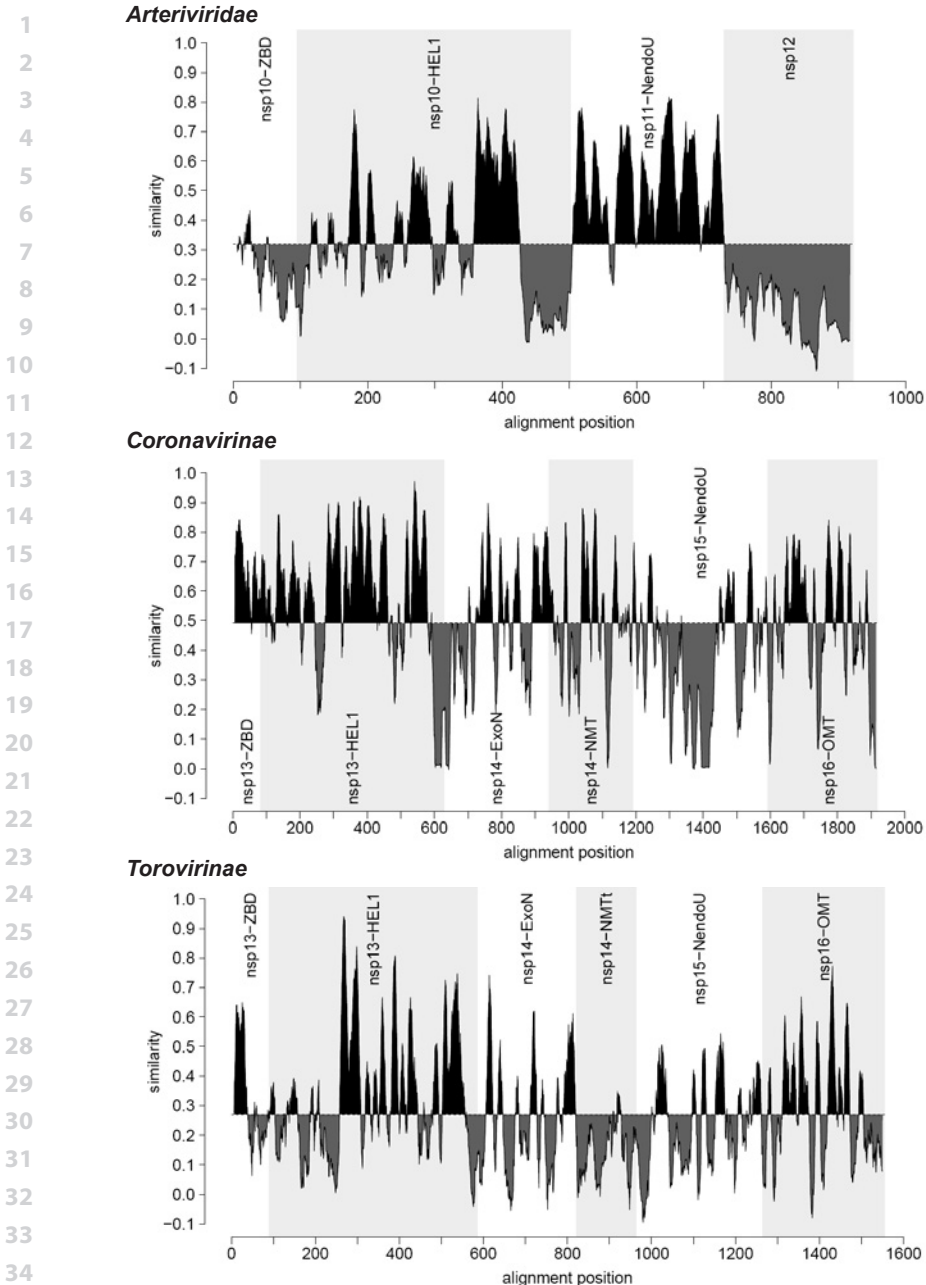
27  
28 Unlike AsD, the coronavirus O-MT has been experimentally characterized (14-17) and  
29 was found to provide one of the four activities required for the formation of a so-called  
30 type I cap (cap-1) (mGpppNm) structure at the 5' end of coronaviral mRNAs (18;19). Two  
31 other coronavirus enzymes, HEL1 (nsp13) (20;21) and the N-MT (nsp14) (14;22), are also  
32 known to be involved in capping, whereas the fourth enzyme required (guanylyltrans-  
33 ferase) remains to be identified. *In vitro* the coronavirus N-MT and O-MT were found to  
34 cooperate during cap formation. The latter enzyme also requires the ORF1a-encoded  
35 nsp10 as a co-factor (14). Although arteriviruses were not characterized in detail, the  
36 SHFV genome was reported to be capped (23), and they do encode a HEL1 (24), which  
37 could contribute to capping. Thus, the discovery of arteriviral N-MT and/or O-MT activi-  
38 ties could be readily accommodated in a functionally sensible manner.

1 Based on the above evolutionary and functional considerations, we sought to character-  
2 ize nsp12 of arteriviruses by testing the hypothesis that it may be a methyltransferase.  
3 We show that, unlike the coronavirus O-MT, nsp12 is poorly conserved among known  
4 arteriviruses compared to the proteins carrying the endoribonuclease (nsp11) and heli-  
5 case (nsp10) activities, and that it contains only one evolutionary invariant residue. No  
6 statistically significant similarity was found between arterivirus nsp12 and coronavirus  
7 nsp16 or other proteins although the two nidovirus proteins may belong to the same  
8  $\alpha/\beta$  fold class. Likewise, no MTase activity was detected in carefully controlled assays  
9 using recombinant EAV nsp12 in the absence or presence of several other nsps that were  
10 included as potential co-factors. Using reverse genetics, a large set of EAV nsp12 mutants  
11 was generated and tested for replication, revealing phenotypes ranging from wild-type-  
12 like to replication-deficient, which broadly correlated with the natural variation of the  
13 probed residues. We conclude that nsp12 plays an essential role in EAV replication and  
14 discuss possible directions to elucidate its enigmatic function.

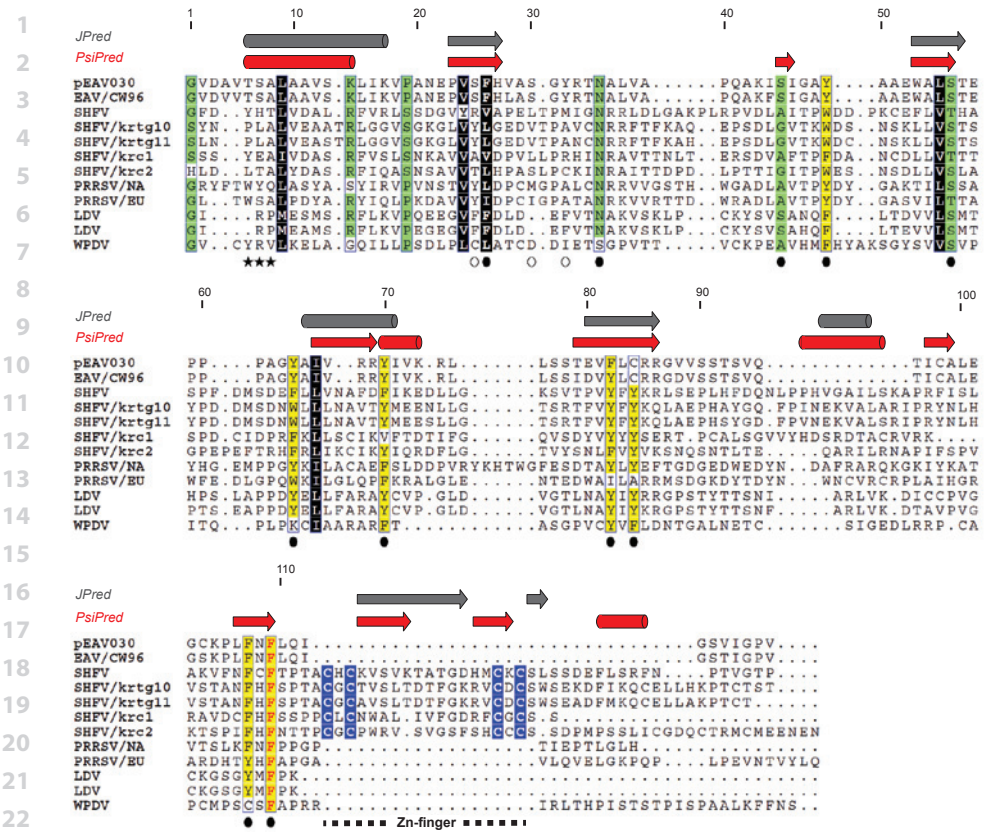
## 15 16 **RESULTS**

### 17 **Sequence similarities and dissimilarities between arterivirus nsp12 and** 18 **(putative) methyltransferases of the *Coronaviridae***

19 We first analyzed the conservation of nsp12 in comparison with that of other proteins  
20 deriving from the C-terminal portion of pp1ab of arteriviruses and the *Coronavirinae*  
21 and *Torovirinae*. Starting at the ZBD, the region analyzed included the three proteins  
22 implicated in 5' cap formation in coronaviruses. We found that nsp12 is conserved in  
23 all established and provisional arterivirus species, including the most distantly related  
24 wobbly possum disease virus (WPDV). Inspection of the arterivirus conservation profile  
25 showed that the entire nsp12 sequence exhibits similarity values that are below average  
26 for this pp1ab region (0.320 on a -0,1-1 scale; Figure 2A). Only the C-terminal domain  
27 of nsp10 and to some extent the ZBD were similarly divergent while the similarity of  
28 the nsp10 helicase core and particularly nsp11 were above average. This remarkably  
29 low conservation distinguishes arterivirus nsp12 also from all proteins in this region  
30 of the *Coronavirinae* (average conservation 0.491) and *Torovirinae* (0.270), including  
31 nsp16 (Figures 2B and C). Accordingly, arterivirus nsp12 contains the smallest number  
32 of conserved residues among the analyzed proteins, with only a single phenylalanine  
33 (F109 in EAV) being evolutionarily invariant (Figure 3). Other notable conserved nsp12  
34 residues (out of 18 in total) are an asparagine, a serine/threonine and six aromatic  
35 residues. We also noted the presence of four conserved cysteines in a pattern typical for  
36 zinc-fingers in the C-terminal part of nsp12 in the five simian arteriviruses, which con-  
37  
38  
39



35 **Figure 2:** Similarity density plots of the C-terminal region of polyprotein 1ab of different nidovirus (sub)  
 36 families. Values above and below average similarities are indicated in black and gray, respectively. nsp,  
 37 nonstructural protein; ZBD, zinc-binding domain; Hel1, helicase core domain; ExoN, exoribonuclease; NMT,  
 38 N7-methyltransferase (t, truncated); NendoU, endoribonuclease; OMT, 2'-O-methyltransferase. For the sake of  
 39 simplicity, we have applied the nsp nomenclature of the *Coronavirinae* subfamily also to the orthologous  
 torovirus domains for which the processing of pp1a/pp1ab is yet to be fully described.



**Figure 3:** Multiple sequence alignment and secondary structure predictions of representative arterivirus nsp12 sequences. Partially and fully conserved amino acids are highlighted in colored boxes. Colors represent residues with similar biophysical properties; yellow, aromatic; black, hydrophobic; blue, (putatively) zinc-binding; green, other. Secondary structures (barrel,  $\alpha$ -helix; arrow,  $\beta$ -strand) were predicted with JPred (50) (gray) or PsiPred (51) (red) based on the multiple sequence alignment. Residue numbers are indicated for nsp12 of the EAV-Bucyrus isolate (pEAV030) (57), the parental strain of pEAV211 used for the reverse genetics experiments. Replaced residues are indicated below the alignment; black stars, positions where stop codons were introduced; empty circles, control residues; filled circles, conserved residues. A putative zinc-finger in simian arterivirus nsp12 sequences is indicated by a dashed line. EAV, equine arteritis virus (GenBank accession number AY349167); SHFV, simian hemorrhagic fever virus (AF180391, JX473847, JX473848, HQ845737, HQ845738); PRRSV, porcine reproductive and respiratory syndrome virus (JX138233, JF802085); LDV, lactate dehydrogenase-elevating virus (L13298, U15146); WPDV, wobbly possum disease virus (JN116253).

ststitute a phylogenetically compact cluster. Patristic pair-wise distances (PPDs) of nsp12 compared to those of ZBD, HEL1, and NendoU were consistently larger while PPDs of (putative) O-MTs were comparable on average with those of five other domains in the *Coronavirinae* and *Torovirinae* (Figure S1). These results showed that, in comparison to the coronavirus O-MT, nsp12 must have evolved under unusually relaxed constraints or

1 in a changing molecular environment. Secondary structure predictions using JPred and  
2 PsiPred consistently indicated the alternation of  $\alpha$ -helices and  $\beta$ -strands in arterivirus  
3 nsp12 (Figure 3). Interestingly, also the coronavirus MTases belong to the  $\alpha/\beta$  structural  
4 class and contain conserved aromatic residues (15;17). Nevertheless, HH-suite profile-  
5 profile comparison did not reveal sequence similarity above the background between  
6 nsp12 and the O-MT of corona- or toroviruses,  $E=0.41$  and  $0.53$ , respectively (Figure S2).  
7 Furthermore, these proteins are also of different sizes: 119-178 aa (arterivirus nsp12)  
8 versus 263-312 aa (coronavirus nsp16), with the arterivirus proteins being also smaller  
9 than MTases of other origins. The above HH-based negative result contrasted with the  
10 strong similarity signal observed in (control) comparisons between arteriviruses and  
11 corona- or toroviruses for HEL1 and NendoU ( $E=3.5e-17$  or better), or in the control  
12 comparison between corona- and torovirus nsp16,  $E=2.3e-32$  (Figure S2). No statistically  
13 significant similarity was observed between nsp12 and other proteins in an HMM-based  
14 scan of the PFAM-A database (top hit: PF12581,  $E=1.0$ ). We thus concluded that nsp12  
15 has diverged beyond recognition from its homologs and differs considerably from the  
16 O-MT of large nidoviruses. Nevertheless, the obtained results did not rule out the possi-  
17 bility that it could be a deviant MTase, and we therefore set out to test this hypothesis  
18 experimentally by biochemical and molecular virological methods.

## 20 **Purification of recombinant EAV nsp12 and several ORF1a-encoded proteins**

22 We engineered vectors encoding recombinant EAV nsp12 derivatives carrying either  
23 an N-terminal or a C-terminal hexahistidine tag and expressed them in *E. coli*. Only  
24 the N-terminally tagged protein was successfully expressed and purified by metal  
25 affinity chromatography using  $Co^{2+}$  (Talon) beads (Figure 4A). The protein appeared  
26 to be reasonably stable at all conditions tested, including a pH range from 6.0 to 7.5  
27 and protein concentrations of up to 500  $\mu M$ . Yet upon storage the protein increasingly  
28 formed dimers and higher order multimers, even in the presence of 1 mM DTT. In gel  
29 filtration experiments with fresh protein these oligomers were not evident. Instead a  
30 single peak was observed (not shown) that corresponded well to the expected size of an  
31 nsp12 monomer (calculated weight 13 kDa vs. predicted weight based on Stokes radius  
32 16 kDa).

34 In addition to nsp12, we also expressed five small mature proteins and cleavage in-  
35 termediates from the nsp7 region of pp1a (nsp6-7, nsp6-7-8, nsp7 $\alpha$ , nsp7 $\beta$ , and nsp7  
36 (i.e., nsp7 $\alpha$ -7 $\beta$ )) (25;26) (Figure 4B). In coronaviruses, the corresponding part of ORF1a  
37 encodes nsp10, an essential co-factor for the O-MT (14). Consequently, we added these  
38 purified recombinant proteins to nsp12 in MTase activity assays (see below).

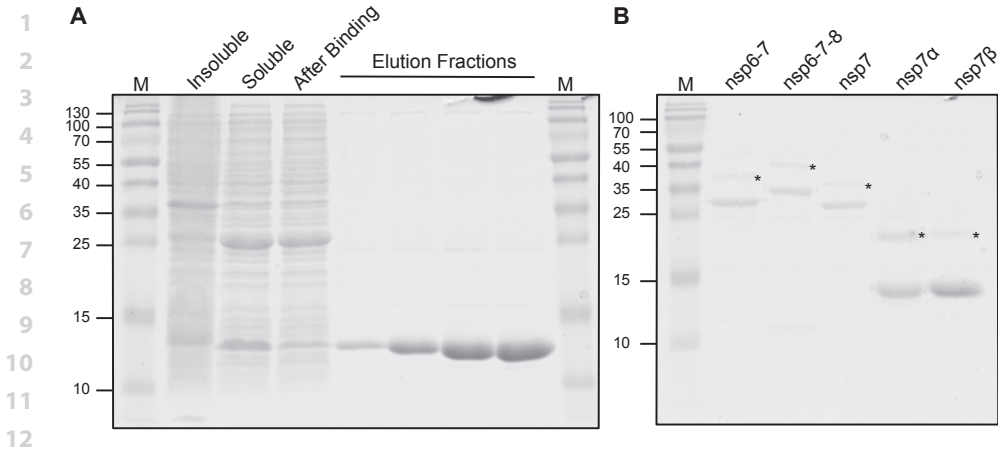
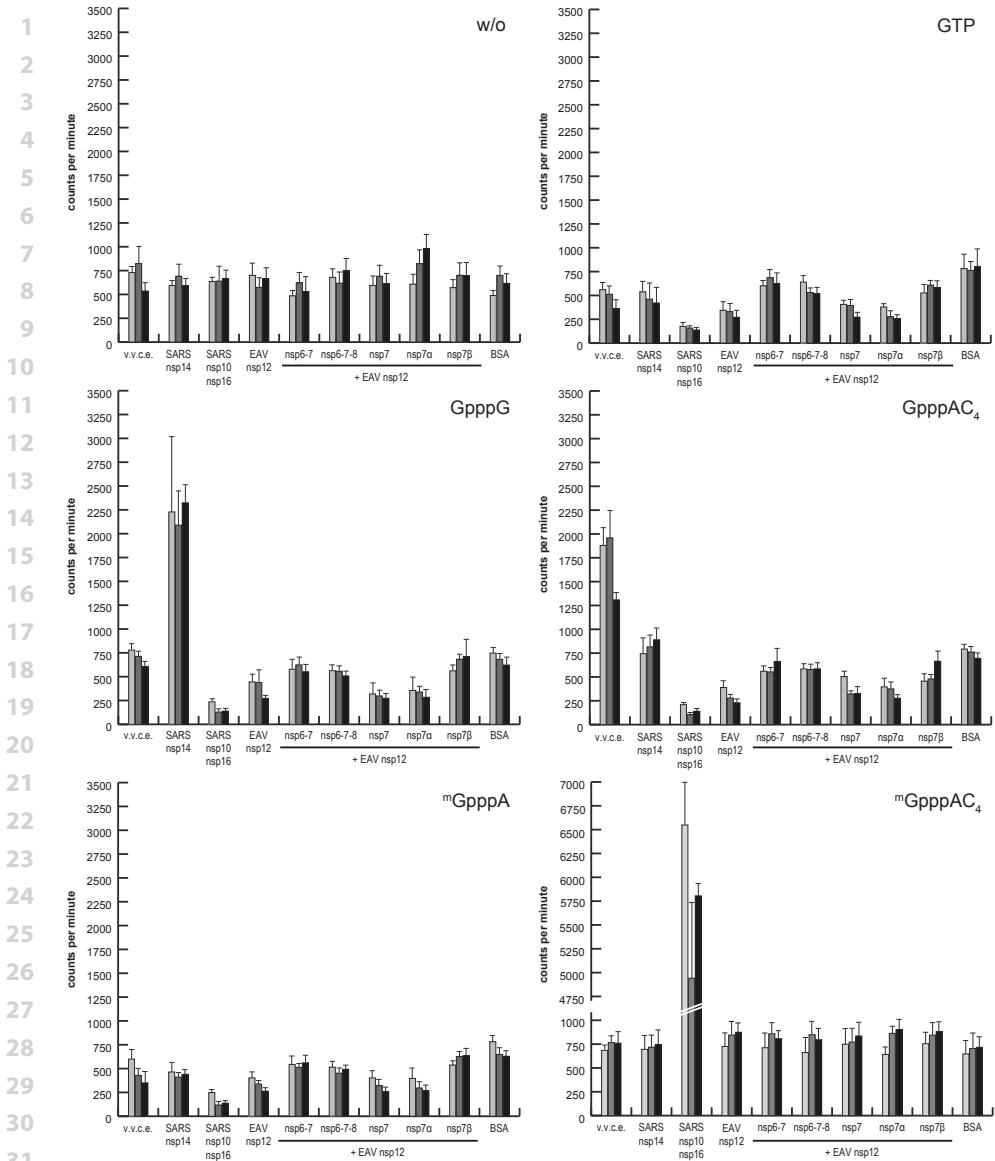


Figure 4: SDS-PAGE analysis of purified EAV nonstructural proteins. **(A)** The progression of metal-ion chromatography of EAV nsp12-containing (MW 13 kDa) E. coli lysates was monitored by Coomassie brilliant blue staining. Insoluble and soluble: proteins retained in pellet or supernatant, respectively, after cell lysis and ultracentrifugation; after binding: proteins in supernatant after removal of Talon beads. **(B)** Elution fractions of EAV ORF1a proteins and intermediates (MW<sub>nsp6-7</sub> 29 kDa, MW<sub>nsp6-7-8</sub> 34 kDa, MW<sub>nsp7</sub> 26 kDa, MW<sub>nsp7α</sub> 15 kDa, MW<sub>nsp7β</sub> 13 kDa). Products marked with an asterisk are remaining ubiquitin-nsp fusion proteins. Size markers are indicated on the left in kDa.

### Recombinant nsp12 does not display *in vitro* MTase activity using a variety of substrates

Using purified arterivirus proteins, we proceeded to test for MTase activity in the presence of different methyl acceptors by employing an *in vitro* assay similar to that previously established for SARS-CoV nsp14 and nsp16 (14). In agreement with published results (14;22;27), both SARS-CoV MTases (kindly provided by Dr. Etienne Decroly, Marseille), which were used as positive controls, transferred the radioactive methyl group from the universal methyl donor S-adenosylmethionine to non-methylated or N7-methylated cap analogs (Figure 5). Likewise, also vaccinia virus capping enzyme, obtained from a commercial source and known to harbor N-MT activity, demonstrated the expected activity. Based on these activities and the results of two negative control reactions (assays using BSA and no acceptor, respectively), we defined an incorporation threshold of 1000 cpm to distinguish the enzyme activity in this assay. According to this definition, EAV nsp12 did not display activity with any of the methyl acceptors in the absence or presence of any of the potential ORF1a-encoded co-factors described above (nsp6-7, nsp6-7-8, nsp7α, nsp7β, and nsp7).



**Figure 5:** Methytransferase (MTase) activity assays using recombinant EAV nsp12 in the presence and absence of possible co-factors. Recombinant EAV nsp12 (1  $\mu$ M) and equimolar amounts of the indicated possible co-factors were incubated for 30 (light gray), 60 (dark gray), or 180 min (black) with S-[methyl- $^3$ H]-adenosylmethionine and the indicated methyl acceptor. Proteins with known MTase activity served as positive controls. v.v.c.e., vaccinia virus capping enzyme (0.1 U/ $\mu$ l, N-MT); SARS nsp14 (75 nM, N-MT); SARS nsp10/nsp16 (2  $\mu$ M complex, O-MT); SARS, SARS coronavirus; BSA served as negative control; error bars indicate the standard deviation of the mean of two independent experiments. The background variation evident for several of the protein combinations using GTP, GpppG, GpppAC<sub>4</sub>, or <sup>m</sup>GpppA most likely represents an artifact originating from a position effect, which was observed repeatedly in the employed 96-well format.

## 1 **The tolerance of EAV replication to nsp12 mutagenesis correlates with the** 2 **natural variation of probed residues**

3  
4 To establish the general importance of nsp12 for EAV replication, we used reverse genet-  
5 ics to assess whether EAV tolerates replacements at conserved positions, including the  
6 single absolutely (F109) and ten partially (F26, N35, S45, Y49, S56, Y64, Y70, F82, C84,  
7 and F107) conserved residues (Figure 3). We also tested replacements of three poorly  
8 conserved residues (S25, S30, and Y32) that served as controls. Furthermore, we also  
9 abolished nsp12 expression by replacing its codons 6 to 8 with three consecutive trans-  
10 lation termination codons (STOP mutant). The engineered cDNA clones were used for *in*  
11 *vitro* transcription, yielding full-length RNA that was subsequently electroporated into  
12 BHK-21 cells. The effects of the replacements were first assessed on the level of viral pro-  
13 tein expression by immunofluorescence microscopy utilizing antibodies against nsp3  
14 and the structural nucleocapsid (N) protein. Furthermore, we monitored the production  
15 of virus progeny by harvesting transfected cell culture supernatants and performing  
16 plaque assays (Table 1).

17  
18 For the STOP mutant neither protein expression nor progeny production was observed,  
19 indicating that nsp12 performs an indispensable function during virus replication.  
20 Alternatively, the truncation of nps12 may have affected virus viability indirectly, e.g.  
21 by impairing proteolytic cleavage of the nsp11/nsp12 junction, which might be detri-  
22 mental to the activity of the nsp11 endoribonuclease. This concern was addressed by  
23 replacing individual nsp12 residues.

24  
25 The 14 residues probed by making 25 mutants could be classified into four groups based  
26 on the impact of their replacement. The first group included residues F107 and F109,  
27 with the four mutants carrying alanine or (more conservative) tyrosine substitutions at  
28 these positions not producing any virus progeny. Interestingly, in contrast to both ala-  
29 nine mutants and F109Y, which also did not produce viral proteins, immunofluorescence  
30 signal for nsp3 and N protein was detected for F107Y at 24 h and 48 h post transfection  
31 (p.t.), with a stronger signal being observed at the earlier time point. Collectively, these  
32 results show that F107 or F109 are most strongly constrained in EAV and indicate a vital  
33 role of these residues in virus viability.

34  
35 The second group comprised residues F26, N35, and C84, which appeared to be only  
36 slightly less important than the aforementioned F107 and F109, based on the pheno-  
37 type of five mutants. Alanine substitutions at position F26 and N35 were either lethal  
38 (F26A) or severely detrimental (N35A), whereas tyrosine or aspartate substitutions of  
39 these residues (F26Y and N35D) were compatible with at least some residual replica-



Table 1: EAV nsp12 mutants and their phenotypes

group <sup>‡</sup>	mutant	observed amino acid variation <sup>§</sup>	wild-type sequence	mutated sequence	immunofluorescence assay			titer (PFU/ml)	nsp12 sequence of P1 virus <sup>†</sup>	
					14 h p.t.	48 h p.t.	68 h p.t.			
	wt				+	+	+	3·10 <sup>6</sup>	2·10 <sup>8</sup>	n.d.
4	S25A	FYTSARC	UCA	<u>G</u> CU	+	+	+	3·10 <sup>6</sup>	2·10 <sup>7</sup>	mutation retained
2	F26A	FILV	UUC	<u>G</u> CA	-	-	-	<20	<20	n.d.
	F26Y	FILV	UUC	UAU	-	-	+	<20	4·10 <sup>7</sup>	reversion
4	S30A	SLVMD	UCA	<u>G</u> CU	+	+	+	6·10 <sup>5</sup>	6·10 <sup>7</sup>	mutation retained
4	Y32A	IMAFRCY	UAC	<u>G</u> CA	-	+	+	20	5·10 <sup>5</sup>	Y32V
	Y32F	IMAFRCY	UAC	UUU	+	+	+	5·10 <sup>8</sup>	2·10 <sup>8</sup>	mutation retained
2	N35A	NS	AAC	<u>G</u> CU	-	-	+	<20	<20	reversion
	N35D	NS	AAC	<u>G</u> AU	+	+	+	1·10 <sup>3</sup>	1·10 <sup>8</sup>	reversion
4	S45A	SAG	UCA	<u>G</u> CU	+	+	+	2·10 <sup>7</sup>	5·10 <sup>7</sup>	mutation retained
	S45T	SAG	UCA	<u>A</u> CC	+	+	+	1·10 <sup>3</sup>	1·10 <sup>8</sup>	reversion
3	Y49A	YFW	UAC	<u>G</u> CA	-	-	-	<20	<20	n.d.
	Y49F	YFW	UAC	UUU	+	+	+	6·10 <sup>5</sup>	4·10 <sup>7</sup>	mutation retained
4	S56A	ST	UCA	<u>G</u> CU	+	+	+	2·10 <sup>7</sup>	4·10 <sup>7</sup>	mutation retained
	S56T	ST	UCA	<u>A</u> CC	+	+	+	4·10 <sup>7</sup>	2·10 <sup>7</sup>	mutation retained
3	Y64A	YFWK	UAU	<u>G</u> CA	-	-	-	<20	<20	n.d.
	Y64F	YFWK	UAU	UUC	+	+	+	6·10 <sup>7</sup>	2·10 <sup>8</sup>	mutation retained
3	Y70A	YFV	UAU	<u>G</u> CA	-	-	-	<20	<20	n.d.
	Y70F	YFV	UAU	UUC	+	+	+	4·10 <sup>7</sup>	4·10 <sup>7</sup>	mutation retained
3	F82A	FYI	UUC	<u>G</u> CA	-	-	-	<20	<20	n.d.
	F82Y	FYI	UUC	UAU	+	+	+	1·10 <sup>8</sup>	1·10 <sup>8</sup>	mutation retained
2	C84Y	CYAF	UGC	UAU	+	+	+	3·10 <sup>2</sup>	3·10 <sup>7</sup>	reversion

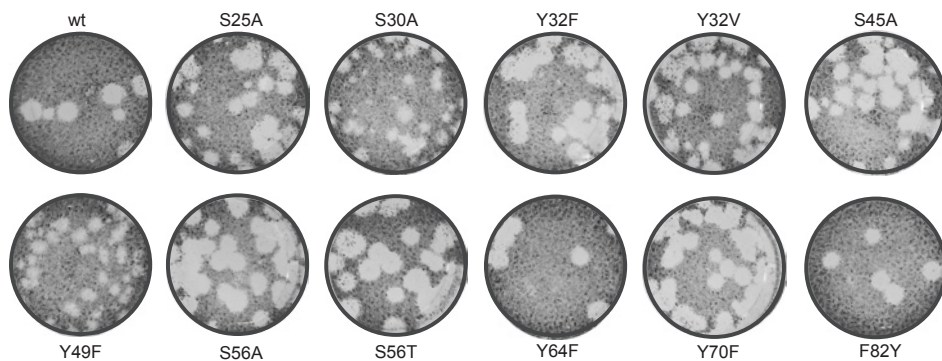
**Table 1:** EAV nsp12 mutants and their phenotypes (continued)

group <sup>‡</sup>	mutant	observed amino acid variation <sup>§</sup>	wild-type sequence	mutated sequence	immunofluorescence assay			nsp12 sequence of P1 virus <sup>†</sup>	
					14 h p.t.	48 h p.t.	68 h p.t.		titer (PFU/ml)
1	F107A	FYC	UUC	<u>GCA</u>	-	-	-	<20	n.d.
	F107Y	FYC	UUC	<u>UAU</u>	-	+*	-	<20	n.d.
1	F109A	F	UUC	<u>GCA</u>	-	-	-	<20	n.d.
	F109Y	F	UUC	<u>UAU</u>	-	-	-	<20	n.d.
	STOP		UCA GCA CUA	UGA <u>UGA</u> <u>UGA</u>	-	-	-	<20	n.d.

<sup>‡</sup>groups as defined in the text; <sup>§</sup>see Figure 3; \*non-spreading, <sup>†</sup>P1 virus was generated by infection of fresh BHK-21 cells with supernatant harvested at 68 h p.t. (for stable mutants) or the earliest positive time point in immunofluorescence microscopy (for reverting mutants and Y32V); n.d. not done; representative results of two independent data sets obtained by two different researchers.

tion, which allowed early reversion of these mutants. Similarly, also the C84Y mutant reverted, which is notable given the presence of a tyrosine at this position in most other arteriviruses.

In contrast to the above results, EAV tolerated replacements by another aromatic residue at four other partially conserved aromatic residues, Y49, Y64, Y70, and F82, which form group 3. These virus mutants were stable and yielded progeny titers up to 1 log below that of the wild-type control. Interestingly, although the titer of Y49F was not very different from that of the parental virus, this mutant exhibited a small-plaque phenotype (Figure 6). In contrast alanine substitutions at these positions were again lethal.



**Figure 6:** Plaque phenotypes of viable EAV nsp12 mutants. Virus-containing supernatants obtained 48 h post transfection were serially diluted and used to infect BHK-21 cells. After 72 h cells were fixed with 4% formaldehyde and stained with crystal violet.

The replacement – more or less conservative – of all residues mentioned thus far had a moderate to severe impact on virus replication. In contrast, the fourth group included five residues whose replacement did neither affect viral protein production nor progeny titers. As expected this group included the three poorly conserved control residues (S25, S30, and Y32). Nevertheless, S30A exhibited a small-plaque phenotype (Figure 6). Unexpectedly, we also repeatedly observed the pseudo-reversion of Y32A to Y32V, which required only a single nucleotide change. Although valine is not among the naturally occurring amino acid residues at this position (Figure 3), a hydrophobic residue is observed in several arteriviruses other than EAV. Besides substitutions of these control residues, EAV also tolerated the substitution of S56 with alanine or threonine. Given the strict conservation of serine and threonine, this lack of impact was the expected outcome for S56T, but was rather surprising for S56A. Finally, S45A was stable and indistinguishable from the parental virus, while S45T reverted. Together with the sequence variation at this position, which is limited to the small amino acids glycine, alanine, and serine, this

1 probably indicates a certain degree of steric hindrance by any residue larger than serine.  
2 Overall the observed mutant phenotypes were compatible with the natural variation  
3 observed at the respective positions, with the possible exception of the C84Y mutant.  
4 These correlations support the multiple sequence alignment of the highly variable  
5 nsp12 and suggest that EAV replication in BHK-21 is a faithful model system for probing  
6 nsp12 function by mutagenesis.

7  
8 Both mutants displaying a small-plaque phenotype (S30A and Y49F), as well as the  
9 unexpected Y32V pseudo-revertant, were further investigated in terms of growth kinet-  
10 ics and accumulation of intracellular viral RNA (not shown). Compared to the wild-type  
11 control, S30A and Y49F demonstrated a slight delay in replication early during infection  
12 (8 h post infection (p.i.)) but eventually reached comparable titers by 24 h p.i.. In line  
13 with this finding, the amounts of genomic and sg mRNA at 8 h p.i. were reduced for  
14 both mutants. Whether this was due to a decreased synthesis or lower stability of their  
15 RNAs remains to be investigated. In contrast, the stable Y32V mutant was essentially  
16 indistinguishable from the wild-type control both in growth kinetics and amounts of  
17 RNA produced.

## 18 19 20 **DISCUSSION**

21  
22 The most conserved ORF1b of nidoviruses encodes only two proteins that have not  
23 been studied before in any virus. Our study aimed to address this knowledge gap for  
24 one of these proteins, arterivirus nsp12. It established (i) the exceptional divergence of  
25 nsp12, (ii) the lack of strong bioinformatics and biochemical support for nsp12 being an  
26 MTase, and (iii) the fact that nsp12 is essential for arterivirus replication.

27  
28 So far, none of the four enzymatic activities required for conventional cap-1 synthesis,  
29 or any of the known alternative capping strategies, was uncovered for arteriviruses  
30 although arteriviral mRNAs are presumed to be capped. In the conserved relative ar-  
31 rangement of replicative enzymes within nidovirus polyproteins 1a and 1ab, the unique  
32 arterivirus nsp12 is encoded in a genome position equivalent to that of the coronavirus  
33 O-MT, which is conserved also in invertebrate nidoviruses (Figure 1). We thus asked  
34 whether this so far uncharacterized subunit may represent an MTase, potentially capable  
35 to perform both methylation reactions as, for example, the flavivirus NS5 MTase domain  
36 is (28). Upon our bioinformatics analysis of nsp12 sequences, we found that this subunit,  
37 similar to the N-MT residing in coronavirus nsp14, is enriched with (partially) conserved  
38 aromatic amino acids and is predicted to fold in alternating  $\alpha$ -helices and  $\beta$ -strands  
39

1 (Figure 3). Nevertheless, no statistically significant similarity was found between nsp12  
2 and other MTases of viral or cellular origin.

3

4 When we subsequently sought to verify our hypothesis using an *in vitro* MTase assay, we  
5 could not detect any activity for recombinant EAV nsp12, whereas our positive controls  
6 clearly confirmed the functionality of the assay. To explain this lack of activity, we argued  
7 that, as for coronavirus nsp16, a second EAV protein may be required to form a func-  
8 tional MTase complex. By analogy with the coronavirus nsp10 co-factor, we tested the  
9 possibility that this second protein might be encoded just upstream of the ORF1a/1b  
10 ribosomal frameshift site. We thus expressed and purified nsp7 $\alpha$  and nsp7 $\beta$ , as well as  
11 three polyprotein cleavage intermediates containing these two proteins, and included  
12 them in our assays (Figure 5). However, also in these extended assays we could not de-  
13 tect any MTase activity. This could have multiple reasons. First, the proteins tested here  
14 may not be the correct co-factors or may be unable to properly associate with nsp12  
15 under the conditions employed. Second, more than one co-factor may be needed to  
16 spur nsp12's MTase activity, or different RNA substrates containing specific sequences  
17 may be required. Finally, our results are compatible with a scenario in which nsp12,  
18 which is smaller than other viral MTases, does not possess MTase activity, in which case  
19 other hypotheses about its function should be considered (see below).

20

21 To explore nsp12's relevance for arterivirus replication, we engineered one truncation  
22 and 25 point mutations of EAV nsp12 and launched the corresponding mutant genomes  
23 in BHK-21 cells. Reflecting the conservation of several aromatic residues in arteriviruses,  
24 substitution with alanine was tolerated in none of the cases, whereas more conservative  
25 substitutions maintaining the residue's aromatic nature were tolerated in most of the  
26 partially conserved positions (Table 1). The only exception was F107Y, which interesting-  
27 ly showed a certain level of protein expression but did not produce infectious progeny.  
28 Since two arteriviruses distantly related to EAV, LDV and PRRSV genotype 1, naturally  
29 encode a tyrosine at this position (Figure 3), this result suggests an epistatic interaction  
30 between residue 107 and other unknown residue(s). EAV also did not tolerate a block of  
31 nsp12 expression (STOP mutant) or the replacement of its single absolutely conserved  
32 nsp12 residue, F109, with alanine or tyrosine. This phenotype could be explained by a  
33 trans-dominant negative effect of the nsp12 substitutions on an interaction partner of  
34 nsp12, if this partner is essential for EAV replication. This explanation is also compat-  
35 ible with the nonviable phenotype of several other mutants and suggests a particularly  
36 important role of the most constrained and proximal F107 and F109 in such a putative  
37 interaction.

38

39

1 The fact that EAV does not tolerate substitution of its single invariant nsp12 residue  
2 stands in remarkable contrast to phenotypes described for mutants of the invariant  
3 residues of the NendoU or O-MT of nidoviruses (29-32), which are both more strongly  
4 conserved than nsp12. In these studies alanine substitutions of absolutely conserved  
5 putative active site residues resulted in lower virus progeny titers and in part in small-  
6 plaque phenotypes in cell culture but did not entirely abolish virus replication.

7  
8 In conclusion, our combined results may be most compatible with the notion that nsp12  
9 is not an MTase and possibly not even an enzyme but rather a co-factor of an essential  
10 component of the arterivirus replicase. In this context, a future in-depth analysis of the  
11 nsp12 interaction network could be most informative. If nsp12 is not an MTase, this ac-  
12 tivity must be provided by another protein, but it is unlikely to be one of the three other  
13 ORF1b proteins, which are known to possess different enzymatic domains. This implies  
14 that arteriviruses may be (very) different from other nidoviruses with respect to either  
15 the nature of the 5' end of their mRNAs and/or the mechanism generating it. We note  
16 that the presence of a 5'-terminal cap-1 structure was reported for the SHFV genome  
17 (23), but that monophosphates were claimed to present at the 5' end of LDV mRNAs  
18 (33), calling for additional studies to resolve the apparent conflict. Finally, the possibility  
19 of cap-snatching, the strategy employed by some families of negative-stranded RNA  
20 viruses (34-36), may be explored for arteriviruses. This mechanism might accommodate  
21 the nsp11 NendoU as endoribonuclease and nsp12 as a cap-binding protein, which  
22 would connect coronavirus nsp16 and arterivirus nsp12 to a common target in an  
23 unorthodox way.

## 24 25 26 **MATERIAL AND METHODS**

### 27 28 **Bioinformatics**

29  
30 Genomes of members of the *Arteriviridae* and *Coronaviridae* families were retrieved  
31 from GenBank (37) and RefSeq (38) using the Homology-Annotation hYbrid retrieval of  
32 GENetic Sequences (Haygens) tool <http://veb.lumc.nl/HAYGENS>. Codon-based multiple  
33 sequence alignments (MSAs) of virus genomes were produced using the ViralIS platform  
34 (39) and assisted by the HMMER 3.1 (40), Muscle 3.8.31 (41), and ClustalW 2.012 (42)  
35 programs. Only one virus per established or tentative species, which were defined  
36 with the help of DEmARC1.3 (43), was retained for bioinformatics analyses. To retrieve  
37 information about genomes, the SNAD program (44) was used. To reveal the full extent  
38 of similarity between pairs of alignments, they were converted into HMM profiles, which  
39 were compared and visualized in a dot-plot fashion using a routine in HH-suite 2.0.15

1 (45;46). Distribution of similarity density in alignments was plotted using R package  
2 Bio3D (47) under the conservation assessment method “similarity”, substitution matrix  
3 Blosum62 (48), and a sliding window of 11 alignment columns. To search for homologs  
4 among profiles in the PFAM A database (49), the HH-suite 2.0.15 software (45;46) was  
5 used. Secondary structure of proteins was predicted by applying JPred 3 (50) and PsiPred  
6 (51) to MSAs, with the prediction being applied to the top sequence in the MSA. The  
7 MSAs were converted into figures using ESPript (52). Reconstruction of phylogenetic  
8 trees was performed using PhyML 3.0, with the WAG amino acid substitution matrix,  
9 allowing substitution rate heterogeneity among sites (4 categories), and 1000 iterations  
10 of non-parametric bootstrapping (53). Pairwise patristic distances (PPDs) between  
11 viruses were calculated from protein trees using R package “ape” (54). Linear regression  
12 was calculated using R package “stats” (55).

13

#### 14 **Reverse genetics of EAV**

15

16 Mutations specifying alanine and conservative replacements of (partially) conserved  
17 and control residues in nsp12 were generated using the QuikChange protocol. In all  
18 cases translationally silent marker mutations were introduced to allow discrimination  
19 between (partial) reversion of mutants after transfection and (possible) contamination  
20 with wild-type virus. Mutated gene fragments were introduced into full-length cDNA  
21 clone pEAV211 (56), a pEAV030 derivative (57), using appropriate shuttle vectors and  
22 restriction enzymes. The presence of the mutations was confirmed by sequencing.  
23 pEAV211 DNA was *in vitro* transcribed and RNA was purified by LiCl precipitation. RNA  
24 was transfected into BHK-21 cells as described previously (58). Transfected cells were  
25 monitored by immunofluorescence microscopy until 68 h post transfection (p.t.), using  
26 antibodies directed against EAV nsp3 and N protein as described (59). To monitor the  
27 production of viral progeny, plaque assays were performed with supernatants collected  
28 at 14 and 48 h p.t. or during the first 24 hours p.i. to determine growth kinetics, as de-  
29 scribed (58). To verify the presence of the introduced mutations or reversions in viable  
30 mutants, fresh BHK-21 cells were infected with supernatants harvested at time points  
31 at which transfected cells were positive in immunofluorescence microscopy. RNA was  
32 isolated after 18 h or when cytopathic effect was detected. Finally, the nsp12-coding  
33 region was amplified by RT-PCR using random hexameric primers in the RT step and  
34 EAV-specific primers for the PCR. PCR fragments were purified and sequenced.

35

#### 36 **Protein expression and purification**

37

38 N-terminal and C-terminal His-tag fusion proteins of wild-type nsp12 were expressed  
39 from a pDEST vector. Plasmids were transformed into *E. coli* BL21 (DE3) and cells were

1 grown in Luria Broth with 100 µg/ml ampicillin at 37°C until OD<sub>600</sub> reached 0.7. Express-  
2 sion was induced after addition of 0.5 mM IPTG and cells were grown for further 4 h at  
3 37°C.

4  
5 EAV ORF1a-encoded proteins were expressed with N-terminal ubiquitin and C-terminal  
6 His tags from pASK vectors (60). Plasmids were transformed into *E. coli* C2523 contain-  
7 ing the pCG1 plasmid, which leads to constitutive expression of the ubiquitin-specific  
8 protease UBP1. Cells were grown in Luria Broth with 100 µg/ml ampicillin and 34 µg/ml  
9 chloramphenicol at 37°C until OD<sub>600</sub> reached 0.7. Expression was induced after addition  
10 of 200 ng/ml anhydrotetracycline and cells were grown for another 18 h at 20°C. All  
11 pellets were harvested by centrifugation and stored at -20°C until further use.

12  
13 Proteins were batch purified by metal affinity chromatography using Co<sup>2+</sup> (Talon beads).  
14 All steps were performed at 4°C or on ice. Cells were resuspended in nsp12 resuspen-  
15 sion buffer (20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol) or co-factor resuspension  
16 buffer (20 mM HEPES, pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanol) contain-  
17 ing 500 mM NaCl and Roche complete EDTA-free protease inhibitor cocktail. Lysis was  
18 achieved by 30 min incubation with lysozyme (0.1 mg/ml). Genomic DNA was sheared  
19 during four sonication cycles of 10 s with intermittent cooling. Cell debris was removed  
20 by centrifugation at 20.000 g for 20 min. Cleared supernatants were incubated with an  
21 appropriate amount of Talon beads for 1 h under slow rolling. Beads were collected  
22 and washed four times for 15 min with a 20-fold volume of the respective resuspen-  
23 sion buffer supplemented with 10 mM imidazole and first 500 mM, then 250 mM, and  
24 finally twice 100 mM NaCl. Proteins were eluted with the respective resuspension buf-  
25 fer containing 300 mM imidazole and 100 mM NaCl. Elution fractions were examined  
26 by SDS-PAGE, pooled, and dialyzed against 20 mM HEPES, pH 7.5, 100 mM NaCl, 25%  
27 glycerol, 1 mM DTT. All proteins were stored at -20°C. Typical yields were 1-2 mg/l culture  
28 for all proteins. Protein concentrations were calculated based on theoretical extinction  
29 coefficients and absorption at 280 nm.

30  
31 Gel filtration of nsp12 was performed on a Superdex 75 10/300 GL gel filtration column  
32 with 10 mM Na-phosphate buffer, pH 6.0, 100 mM NaCl, 1 mM DTT at 4°C and a flow rate  
33 of 0.5 ml/min.

### 34 **Methyltransferase assay**

35  
36  
37 Methyltransferase assays were performed essentially as described previously (14). Pro-  
38 teins at the indicated final concentrations were incubated at 30°C for 30, 60, or 180 min  
39 in a buffer containing 20 mM HEPES, pH 7.5, 5 mM DTT, 0.5 mM MgCl<sub>2</sub>, 0.5 mM MnCl<sub>2</sub>,

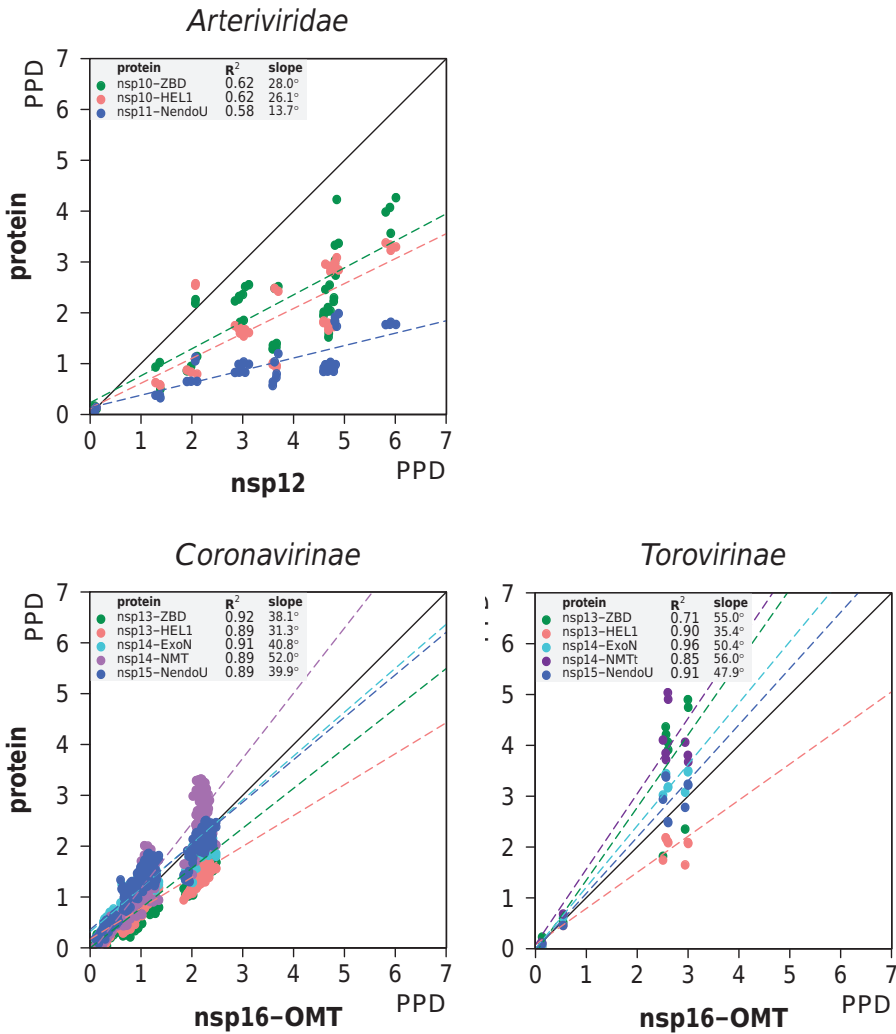


1 10  $\mu\text{M}$  S-adenosylmethionine, 2  $\mu\text{M}$  capping substrate, and  $1 \times 10^3$  Bq/ $\mu\text{l}$  S-[methyl- $^3\text{H}$ ]-  
2 adenosylmethionine. Additionally 7.5 mM NaCl were carried over from the protein  
3 storage buffer. Vaccinia virus capping enzyme (New England Biolabs) was incubated in  
4 the buffer supplied by the vendor. To stop the reaction a 10-fold volume of ice-cold  
5 S-adenosylhomocysteine (100  $\mu\text{M}$ ) was added. Samples were spotted on DEAE filter-  
6 mats (Perkin Elmer), which were subsequently washed twice with 10 mM ammonium  
7 formate, pH 8.0, then twice with water, and finally with ethanol. Filtermats were cut, and  
8 radioactivity was measured by scintillation counting.

## 11 **ACKNOWLEDGEMENTS**

13 This work was supported by the European Union's Seventh Framework program  
14 (FP7/2007-2013) through the EUVIRNA project (European Training Network on (+) RNA  
15 virus replication and antiviral drug development, grant agreement no. 264286) and  
16 the SILVER project (grant agreement no. 260644), by the Collaborative Agreement on  
17 Bioinformatics between Leiden University Medical Center and Moscow State University  
18 (MoBiLe), and by the Leiden University Fund. The authors wish to acknowledge Dr. Eti-  
19 enne Decroly (AFMB, Marseille, France) for helpful discussions and for providing purified  
20 SARS-CoV nsp14 and nsp10/16 as well as MTase substrates; Alexander Kravchenko and  
21 Igor Sidorov for maintaining and advancing the Viralis platform and its databases; and  
22 Linda Boomaars and Irina Albulescu for technical assistance.

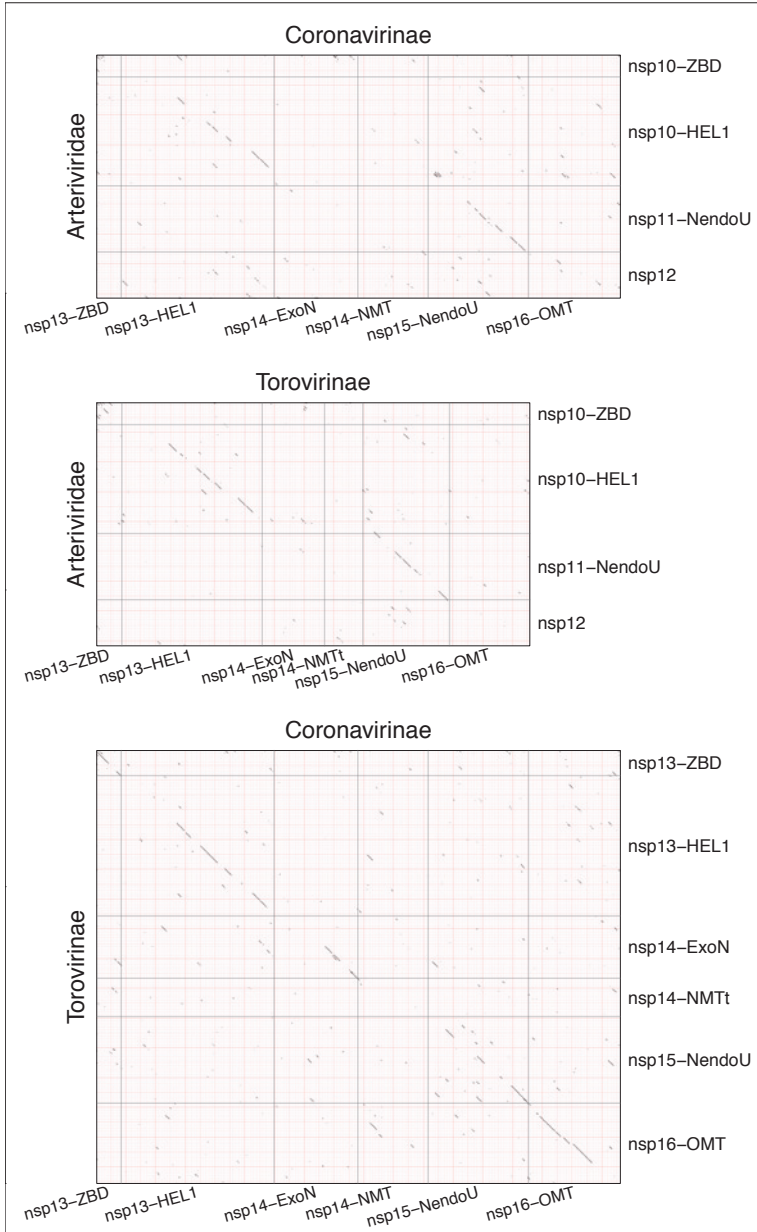
## 1 SUPPLEMENTARY DATA



32 **Figure S1.** Relative scale of divergence of nsp12 of the *Arteriviridae* and (putative) nsp16 of the *Coronavirinae* and *Torovirinae*. Shown are three (sub)family-specific two-dimensional scatter plots that compare PPDs of the most C-terminal protein (nsp12 or nsp16-OMT, x-axis) versus PPDs of other proteins/domains of ORF1b starting from the ZBD (detailed in inset, y-axis). PPDs were calculated from PhyML trees for separate proteins. Dashed lines, linear regressions fit in respective (color matching) dot distributions with  $R^2$  and slope values being detailed in the inset panels.

33  
34  
35  
36  
37  
38  
39

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39



**Figure S2.** Analysis of co-conservation of C-terminal parts of ORF1b of the *Arteriviridae*, *Coronavirinae*, and *Torovirinae*. Shown are three pair-wise MSA-based HMM-HMM plots comparing parts of ORF1b starting from the ZBD of three origins. The position of proteins and some domains are indicated. Each MSA was converted to an HMM profile, three possible pairs of obtained HMMs were aligned with the help of HH-suite 2.0.15 software (45,46). The presence of similarity above the threshold of 0.3 is recorded with a dot. Diagonal persistence of dots is strong evidence for statistically significant similarity (homology) of a protein pair.

## 1 REFERENCE LIST

- 2
- 3 1. Snijder EJ, Kikkert M, Fang Y. Arterivirus molecular biology and pathogenesis. *J.Gen.Virol.* 2013;
- 4 94(Pt 10):2141-2163.
- 5 2. Faaberg KS, Balasuriya UB, Brinton MA, *et al.* Family *Arteriviridae*. In King AMQ, Adams MJ, Carstens
- 6 EB *et al.* editors, *Virus taxonomy. Ninth report of the international committee on taxonomy of*
- 7 *viruses*, Amsterdam, Elsevier Academic Press, 2012;796-805.
- 8 3. Sang Y, Rowland RR, Blecha F. Antiviral regulation in porcine monocytic cells at different activa-
- 9 tion States. *J.Virol.* 2014;88(19):11395-11410.
- 10 4. Dunowska M, Biggs PJ, Zheng T, *et al.* Identification of a novel nidovirus associated with a neuro-
- 11 logical disease of the Australian brushtail possum (*Trichosurus vulpecula*). *Vet.Microbiol.* 2012;
- 12 156(3-4):418-424.
- 13 5. Bailey AL, Lauck M, Sibley SD, *et al.* Two novel simian arteriviruses in captive and wild baboons
- 14 (*Papio spp.*). *J.Virol.* 2014;88(22):13231-13239.
- 15 6. Lauck M, Hyeroba D, Tumukunde A, *et al.* Novel, divergent simian hemorrhagic fever viruses in a
- 16 wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS.One.* 2011;6(4):
- 17 e19056.
- 18 7. Lauck M, Sibley SD, Hyeroba D, *et al.* Exceptional simian hemorrhagic fever virus diversity in a wild
- 19 African primate community. *J.Virol.* 2013;87(1):688-691.
- 20 8. Molenkamp R, Greve S, Spaan WJ, *et al.* Efficient homologous RNA recombination and require-
- 21 ment for an open reading frame during replication of equine arteritis virus defective interfering
- 22 RNAs. *J.Virol.* 2000;74(19):9062-9070.
- 23 9. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J.Gen.Virol.*
- 24 2006;87(Pt 6):1403-1421.
- 25 10. de Groot RJ, Baker SC, Baric R, *et al.* Family *Coronaviridae*. In King AMQ, Adams MJ, Carstens EB *et*
- 26 *al.* editors, *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*,
- 27 Amsterdam, Elsevier Academic Press, 2012;806-828.
- 28 11. de Groot RJ, Cowley JA, Enjuanes L, *et al.* Order *Nidovirales*. In King AMQ, Adams MJ, Carstens EB *et*
- 29 *al.* editors, *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*,
- 30 Amsterdam, Elsevier Academic Press, 2012.
- 31 12. Nga PT, Parquet MC, Lauber C, *et al.* Discovery of the first insect nidovirus, a missing evolutionary
- 32 link in the emergence of the largest RNA virus genomes. *PLoS.Pathog.* 2011;7(9):e1002215.
- 33 13. Gorbalenya AE. Big nidovirus genome. When count and order of domains matter. *Adv.Exp.Med.*
- 34 *Biol.* 2001;49:41-17.
- 35
- 36
- 37
- 38
- 39

- 1 14. Bouvet M, Debarnot C, Imbert I, *et al.* *In vitro* reconstitution of SARS-coronavirus mRNA cap  
2 methylation. *PLoS.Pathog.* 2010;6(4):e1000863.
- 3 15. Chen Y, Su C, Ke M, *et al.* Biochemical and structural insights into the mechanisms of SARS coro-  
4 navirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS.Pathog.* 2011;7(10):  
5 e1002294.
- 6 16. Decroly E, Imbert I, Coutard B, *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding  
7 enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J.Virol.* 2008;82(16):8071-8084.
- 8 17. Decroly E, Debarnot C, Ferron F, *et al.* Crystal structure and functional analysis of the SARS-  
9 coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS.Pathog.* 2011;7(5):  
10 e1002059.
- 11 18. Lai MM, Stohlman SA. Comparative analysis of RNA genomes of mouse hepatitis viruses. *J.Virol.*  
12 1981;38(2):661-670.
- 13 19. van Vliet AL, Smits SL, Rottier PJ, *et al.* Discontinuous and non-discontinuous subgenomic RNA  
14 transcription in a nidovirus. *EMBO J.* 2002;21(23):6571-6580.
- 15 20. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of  
16 duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. *J.Virol.* 2004;  
17 78(14):7833-7838.
- 18 21. Ivanov KA, Thiel V, Dobbe JC, *et al.* Multiple enzymatic activities associated with severe acute  
19 respiratory syndrome coronavirus helicase. *J.Virol.* 2004;78(11):5619-5632.
- 20 22. Chen Y, Cai H, Pan J, *et al.* Functional screen reveals SARS coronavirus nonstructural protein nsp14  
21 as a novel cap N7 methyltransferase. *Proc.Natl.Acad.Sci.U.S.A* 2009;106(9):3484-3489.
- 22 23. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion  
23 RNA. *Virology* 1986;151(1):146-150.
- 24 24. Seybert A, van Dinten LC, Snijder EJ, *et al.* Biochemical characterization of the equine arteritis  
25 virus helicase suggests a close functional relationship between arterivirus and coronavirus heli-  
26 cases. *J.Virol.* 2000;74(20):9586-9593.
- 27 25. van Aken D, Zevenhoven-Dobbe J, Gorbalenya AE, *et al.* Proteolytic maturation of replicase  
28 polyprotein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication  
29 and includes internal cleavage of nsp7. *J.Gen.Virol.* 2006;87(Pt 12):3473-3482.
- 30 26. Wassenaar AL, Spaan WJ, Gorbalenya AE, *et al.* Alternative proteolytic processing of the arterivirus  
31 replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease.  
32 *J.Virol.* 1997;71(12):9313-9322.
- 33 27. Jin X, Chen Y, Sun Y, *et al.* Characterization of the guanine-N7 methyltransferase activity of coro-  
34 navirus nsp14 on nucleotide GTP. *Virus Res.* 2013;176(1-2):45-52.
- 35  
36  
37  
38  
39

- 1 28. Zhou Y, Ray D, Zhao Y, *et al.* Structure and function of flavivirus NS5 methyltransferase. *J.Virol.* 2007;81(8):3891-3903.
- 2
- 3 29. Kang H, Bhardwaj K, Li Y, *et al.* Biochemical and genetic analyses of murine hepatitis virus Nsp15
- 4 endoribonuclease. *J.Virol.* 2007;81(24):13587-13597.
- 5
- 6 30. Menachery VD, Yount BL, Jr., Josset L, *et al.* Attenuation and restoration of severe acute respiratory
- 7 syndrome coronavirus mutant lacking 2'-o-methyltransferase activity. *J.Virol.* 2014;88(8):4251-
- 8 4264.
- 9
- 10 31. Posthuma CC, Nedialkova DD, Zevenhoven-Dobbe JC, *et al.* Site-directed mutagenesis of the
- 11 Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life
- 12 cycle. *J.Virol.* 2006;80(4):1653-1661.
- 13
- 14 32. Züst R, Cervantes-Barragan L, Habjan M, *et al.* Ribose 2'-O-methylation provides a molecular
- 15 signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat.*
- 16 *Immunol.* 2011;12(2):137-143.
- 17
- 18 33. Chen Z, Faaberg KS, Plagemann PG. Determination of the 5' end of the lactate dehydrogenase-
- 19 elevating virus genome by two independent approaches. *J.Gen.Virol.* 1994;75 ( Pt 4):925-930.
- 20
- 21 34. Fujimura T, Esteban R. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. *Proc.*
- 22 *Natl.Acad.Sci.U.S.A* 2011;108(43):17667-17671.
- 23
- 24 35. Mir MA, Duran WA, Hjelle BL, *et al.* Storage of cellular 5' mRNA caps in P bodies for viral cap-
- 25 snatching. *Proc.Natl.Acad.Sci.U.S.A* 2008;105(49):19294-19299.
- 26
- 27 36. Reich S, Guilligay D, Pflug A, *et al.* Structural insight into cap-snatching and RNA synthesis by
- 28 influenza polymerase. *Nature* 2014;516(7531):361-366
- 29
- 30 37. Benson DA, Cavanaugh M, Clark K, *et al.* GenBank. *Nucleic Acids Res.* 2013;41(Database issue):
- 31 D36-D42.
- 32
- 33 38. Pruitt KD, Brown GR, Hiatt SM, *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 2014;42(Database issue):D756-763.
- 34
- 35 39. Gorbalenya AE, Lieutaud P, Harris MR, *et al.* Practical application of bioinformatics by the multidis-
- 36 ciplinary VIZIER consortium. *Antiviral Res.* 2010;87(2):95-110.
- 37
- 38 40. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39(Web Server issue):W29-37.
- 39
41. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-1797.
42. Larkin MA, Blackshields G, Brown NP, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947-2948.

- 1 43. Lauber C, Gorbalenya AE. Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. *J.Virol.* 2012;86(7):3890-3904.
- 2
- 3 44. Sidorov IA, Reshetov DA, Gorbalenya AE. SNAD: Sequence Name Annotation-based Designer. *BMC.Bioinformatics.* 2009;10:251.
- 4
- 5
- 6 45. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;21(7):951-960.
- 7
- 8 46. Remmert M, Biegert A, Hauser A, *et al.* HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat.Methods* 2012;9(2):173-175.
- 9
- 10 47. Grant BJ, Rodrigues AP, ElSawy KM, *et al.* Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics.* 2006;22(21):2695-2696.
- 11
- 12
- 13 48. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc.Natl.Acad. Sci.U.S.A* 1992;89(22):10915-10919.
- 14
- 15 49. Finn RD, Bateman A, Clements J, *et al.* Pfam: the protein families database. *Nucleic Acids Res.* 2014;42(Database issue):D222-230.
- 16
- 17
- 18 50. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 2008;36(Web Server issue):W197-201.
- 19
- 20 51. Buchan DW, Minneci F, Nugent TC, *et al.* Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* 2013;41(Web Server issue):W349-357.
- 21
- 22 52. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 2014;42(Web Server issue):W320-324.
- 23
- 24 53. Guindon S, Dufayard JF, Lefort V, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst.Biol.* 2010;59(3):307-321.
- 25
- 26 54. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics.* 2004;20(2):289-290.
- 27
- 28
- 29 55. R Development Core Team. R: A language and environment for statistical computing. 2011. Vienna, Austria, R Foundation for statistical computing.
- 30
- 31 56. van den Born E, Gultyaev AP, Snijder EJ. Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. *RNA.* 2004;10(3):424-437.
- 32
- 33
- 34 57. van Dinten LC, Den Boon JA, Wassenaar AL, *et al.* An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc.Natl. Acad.Sci.U.S.A* 1997;94(3):991-996.
- 35
- 36
- 37
- 38
- 39

- 1 58. Nedialkova DD, Goralenya AE, Snijder EJ. Arterivirus Nsp1 modulates the accumulation of  
2 minus-strand templates to control the relative abundance of viral mRNAs. *PLoS.Pathog.* 2010;  
3 6(2):e1000772.
- 4 59. van der Meer Y, van TH, Locker JK, *et al.* ORF1a-encoded replicase subunits are involved in the  
5 membrane association of the arterivirus replication complex. *J.Virol.* 1998;72(8):6689-6698.
- 6 60. Gohara DW, Ha CS, Kumar S, *et al.* Production of "authentic" poliovirus RNA-dependent RNA  
7 polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in *Escherichia coli*. *Protein Expr.*  
8 *Purif.* 1999;17(1):128-138.
- 9 61. Lauber C, Goeman JJ, Parquet MC, *et al.* The footprint of genome architecture in the largest  
10 genome expansion in RNA viruses. *PLoS.Pathog.* 2013;9(7):e1003500.
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

Development of an anti-coronavirus  
drug – a private sector opportunity?

## CHAPTER 7

## 1 PREFACE

2  
3 In contrast to all other chapters of this thesis, this chapter does not present scientific  
4 data on single nidovirus proteins that may serve as drug targets and ultimately may  
5 lead to the discovery of antiviral compounds. Instead it gives a rough overview of some  
6 of the economic and legal questions that need to be addressed before a pharmaceuti-  
7 cal company would engage in the development of such a compound. This project was  
8 realized as part of the Marie Curie Initial Training Network EUVIRNA, which aimed to  
9 provide multidisciplinary training at the interface between academia and industry.  
10 Under guidance of a mentor from one of the industrial partners of the consortium,  
11 Janssen Infectious Diseases, the presented information was gathered from internet re-  
12 search, telephone conference calls, and face-to-face meetings with experts with diverse  
13 backgrounds working in different relevant fields, such as regulatory affairs, global public  
14 health, strategic marketing, and market access. Although the duration of this project  
15 was too short for a more detailed analysis, it still was an interesting experience writing  
16 it, and it is hoped that it at least provides some food for thought that is otherwise rarely  
17 obtained in a purely academic setting.

## 19 SUMMARY

20  
21  
22 In the past ten years the emergence of two new zoonotic coronaviruses (SARS-CoV and  
23 MERS-CoV) has prompted concerns about the possibility of a serious human pandemic  
24 caused by such a previously unknown virus from an animal source. In response to this  
25 threat, several academic groups started to design or search for inhibitors of coronavirus  
26 replication. However, to date none of the identified molecules has been further devel-  
27 oped. This essay summarizes and discusses the reasons for this slow development. From  
28 an industrial perspective, questions relating to regulatory requirements, drug proper-  
29 ties, and economic incentives are addressed. Due to market considerations, the only  
30 potentially profitable purpose of an anti-coronavirus drug appeared to be its inclusion  
31 in (mostly public) stockpiles as a countermeasure in case of an outbreak of a newly  
32 emerging coronavirus. Since the actual target virus will thus not be available during  
33 the development phase, a broad-spectrum inhibitor would need to be discovered to in-  
34 crease the chances that the drug will be effective against such a virus. Similarly, this lack  
35 of a target definition during clinical testing would require a substantial deviation from  
36 the standard approval process. Nevertheless, special FDA (United States Food and Drug  
37 Administration) regulations are in place to allow approval for the intended use. Although  
38 early countermeasures may mitigate the impact of an outbreak on public health and  
39 economy, it would be unclear which countries would indeed prepare for a coronavirus

1 pandemic and on which scale they would do so. Given this major uncertainty, it seems  
2 unlikely that a private company will invest in the development of an anti-coronavirus  
3 drug. Yet, the establishment of a public-private partnership with a clear public commit-  
4 ment to support research and development, coupled to a guarantee to purchase certain  
5 quantities of the product, may compensate for this risk. However, where coronaviruses  
6 are concerned the sense of urgency required to realize such a commitment is currently  
7 not evident.

## 8 9 10 **1. INTRODUCTION**

11  
12 In late 2002, the sudden appearance of severe acute respiratory syndrome coronavirus  
13 (SARS-CoV) and the first pandemic of the 21<sup>st</sup> century that it caused exemplified the  
14 potential threat newly emerging viruses may pose to public health, society, and the  
15 global economy. To confine the spread of such viruses, it would be desirable to have  
16 potent antiviral drugs for treatment and prophylaxis available when the first cases are  
17 recognized. However, meeting this goal would require the arrangement of national or  
18 global drug stockpiles and therefore the prior approval to use such a drug before any  
19 information on the actual pandemic agent is available.

20  
21 To exemplify the spectrum of considerations a company needs to take into account  
22 before engaging in such an endeavor, this essay will explore the potential development  
23 and use of drugs directed against coronaviruses. Although this assumption cannot be  
24 supported by sufficient scientific data at this time, it shall be assumed that it is, at least in  
25 principle, possible to design a broad-spectrum antiviral drug effective against all or the  
26 majority of currently known mammalian coronaviruses. However, I would like to point  
27 out that a detailed scientific evaluation of this undertaking is beyond the scope of this  
28 essay, which will be mainly focused on economic facets of developing a potential new  
29 drug. The central question will be:

30  
31 would it, from an economic point of view, be reasonable to invest in the dis-  
32 covery and/or development of an anti-coronavirus compound, even though  
33 there is neither a guarantee that a coronavirus will cause an outbreak in the  
34 near future nor the certainty that the developed drug will actually be active  
35 against this particular coronavirus?  
36  
37  
38  
39

1 To reach a conclusion on this question, aspects as regulatory requirements, market and  
2 drug properties, as well as competing products will be discussed below<sup>1</sup>.

## 5 2. BACKGROUND INFORMATION ON CORONAVIRUSES

7 Coronaviruses are a family of RNA viruses with a single-stranded genome of positive  
8 polarity, i.e., their genomes serve immediately as mRNAs for protein production. Their  
9 name originates from the characteristic crown-like appearance of their surface projec-  
10 tions under the electron microscope. Coronaviruses typically replicate in the respiratory  
11 or enteric tract of humans and a wide range of animal hosts and cause clinical symptoms  
12 accordingly (see Table 1). Notably, infection with one of the four human coronaviruses  
13 (HCoVs) is the second leading cause of the common cold, exceeded only by rhinovirus  
14 infections. Besides these widespread HCoVs with a seroprevalence in adults of up to  
15 80% (1), which were in part originally described in the mid-1960s, two zoonotic<sup>2</sup> coro-  
16 naviruses have jumped to humans more recently, SARS-CoV and MERS-CoV (Middle East  
17 respiratory syndrome coronavirus). In contrast to the “established” HCoVs, both zoonotic  
18 agents are associated with severe respiratory disease and a high case fatality rate. For  
19 SARS 8,439 cases leading to 812 deaths were reported between November 2002 and July  
20 2003 (2). For the ongoing MERS outbreak, the number of cases since April 2012 currently  
21 approaches 1,000, of which 35-40% had a fatal outcome, often in patients with underly-  
22 ing medical conditions (3). Reasons for the limited spread of these viruses appear to lie in  
23 their late onset of transmission in comparison to the onset of symptoms and a relatively  
24 poor transmissibility among humans. Nevertheless, direct spread upon close contact,  
25 for instance between family members or in healthcare facilities, has been reported for  
26 both viruses (4;5). In comparison, a highly contagious virus, like influenza virus, was  
27 estimated to be responsible for 43-88 million infections, including 8,700-18,000 deaths,  
28 during the 2009 influenza A(H1N1) pandemic in the United States alone (6). Obviously,  
29 if a novel coronavirus would surface that would be transmitted with similar efficiency  
30 as the established HCoVs appear to be while sharing the high pathogenicity with the  
31 zoonotic coronaviruses, the impact on society and economy could be devastating.

---

35  
36 1 The discussed information is, for the most part, derived from regulations and prepared-  
37 ness plans issued by the United States (U.S.) federal government. This is due to the consid-  
38 eration that the U.S. are one of the most likely and most relevant partners for stockpiling  
39 an anti-coronavirus drug given their previous stockpiling decisions and the scale of these  
stockpiles.

**Table 1:** Genera, hosts, and disease of and caused by coronaviruses of humans and companion animals

genus	virus	host	respiratory infection	enteric infection	hepatitis	neurologic infection	other
α	HCov-229E	Human	+			?	
	HCov-NL63	Human	+				
	PEDV	Pig		+			
	TGEV	Pig	+	+			+
	CCoV	Dog		+			
	FECov	Cat		+			
	FIPV	Cat	+	+	+	+	+
	RbCoV	Rabbit		+			+
β	HCov-OC43	Human	+	?		?	
	HCov-HKU1	Human	+				
	SARS-CoV	Human	+	+			+
	MERS-CoV	Human	+	+			+
	PHEV	Pig	+	+		+	
	BCoV	Cow	+	+			
	ECoV	Horse		+		+	
γ	IBV	Chicken	+		+		+
	TCoV	Turkey	+	+			
δ	SDCV	Pig		+			

HCov, human coronavirus; PEDV, porcine epidemic diarrhea virus; TGEV, transmissible gastroenteritis virus; CCoV, canine coronavirus; FECov, feline enteric coronavirus; FIPV, feline infectious peritonitis virus; RbCoV, rabbit coronavirus; SARS-CoV, severe acute respiratory syndrome coronavirus; MERS-CoV, Middle East respiratory syndrome coronavirus; PHEV, porcine hemagglutinating encephalomyelitis virus; BCoV, bovine coronavirus; ECoV, equine coronavirus; IBV, infectious bronchitis virus; TCoV, turkey coronavirus; SDCV, swine delta coronavirus. Adapted from (1).

### 3. NEW DRUG APPROVAL

Obtaining approval to use a new drug and subsequently bringing it to market typically requires more than ten years of pre-clinical development and clinical research. Obviously, this time frame precludes the development of a drug as part of the direct response to an outbreak of a novel infectious disease and calls for an alternative route towards approval in those situations. The most promising of these alternative routes permitted by FDA (United States Food and Drug Administration) regulations is the approval of a broad-spectrum antiviral drug before a pandemic, followed by an extension of the approved indication to include the newly emerged pathogen. However, when employing this quicker approach special criteria and prerequisites have to be met, which are outlined below. Furthermore, although these approval and emergency processes are based

1 on U.S. law, equivalent arrangements likely exist under European Union and national  
2 regulations as well, enabling a similar approval strategy in all high-income countries.

3  
4 As the etiological agent of the postulated future coronavirus pandemic is probably  
5 unknown, it first needs to be decided which related viruses will be the basis for the pre-  
6 clinical screening process. Inclusion criteria could be our knowledge-base for those par-  
7 ticular agents, the biosafety level required to handle them, the availability of cell culture  
8 and animal infection models, and preferably also an animal model reproducing disease.  
9 Furthermore, to increase the likelihood of the drug being efficacious against any new  
10 coronavirus, the viruses tested should ideally represent all genera of the coronavirus  
11 family but at least alpha- (e.g., HCoV-229E) and beta-coronaviruses (e.g., HCoV-OC43),  
12 which include the majority of currently known coronaviruses with a mammalian host.

13  
14 After obtaining the investigational new drug-status (IND status) and demonstrating  
15 safety in a clinical phase I study, efficacy needs to be shown either in patients with a  
16 confirmed coronavirus infection or in challenge studies where healthy volunteers are  
17 exposed to the virus within a closed, monitored setting. For the former option MERS  
18 patients may represent a small but suitable target group, should the virus continue to  
19 circulate in the Middle East. Alternatively, any of the four HCoVs may be considered for  
20 use in clinical trials. However, due to the mildness of symptoms of most HCoV infections,  
21 which typically do not prompt any clinical testing or differentiation against, for example,  
22 rhinoviruses, technical problems will likely arise during the recruitment process of  
23 patients. Therefore, challenge studies with HCoVs may be the only feasible option for  
24 phase II trials in which also the dose for treatment or suitability for prophylaxis can be  
25 established. Finally, HCoV patients already hospitalized with other illnesses affecting the  
26 immune system or respiratory tract could be a target group to complete the require-  
27 ments for approval in phase III trials by demonstrating the immediate benefit of an  
28 anti-coronavirus drug.

29  
30 If an outbreak occurs that is caused by a new coronavirus, the above drug will not be  
31 automatically approved for treatment of this new disease since its indication will at that  
32 time be limited to treatment of those CoVs evaluated in phase III trials. However, it is pos-  
33 sible to amend the original approval with additional indications as soon as they become  
34 apparent. Typically such a so-called prior approval supplement needs to be supported  
35 by new efficacy studies in animals as well as phase II and III trials (7). Yet, since the virus  
36 itself and the disease it causes will not be well defined during the early stages of an out-  
37 break, any studies involving human subjects would be considered unethical. Therefore,  
38 it is likely that the extension will have to be granted based on the two animal rule (8),  
39 which is meant for approval of drugs that for safety, ethical, or other reasons cannot

1 be tested for effectiveness in humans but are expected to be beneficial for patients.  
2 This alternative drug approval track requires a demonstration of efficacy in at least two  
3 animal models that are assumed to respond similar to humans. Once this data is deemed  
4 satisfactory, dispensing could start quickly given the fact that safety of the compound  
5 has already been demonstrated and a novel treatment is provided which will likely have  
6 an impact on survival. Ideally, data generated from treated patients should be collected  
7 and analyzed in real time to serve as uncontrolled clinical study and to enable a better  
8 determination of the doses needed for treatment but also for prophylaxis.

9  
10 Depending on the dynamics of the outbreak it may still be necessary to act faster than  
11 this approval process would allow. For these cases, it has to be possible to bypass the  
12 strict FDA requirements for a limited amount of time during emergency situations. Thus,  
13 it is generally always possible for each physician to prescribe any drug for any indication  
14 on a per patient basis. Therefore, it can be expected, especially in the absence of any  
15 other known treatments for life-threatening coronavirus infections, that the antiviral  
16 drug, even if it were only approved for treatment of HCoVs (or MERS-CoV), would be  
17 used for this new indication by informed physicians. However, as pharmaceutical  
18 companies themselves are forbidden to advertise the off-label use of their drugs, WHO  
19 (World Health Organization) or national bodies would need to promote this application.  
20 For this purpose a special emergency act is in place allowing the FDA commissioner to  
21 temporarily permit and promote the use of drugs for an unapproved indication if they  
22 are intended to treat or prevent life-threatening diseases and no approved alternative  
23 exists (9).

24  
25 In conclusion, although U.S. federal regulations for approval of new medical products are  
26 stringent to ensure efficacy but foremost consumer safety, there is still room to respond  
27 quickly to newly emerging public health threats. Nevertheless, the time from initial  
28 discovery of an active compound to approval will never be short enough to prevent a  
29 pandemic if the development is only started once first cases appear. Thus, preparation  
30 and planning ahead, even for apparently unlikely events that may have a huge impact  
31 on the global population and economy, is essential.

#### 32 33 34 **4. MARKET CONSIDERATIONS**

35  
36 In the absence of preparedness plans to combat a potential coronavirus pandemic, the  
37 following market considerations are in part based on information available for pandemic  
38 influenza.  
39



## 4.1 Market need

Whenever and wherever an outbreak of a contagious agent occurs, the primary goal has to be to prevent further spread and to contain the disease within a certain area. Since modern modes of transportation have made this increasingly difficult to achieve, rapid initiation and enforcement of countermeasures, as social distancing, travel restrictions, and pharmaceutical intervention, are all the more important. Although these measures usually are applied in combination, a recent study using pandemic influenza as model indicated that, in the absence of a vaccine, the curative and prophylactic use of antivirals has a bigger impact on the progression of an outbreak than any of the other countermeasures. In short, the authors concluded that under all pandemic severity categories an intervention strategy including antivirals leads to the biggest reduction of costs, both in terms of loss of life and economic loss, compared to a scenario without any intervention (10). Furthermore, antivirals are, next to a vaccine, the only measure providing protection for healthcare workers and other employees in essential services. Unfortunately, there is currently no drug specifically against any coronavirus on the market that would at least offer the prospect of treatment in case of a coronavirus pandemic. This again emphasizes the need for the development of an anti-coronavirus drug before a major outbreak occurs.

## 4.2 Purpose of the antiviral drug treatment

The anti-coronavirus drug is intended to be used as an outbreak control measure complementary to non-pharmaceutical interventions. Accordingly, the drug will be used both curatively as well as prophylactically (if allowed by the infection kinetics), with the latter presumably being of greater importance in respect to sales volumes. To serve this purpose, the drug will need to be included in public and private stockpiles. Additionally, the development of a diagnostic test may be a valuable complementary tool to support the drug's use.

Although human coronaviruses are estimated to cause 10 to 15 percent of common cold cases (11), this field is not expected to be a significant market considering the mild nature and short duration of symptoms and the consequential lack of clinical testing. However, there could be a small market for treatment of HCoV infections in patients with certain co-morbidities.

Besides humans many animals are hosts to one or multiple coronaviruses, suggesting the possibility of veterinary applications. However, vaccines are on the market against the three major coronaviruses that affect livestock, BCoV (bovine coronavirus) (12), TGEV

1 (porcine transmissible gastroenteritis virus) (13), and IBV (avian infectious bronchitis  
2 virus) (14). Due to this limitation, usage of an antiviral in an animal health setting is not  
3 expected to provide a significant market and will not be considered further.

#### 4 5 **4.3 Customers and market size** 6

7 Currently none of the three major healthcare authorities CDC (United States Centers  
8 for Disease Control and Prevention), ECDC (European Center for Disease Prevention and  
9 Control), or WHO have developed a specific response plan for the event of a severe coro-  
10 navirus pandemic. Thus, any of the following considerations are based on recommenda-  
11 tions for an influenza pandemic made by the U.S. Department of Health and Human  
12 Services (15). Since both influenza and coronaviruses cause respiratory symptoms and  
13 appear to be transmitted via respiratory droplets and contaminated surfaces (16;17),  
14 these considerations likely also apply in scope and size for a potential coronavirus pan-  
15 demic.

16  
17 In order to bridge the gap between the beginning of an influenza outbreak and the start  
18 of antiviral drug or vaccine production, the preparedness plan recommends the buildup  
19 of public stockpiles with amounts of antivirals that will suffice to treat about 25 percent  
20 of the U.S. population (80 million people). If we apply the same percentage to the EU,  
21 125 million regimens would need to be available. However, since EU member states do  
22 not necessarily follow the same policy, the actual coverage of each individual state may  
23 deviate considerably. For instance, a survey made by the ECDC in 2006 counted national  
24 stockpiles of influenza antivirals sufficient for treatment of 5 to 50 percent of the coun-  
25 tries' respective populations and listed many nations intending to increase this level  
26 further (18). If we assume a similar coverage for the event of a coronavirus pandemic, the  
27 total demand in high-income countries with about 1.1 billion inhabitants may add up to  
28 150-250 million regimens. Additionally, it may be considered to increase a coronavirus  
29 stockpile further as it is unlikely that a vaccine will be available as fast as one against  
30 influenza virus (typically six months). However, given the fact that the anti-coronavirus  
31 and anti-influenza virus drug stockpiles would compete for the same budget and in-  
32 frastructure, it seems to be more realistic to expect a significant downward deviation  
33 from the set targets. In contrast to drugs against influenza, which re-occurs seasonally,  
34 a coronavirus drug would be restricted to those stockpiles, resulting in a limited regular  
35 turnover during non-outbreak phases, primarily due to shelf life limitations. On the  
36 other hand, if a pandemic would occur, demand could increase dramatically. However,  
37 the extent to which a company could profit from this situation would depend on its  
38 potential for a rapid scale-up of production and distribution. Yet, in order to obtain this  
39

1 potential, production capacity would have to match the hypothetical future demand  
2 and not the regular turnover, thus causing additional costs.

3  
4 Next to treatment, the recommendations include expansion of the U.S. national stock-  
5 pile to provide multiple drug regimens for outbreak prophylaxis for people working in  
6 healthcare and emergency services, as well as for people with compromised immunity  
7 but not for the general public as undifferentiated mass prophylaxis is unfeasible. Addi-  
8 tionally, people living in group settings, for instance nursing homes, should be provided  
9 with single regimen post-exposure prophylaxis if an outbreak occurs at their facility. Ide-  
10 ally, an anti-coronavirus drug should also be suitable for prophylactic use. The number  
11 of additionally required regimens is, however, hard to predict as it depends, for instance,  
12 on the specific estimates on the progression of a local outbreak, drug efficacy, and the  
13 efficiency of human-to-human transmission of the newly emerging virus. Alternatively,  
14 the development of SARS symptoms in relation to the viral titer gives rise to the assump-  
15 tion that early intervention may be as suitable in containing an outbreak as infection  
16 prevention. In this particular case, patients started to spread the disease only from  
17 day seven after onset of symptoms and peak viral loads were reached three days later,  
18 leaving a window of several days to intervene with antivirals in order to suppress the  
19 infection (19). Nevertheless, as it cannot be predicted whether these dynamics will be  
20 shared by the new pandemic coronavirus, it seems likely that a stockpile would at least  
21 to a certain extent allow for prophylactic drug use.

22  
23 To complement public stockpiles, employers are encouraged to ensure private stocks to  
24 offer protection to workers who are critical to maintain operations in businesses essen-  
25 tial to the community, as for example, power grid or money and food supply. Moreover,  
26 employers may consider to arrange additional stocks for employees overseas, who may  
27 not have access to any publicly supplied drugs, or, if concerns regarding the availability  
28 and timely dispensing of public stockpiles exist, for the early treatment of employees  
29 falling ill. Additionally, an employer has to consider which dispensing model should be  
30 used. The options are: triggered dispensing from a central pharmacy when an outbreak  
31 starts and pre-pandemic dispensing to the actual consumer. The most fundamental  
32 difference, as it relates to market size, lies in the applicable expiration date of the drug.  
33 If the drug is to be stored in a licensed pharmacy, the FDA-approved expiration date  
34 applies, for example seven years for the anti-influenza drug tamiflu (20), while the same  
35 is limited to six to twelve months once the drug has been distributed to the consumer.  
36 Obviously, the latter would considerably increase the turnover of the antiviral within  
37 private stockpiles. To conclude, depending on the number of employers following these  
38 recommendations, the number of employees to protect, and the choice regarding dis-  
39

1 pensing, private stockpiles could represent a significant market segment next to public  
2 stockpiles.

#### 3 4 **4.4 Expected customer advantage**

5  
6 The experience of the SARS pandemic has reminded us that these kinds of events do  
7 not only cost lives but also have a huge impact on economy. Although this pandemic  
8 was relatively small with less than 9,000 cases including about 800 deaths and was con-  
9 tained within half a year, its costs to the global economy were estimated at about \$30  
10 billion. In comparison, the costs of a global influenza pandemic, which may more closely  
11 resemble a future pandemic with a more contagious coronavirus, may well exceed \$500  
12 billion (21). These estimates, including costs for healthcare and countermeasures, loss  
13 of productivity due to prophylactic absenteeism, illness or death, loss of productivity  
14 due to supply chain effects, decreased demand for products and services, and decrease  
15 of tourism and travel, make clear how important the rapid elimination of a contagious  
16 disease is. An antiviral can help in this process and considerably shorten the duration of  
17 an outbreak and can thereby reduce the costs of the disease for individual companies  
18 and society.

#### 19 20 **4.5 Target product profile**

21  
22 As the drug is, on the one hand, intended to provide protection for people coming into  
23 contact with patients and, on the other hand, intended to enable treatment of these  
24 patients, two different subpopulations with quite different requirements need to be  
25 considered. According to the perceived relative importance of these two subgroups,  
26 drug requirements arising from prophylactic use will be given priority and shall be ad-  
27 dressed first.

28  
29 Besides the disruption of the chain of infection, continuation of essential community  
30 services is the main goal of the prophylactic administration of the drug. Therefore, the  
31 antiviral cannot cause side effects that interfere with the provision of such services. Ad-  
32 ditionally, the absence of noteworthy side effects will increase the general compliance  
33 to take the antiviral. For the same purpose, the drug should be easy to take, preferen-  
34 tially as an oral, single daily dose. Next to side effects also cytotoxicity, especially for  
35 the liver, has to be low over the duration of intake. However, the length of this period  
36 is hard to predict and would essentially depend on the severity of the local outbreak  
37 and the number of days the virus is transmitted by patients but may fall in a window  
38 of several weeks up to a few months. Furthermore, since the recipients of prophylactic  
39 treatment are healthy individuals, full inhibition of viral replication may not be necessary

1 to suppress the development of symptoms and spread of the disease. Thus, the dose for  
2 effective prophylaxis could potentially be lower than the one for intervention. This fact  
3 in turn may help to reduce side effects during the indicated period.

4  
5 In contrast, more severe side effects would be expected for the higher dose of the anti-vi-  
6 ral used for treatment of patients, but this would also be more acceptable for this group.  
7 However, to relieve the strain on the healthcare infrastructure, which has seen average  
8 admission periods exceeding 25 days for SARS (21), it would be desirable to be able to  
9 release mild cases to their homes. Therefore, side effects requiring permanent medi-  
10 cal attention should not occur. Another necessity, particular for one group of patients,  
11 became apparent from our experiences with MERS CoV infection, which causes the most  
12 severe symptoms, often eventually leading to the death of the patient, in patients with  
13 medical preconditions (22). Consequently, the antiviral has to be compatible with other  
14 drugs frequently used to treat common preconditions.

15  
16 Finally, the last drug requirement that should be met is arising from the intention to  
17 stockpile rather than from the targeted user group. As it is unpredictable when the  
18 antiviral will be needed, the necessity for drug turnover due to shelf life expiration as  
19 well as the costs for long term storage and distribution should be taken into account.  
20 For instance, although cooled warehouses exist for certain parts of the U.S. emergency  
21 stockpile, cheaper storage and transport at ambient temperature is surely preferred and  
22 may be a factor in the decision for or against stockpiling a certain drug. Thus, long-term  
23 stability at ambient temperature is another desirable drug property.

## 24 25 **4.6 Competing products**

### 26 27 *4.6.1 Vaccine*

28  
29 Efforts have been initiated to develop vaccines against SARS and MERS coronaviruses  
30 (23;24), which may, if approved, provide cross-protection against other coronaviruses  
31 and could potentially be distributed with little response time. However, given the high  
32 antigenic diversity of coronaviruses, such a case of effective cross-protection may be the  
33 exception rather than the rule (1).

34  
35 In contrast, an effective vaccine specifically tailored against the novel pandemic agent  
36 will most likely not be available within the first year of a pandemic since multiple difficul-  
37 ties in the vaccine's development could arise. First, natural infections with HCoV's seem  
38 to fail to induce long-lasting immunity as test subjects could not only be re-infected  
39 with the same viral strain within one year of the first challenge but also redeveloped

1 symptoms. If this observation represents a general feature of the human immune re-  
2 sponse towards coronaviruses, strategies that will overcome this shortcoming need to  
3 be developed. Additionally, vaccination against feline enteric coronavirus (FECoV) has  
4 been associated with antibody-dependent enhancement of disease. Next to these two  
5 antibody-related complications, additional challenges will likely arise from the genetic  
6 properties of coronaviruses. Foremost their ability to mutate quickly and to recombine  
7 presents a problem. For instance, a case of recombination between a live attenuated IBV  
8 vaccine strain and a field strain was reported, which had caused a local outbreak and  
9 raised some safety concerns. Furthermore, recombination likely increases the range of  
10 antigenic variants against which a vaccine would need to provide protection (1). Also, as  
11 the primary target of a vaccine will be the spike (S) surface protein, mutations affecting  
12 the efficacy of a vaccine may arise more readily than those providing resistance against  
13 an antiviral, which may target a more conserved enzyme.

14  
15 To conclude, although a potent and safe vaccine is as an outbreak control measure surely  
16 superior to an antiviral drug, the probability that a vaccine actually can be developed in a  
17 timely fashion when an outbreak occurs is low. With this said, it also has to be made clear  
18 that eventually a vaccine will be approved if the virus stays in circulation long enough to  
19 warrant the efforts to overcome the described complications. As soon as this happens,  
20 the market share of the putative, novel antiviral will gradually decrease. Nevertheless, in  
21 view of the intended use of this drug as part of an emergency stockpile, a coronavirus  
22 vaccine does not represent a source of competition during the initial outbreak situation.

#### 23 24 4.6.2 Antiviral drug candidates

25  
26 Over the past decade, several academic groups screened for and identified compounds  
27 active against SARS-CoV. These compounds included inhibitors of viral fusion and entry,  
28 as well as of essential viral enzymes as proteases, RNA-dependent RNA polymerase, and  
29 helicase. Furthermore, siRNAs targeting the expression of structural or accessory pro-  
30 teins were evaluated. Next to these virus-oriented strategies, also compounds targeting  
31 host factors, which may be essential for virus replication but temporarily dispensable for  
32 the human host, were explored (25). Currently, efforts are made to test the efficacy of  
33 some of these drug candidates against MERS-CoV, which may lead to the identification  
34 of compounds with a broader activity against coronaviruses (26).

35  
36 Although it is encouraging that inhibitory compounds have been found in *in vitro* assays  
37 and animal models, none of these compounds have been brought to clinical phase I  
38 studies yet, and they are unlikely to be tested in such a setting by any of the academic  
39

1 patent holders. However, licenses may have been or may be acquired in the future by  
2 any competitor.

3

#### 4 4.6.3 Repurposed drugs

5

6 In the absence of an approved drug, different strategies of treatment were followed  
7 in response to the SARS pandemic. These included administration of interferon, cor-  
8 ticosteroids, ribavirin, and the HIV protease inhibitors Lopinavir/ritonavir. As the case  
9 numbers of SARS have been generally low, no sufficient evidence supporting or refuting  
10 the efficacy of any of these treatments could be gathered from patient data. Therefore,  
11 none of these experimental interventions are currently recommended as treatment for  
12 MERS (27). However, these or other rededicated approved drugs may compete with or  
13 at least may define the minimal efficacy required for a new anti-coronavirus drug if their  
14 efficacies are better characterized.

15

16

## 17 5. CONCLUSION

18

19 Bringing a new drug to the market requires on average an investment of \$1.2 billion (28).  
20 To ensure the return of this significant investment, any factors influencing the future  
21 marketability need to be carefully assessed and opportunities weighed against risks.  
22 For this analysis of the potential of an antiviral drug against coronaviruses, a scenario  
23 of an imminent pandemic caused by an at present unknown coronavirus was assumed.  
24 Implicitly this scenario acknowledges the fact that currently circulating coronaviruses  
25 either do not pose a threat to public health due to the mild clinical symptoms they cause  
26 (HCoVs) or are associated with low case numbers (MERS-CoV). Therefore, pre-pandemic  
27 stockpiling is, from a current perspective, the only purpose providing a sufficient market  
28 need and size to generate profit. However, this intended use also poses some inherent  
29 risks for development and marketing. First, as a broad-spectrum antiviral is required, it  
30 can be expected that the discovery and development process will be longer and more  
31 difficult, and thus more expensive, than for drugs with more restricted therapeutic use.  
32 Additionally, any efforts to demonstrate the value of the antiviral will be hampered by  
33 the fact that the pathogen that ultimately may become the trigger for dispensing of  
34 the stockpile is not present at the time of the purchase decision. Thus, in a competitive  
35 environment where other drugs are developed for which a market need and health  
36 benefit can clearly be proven, a coronavirus antiviral may not be considered to have  
37 a positive cost-benefit ratio to make it a reasonable choice for stockpiling, especially  
38 under continuing austerity measures. Therefore, an essential question would be how the  
39 risk of an emerging coronavirus is perceived by potential buyers. For example it could,

1 on the one hand, be expected that the number and intensity of virus outbreaks but also  
2 the frequency of emergence events of zoonotic viruses will increase in the future due  
3 to increased mobility and population density in urban areas, as well as the expansion of  
4 settlements into previously not inhabited areas. On the other hand, it will remain unpre-  
5 dictable to which family a newly emerging virus will belong. How these two arguments  
6 are weighed and may affect the decision to stockpile will likely differ from country to  
7 country, also depending on their individual financial situation as well as their willingness  
8 to take risks. Given their previous commitment to prepare for the event of future out-  
9 breaks caused by different biological agents and the recent classification of SARS-CoV  
10 as a Select Agent (29), recognizing a potential public health threat, the United States  
11 are one of the most likely customers for an anti-coronavirus drug. This assumption is  
12 additionally supported by their recent purchase of 2 million doses of a smallpox antiviral  
13 for inclusion into the national stockpile (30). Although the volume of this purchase was  
14 far smaller (because 300 million doses of the cheaper, complementary vaccine are stock-  
15 piled) than would be expected for an anti-coronavirus drug, the purchase of this antiviral  
16 demonstrates a mode of public-private partnership that could also be applicable for a  
17 coronavirus antiviral. In essence the smallpox antiviral was developed by order of the  
18 U.S. federal government, which granted a total of \$100 million in research grants and  
19 contracts through several governmental agencies and agreed to purchase 2 million to  
20 14 million courses for a minimum of \$433 million and up to \$2.8 billion (31;32). In the  
21 absence of regular retail sales and the presence of high uncertainty in respect to sales  
22 volumes, seeking such partnerships to mitigate financial risks may be the only feasible  
23 option to realize such projects. To conclude, if such a public commitment for the devel-  
24 opment of a coronavirus drug is not evident, the inherently high risk of this project will  
25 likely preclude any private initiative, even though market need and size would warrant  
26 an investment.

## 27 28 29 **6. ACKNOWLEDGEMENTS** 30

31 I like to thank Leon Martens and Greg Fanning for their dedication, enthusiasm, and sup-  
32 port that made this secondment possible. I am also grateful for the insights and informa-  
33 tion provided by Wendy Balemans, Els van Beirendonck, Jacques Bollekens, Olivier Brun,  
34 Walter van den Broeck, Henk Laanen, Johan Peters, George Wan, and Lara Wolfson. I wish  
35 to extent my thanks to Caroline Sage and Evelyne Vanneste for organizing the meetings  
36 at Janssen in Beerse. Finally, I want to acknowledge Leon Martens, Olivier Brun, Clara  
37 Posthuma, Eric Snijder, and Alexander Gorbalenya for commenting on the manuscript.  
38  
39



## 1 REFERENCE LIST

- 2  
3 1. Lai, MMC, Perlman S, Anderson LJ. *Coronaviridae*. In Fields' Virology Volume One, 5<sup>th</sup> edition.  
4 Philadelphia: Lippincott, Williams & Wilkins, 2007. p. 1305-1324
- 5 2. World Health Organization. SARS outbreak contained worldwide. [press release], 5 July 2003.  
6 Available at: <http://www.who.int/mediacentre/news/releases/2003/pr56/en> [accessed 23 May  
7 2013]
- 8 3. World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV) – update.  
9 2014 [online] Available at: [http://www.who.int/csr/don/2014\\_07\\_14\\_mers/en/](http://www.who.int/csr/don/2014_07_14_mers/en/) [accessed 21 July  
10 2014]
- 11 4. Centers for Disease Control and Prevention. Middle East respiratory syndrome – transmission.  
12 2014 [online] Available at: <http://www.cdc.gov/coronavirus/mers/about/transmission.html> [ac-  
13 cessed 23 July 2014]
- 14 5. Centers for Disease Control and Prevention. Severe acute respiratory syndrome – frequently  
15 asked questions. 2012 [online] Available at: <http://www.cdc.gov/sars/about/faq.html> [accessed  
16 23 July 2014]
- 17 6. Centers for Disease Control and Prevention. CDC estimates of 2009 H1N1 influenza cases, hos-  
18 pitalizations, and deaths in the United States, April 2009 – March 13, 2010. [online] Available at:  
19 [http://www.cdc.gov/h1n1flu/estimates/April\\_March\\_13.htm](http://www.cdc.gov/h1n1flu/estimates/April_March_13.htm) [accessed 23 May 2013]
- 20 7. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug  
21 Evaluation and Research. Guidance for industry - changes to an approved NDA or ANDA. 2004  
22 [pdf] Available at: [http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInfor-  
23 mation/Guidances/ucm077097.pdf](http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm077097.pdf) [accessed 2 December 2013]
- 24 8. Approval based on evidence of effectiveness from studies in animals, 21 C.F.R. §314.610. (2013).
- 25 9. Food and Drug Administration. Emergency preparedness and response - Emergency use autho-  
26 rization. 2013 [online] Available at: [http://www.fda.gov/emergencypreparedness/counterterror-  
27 ism/ucm182568.htm](http://www.fda.gov/emergencypreparedness/counterterrorism/ucm182568.htm) [accessed 2 December 2013]
- 28 10. Kelso JK, Halder N, Postma MJ, *et al*. Economic analysis of pandemic influenza mitigation strate-  
29 gies for five pandemic severity categories. BMC Public Health. 2013; 13:211
- 30 11. Public Health Agency of Canada. Pathogen safety data sheet – human coronavirus. 2010 [online]  
31 Available at: <http://www.phac-aspc.gc.ca/lab-bio/res/psds-ftss/coronavirus-eng.php> [accessed 8  
32 June 2013]
- 33 12. Merck Animal Health. Guardian – Overview. 2009 [online] Available at: [http://www.merck-animal-  
34 health-usa.com/products/130\\_163325/productdetails\\_130\\_163625.aspx](http://www.merck-animal-health-usa.com/products/130_163325/productdetails_130_163625.aspx) [accessed 11 June  
35 2013]
- 36  
37  
38  
39

- 1 13. Merck Animal Health. Transmissible gastroenteritis – overview. 2009 [online] Available at: [http://www.merck-animal-health-usa.com/diseases/130\\_120816/ProductDetails\\_130\\_121064.aspx](http://www.merck-animal-health-usa.com/diseases/130_120816/ProductDetails_130_121064.aspx) [accessed 11 June 2013]
- 2
- 3
- 4 14. Merck Animal Health. Bron-Newcavac-SE – overview. 2009 [online] Available at: [http://www.merck-animal-health-usa.com/products/130\\_120659/productdetails\\_130\\_121097.aspx](http://www.merck-animal-health-usa.com/products/130_120659/productdetails_130_121097.aspx) [ac-
- 5
- 6
- 7
- 8 15. U.S. Department of Health and Human Services. Considerations for antiviral drug stockpiling by
- 9
- 10
- 11
- 12
- 13 16. Centers for Disease Control and Prevention. How flu spreads. 2013 [online] Available at: <http://www.cdc.gov/flu/about/disease/spread.htm> [accessed 3 November 2013]
- 14
- 15
- 16 17. Centers for Disease Control and Prevention. Interim domestic guidance on the use of respirators
- 17
- 18
- 19
- 20 18. European Center for Disease Prevention and Control. Technical Report: Pandemic influenza
- 21
- 22
- 23 19. Anderson RM, Fraser C, Ghani AC, *et al.* Epidemiology, transmission dynamics and control of SARS:
- 24
- 25
- 26 20. World Health Organization. Policy on Oseltamivir shelf-life expiry. 2010 [PDF] Available at: [http://www.who.int/medicines/areas/quality\\_safety/quality\\_assurance/Oseltamivir\\_](http://www.who.int/medicines/areas/quality_safety/quality_assurance/Oseltamivir_shelf-life_QAS11-398_30032011.pdf)
- 27
- 28
- 29
- 30 21. Begley S. Flu-economics: the next pandemic could trigger global recession. 2013 [online] Avail-
- 31
- 32
- 33
- 34 22. WHO MERS-CoV Research Group. State of knowledge and data gaps of Middle East respiratory
- 35
- 36
- 37
- 38 23. Roberts A, Lamirande EW, Vogel L, *et al.* Animal models and vaccines for SARS-CoV infection. *Virus Res.* 2008; 133(1):20-32.
- 39
24. Almazan F, DeDiego ML, Sola I, *et al.* Engineering a replication-competent, propagation-defective Middle East respiratory syndrome coronavirus as a vaccine candidate. *MBio.* 2013; 4(5):e00650-13
25. Barnard DL, Kumaki Y. Recent developments in anti-severe acute respiratory syndrome coronavirus chemotherapy. *Future Virol.* 2011; 6(5):615-31

- 1 26. Adedeji AO, Sarafianos SG. Antiviral drugs specific for coronaviruses in preclinical development.  
2 Curr Opin Virol. 2014; 8:45-53.
- 3 27. Public Health England, International Severe Acute Respiratory & Emerging Infection Consortium.  
4 Treatment of MERS-CoV: Decision support tool; Clinical Decision Making Tool for Treatment  
5 of MERS-CoV v.1.1. 2013 [online] Available at: [http://www.hpa.org.uk/webc/HPAwebFile/](http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317139281416)  
6 [HPAweb\\_C/1317139281416](http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317139281416) [accessed 29 October 2013]
- 7 28. Pharmaceutical Research and Manufacturers of America. Pharmaceutical research industry pro-  
8 file. 2013 [pdf] Available at: [http://phrma.org/sites/default/files/pdf/](http://phrma.org/sites/default/files/pdf/PhRMA%20Profile%202013.pdf) [PhRMA%20Profile%202013.](http://phrma.org/sites/default/files/pdf/PhRMA%20Profile%202013.pdf)  
9 [pdf](http://phrma.org/sites/default/files/pdf/PhRMA%20Profile%202013.pdf) [accessed 24 July 2014]
- 10 29. U.S. Department of Health and Human Services select agents and toxins, 42 C.F.R. §73.3. (2013).  
11
- 12 30. McNeil Jr. D. Wary of attacks with smallpox, U.S. buys up a costly drug. The New York Times,  
13 [online] 12 march 2013. Available at: [http://www.nytimes.com/2013/03/13/](http://www.nytimes.com/2013/03/13/health/us-stockpiles-smallpox-drug-in-case-of-bioterror-attack.html?_r=1) [health/us-stockpiles-](http://www.nytimes.com/2013/03/13/health/us-stockpiles-smallpox-drug-in-case-of-bioterror-attack.html?_r=1)  
14 [smallpox-drug-in-case-of-bioterror-attack.html?\\_r=1](http://www.nytimes.com/2013/03/13/health/us-stockpiles-smallpox-drug-in-case-of-bioterror-attack.html?_r=1) & [accessed at 25 July 2014]
- 15 31. SIGA Technologies, Inc., 2011. SIGA Technologies awarded U.S. government contract valued at up  
16 to \$2.8 billion. [press release], 13 May 2011. Available at: [http://investor.siga.com/releasedetail.](http://investor.siga.com/releasedetail.cfm?ReleaseID=577406)  
17 [cfm?ReleaseID=577406](http://investor.siga.com/releasedetail.cfm?ReleaseID=577406) [accessed 25 July 2014]
- 18 32. SIGA Technologies, Inc., 2014. Annual report 2013. [pdf] Available at: [http://files.shareholder.com/](http://files.shareholder.com/downloads/SIGA/3353439484x0xS1010086-14-6/1010086/filing.pdf)  
19 [downloads/SIGA/3353439484x0xS1010086-14-6/1010086/filing.pdf](http://files.shareholder.com/downloads/SIGA/3353439484x0xS1010086-14-6/1010086/filing.pdf) [accessed 25 July 2014]  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

General discussion

# CHAPTER 8



## 1 ONE GOAL, MANY SOLUTIONS: MECHANISTIC ALTERNATIVES FOR 2 ARTERIVIRUS RNA SYNTHESIS AND CAPPING

### 3 4 A few words on nidovirus diversity and genome architecture

5  
6 The order *Nidovirales* with its families *Arteriviridae*, *Coronaviridae* (subfamilies *Corona-*  
7 *virinae* and *Torovirinae*), *Mesoniviridae*, and *Roniviridae* comprises members that are  
8 genetically more distant from each other than the most diverged organisms of the Tree  
9 of Life (1-6). It may thus be asked, what defines a nidovirus and what unites members  
10 of the order. On the protein level three enzymatic subunits, the chymotrypsin-like main  
11 protease 3CL<sup>pro</sup>, the RNA-dependent RNA polymerase (RdRp), and the superfamily (SF) 1  
12 helicase, are the most conserved (1). However, none of these enzymatic domains can be  
13 considered nidovirus-specific as they are also common in many other positive-stranded  
14 (+) RNA viruses. Instead, bioinformatics studies identified the N-terminal domains of the  
15 proteins including also the RdRp or helicase domains as genetic markers for the order  
16 (chapter 5 and (5)). On a higher level, nidoviruses share a unique genome organization  
17 comprising a conserved array of features, encoded in the two 5' replicase open reading  
18 frames (ORFs), and multiple 3' ORFs that are translated from subgenomic (sg) mRNAs. In  
19 the 5' to 3' direction, this array includes the 3CL<sup>pro</sup> flanked by two transmembrane pro-  
20 teins, a ribosomal frameshift site (RFS), the RdRp, and the helicase. The positioning of the  
21 helicase subunit downstream of the RdRp is not observed for any other group of (+) RNA  
22 viruses, and the implications of this organization have remained elusive so far. However, it  
23 is assumed that the particular arrangement of the array reflects strong constraints due to  
24 certain essential and universal requirements of the nidovirus replication cycle, which are  
25 poorly understood (5). Next to these core attributes, additional less conserved domains  
26 may be integrated into the replicase of specific members or subgroups of the order (see  
27 Figure 1 of chapter 6, p. 199). These additional domains, as well as a further expansion  
28 of the size of conserved proteins, were the major contributors to the lineage-specific  
29 increase in genome length of intermediate (*Mesoniviridae*) and large (*Coronaviridae*,  
30 *Roniviridae*) nidoviruses compared to that of the smaller arteriviruses (7). The profound  
31 divergence of nidovirus families on the protein level must have functional implications  
32 and calls for caution when attempting to generalize our limited biochemical knowledge  
33 on nidoviruses, which is based on studies of only a few corona- and arteriviruses. In this  
34 context, the following paragraphs discuss alternative interpretations of the arterivirus  
35 data presented in this thesis and elsewhere and formulate plausible hypotheses, which  
36 are worth pursuing in future studies on the molecular details of arterivirus replication.

## 1 Initiation of RNA synthesis

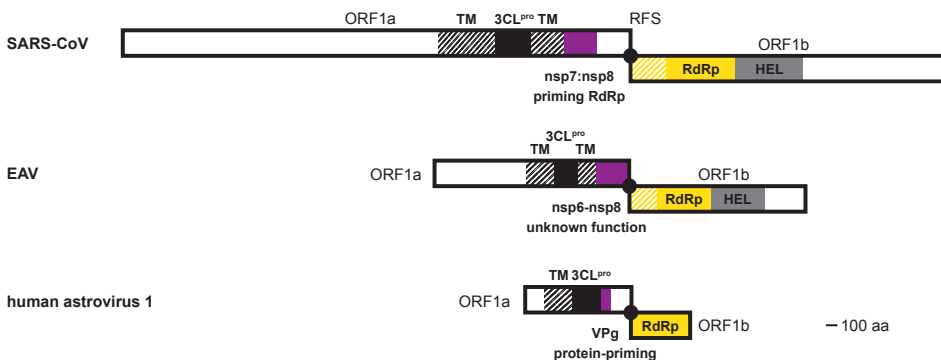
2

3 The copying of genetic information with the intent to produce progeny and blueprints  
4 for protein production is one of the few characteristics viruses and living organisms  
5 have in common. It is therefore not surprising that the “cupped right-hand” structure  
6 of canonical polymerases and their mechanisms are heavily conserved, even among  
7 otherwise genetically distant viruses and organisms (8;9). Nevertheless, some variation  
8 exists in the mechanisms used to initiate nucleic acid synthesis. In general two types of  
9 initiation can be distinguished: primer-dependent and primer-independent (also called  
10 *de novo*). Polymerases capable of the latter start nucleic acid synthesis by joining two  
11 NTPs, either independently or in association with a template but not necessarily with its  
12 3' end. This dinucleotide then gets elongated in the 5'-3' direction in a strictly template-  
13 dependent manner. In contrast, primer-dependent polymerases rely on another enzyme  
14 to generate this starting dinucleotide (9). Which type of initiation a polymerase utilizes  
15 has been associated with the presence (primer-dependent) or absence (*de novo*) of the  
16 conserved so-called G motif in the enzyme (10). Particularly, this motif was found in  
17 the putative RdRps of the *Arteriviridae* and *Coronaviridae* (11;12), and of the other two  
18 nidovirus families (Gorbalenya, personal communication). In the RdRp of severe acute  
19 respiratory syndrome coronavirus (SARS-CoV), designated nsp12, the presence of this  
20 motif was correlated with primer-dependent RdRp activity (13). These observations  
21 thus provided a rationale for the earlier described, but still debated, non-processive *de*  
22 *novo* RdRp activity of a smaller coronavirus protein, nsp8, encoded upstream of the RFS  
23 and lacking canonical polymerase motifs (14-16), which was proposed to generate the  
24 primers for the nsp12-based RdRp. In contrast to the situation for coronavirus nsp12,  
25 the G motif was found to be incomplete in the arterivirus equine arteritis virus (EAV)  
26 RdRp designated nsp9 (12). Additionally, the protein with the proposed “accessory  
27 RdRp” activity was found to be conserved in corona- and toroviruses, but its presence in  
28 other nidoviruses is uncertain due to very low sequence similarity in the corresponding  
29 part of the replicase protein (Gorbalenya, personal communication). It may therefore be  
30 speculated that, contrary to its coronavirus homolog, the RdRp encoded downstream  
31 of the RFS in arteriviruses is itself a *de novo*-initiating polymerase. This hypothesis was  
32 seemingly confirmed in *in vitro* assays using recombinant EAV nsp9 (12). Yet, this finding  
33 could not be extended to activity on natural EAV RNA templates and could furthermore  
34 not be reproduced in the present study (chapter 4).

35

36 Interestingly, EAV nsp9 was shown to possess a second enzymatic activity (chapter 5),  
37 which may be part of an alternative priming mechanism that is independent of the  
38 recruitment of a second cognate RdRp acting as a primase. This mechanism, which is  
39 exclusively employed by viral polymerases, is characterized by the formation of short,

1 protein-linked polynucleotides that are subsequently positioned at the template's 3' end  
 2 to allow elongation by the polymerase. The enzyme activity catalyzing this reaction is,  
 3 where known, exerted by the polymerase itself with the help of a substrate protein  
 4 that is typically named VPg (viral protein genome-linked) and ranges in size from 20 to  
 5 more than 200 amino acids (17;18). So far this type of "protein priming" was described  
 6 for double-stranded (ds) DNA viruses of the *Hepadnaviridae* (19), dsRNA viruses of the  
 7 *Birnaviridae* (20), and (+) RNA viruses of the *Picornavirales* (21-25), *Caliciviridae* (26),  
 8 *Potyviridae* (27;28), *Permutotetraviridae* (29), and *Astroviridae* (30). Intriguingly, despite  
 9 having genomes of only half the size of those of arteriviruses, the latter share the basic  
 10 genome organization of nidoviruses including the 3CL<sup>pro</sup>-RFS-RdRp array and expres-  
 11 sion of structural proteins from a 3'-coterminal sg mRNA (Figure 1). It was therefore  
 12 speculated that an early nidovirus ancestor might have resembled the contemporary  
 13 astroviruses (5). Although this genetic similarity is no guarantee that also mechanistic  
 14 details of viral replication correspond, it is tempting to speculate that the mode of initia-  
 15 tion of RNA synthesis might only have been altered upon acquisition of a second RdRp  
 16 in the course of the genome expansion of the large (and intermediate) nidoviruses.



28 **Figure 1:** Replicase organization of nidoviruses and human astrovirus. Protein domains belonging to the  
 29 conserved functional array of nidoviruses and their counterparts in astrovirus are indicated. Proteins that  
 30 are known or hypothesized to take part in the initiation of RNA synthesis are depicted in purple. TM, trans-  
 31 membrane domain; 3CL<sup>pro</sup>, 3C-like protease; black dot and RFS, ribosomal frameshift site; RdRp, RNA-de-  
 32 pendent RNA polymerase; HEL, helicase/RNA triphosphatase (adapted from (5)).

33 In the first step of protein-priming a nucleotide monophosphate is covalently attached  
 34 to the substrate protein under release of pyrophosphate, a reaction classified as nucleo-  
 35 tidylation. In most cases this substrate is not part of the polymerase subunit itself, but  
 36 the extent of auto-nucleotidylation may vary in response to reaction conditions, as it  
 37 was described for the RdRp of poliovirus (*Picornavirales*) (23). Currently, the only known  
 38 notable exceptions to this general trend are proteins of the hepadnavirus hepatitis B  
 39 virus and birnaviruses, which simultaneously serve as enzyme and substrate (19;20).



1 Furthermore, for infectious bursa disease virus (*Birnaviridae*) VP1, which contains the  
2 RdRp domain, it was shown that auto-nucleotidylation at the site located upstream of  
3 the RdRp domain does not depend on the conserved polymerase active site (20). In  
4 contrast, such a dependence was observed for VP1 of infectious pancreatic necrosis  
5 virus, another distantly related birnavirus (31). That study demonstrated a template-  
6 independent auto-guanlylation activity that modified a serine residue conserved in  
7 birnaviruses. Interestingly, only the fraction of VP1 molecules that served as primers was  
8 guanylated while other, non-modified RdRp molecules served in elongation (32). This  
9 mechanism may also be conserved in the (+) RNA *Permutotetraviridae* (29).

10  
11 In these respects nucleotidylation of EAV nsp9 (chapter 5) may behave quite similar to  
12 that seen in birnaviruses. However, a rough estimation indicated that only a very small  
13 fraction (<1%) of nsp9 proteins was labeled with UMP or GMP – not taking into account  
14 the potential presence of inactive nsp9 molecules, which may be numerous given the  
15 instability of this recombinant protein. It may thus be questionable if a birnavirus-like  
16 division between priming and elongating protein fractions can be envisioned for EAV  
17 nsp9. In addition, we noticed a tendency to transfer this label to other EAV proteins  
18 but also unrelated polypeptides. Yet, as we were so far unable to identify a protein that  
19 serves as the preferred acceptor for UMP or GMP, we currently consider this transfer  
20 activity an artifact of the *in vitro* assay, maybe due to the general instability of the  
21 phosphoamide bond that is formed between the nucleotides and nsp9. Neverthe-  
22 less, the low nucleotidylation efficiency and nonspecific transfer may be indicators of  
23 suboptimal reaction conditions, especially a lack of co-factors that may enhance nsp9  
24 activity or serve themselves as VPg. Given the numerous replicase subunits and long-  
25 lived cleavage intermediates without an assigned function in arteriviruses (33), it can  
26 only be speculated which subunits might fulfill such a function. Strikingly, all currently  
27 known co-factors of RNA-processing enzymes in nidoviruses, specifically coronavirus  
28 nsp7, 8, and 10 (15;16;34), derive from the region between the transmembrane domain  
29 downstream of the 3CL<sup>pro</sup> and the RFS. Could an arterivirus VPg also be derived from that  
30 region, which comprises nsp6 to nsp8? Interestingly, arterivirus nsp6, a conserved 11 to  
31 22 amino acid peptide, is known to be part of a number of uncharacterized cleavage  
32 intermediates that are subject to alternative processing pathways in EAV (33). Next to  
33 the fully cleaved nsp6, one of these nsp6-containing intermediates may be considered  
34 as initial nucleotidylation substrate, whose regulated cleavage may be a convenient way  
35 to reduce affinity for and thus prevent retention of the polymerase at the RNA 5' end  
36 once initiation has occurred. Finally, to invoke again the above mentioned similarity to  
37 astroviruses, it remains to be noted that the VPg of this virus family is indeed located  
38 between 3CL<sup>pro</sup> and the RFS (30) (Figure 1).

1 After auto-nucleotidylation of the RdRp-containing protein or its nucleotidylation of the  
2 VPg, the first nucleotide is extended by one or more additional nucleotides to generate  
3 a sufficient platform for annealing to the template strand. Thus, if the specificity of the  
4 nucleotidylation reaction *in vitro* faithfully reflects that *in vivo*, it must match the con-  
5 servation of the 5' ends of genome and/or antigenome. In the case of the EAV Bucyrus  
6 strain, which was the source of the nsp9 characterized in chapter 5, this sequence is GCU  
7 for the genome and GGU for the antigenome. Due to the unique replication mecha-  
8 nism of nidoviruses, all sg mRNAs and subgenome-size negative-stranded RNAs would  
9 contain identical 5' ends as the genome or antigenome, respectively (5;35). Although  
10 no evidence for an elongation of the first nucleotide was obtained in our experiments  
11 (chapter 5), it is noteworthy that for both RNA polarities nucleotidylation using GTP as  
12 substrate would be consistent with the observed nucleotide sequences. Intriguingly,  
13 bovine coronavirus was reported to encode a short poly uridine (poly(U)) tract at the an-  
14 tigenome's 5' end, which is thought to serve as template for poly(A) tailing of all mRNAs  
15 (36). The antigenome sequence has not been characterized for any arterivirus, but if it in-  
16 cludes poly(U) at the 5' end, it could explain the dual specificity of nsp9 nucleotidylation.  
17 In contrast, neither the presence of this poly(U) tract nor the sequence of the genomic  
18 5' end can be easily reconciled with *de novo* synthesis, which is assumed to generally  
19 require two purines as start nucleotides based on the biochemical characterization of a  
20 number of polymerases (9). To accommodate *de novo* synthesis, arteriviruses would thus  
21 require additional editing of one or both 5' ends. In contrast to large nidoviruses, encod-  
22 ing dedicated enzyme domains for this function may represent a significant burden to  
23 arteriviruses.

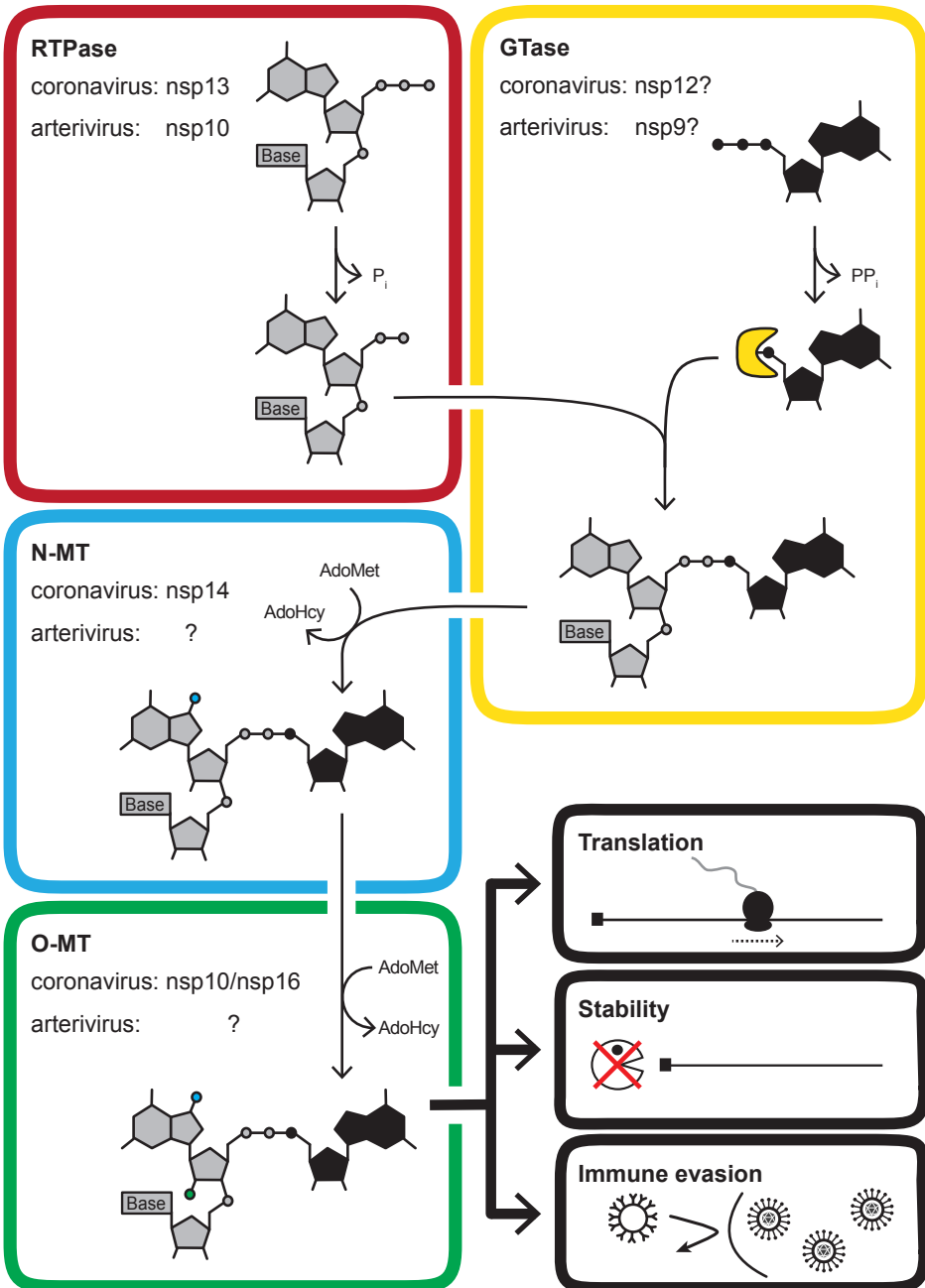
24  
25 Despite these arguments for protein-primed RNA synthesis in EAV, there is one compli-  
26 cation with this hypothesis since the chemical nature of the protein-nucleotide bond in  
27 the previously characterized VPgs does not match that between nsp9 and GMP/UMP. In  
28 all viruses with protein-primed replication that were investigated in detail the protein-  
29 nucleotide bond was established with the help of either a tyrosine or serine, in other  
30 words via a hydroxyl moiety (17;18;20;23). In EAV nsp9 the bond to the GMP/UMP is  
31 formed via the side chain amino group of a lysine or less likely histidine (chapter 5),  
32 and thus the situation would be more similar to that observed in nucleotidylating  
33 enzymes involved in nucleic acid ligation or mRNA capping (both discussed in detail in  
34 chapter 5) (37-39). This difference could have profound functional implications, because  
35 phosphoesters are known to be chemically more stable under physiological conditions  
36 than phosphoamides (40). However, arteriviruses may have evolved to have a labile  
37 bond between nsp9 and the nucleotide to allow the subsequent nucleotide transfer to  
38 a genuine VPg. This two-stage mechanism would provide an additional level of control  
39 and plasticity, with nsp9 being the designated carrier of the transferrable nucleoside

1 monophosphate. Furthermore, the instability and therefore transient nature of the bond  
2 would only impact EAV replication if the protein modification at the 5' end would serve  
3 functions beyond initiation of RNA synthesis that are required throughout the entire  
4 lifetime of the RNA, for example nuclease protection or translation initiation. The lat-  
5 ter function is fulfilled by calicivirus VPgs, which substitute for 5' mRNA nucleotide cap  
6 structures and bind directly to the translation initiation factor eIF4E (41;42). The fact that  
7 the putative bond between EAV RNA and a terminal protein might be unstable, may  
8 thus merely indicate that this virus utilizes a secondary mechanism to modify its mRNAs  
9 with regular cap structures after the transiently bound VPg has been removed.

### 10 11 **mRNA 5'-terminal modifications**

12  
13 Besides protein-priming, nucleotidylation could be implicated in the formation of  
14 the cap structure of mRNAs. Nidovirus mRNAs are thought to contain a type 1 cap  
15 structure (cap-1) ( $^m\text{GpppN}_m$ ) at their 5' end that enables translation in the absence of  
16 special RNA secondary structure elements as internal ribosome entry sites (IRESs) or 3'  
17 cap-independent translation enhancers (CITEs). However, some experimental evidence  
18 supporting that assumption has only been obtained for equine torovirus (43), the coro-  
19 navirus mouse hepatitis virus (MHV) (44;45), and the arterivirus simian hemorrhagic fe-  
20 ver virus (46). As the question regarding the 5' modification was not addressed for either  
21 roniviruses or mesoniviruses and a contradicting report exists for another arterivirus  
22 (47), it is far from proven that the presence of a cap-1 is a universal feature of the order.  
23 Additionally, there is considerable uncertainty about nidoviruses universally encoding a  
24 set of specific enzymes that were shown to be required to produce the cap-1 structure  
25 in better characterized (+) RNA viruses (Figure 2). While two of the enzymatic activities  
26 that are essential for the four-step synthesis of the cap-1 may reside in the nidovirus-  
27 wide conserved N-terminal domain of the RdRp subunit (chapter 5) and the helicase  
28 subunit (chapter 3), the two methyltransferase domains were so far only identified in  
29 coronaviruses, roniviruses, and mesoniviruses (6;48-52) but not in arteriviruses (7). Yet,  
30 methylation of the cap serves, on the one hand, translation initiation (N7-methylation)  
31 via the recruitment of eIF4E and, on the other hand, immune evasion (2'-O-methylation)  
32 (53-55). Therefore in theory, at least N7-methylation should be an essential step in the  
33 transcription of nidoviral mRNAs if it follows mechanisms established for other viruses. In  
34 chapter 6 we investigated whether the arterivirus-specific protein nsp12 might contain  
35 methyltransferase activity but were unsuccessful in verifying this hypothesis, potentially  
36 due to purely technical reasons. Thus, it remains an open question how arteriviruses  
37 would achieve the complete synthesis of a cap-1.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39



**Figure 2:** Conventional mRNA capping mechanism, nidovirus proteins (putatively) involved, and functional roles of the cap. RTPase, RNA 5'-triphosphatase; GTase, guanylyltransferase; N-MT, N7-methyltransferase; O-MT, 2'-O-methyltransferase; AdoMet, S-adenosylmethionine; AdoHcy, S-adenosylhomocysteine.

1 Besides nsp12 supplying the N7-methyltransferase activity in arteriviruses, several other  
2 options could be considered. First, the recruitment of cellular enzymes, which are lo-  
3 cated in the nucleus (56), seems to be a possibility potentially through the involvement  
4 of dedicated viral proteins that shuttle to the nucleus. Another option is snatching a cap  
5 structure from cellular mRNAs. For this purpose, several groups of negative-stranded  
6 (-) RNA viruses that employ this mechanism have evolved specific cap-binding and  
7 endoribonuclease domains (57-60). Also arteriviruses are known to encode an endori-  
8 bonuclease that is associated with nsp11 (61). However, upon analysis of genomes and  
9 mRNAs of established cap-snatching viruses of the *Orthomyxoviridae*, *Bunyaviridae*, and  
10 *Arenaviridae*, it became evident that all of these viruses harbor at least one (arenaviruses)  
11 and up to 17 (bunyaviruses) nucleotides at their mRNA 5' ends that are variable and not  
12 virus-encoded (62;63). Since such host-derived sequences have not been discovered in  
13 the extensive study of arterivirus 5' untranslated regions (UTRs) (47;64-66), the utiliza-  
14 tion of an analogous cap-snatching mechanism by arteriviruses seems very unlikely.

15  
16 Noteworthy, a variant of cap-snatching in which only the terminal <sup>m</sup>GMP moiety is  
17 removed from cellular substrates is employed by the dsRNA viruses of the *Totiviridae*  
18 (67). Still, if this mechanism, or a variation thereof, is considered for arteriviruses, other  
19 incompatibilities between its characteristics and our knowledge about arteriviruses be-  
20 come apparent. Particularly, cleavage within the peculiar 5'-5' linkage of the cap is usu-  
21 ally not catalyzed by ribonucleases that are able to cleave regular 5'-3' bonds. Although  
22 the physiological substrate for nsp11 has not been established yet, its demonstrated  
23 *in vitro* specificity for pyrimidine-containing single- and double-stranded RNAs makes  
24 it unlikely that this unusual bond would fall within the enzyme's substrate range (61).  
25 Finally, one of the arterivirus proteins may specifically recognize the 5' end of arterivirus  
26 mRNAs and facilitate translation initiation without the need for eIF4E involvement and  
27 hence potentially independent of N7-methylation.

28  
29 To conclude, the remaining significant gaps in our understanding of arterivirus RNA  
30 synthesis and 5' modification(s) currently leave space to formulate a number of parallel  
31 hypotheses. Given that arteriviruses are even considerably well characterized compared  
32 to all nidovirus families other than coronaviruses, a significantly bigger effort regarding  
33 biochemical and structural studies is required to establish universal and lineage-specific  
34 mechanisms in different families of the *Nidovirales*.

35  
36  
37  
38  
39

## 1 THE IMPORTANCE OF HOST HELICASES FOR NIDOVIRUS REPLICATION

2  
3 In chapters 2 and 3 potential functions of the conserved UPF1-like nidovirus helicase  
4 have been discussed. Reflecting the enzyme's versatile nature, those include roles during  
5 replication, transcription, RNA modification and processing, as well as virion biogenesis,  
6 which place this protein at the very center of the nidoviral replication cycle. Despite  
7 these prominent roles in nidoviruses and other (+) RNA viruses with a genome larger  
8 than 7 kilobases, helicases are not encoded by retroviruses or (-) RNA viruses, which are,  
9 for the most part, considerably larger than 7 kilobases (68). This may be explained by  
10 significant differences in their respective replication cycles compared to (+) RNA viruses,  
11 which, on the one hand, may have rendered those viruses helicase-independent or, on  
12 the other hand, may have enabled them to recruit cellular helicases. A case in point is  
13 the retrovirus human immunodeficiency virus, whose proteins or RNA elements interact  
14 with at least seven host RNA helicases – MOV10 (Upf1-like, SF1), DDX1, DDX3, DDX24 (all  
15 DEAD-box family, SF2), DHX9, DHX30 (both DEAH-box family, SF2), and RH116 (DExH-  
16 box family, SF2) – to promote, amongst others, transcription initiation, translation, and  
17 virion infectivity (reviewed in (69)). Furthermore, recently DDX21 was implicated in the  
18 temporal regulation of influenza A virus gene expression (70).

19  
20 Interestingly, utilization of host helicases for specific functions during their replication  
21 cycle is not exclusively a feature of viruses that do not encode their own helicase. For  
22 example, also the helicase-encoding bovine viral diarrhea virus (BVDV), hepatitis C virus  
23 (both *Flaviviridae*), and foot-and-mouth disease virus (FMDV, *Picornaviridae*) were found  
24 to depend on DHX9 for genome replication. While the exact role of this protein in BVDV  
25 infection remains unknown (71), the latter two viruses likely require this host factor for  
26 circularization of viral RNAs, as the protein was shown to bind to the 5' and 3' UTRs of the  
27 viral genome (72;73). Furthermore, in FMDV the same protein co-immunoprecipitated  
28 with the viral SF3 helicase 2C and nonstructural protein 3A (72). Together these results  
29 may indicate that cellular helicases are not only required to exert their enzymatic  
30 activities but may also serve as scaffolds for the assembly of multimeric protein-RNA  
31 complexes via their accessory domains.

32  
33 Although the number of RNA viruses with proven dependence on host cell helicases  
34 is currently small, it can be expected that more and more of these host factors will  
35 be identified due to the rising popularity of large-scale siRNA, yeast two-hybrid, and  
36 proteomics screens. For example, these approaches recently led to the discovery of an  
37 interaction between DDX1 and two coronavirus proteins, nsp14 and the nucleocapsid  
38 protein N (74-76). The results of two independent studies addressing these interactions  
39 are discussed in more detail below.

1 Moreover, in recent years a number of RNA helicases other than the well characterized  
2 RIG-I and MDA5 have been implicated in cellular antiviral defense mechanisms (77).  
3 Curiously, these again include DDX1 and Upf1. Thus, the recruitment of or structural  
4 similarity to these cellular helicases possibly serve a dual role in the nidovirus replica-  
5 tion cycle. This section is therefore concluded with a short summary on the defense  
6 mechanisms mediated by these two proteins.

## 7 8 **The cellular helicase DDX1 and nidovirus transcription regulation**

9  
10 As discussed in detail in chapter 3, the nidovirus helicase is one of the few proteins that  
11 has been directly implicated in the mechanism of discontinuous negative-strand syn-  
12 thesis that produces the subgenome-length templates for sg mRNA synthesis (78;79).  
13 Surprisingly, two studies now also linked the cellular helicase DDX1 to (sg) mRNA  
14 synthesis. Originally this host factor, which seems to be involved in 3' mRNA processing  
15 and tRNA splicing in the nucleus of uninfected cells (80;81), was identified in a large-  
16 scale yeast two-hybrid screen as an interaction partner of nsp14 of the coronaviruses  
17 infectious bronchitis virus (IBV) (76). Further directed investigation by the same group  
18 extended this interaction also to SARS-CoV nsp14 and mapped the interaction surface  
19 to the N-terminal exoribonuclease domain of this protein. In line with this finding, upon  
20 IBV infection of Vero cells the mostly nuclear localization of DDX1 was altered into a  
21 cytoplasmic punctuate pattern, similar to that observed for coronavirus replicase pro-  
22 teins. Given this apparent recruitment to replication-transcription complexes, it was not  
23 surprising that stable or transient knock-down of DDX1 led to a tenfold decrease of virus  
24 peak titers. Interestingly, when examining the levels of N and S protein expression – pro-  
25 duced from the shortest and longest sg mRNAs, respectively, in IBV – the amount of S  
26 protein appeared to be significantly reduced upon DDX1 knock-down while the amount  
27 of N protein was not affected. This finding correlated with the preferential decrease of  
28 the transcription level of longer mRNAs (subgenomic and genomic) compared to shorter  
29 ones, which was also observed. Since this effect had the same relative magnitude for  
30 RNAs of both polarities, it was speculated that DDX1 might be involved in the regulation  
31 of the relative abundance of individual negative-stranded RNAs (76). Notwithstanding  
32 the fact that this hypothesis was devised to explain the features of coronavirus repli-  
33 cation, this host factor may thus act on a different level or by a different mechanism  
34 than the endogenous nidoviral helicase, whose mutation may impair the synthesis of  
35 all sg mRNAs uniformly and selectively relative to genomic RNA production in EAV (79).  
36 Whether or not this regulatory mechanism actually involves the proven interaction with  
37 IBV nsp14 was not established. However, it was noted that continued passaging of IBV  
38 in DDX1 knock-down cells did neither induce mutations in nsp14 nor had any impact  
39 on viral fitness, which would be expected if nsp14 proofreading would be affected by

1 DDX1. Based on these results, it seems more likely that the nsp14-DDX1 interaction plays  
2 a different, as yet unidentified, role in the coronavirus replication cycle (76).

3  
4 A few years after this study, the same host helicase was identified as interaction partner  
5 of the IBV and MHV N proteins both in the absence and presence of cellular RNA (74;75).  
6 As for IBV earlier, also the MHV study demonstrated the selective reduction of longer  
7 RNA species upon DDX1 knock-down. Additionally, subsequent ectopic over-expression  
8 of knock-down resistant, functional DDX1 but not of a helicase active site mutant led to a  
9 reversal of this effect, establishing its dependence on helicase activity. In contrast to the  
10 study detailed above, it was furthermore demonstrated that abolishing the interaction  
11 between N and DDX1 by preventing phosphorylation of N at serine 197 had the same  
12 impact on RNA abundance as DDX1 knock-down. It was thus concluded that complex  
13 formation between these two proteins may promote read-through at transcription  
14 regulatory sequences during discontinuous negative-stranded RNA synthesis (75). How-  
15 ever, this would imply that abolishing complex formation should not only specifically  
16 diminish the quantity of longer RNA species but at the same time also increase that of  
17 shorter RNAs if no other limiting factor plays a role. Neither of the two studies reported  
18 such an outcome (75;76). Instead total RNA amounts were reduced while that of short  
19 sg mRNAs remained largely constant. The most obvious alternative explanation for this  
20 pattern would therefore be a direct stimulation of RdRp processivity by the N:DDX1  
21 complex. This, however, is unlikely to be true as the synthesis of genomic RNA, which  
22 is more than three times as long as the longest sg mRNA, was affected to the same  
23 extent as that of this latter sg mRNA. Next to a direct involvement in RNA synthesis, the  
24 reported data would also be consistent with a role for the N:DDX1 complex in a selective  
25 stabilization of certain negative-stranded RNAs before mRNA synthesis commences.  
26 In order to elucidate the exact role of DDX1, a deeper understanding of the nidovirus  
27 replication mechanism and potential downstream regulatory pathways influencing the  
28 stability of the negative-stranded subgenome-length RNAs would be required. Yet, this  
29 appears to become an ever more daunting task with every additional protein, viral or  
30 cellular, that is implicated in the nidovirus replication cycle.

### 31 **Cellular helicases and antiviral defense**

32  
33  
34 In the course of evolution eukaryotic cells have developed an intricate defense system  
35 to counteract infections by bacterial, eukaryotic, and viral pathogens. A central role in  
36 one arm of this system, the innate immune system, is played by conserved pattern-  
37 recognition receptors that recognize a certain signature molecule of defined groups of  
38 pathogens. In the case of RNA viruses this recognition largely depends on the sensing  
39 of viral nucleic acids inside the cell. The proteins responsible for detection are the endo-



1 somal Toll-like receptors 3, 7, and 8, as well as cytosolic NOD-like and RIG-I-like receptors.  
2 The latter group is comprised of the three DExH-box SF2 RNA helicases RIG-I, MDA5, and  
3 LGP2, of which MDA5 was shown to sense MHV RNA (82). Once these receptors bind  
4 their ligand, a complex cascade of downstream effectors is activated, eventually leading  
5 to the transcription of genes involved in inflammatory response and cross-talk with the  
6 adaptive immune system. Most notably is the production of type I interferons, which in  
7 turn indirectly regulate protein synthesis, cell growth, and apoptosis (reviewed in (83)).

8  
9 As would be expected, during an extensive period of co-evolution with their hosts, vi-  
10 ruses have developed a variety of means to avoid, inhibit, or redirect essential factors of  
11 the innate immune system. For instance, nidoviruses, as all other characterized (+) RNA  
12 viruses infecting eukaryotes, are known to induce extensive membrane modifications  
13 inside the host cell (84). Since these are thought to be the site of viral RNA synthesis,  
14 they may serve to hide viral nucleic acids, in particular the highly immunogenic double-  
15 stranded replication intermediates, from cytosolic sensors. Additional avoidance  
16 strategies that could be employed by at least a subset of nidoviruses are the disguising  
17 of viral RNAs by attaching the typical eukaryotic double-methylated cap structure to  
18 mRNA 5' ends or the degradation of an excess of viral RNAs by either of the two viral  
19 ribonucleases (85). Furthermore, the nsp1 $\alpha$ , nsp1 $\beta$ , and nsp4 proteases of the arterivirus  
20 porcine reproductive and respiratory syndrome virus (PRRSV), as well as the PLP2 prote-  
21 ase domains of EAV and PRRSV nsp2 have been implicated in the inhibition of immune  
22 signaling although interestingly not in all cases through their proteolytic activity (86).  
23 Also coronavirus nsp1 and nsp3 appear to be engaged in immune suppression (87-89).  
24 It remains to be seen whether similar strategies have also evolved in the other nidovirus  
25 (sub)families. Yet, these examples show that a significant number of proteins may be  
26 dedicated to counteract the host's defense mechanisms and to shape a more beneficial  
27 environment for virus replication. Thus, given the extensive divergence between indi-  
28 vidual members of the order, it is currently difficult to estimate how large the repertoire  
29 of nidovirus evasion strategies really is.

### 30 *DEAD/H-box helicases and RNA detection*

31  
32  
33 Interestingly, a number of SF2 RNA helicases besides RIG-1 and MDA-5 were identified  
34 as additional players in virus sensing and immune signaling in recent years (77). One  
35 of those is the above mentioned DDX1, which was shown to bind to poly(I:C) RNA and  
36 may recognize any RNA species. Binding of DDX1 to a substrate promotes the complex  
37 formation with two other helicases, DDX21 and DHX36. Both of these helicases can sub-  
38 sequently interact with the innate immunity signaling protein TRIF and thereby induce  
39 an inflammatory response. The importance of this pathway was demonstrated in studies

1 on influenza virus and reovirus, in which interferon production was reduced once any  
2 of the three helicases was knocked-down (90). Additionally, DDX1 is able to directly  
3 bind to the RelA subunit of the pro-inflammatory transcription factor NF- $\kappa$ B, thereby  
4 stimulating transcription activation by this factor (91). Intriguingly, next to DDX1, also  
5 DDX21 and DHX36 were identified as putative interaction partners of the IBV N protein  
6 by co-immunoprecipitation (74). The latter interaction was also found for PRRSV (92).

7  
8 Further RNA helicases that are involved in immunity and that were identified as binding  
9 partners of N in IBV and PRRSV are DDX3 and DHX9 (74;77;92). The former, which also  
10 serves in translation regulation especially of mRNAs with complex 5' UTRs (93), acts as  
11 a sensor for dsRNA. However, unlike signaling by the DDX1 complex, the pathway for  
12 this helicase is identical to that of RIG-I and involves the downstream effector MAVS  
13 (94). Similarly, also the transcriptional regulator DHX9 was shown to interact with MAVS  
14 upon encountering dsRNA (95). It was therefore speculated that both DDX3 and DHX9  
15 may be of particular importance early in infection when the RIG-I concentration is still  
16 low. Whether any of these cellular helicases actually plays a role in nidovirus sensing,  
17 immune evasion, or replication remains to be seen. Nevertheless, it is an interesting  
18 possibility that the interaction of these host proteins with N might interfere with their  
19 immune signaling responsibilities.

#### 20 21 *Upf1 and NMD-mediated defense*

22  
23 In eukaryotic cells the relative abundance of mRNAs is heavily regulated. One of the  
24 involved mechanisms, which controls the quantities of up to 10% of all transcribed  
25 mRNAs, is termed nonsense-mediated decay (NMD). Besides this function, the same  
26 mechanism also controls ribosome release from, as well as translation repression  
27 and – ultimately – decay of aberrant transcripts with, for example, premature stop  
28 codons that may arise due to wrong or incomplete splicing and nonsense or frameshift  
29 mutations. Although NMD has been studied extensively in different species, neither  
30 its RNA or protein triggers nor the exact sequence of events involved in this process  
31 are well understood. Notwithstanding this uncertainty, the SF1 helicase Upf1 and the  
32 poly(A) binding protein (PABP) – or more precisely the competition between these  
33 proteins – appear to be of special importance. During translation of wild-type host  
34 mRNAs, PABP is bound sufficiently close to the terminating ribosome to establish an  
35 interaction with the termination factor eRF3, which in turn stimulates termination  
36 and triggers ribosomal release. Conversely, if the distance between eRF3 and PABP is  
37 artificially elongated by, for instance, the introduction of an upstream stop codon or  
38 the presence of a second downstream ORF, termination becomes less efficient. In this  
39 situation Upf1 is able to compete with PABP for eRF3, triggering the formation of a

1 larger protein complex, which marks this mRNA for decay (reviewed in (96;97)). While  
2 cellular mRNAs have evolved to contain the correct spacing – on average 700 to 800  
3 nucleotides in humans – between stop codon and poly(A) tail (98), some viral RNAs  
4 comprising multiple ORFs and elongated 3' UTRs may be particularly vulnerable to this  
5 quality control mechanism. This assumption was recently confirmed with the help of the  
6 (+) RNA viruses potato virus X (*Alphaflexiviridae*) and Semliki Forest virus (SFV) (*Togaviri-*  
7 *dae*) (99;100), which both utilize 3' co-terminal sg mRNAs. For the plant virus a mutation  
8 within the Upf1-gene was shown to lead to an increase in the amount of sg mRNAs with  
9 long 3' UTRs compared to the wild-type situation. At the same time, the abundance of  
10 the shortest sg mRNA comprising only a very short 3' UTR was unaffected (99). Similarly,  
11 knock-down of Upf1 extended the half-life of the SFV genomic RNA in HeLa cells from  
12 63 min to 89 min. Consequently, viral titers increased by threefold. This effect could be  
13 reversed by ectopic expression of a knock-down resistant functional Upf1 but not an  
14 helicase active site mutant. Surprisingly, shortening of the ~4000 nucleotide 3' UTR of  
15 the SFV genomic RNA to 62 nucleotides did not abolish Upf1-mediated decay, a find-  
16 ing which stands in sharp contrast to the observations for the plant virus (100). Even  
17 more puzzling is the proven independence of the antiviral mechanism from 5' cap but  
18 especially 3' poly(A) tail (99). These observations essentially argue for a non-canonical  
19 NMD mechanism being involved during antiviral defense. Since two (-) RNA viruses,  
20 respiratory syncytial virus (*Paramyxoviridae*) and Uukuniemi virus (*Bunyaviridae*) that  
21 strictly encode monocistronic mRNAs, were not affected by Upf1 knock-down (100), the  
22 polycistronic nature of mRNAs appears to be the common denominator of this pathway  
23 at the moment. Future research may reveal whether different RNA features emerge as  
24 triggering factors.

25

26 As already speculated for cellular NMD, specific RNA sequences or secondary structures  
27 might have evolved to recruit NMD antagonists or inhibit NMD in other ways (96). Given  
28 this context, if this defense mechanism actually plays a role for at least a subset of viruses,  
29 it would be expected that these viruses employ certain mechanisms to counteract Upf1-  
30 mediated decay (101). In view of this assumption, it is intriguing that all nidoviruses  
31 encode a helicase that structurally resembles Upf1 (chapters 2 and 3). Although a direct  
32 role of this viral protein in RNA quality control might be a preferable explanation for its  
33 fixation in the ancestral nidovirus genome, one could envision a secondary function  
34 of the nidovirus helicase in counteracting cellular defenses. However, due to its pre-  
35 sumed multiple roles during viral replication, confirming this immune evasion strategy  
36 will not be an easy task. In addition, Upf1 is known to be a central player in a number  
37 of pathways that involve the manipulation of nucleic acids and each employ different  
38 protein complexes (96;97). Further research in this area should therefore initially focus  
39 on establishing and rigorously validating host interactions with the nidovirus helicase.

## 1 REFERENCE LIST

- 2 1. Lauber C, Ziebuhr J, Junglen S, *et al.* *Mesoniviridae*: a proposed new family in the order *Nidovirales*  
3 formed by a single species of mosquito-borne viruses. *Arch.Virol.* 2012;157(8):1623-1628.  
4
- 5 2. Cowley JA, Walker PJ, Flegel TW, *et al.* Family *Roniviridae*. In King AMQ, Adams MJ, Carstens EB *et al.*  
6 *et al.* editors, *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*,  
7 Amsterdam, Elsevier Academic Press, 2012;829-834.
- 8 3. de Groot RJ, Baker SC, Baric R, *et al.* Family *Coronaviridae*. In King AMQ, Adams MJ, Carstens EB *et al.*  
9 *et al.* editors, *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*,  
10 Amsterdam, Elsevier Academic Press, 2012;806-828.
- 11 4. Faaberg KS, Balasuriya UB, Brinton MA, *et al.* Family *Arteriviridae*. In King AMQ, Adams MJ, Carstens  
12 EB *et al.* editors, *Virus Taxonomy. Ninth report of the international committee on taxonomy of*  
13 *viruses*, Amsterdam, Elsevier Academic Press, 2012;796-805.
- 14 5. Gorbalenya AE, Enjuanes L, Ziebuhr J, *et al.* *Nidovirales*: evolving the largest RNA virus genome.  
15 *Virus Res.* 2006;117(1):17-37.  
16
- 17 6. Nga PT, Parquet MC, Lauber C, *et al.* Discovery of the first insect nidovirus, a missing evolutionary  
18 link in the emergence of the largest RNA virus genomes. *PLoS.Pathog.* 2011;7(9):e1002215.
- 19 7. Lauber C, Goeman JJ, Parquet MC, *et al.* The footprint of genome architecture in the largest  
20 genome expansion in RNA viruses. *PLoS.Pathog.* 2013;9(7):e1003500.  
21
- 22 8. Iyer LM, Koonin EV, Aravind L. Evolutionary connection between the catalytic subunits of DNA-  
23 dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of  
24 RNA polymerases. *BMC.Struct.Biol.* 2003;3:1.
- 25 9. Ng KK, Arnold JJ, Cameron CE. Structure-function relationships among RNA-dependent RNA  
26 polymerases. *Curr.Top.Microbiol.Immunol.* 2008;320:137-156.  
27
- 28 10. Gorbalenya AE, Pringle FM, Zeddarn JL, *et al.* The palm subdomain-based active site is internally  
29 permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J.Mol.Biol.* 2002;324(1):  
30 47-62.
- 31 11. Xu Z, Zhang H, Tian X, *et al.* The R protein of SARS-CoV: analyses of structure and function based  
32 on four complete genome sequences of isolates BJ01-BJ04. *Genomics Proteomics.Bioinformatics.*  
33 2003;1(2):155-165.
- 34 12. Beerens N, Selisko B, Ricagno S, *et al.* De novo initiation of RNA synthesis by the arterivirus RNA-  
35 dependent RNA polymerase. *J.Virol.* 2007;81(16):8384-8395.  
36
- 37 13. te Velhuis AJ, Arnold JJ, Cameron CE, *et al.* The RNA polymerase activity of SARS-coronavirus  
38 nsp12 is primer dependent. *Nucleic Acids Res.* 2010;38(1):203-214.  
39

- 1 14. Imbert I, Guillemot JC, Bourhis JM, *et al.* A second, non-canonical RNA-dependent RNA poly-  
2 merase in SARS coronavirus. *EMBO J.* 2006;25(20):4933-4942.
- 3 15. Subissi L, Posthuma CC, Collet A, *et al.* One severe acute respiratory syndrome coronavirus pro-  
4 tein complex integrates processive RNA polymerase and exonuclease activities. *Proc.Natl.Acad.*  
5 *Sci.U.S.A* 2014; 111(37):e3900
- 6 16. teVelthuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+nsp8complex is a unique  
7 multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic*  
8 *Acids Res.* 2012;40(4):1737-1747.
- 9 17. Jiang L, Laliberte JF. The genome-linked protein VPg of plant viruses-a protein with many part-  
10 ners. *Curr.Opin.Virol.* 2011;1(5):347-354.
- 11 18. Goodfellow I. The genome-linked protein VPg of vertebrate viruses - a multifaceted protein. *Curr.*  
12 *Opin.Virol.* 2011;1(5):355-362.
- 13 19. Wang GH, Seeger C. The reverse transcriptase of hepatitis B virus acts as a protein primer for viral  
14 DNA synthesis. *Cell* 1992;71(4):663-670.
- 15 20. Pan J, Lin L, Tao YJ. Self-guanlylation of birnavirus VP1 does not require an intact polymerase  
16 activity site. *Virology* 2009;395(1):87-96.
- 17 21. Paul AV, Yin J, Mugavero J, *et al.* A "slide-back" mechanism for the initiation of protein-primed RNA  
18 synthesis by the RNA polymerase of poliovirus. *J.Biol.Chem.* 2003;278(45):43951-43960.
- 19 22. Paul AV, Peters J, Mugavero J, *et al.* Biochemical and genetic studies of the VPg uridylylation reac-  
20 tion catalyzed by the RNA polymerase of poliovirus. *J.Virol.* 2003;77(2):891-904.
- 21 23. Richards OC, Spagnolo JF, Lyle JM, *et al.* Intramolecular and intermolecular uridylylation by polio-  
22 virus RNA-dependent RNA polymerase. *J.Virol.* 2006;80(15):7405-7415.
- 23 24. Digiaro M, Nahdi S, Elbeaino T. Complete sequence of RNA1 of grapevine Anatolian ringspot  
24 virus. *Arch.Virol.* 2012;157(10):2013-2016.
- 25 25. Weitz M, Baroudy BM, Maloy WL, *et al.* Detection of a genome-linked protein (VPg) of hepatitis A  
26 virus and its comparison with other picornaviral VPgs. *J.Virol.* 1986;60(1):124-130.
- 27 26. Rohayem J, Robel I, Jager K, *et al.* Protein-primed and de novo initiation of RNA synthesis by  
28 norovirus 3Dpol. *J.Virol.* 2006;80(14):7060-7069.
- 29 27. Grzela R, Szolajska E, Ebel C, *et al.* Virulence factor of potato virus Y, genome-attached terminal  
30 protein VPg, is a highly disordered protein. *J.Biol.Chem.* 2008;283(1):213-221.
- 31 28. Rantalainen KI, Uversky VN, Permi P, *et al.* Potato virus A genome-linked protein VPg is an in-  
32 trinsically disordered molten globule-like protein with a hydrophobic core. *Virology* 2008;377(2):  
33 280-288.
- 34  
35  
36  
37  
38  
39

- 1 29. Zeddiam JL, Gordon KH, Lauber C, *et al.* Euprosterona elaeasa virus genome sequence and evolu-  
2 tion of the Tetraviridae family: emergence of bipartite genomes and conservation of the VPg  
3 signal with the dsRNA Birnaviridae family. *Virology* 2010;397(1):145-154.
- 4 30. Fuentes C, Bosch A, Pinto RM, *et al.* Identification of human astrovirus genome-linked protein  
5 (VPg) essential for virus infectivity. *J.Virol.* 2012;86(18):10070-10078.
- 6 31. Graham SC, Sarin LP, Bahar MW, *et al.* The N-terminus of the RNA polymerase from infectious  
7 pancreatic necrosis virus is the determinant of genome attachment. *PLoS.Pathog.* 2011;7(6):  
8 e1002085.
- 9 32. Dobos P. Protein-primed RNA synthesis *in vitro* by the virion-associated RNA polymerase of infec-  
10 tious pancreatic necrosis virus. *Virology* 1995;208(1):19-25.
- 11 33. van Aken D, Snijder EJ, Gorbalenya AE. Mutagenesis analysis of the nsp4 main proteinase reveals  
12 determinants of arterivirus replicase polyprotein autoprocessing. *J.Virol.* 2006;80(7):3428-3437.
- 13 34. Bouvet M, Lugari A, Posthuma CC, *et al.* Coronavirus Nsp10, a critical co-factor for activation of  
14 multiple replicative enzymes. *J.Biol.Chem.* 2014;289(37):25783-25796.
- 15 35. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J.Gen.Virol.*  
16 2006;87(Pt 6):1403-1421.
- 17 36. Hofmann MA, Brian DA. The 5' end of coronavirus minus-strand RNAs contains a short poly(U)  
18 tract. *J.Virol.* 1991;65(11):6331-6333.
- 19 37. Shuman S, Schwer B. RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl  
20 transferases. *Mol.Microbiol.* 1995;17(3):405-410.
- 21 38. Nandakumar J, Shuman S, Lima CD. RNA ligase structures reveal the basis for RNA specificity and  
22 conformational changes that drive ligation forward. *Cell* 2006;127(1):71-84.
- 23 39. Ghosh A, Lima CD. Enzymology of RNA cap synthesis. *Wiley.Interdiscip.Rev.RNA.* 2010;1(1):152-  
24 172.
- 25 40. Duclos B, Marcandier S, Cozzone AJ. Chemical properties and separation of phosphoamino acids  
26 by thin-layer chromatography and/or electrophoresis. *Methods Enzymol.* 1991;201:10-21.
- 27 41. Goodfellow I, Chaudhry Y, Gioldasi I, *et al.* Calicivirus translation initiation requires an interaction  
28 between VPg and eIF 4 E. *EMBO Rep.* 2005;6(10):968-972.
- 29 42. Hosmillo M, Chaudhry Y, Kim DS, *et al.* Sapovirus Translation Requires an Interaction between VPg  
30 and the Cap Binding Protein eIF4E. *J.Virol.* 2014;88(21):12213-12221.
- 31 43. van Vliet AL, Smits SL, Rottier PJ, *et al.* Discontinuous and non-discontinuous subgenomic RNA  
32 transcription in a nidovirus. *EMBO J.* 2002;21(23):6571-6580.
- 33  
34  
35  
36  
37  
38  
39

- 1 44. Lai MM, Stohlman SA. Comparative analysis of RNA genomes of mouse hepatitis viruses. *J.Virol.* 1981;38(2):661-670.
- 2
- 3 45. Lai MM, Patton CD, Stohlman SA. Further characterization of mRNA's of mouse hepatitis virus: 4 presence of common 5'-end nucleotides. *J.Virol.* 1982;41(2):557-565.
- 5
- 6 46. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion 7 RNA. *Virology* 1986;151(1):146-150.
- 8
- 9 47. Chen Z, Faaberg KS, Plagemann PG. Determination of the 5' end of the lactate dehydrogenase- 10 elevating virus genome by two independent approaches. *J.Gen.Virol.* 1994;75 ( Pt 4):925-930.
- 11
- 12 48. Decroly E, Imbert I, Coutard B, *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding 13 enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J.Virol.* 2008;82(16):8071-8084.
- 14
- 15 49. Decroly E, Debarnot C, Ferron F, *et al.* Crystal structure and functional analysis of the SARS- 16 coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS.Pathog.* 2011;7(5): 17 e1002059.
- 18
- 19 50. Chen Y, Cai H, Pan J, *et al.* Functional screen reveals SARS coronavirus nonstructural protein nsp14 20 as a novel cap N7 methyltransferase. *Proc.Natl.Acad.Sci.U.S.A* 2009;106(9):3484-3489.
- 21
- 22 51. Chen Y, Su C, Ke M, *et al.* Biochemical and structural insights into the mechanisms of SARS coro- 23 navirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS.Pathog.* 2011;7(10): 24 e1002294.
- 25
- 26 52. Bouvet M, Debarnot C, Imbert I, *et al.* *In vitro* reconstitution of SARS-coronavirus mRNA cap 27 methylation. *PLoS.Pathog.* 2010;6(4):e1000863.
- 28
- 29 53. Zust R, Cervantes-Barragan L, Habjan M, *et al.* Ribose 2'-O-methylation provides a molecular 30 signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat.* 31 *Immunol.* 2011;12(2):137-143.
- 32
- 33 54. Rhoads RE. eIF4E: new family members, new binding partners, new roles. *J.Biol.Chem.* 2009; 34 284(25):16711-16715.
- 35
- 36 55. Daffis S, Szretter KJ, Schriewer J, *et al.* 2'-O methylation of the viral mRNA cap evades host restric- 37 tion by IFIT family members. *Nature* 2010;468(7322):452-456.
- 38
- 39 56. Neugebauer KM. On the importance of being co-transcriptional. *J.Cell Sci.* 2002;115(Pt 20):3865- 3871.
57. Morin B, Coutard B, Lelke M, *et al.* The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLoS.Pathog.* 2010;6(9):e1001038.
58. Reich S, Guilligay D, Pflug A, *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 2014; 516(7531):361-366

- 1 59. Guilligay D, Kadlec J, Crepin T, *et al.* Comparative structural and functional analysis of orthomyxo-  
2 virus polymerase cap-snatching domains. *PLoS.One.* 2014;9(1):e84973.
- 3 60. Hopkins K, Cherry S. Bunyaviral cap-snatching vs. decapping: recycling cell cycle mRNAs. *Cell*  
4 *Cycle* 2013;12(24):3711-3712.
- 5 61. Nedialkova DD, Ulferts R, van den Born E, *et al.* Biochemical characterization of arterivirus non-  
6 structural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease.  
7 *J.Virol.* 2009;83(11):5671-5682.
- 8 62. Mir MA, Duran WA, Hjelle BL, *et al.* Storage of cellular 5' mRNA caps in P bodies for viral cap-  
9 snatching. *Proc.Natl.Acad.Sci.U.S.A* 2008;105(49):19294-19299.
- 10 63. Raju R, Raju L, Hacker D, *et al.* Nontemplated bases at the 5' ends of Tacaribe virus mRNAs. *Virology*  
11 1990;174(1):53-59.
- 12 64. van den Born E, Gulytaev AP, Snijder EJ. Secondary structure and function of the 5'-proximal  
13 region of the equine arteritis virus RNA genome. *RNA.* 2004;10(3):424-437.
- 14 65. Oleksiewicz MB, Botner A, Nielsen J, *et al.* Determination of 5'-leader sequences from radically  
15 disparate strains of porcine reproductive and respiratory syndrome virus reveals the presence of  
16 highly conserved sequence motifs. *Arch.Virol.* 1999;144(5):981-987.
- 17 66. Kheyar A, St-Laurent G, Archambault D. Sequence determination of the extreme 5' end of equine  
18 arteritis virus leader region. *Virus Genes* 1996;12(3):291-295.
- 19 67. Fujimura T, Esteban R. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. *Proc.*  
20 *Natl.Acad.Sci.U.S.A* 2011;108(43):17667-17671.
- 21 68. Lauber C. On the evolution of genetic diversity in RNA virus species - Uncovering barriers to  
22 genetic divergence and gene length in picorna- and nidoviruses. 2012. [dissertation]
- 23 69. Ranji A, Boris-Lawrie K. RNA helicases: emerging roles in viral replication and the host innate  
24 response. *RNA.Biol.* 2010;7(6):775-787.
- 25 70. Chen G, Liu CH, Zhou L, *et al.* Cellular DDX21 RNA helicase inhibits influenza A virus replication  
26 but is counteracted by the viral NS1 protein. *Cell Host.Microbe* 2014;15(4):484-493.
- 27 71. Jefferson M, Donaszi-Ivanov A, Pollen S, *et al.* Host factors that interact with the pestivirus N-  
28 terminal protease, Npro, are components of the ribonucleoprotein complex. *J.Virol.* 2014;88(18):  
29 10340-10353.
- 30 72. Lawrence P, Rieder E. Identification of RNA helicase A as a new host factor in the replication cycle  
31 of foot-and-mouth disease virus. *J.Virol.* 2009;83(21):11356-11366.
- 32 73. Isken O, Baroth M, Grassmann CW, *et al.* Nuclear factors are involved in hepatitis C virus RNA  
33 replication. *RNA.* 2007;13(10):1675-1692.
- 34  
35  
36  
37  
38  
39



- 1 74. Emmott E, Munday D, Bickerton E, *et al.* The cellular interactome of the coronavirus infectious  
2 bronchitis virus nucleocapsid protein and functional implications for virus biology. *J.Virol.* 2013;  
3 87(17):9486-9500.
- 4 75. Wu CH, Chen PJ, Yeh SH. Nucleocapsid Phosphorylation and RNA Helicase DDX1 Recruitment En-  
5 ables Coronavirus Transition from Discontinuous to Continuous Transcription. *Cell Host.Microbe*  
6 2014;16(4):462-472.
- 7 76. Xu L, Khadijah S, Fang S, *et al.* The cellular RNA helicase DDX1 interacts with coronavirus nonstruc-  
8 tural protein 14 and enhances viral replication. *J.Virol.* 2010;84(17):8571-8583.
- 9 77. Fullam A, Schroder M. DExD/H-box RNA helicases as mediators of anti-viral innate immunity and  
10 essential host factors for viral replication. *Biochim.Biophys.Acta* 2013;1829(8):854-865.
- 11 78. van Dinten LC, van Tol H, Gorbalenya AE, *et al.* The predicted metal-binding region of the arteri-  
12 virus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion  
13 biogenesis. *J.Virol.* 2000;74(11):5213-5223.
- 14 79. van Marle G, van Dinten LC, Spaan WJ, *et al.* Characterization of an equine arteritis virus replicase  
15 mutant defective in subgenomic mRNA synthesis. *J.Virol.* 1999;73(7):5274-5281.
- 16 80. Chen HC, Lin WC, Tsay YG, *et al.* An RNA helicase, DDX1, interacting with poly(A) RNA and hetero-  
17 geneous nuclear ribonucleoprotein K. *J.Biol.Chem.* 2002;277(43):40403-40409.
- 18 81. Popow J, Jurkin J, Schleiffer A, *et al.* Analysis of orthologous groups reveals archease and DDX1 as  
19 tRNA splicing factors. *Nature* 2014;511(7507):104-107.
- 20 82. Roth-Cross JK, Bender SJ, Weiss SR. Murine coronavirus mouse hepatitis virus is recognized by  
21 MDA5 and induces type I interferon in brain macrophages/microglia. *J.Virol.* 2008;82(20):9829-  
22 9838.
- 23 83. Jensen S, Thomsen AR. Sensing of RNA viruses: a review of innate immune receptors involved in  
24 recognizing RNA virus invasion. *J.Virol.* 2012;86(6):2900-2910.
- 25 84. Den Boon JA, Ahlquist P. Organelle-like membrane compartmentalization of positive-strand RNA  
26 virus replication factories. *Annu.Rev.Microbiol.* 2010;64:241-256.
- 27 85. Kindler E, Thiel V. To sense or not to sense viral RNA—essentials of coronavirus innate immune  
28 evasion. *Curr.Opin.Microbiol.* 2014;20:69-75.
- 29 86. van Kasteren P. Arterivirus PLP2 - an OTU deubiquitinase that counteracts innate immunity. 2014.  
30 [dissertation]
- 31 87. Jauregui AR, Savalia D, Lowry VK, *et al.* Identification of residues of SARS-CoV nsp1 that differen-  
32 tially affect inhibition of gene expression and antiviral signaling. *PLoS.One.* 2013;8(4):e62416.
- 33
- 34
- 35
- 36
- 37
- 38
- 39

- 1 88. Narayanan K, Huang C, Lokugamage K, *et al.* Severe acute respiratory syndrome coronavirus nsp1  
2 suppresses host gene expression, including that of type I interferon, in infected cells. *J.Virol.* 2008;  
3 82(9):4471-4479.
- 4 89. Wang G, Chen G, Zheng D, *et al.* PLP2 of mouse hepatitis virus A59 (MHV-A59) targets TBK1 to  
5 negatively regulate cellular type I interferon signaling pathway. *PLoS.One.* 2011;6(2):e17192.
- 6 90. Zhang Z, Kim T, Bao M, *et al.* DDX1, DDX21, and DHX36 helicases form a complex with the adaptor  
7 molecule TRIF to sense dsRNA in dendritic cells. *Immunity.* 2011;34(6):866-878.
- 8 91. Ishaq M, Ma L, Wu X, *et al.* The DEAD-box RNA helicase DDX1 interacts with RelA and enhances  
9 nuclear factor kappaB-mediated transcription. *J.Cell Biochem.* 2009;106(2):296-305.
- 10 92. Jourdan SS, Osorio F, Hiscox JA. An interactome map of the nucleocapsid protein from a highly  
11 pathogenic North American porcine reproductive and respiratory syndrome virus strain gener-  
12 ated using SILAC-based quantitative proteomics. *Proteomics.* 2012;12(7):1015-1023.
- 13 93. Soto-Rifo R, Rubilar PS, Limousin T, *et al.* DEAD-box protein DDX3 associates with eIF4F to pro-  
14 mote translation of selected mRNAs. *EMBO J.* 2012;31(18):3745-3756.
- 15 94. Oshiumi H, Sakai K, Matsumoto M, *et al.* DEAD/H BOX 3 (DDX3) helicase binds the RIG-I adaptor  
16 IPS-1 to up-regulate IFN-beta-inducing potential. *Eur.J.Immunol.* 2010;40(4):940-948.
- 17 95. Zhang Z, Yuan B, Lu N, *et al.* DHX9 pairs with IPS-1 to sense double-stranded RNA in myeloid  
18 dendritic cells. *J.Immunol.* 2011;187(9):4501-4508.
- 19 96. Nicholson P, Yepiskoposyan H, Metze S, *et al.* Nonsense-mediated mRNA decay in human cells:  
20 mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell*  
21 *Mol.Life Sci.* 2010;67(5):677-700.
- 22 97. Kervestin S, Jacobson A. NMD: a multifaceted response to premature translational termination.  
23 *Nat.Rev.Mol.Cell Biol.* 2012;13(11):700-712.
- 24 98. Eberle AB, Stalder L, Mathys H, *et al.* Posttranscriptional gene regulation by spatial rearrangement  
25 of the 3' untranslated region. *PLoS.Biol.* 2008;6(4):e92.
- 26 99. Garcia D, Garcia S, Voinnet O. Nonsense-mediated decay serves as a general viral restriction  
27 mechanism in plants. *Cell Host.Microbe* 2014;16(3):391-402.
- 28 100. Balistreri G, Horvath P, Schweingruber C, *et al.* The host nonsense-mediated mRNA decay pathway  
29 restricts Mammalian RNA virus replication. *Cell Host.Microbe* 2014;16(3):403-411.
- 30 101. Wachter A, Hartmann L. NMD: nonsense-mediated defense. *Cell Host.Microbe* 2014;16(3):273-  
31 275.
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39



## 1 LIST OF ABBREVIATIONS

2		
3		
4	(-)	negative-stranded
5	(+)	positive-stranded, i.e., of mRNA polarity
6	3' CITE	3' cap-independent translation enhancer
7	AsD	Arterivirus-specific domain
8	ATP	adenosine triphosphate
9	BCoV	bovine coronavirus ( <i>Coronaviridae</i> )
10	BHK	baby hamster kidney
11	bp	base pair
12	BSA	bovine serum albumin
13	BVDV	bovine viral diarrhea virus ( <i>Flaviviridae</i> )
14	cap-0	type-0 cap structure, <sup>m</sup> GpppN
15	cap-1	type-1 cap structure, <sup>m</sup> GpppN <sub>m</sub>
16	CDC	United States Centers for Disease Control and Prevention
17	cDNA	complementary DNA
18	CHES	N-cyclohexyl-2-aminoethanesulfonic acid
19	CHIKV	chikungunya virus ( <i>Alphaviridae</i> )
20	CM	convoluted membrane
21	CPE	cytopathic effect
22	cpm	counts per minute
23	CTP	cytidine triphosphate
24	DMV	double-membrane vesicle
25	DNA	deoxyribonucleic acid
26	ds	double-stranded
27	DTT	dithiothreitol
28	E	nidovirus envelop protein
29	<i>E. coli</i>	<i>Escherichia coli</i>
30	EAV	equine arteritis virus ( <i>Arteriviridae</i> )
31	ECDC	European Center for Disease Prevention and Control
32	EDTA	ethylenediaminetetraacetic acid
33	ER	endoplasmic reticulum
34	ExoN	nidovirus 3'-5' exoribonuclease domain
35	FDA	United States Food and Drug Administration
36	FECoV	feline enteric coronavirus ( <i>Coronaviridae</i> )
37	FMDV	foot-and-mouth disease virus ( <i>Picornaviridae</i> )
38	FSBG	5'-(4-fluorosulfonylbenzoyl)guanosine
39	GP	glycoprotein

1	GST	glutathione S-transferase
2	GTase	guanylyltransferase
3	GTP	guanosine triphosphate
4	HCoV	human coronavirus ( <i>Coronaviridae</i> )
5	HCoV-229E	human coronavirus 229E ( <i>Coronaviridae</i> )
6	HCV	hepatitis C virus ( <i>Flaviviridae</i> )
7	HEL1	nidovirus helicase domain
8	HEPES	hydroxyethyl piperazineethanesulfonic acid
9	HEV	hepatitis E virus ( <i>Hepeviridae</i> )
10	HMM	Hidden Markov Model
11	IBV	infectious bronchitis virus ( <i>Coronaviridae</i> )
12	IC <sub>50</sub>	inhibitory concentration
13	ICTV	International Committee on Taxonomy of Viruses
14	IFA	immunofluorescence assay
15	IND	investigational new drug
16	IRES	internal ribosome entry site
17	kb	kilobase
18	k <sub>cat</sub>	catalytic constant
19	kDa	kilodalton
20	K <sub>M</sub>	Michaelis constant
21	LDV	lactate dehydrogenase-elevating virus ( <i>Arteriviridae</i> )
22	M	nidovirus membrane protein
23	MAD	multiple-wavelength anomalous diffraction
24	MBP	maltose-binding protein
25	Mbp	Megabase pair
26	MERS-CoV	Middle East respiratory syndrome coronavirus ( <i>Coronaviridae</i> )
27	MES	2-(N-morpholino)ethanesulfonic acid
28	MHV	mouse hepatitis virus ( <i>Coronaviridae</i> )
29	MOPS	3-(N-morpholino)propanesulfonic acid
30	mRNA	messenger RNA
31	MSA	multiple sequence alignment
32	MTase	methyltransferase
33	MWCO	molecular weight cut-off
34	N	nidovirus nucleocapsid protein
35	n.a.	not applicable
36	n.d.	not done
37	NendoU	nidovirus uridylate-specific endoribonuclease domain
38	NiRAN	nidovirus RdRp-associated nucleotidyltransferase
39	NMD	nonsense-mediated decay

1	N-MT	N7-methyltransferase
2	nsp	nonstructural protein
3	nt	nucleotide
4	NTP	nucleoside triphosphate
5	O-MT	2'-O-methyltransferase
6	ORF	open reading frame
7	p.i.	post infection
8	p.t.	post transfection
9	P1	first passage
10	PABP	poly(A) binding protein
11	PAGE	polyacrylamide gel electrophoresis
12	pC	poly-cytidine
13	PCR	polymerase chain reaction
14	PDB	Protein Data Bank
15	PEG	polyethylene glycol
16	PFU	plaque-forming unit
17	Pol	polymerase
18	poly(A)	polyadenylate
19	pp1ab	polyprotein 1ab
20	PRRSV	porcine reproductive and respiratory syndrome virus ( <i>Arteriviridae</i> )
21	PSSM	position-specific scoring matrix
22	pU	poly-uridine
23	RdRp	RNA-dependent RNA polymerase
24	RF	replicative form
25	RFS	ribosomal frameshift site
26	RI	replicative intermediate
27	RMSD	root-mean-square deviation
28	RNA	ribonucleic acid
29	RNase	ribonuclease
30	RsD	ronivirus-specific domain
31	RTC	replication-transcription complex
32	RTPase	RNA 5'-triphosphatase
33	RT-PCR	reverse transcription polymerase chain reaction
34	SARS-CoV	severe acute respiratory syndrome coronavirus ( <i>Coronaviridae</i> )
35	SDS	sodium dodecyl sulfate
36	SF	helicase superfamily
37	SFV	Semliki Forest virus ( <i>Togaviridae</i> )
38	sg	subgenomic
39	SHFV	simian hemorrhagic fever virus ( <i>Arteriviridae</i> )

1	TBE	tris-borate-EDTA buffer
2	TEV	tobacco etch virus ( <i>Potyviridae</i> )
3	TGEV	transmissible gastroenteritis virus ( <i>Coronaviridae</i> )
4	TLR	Toll-like receptor
5	TM	transmembrane domain
6	ToMV	tomato mosaic virus ( <i>Virgaviridae</i> )
7	tRNA	transfer RNA
8	TRS	transcription-regulating sequence
9	U	unit
10	ub	ubiquitin
11	UTP	uridine triphosphate
12	UTR	untranslated region
13	VPg	viral protein genome-linked
14	WHO	World Health Organization
15	WPDV	wobbly possum disease virus ( <i>Arteriviridae</i> )
16	wt	wild-type
17	ZBD	zinc-binding domain

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

## 1 SUMMARY

2  
3 The order Nidovirales comprises a monophyletic group of viruses with positive-stranded  
4 RNA genomes that are classified in the families *Arteriviridae*, *Coronaviridae*, *Mesoniviridae*,  
5 and *Roniviridae*. They share a conserved genome organization and a characteristic set  
6 of key replicative proteins. Although, in principle, this suggests a conserved replication  
7 mechanism, it is currently unclear how far exactly the resemblance extends on a more  
8 detailed level. This is foremost due to our poor understanding of the role of most viral  
9 proteins in the replication cycle. In addition, most of the knowledge that was obtained  
10 predominantly derives from studies of only few coronaviruses, the nidovirus subgroup  
11 with the largest known genome and therefore presumably employing the most complex  
12 replication strategy. In contrast, thus far only limited attention was given to the RNA rep-  
13 licating and processing enzymes of arteriviruses, and none at all to those of mesoni- and  
14 roniviruses, whose genome sizes are (much) smaller than those of coronaviruses. Given  
15 this disparity, it may be premature to assume that within this divergent group of viruses  
16 essential steps of the viral replication cycle, like for example RNA synthesis and mRNA  
17 5' end modification, strictly follow the same mechanistic pathways.

18  
19 The work described in this thesis addresses some poorly or uncharacterized (domains  
20 of) nonstructural proteins (nsps) that are likely involved in one or multiple steps of  
21 RNA replication and/or transcription of the prototypic arterivirus equine arteritis virus  
22 (EAV). After a short introduction on the nidovirus replication cycle and our knowledge  
23 of the molecular details of the unusual transcription and mRNA processing mechanism  
24 (chapter 1), chapter 2 presents the crystal structure of the enzymatically active EAV  
25 helicase nsp10, which was obtained and analyzed in close collaboration with Chinese  
26 colleagues. Interestingly, a strong resemblance between this viral protein and the con-  
27 served cellular helicase Upf1, in particular with respect to their N-terminal zinc-binding  
28 domains, became obvious. Since this cellular helicase has been implicated in a number  
29 of eukaryotic post-transcriptional quality control mechanisms, a role for nsp10 and its  
30 nidovirus homologs in genome expansion is proposed. This and other potential func-  
31 tions of the nidovirus helicase in RNA replication, transcription, and translation, as well  
32 as virion biogenesis are further discussed in chapter 3, which presents a review of our  
33 current knowledge about nidovirus helicases. Special emphasis is placed on gaps that  
34 still remain, facts that cannot be easily reconciled with our current understanding of the  
35 nidovirus replication mechanisms, and questions that need to be addressed in future.

36  
37 Chapters 4 and 5 focus on one of the central arterivirus replication proteins, nsp9, which  
38 harbors the RNA-dependent RNA polymerase (RdRp) domain. Chapter 4 describes a care-  
39 fully controlled study to investigate different polymerase activities that nsp9 may have,



1 including a previously claimed primer-independent RdRp activity. Despite considerable  
2 efforts, involving experiments with different preparations of nsp9 and assays performed  
3 in the presence of putative polymerase co-factors, no *in vitro* activity was observed that  
4 could be clearly attributed to this protein. Moreover, circumstantial evidence suggested  
5 that the previously reported activity may have been caused by a contamination of the  
6 recombinant nsp9 preparation with the T7 RNA polymerase used to drive its expression  
7 in *E. coli*. In arteriviruses, the RdRp domain is located in the C-terminal two-thirds of nsp9.  
8 In chapter 5, it is now described for the first time that the RdRp domain is flanked at its  
9 N-terminus by another domain that is conserved in all nidoviruses. However, unlike the  
10 situation for the RdRp domain, no homologs of this domain have been found in other  
11 RNA viruses. This domain is thus proposed to be a second marker for the *Nidovirales*  
12 order, besides the N-terminal zinc-binding domain of the helicase subunit. Residues that  
13 are part of three conserved sequence motifs were without exception associated with  
14 a newly discovered nucleotidylation activity of recombinant nsp9. It is thus proposed  
15 that this activity could play a role in the modification of the 5' end of viral RNAs through  
16 either RNA ligation, protein priming of RNA synthesis, or guanylyl transfer during mRNA  
17 capping. Further research is required to definitely tie nsp9 to one of these pathways.  
18 Nevertheless, alanine substitution of any of these conserved residues was either lethal  
19 to EAV and severe acute respiratory syndrome coronavirus (SARS-CoV) or severely  
20 crippled these viruses, eventually resulting in reversion of the mutation. These results  
21 thus demonstrate the essential nature of this domain for virus replication, whatever its  
22 exact function will turn out to be.

23  
24 Two methyltransferase activities, commonly required for capping of mRNAs, were  
25 previously identified in two ORF1b-encoded coronavirus proteins, nsp14 and nsp16.  
26 While the former has no counterpart among the arterivirus nonstructural proteins, the  
27 latter and the arterivirus C-terminal subunit nsp12 occupy equivalent positions in the  
28 ORF1b-encoded part of the replicase although the two proteins share no detectable  
29 sequence similarity. It is thus a long standing question, how arteriviruses may catalyze  
30 the 5' end modification of mRNAs, and we therefore performed a first characterization of  
31 the entirely uncharacterized EAV nsp12 subunit (chapter 6). Based on the genomic posi-  
32 tion of its coding sequence, sequence alignment, and secondary structure prediction, it  
33 is hypothesized that nsp12 might represent a unique arterivirus methyltransferase that  
34 has diverged from its homologs beyond sharing appreciated similarity. To test this hy-  
35 pothesis, recombinant nsp12 was expressed in and purified from *E. coli* and tested alone  
36 and in combination with potential co-factors for N7- and 2'-O-methyltransferase activ-  
37 ity. Although positive controls represented by the SARS-CoV methyltransferases (nsp14  
38 and the nsp10:nsp16 complex) demonstrated the functionality of the assay, no activity  
39 was detected for EAV nsp12. Guided by the sequence alignment, an extensive set of EAV

1 mutants was generated and characterized with respect to their plaque phenotype and  
2 progeny titer, as well as their protein expression. These reverse genetics experiments  
3 revealed a number of phenotypes ranging from wild-type-like via non-spreading to  
4 replication-incompetent, which indicated that nsp12 is essential for viral replication.

5  
6 The above chapters describing biochemical properties of selected proteins may ulti-  
7 mately contribute to the identification of drug targets to combat nidovirus infections. In  
8 chapter 7 the prerequisites under which the marketing of such an antiviral drug would  
9 be economically viable are analyzed. This project was realized under guidance of several  
10 specialists of one of the industrial partners, Janssen Infectious Diseases, of the EUVIRNA  
11 consortium, the Marie Curie Initial Training Network to which my research project be-  
12 longed. This study concludes that, at the moment, none of the circulating nidoviruses  
13 constitutes a sufficiently sized market to warrant the considerable investments required  
14 for drug development. The situation may be different if a new highly-pathogenic virus  
15 would emerge, as exemplified in 2002 by SARS-CoV or 2012 by MERS-CoV. In view of  
16 such threats, pre-pandemic drug stockpiling could be considered. However, also under  
17 those circumstances, it seems likely that the inherent financial risk would preclude an in-  
18 dependent private initiative, even though market parameters and approval procedures  
19 appear to be favorable.

20  
21 Finally, chapter 8 connects some of the main findings described in this thesis with  
22 previously described data. In particular, potential differences between small and large  
23 nidoviruses on the level of the molecular mechanisms of RNA synthesis initiation and  
24 mRNA capping are highlighted. To this end, alternative mechanisms are considered that  
25 would be consistent with the data on arteriviruses presented in this thesis and else-  
26 where. Furthermore, potential roles of cellular helicases in nidovirus replication and the  
27 host's immune response against nidoviruses are discussed.



## 1 SAMENVATTING

2

3 De orde Nidovirales omvat een monofyletische groep van virussen met positiefstrengige  
4 RNA genomen, die geclassificeerd zijn in de families *Arteriviridae*, *Coronaviridae*, *Meso-*  
5 *niviridae* en *Roniviridae*. Deze groepen delen een geconserveerde genomorganisatie  
6 en een karakteristieke set van cruciale replicatie-eiwitten. Alhoewel dit, in principe, een  
7 geconserveerd replicatiemechanisme suggereert, is het op dit moment onduidelijk in  
8 hoeverre deze gelijkenis op een gedetailleerder niveau stand houdt. Dit komt voorna-  
9 melijk door ons beperkte begrip van de rol van het gros van de virale eiwitten in de  
10 replicatiecyclus. Bovendien, het grootste deel van de beschikbare kennis is afkomstig  
11 van studies van slechts enkele coronavirussen, de nidovirus subgroep met het langste  
12 bekende RNA genoom die daarom vermoedelijk gebruik maakt van de meest complexe  
13 replicatiestrategie. In tegenstelling hiermee is tot nu toe slechts beperkt aandacht be-  
14 steed aan de RNA-replicerende en -modificerende enzymen van arterivirussen en geen  
15 enkele aandacht aan die van de mesoni- en ronivirussen, die (veel) kleinere genomen  
16 hebben dan coronavirussen. Gegeven deze ongelijkheid, kan het voorbarig zijn om aan  
17 te nemen dat de essentiële stappen van de virale replicatiecyclus, zoals bijvoorbeeld  
18 RNA synthese en modificatie van het 5' uiteinde van mRNA's, in deze uiteenlopende  
19 groep van verwante virussen strikt dezelfde mechanistische routes volgen.

20

21 Het werk beschreven in dit proefschrift betreft enkele slecht of niet gekarakteriseerde  
22 (domeinen van) niet-structurele proteïnen (nsp) die waarschijnlijk betrokken zijn bij één  
23 of meerdere stappen tijdens RNA replicatie en/of transcriptie van het prototype arterivi-  
24 rus, equine arteritis virus (EAV). Na een korte inleiding over de nidovirus replicatiecyclus  
25 en onze kennis van de moleculaire details van hun ongewone transcriptie- en mRNA  
26 modificatiemechanisme (hoofdstuk 1), wordt in hoofdstuk 2 de kristalstructuur van het  
27 enzymatisch actieve EAV helicase nsp10 gepresenteerd, die verkregen en geanalyseerd  
28 is in nauwe samenwerking met Chinese collega's. Interessant genoeg, werd een sterke  
29 gelijkenis duidelijk tussen dit virale eiwit en het geconserveerde cellulaire helicase Upf1,  
30 vooral wat betreft het N-terminale zink-bindende domein. Aangezien dit cellulaire heli-  
31 case verondersteld wordt betrokken te zijn in een aantal eukaryote mechanismen voor  
32 posttranscriptionele kwaliteitscontrole, wordt voor nsp10 (en homologen in andere  
33 nidovirussen) een rol in genoomexpansie voorgesteld. Deze en andere mogelijke func-  
34 ties van nidovirus helicases in RNA replicatie, transcriptie en translatie, evenals virion  
35 biogenese, worden verder besproken in hoofdstuk 3, dat een overzicht presenteert van  
36 onze huidige kennis over nidovirus helicases. Bijzondere aandacht wordt daarbij ge-  
37 geven aan de resterende lacunes in onze kennis, feiten die minder eenvoudig overeen  
38 lijken te stemmen met ons huidige begrip van de nidovirus replicatiemechanismen en  
39 vragen die in de toekomst aangepakt dienen te worden.

1 Hoofdstukken 4 en 5 richten zich op één van de centrale arterivirus replicatie-eiwitten,  
2 nsp9, dat een RNA-afhankelijke RNA polymerase (RdRp) domein herbergt. Hoofdstuk 4  
3 beschrijft een zorgvuldig uitgevoerde studie naar de verschillende polymerase-activi-  
4 teiten die nsp9 zou kunnen hebben, inclusief een eerder beschreven primer-onafhan-  
5 kelijke RdRp-activiteit. Ondanks aanzienlijke inspanningen, waaronder experimenten  
6 met verschillende preparaten van nsp9 en proeven uitgevoerd in aanwezigheid van  
7 mogelijke polymerase-cofactoren, werd geen *in vitro* activiteit gevonden die duidelijk  
8 toegewezen zou kunnen worden aan dit eiwit. Er werd ook indirect bewijs verkregen  
9 dat suggereert dat de eerder beschreven activiteit wellicht te danken was aan een  
10 contaminatie van het recombinant nsp9 met het T7 RNA polymerase dat was gebruikt  
11 om het eiwit in *E. coli* tot expressie te brengen. In arterivirussen, is het RdRp domein  
12 gelokaliseerd in het C-terminale twee-derde van nsp9. In hoofdstuk 5 wordt voor het  
13 eerst beschreven dat het RdRp-domein N-terminaal geflankeerd wordt door een ander  
14 domein dat is geconserveerd in alle nidovirussen. In tegenstelling tot het RdRp-domein,  
15 zijn van dit domein geen homologen gevonden in andere RNA virussen. Daarom wordt  
16 voorgesteld dat dit domein een tweede universeel kenmerk voor de *Nidovirales* orde is,  
17 naast het N-terminale zink-bindende domein van de helicase subunit. Aminozuren die  
18 onderdeel zijn van drie geconserveerde sequentiemotieven werden zonder uitzonde-  
19 ring geassocieerd met een nieuw ontdekte nucleotidyleringsactiviteit van recombinant  
20 nsp9. Deze activiteit speelt mogelijk een rol in de modificatie van het 5' einde van virale  
21 RNAs middels ofwel RNA ligatie, eiwit priming van RNA synthese of guanylyl overdracht  
22 tijdens RNA capping. Nader onderzoek is vereist om nsp9 definitief met één van deze  
23 mogelijkheden in verband te brengen. Niettemin was alanine-substitutie van ieder van  
24 deze geconserveerde residuen ofwel letaal voor EAV en het severe acute respiratory syn-  
25 drome coronavirus (SARS-CoV) of het beperkte deze virussen ernstig in hun replicatie,  
26 wat uiteindelijk resulteerde in reversie van de mutatie. Deze resultaten demonstreren  
27 dus de essentiële aard van dit domein voor virusreplicatie, ongeacht wat de exacte  
28 functie zal blijken te zijn.

29

30 Twee methyltransferase-activiteiten, vereist voor mRNA capping, werden eerder ge-  
31 identificeerd in twee ORF1b-gecodeerde coronaviruseiwitten, nsp14 en nsp16. Terwijl  
32 de eerste geen tegenhanger heeft onder de niet-structurele eiwitten van arterivirussen,  
33 bezetten de laatstgenoemde en het C-terminale subunit van arterivirussen, nsp12,  
34 vergelijkbare posities in het ORF1b-gecodeerde gedeelte van de replicase, hoewel  
35 de twee eiwitten geen detecteerbare sequentiegelijkenis delen. Het is een langdurig  
36 openstaande vraag hoe arterivirussen de modificatie van het 5' einde van hun mRNAs  
37 katalyseren, en in dat kader hebben we een eerste karakterisering van het nog ongeka-  
38 rakteriseerde nsp12 van EAV uitgevoerd (hoofdstuk 6). Gebaseerd op de genomische  
39 positie van de coderende sequentie, de sequentievergelijking en voorspelling van de

1 secundaire structuur werd verondersteld dat nsp12 mogelijk een unieke arterivirus  
2 methyltransferase zou zijn, dat zich afgesplitst heeft van nidovirus homologen waar-  
3 door noemenswaardige gelijkens ontbreekt. Om deze hypothese te toetsen werd  
4 recombinant nsp12 tot expressie gebracht in en gezuiverd uit *E. coli*, en zowel alleen  
5 als in combinatie met potentiële cofactoren getest op N7- en 2'-O-methyltransferase  
6 activiteit. Hoewel de positieve controles, de beide SARS-CoV methyltransferases (nsp14  
7 en het nsp10:nsp16 complex), de functionaliteit van het experiment demonstreerden,  
8 werd geen activiteit voor EAV nsp12 gedetecteerd. Op basis van de sequentievergelij-  
9 king werd een uitgebreide set van EAV mutanten gegenereerd en gekarakteriseerd op  
10 plauefenotypen en de geproduceerde virus titer, evenals op eiwitexpressie. Deze re-  
11 verse genetics experimenten onthulden een aantal fenotypen, variërend van nagenoeg  
12 wild-type via niet-verspreidend tot replicatie-incompetent, wat aangeeft dat nsp12  
13 essentieel is voor virale replicatie.

14

15 De bovenstaande hoofdstukken die de biochemische eigenschappen van bepaalde  
16 eiwitten beschrijven kunnen uiteindelijk een bijdrage leveren tot de identificatie van  
17 drug targets voor de bestrijding van nidovirus infecties. In hoofdstuk 7 worden de rand-  
18 voorwaarden geanalyseerd die bepalen of de marketing van een dergelijke antivirale  
19 drug economisch levensvatbaar zou kunnen zijn. Dit project werd gerealiseerd onder  
20 de begeleiding van verscheidene specialisten van Janssen Infectious Diseases, één van  
21 de industriële partners van het EUVIRNA consortium, de Marie Curie Initial Training  
22 Network waartoe dit onderzoeksproject behoorde. Deze studie wees uit dat, op dit  
23 moment, geen van de circulerende nidovirussen een voldoende grote markt vormt om  
24 de aanzienlijke investeringen die nodig zijn voor drugontwikkeling te rechtvaardigen.  
25 De situatie kan anders zijn als een nieuw, hoog-pathogeen nidovirus zou opduiken,  
26 zoals geïllustreerd in 2002 door SARS-CoV en in 2012 door MERS-CoV. Met het oog op  
27 zulke bedreigingen, zou het pre-pandemisch aanleggen van drugvoorraden overwogen  
28 kunnen worden. Echter, ook onder die omstandigheden lijkt het waarschijnlijk dat de  
29 inherente financiële risico's een onafhankelijk privaat initiatief uitsluiten, hoewel markt-  
30 parameters en goedkeuringsprocedures gunstig lijken te zijn.

31

32 Tenslotte wordt in hoofdstuk 8 de samenhang beschreven tussen eerder beschreven  
33 data en enkele van de voornaamste bevindingen in dit proefschrift. Vooral de poten-  
34 tiële verschillen tussen kleine en grote nidovirussen op het niveau van de moleculaire  
35 mechanismen van de initiatie van RNA synthese en mRNA capping worden besproken.  
36 Voor dit doel worden ook alternatieve mechanismen in beschouwing genomen die con-  
37 sistent zouden zijn met de arterivirus data gepresenteerd in dit proefschrift en elders.  
38 Bovendien worden de potentiële rollen van cellulaire helicases in nidovirusreplicatie en  
39 de immuunrespons van de gastheer tegen nidovirussen bediscussieerd.



## 1 CURRICULUM VITAE

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

Kathleen Lehmann was born on September 1<sup>st</sup> 1985 in Strausberg, Germany. In 2005 she finished secondary school with a university-entrance diploma (Abitur) awarded by the Herder-Gymnasium Minden, Germany. Afterwards, she went on to Bachelor studies in biochemistry at the University Bayreuth from where she graduated in 2008. To conclude her studies, she moved on to the University of Lübeck to graduate with a Master degree in molecular life science in early 2011. During the first year of this study program, she also worked as tutor for a student practical course in biochemistry and as postgraduate research assistant at the Department of Biochemistry under the supervision of Prof. dr. Holger Steuber. Subsequently, she spent the first half of her second year at two different institutes to complete internships, first at the University of Edinburgh at the Center for Infectious Diseases under the supervision of Dr. Amy Buck and then at the Max Planck Institute for Developmental Biology at the Department of Protein Evolution under the supervision of Dr. Jörg Martin. To conclude the study program, she performed her Master thesis research focused on SARS-CoV nonstructural proteins at the University of California, Irvine at the Department of Molecular Biology and Biochemistry under the supervision of Prof. dr. Michael Buchmeier. In February 2011, she joined the Marie Curie Initial Training Network EUVIRNA and started her doctoral research at the Department of Medical Microbiology of the Leiden University Medical Center under the supervision of Dr. Clara Posthuma, Prof. dr. Alexander Gorbalenya, and Prof. dr. Eric Snijder.



**1 LIST OF PUBLICATIONS**

2

3 **Kathleen C. Lehmann**, Lisa Hooghiemstra, Anastasia Gulyaeva, Dmitry Samborskiy,  
4 Jessika C. Zevenhoven-Dobbe, Eric J. Snijder, Alexander E. Gorbalenya, and Clara C.  
5 Posthuma. Arterivirus nsp12 versus the coronavirus nsp16 2'-O-methyltransferase:  
6 comparison of the C-terminal cleavage products of two nidovirus pp1ab polyproteins  
7 (submitted for publication)

8

9 **Kathleen C. Lehmann**, Anastasia Gulyaeva, Jessika C. Zevenhoven-Dobbe, George M. C.  
10 Janssen, Mark Ruben, Hermen S. Overkleeft, Peter A. van Veelen, Dmitriy V. Samborskiy,  
11 Alexander A. Kravchenko, Andrey M. Leontovich, Igor A. Sidorov, Eric J. Snijder, Clara  
12 C. Posthuma, and Alexander E. Gorbalenya. Discovery of an essential nucleotidylating  
13 activity associated with a newly delineated conserved domain in the RNA polymerase-  
14 containing protein of all nidoviruses (submitted for publication)

15

16 **Kathleen C. Lehmann**, Eric J. Snijder, Clara C. Posthuma, Alexander E. Gorbalenya. What  
17 we know but do not understand about nidovirus helicases; *Virus Res.* (in print)

18

19 Zengqin Deng, **Kathleen C. Lehmann**, Xiaorong Li, Chong Feng, Guoqiang Wang, Qi  
20 Zhang, Xiaoxuan Qi, Lin Yu, Xingliang Zhang, Wenhai Feng, Wei Wu, Peng Gong, Ye Tao,  
21 Clara C. Posthuma, Eric J. Snijder, Alexander E. Gorbalenya, and Zhongzhou Chen. Struc-  
22 tural basis for the regulatory function of a complex zinc-binding domain in a replicative  
23 arterivirus helicase resembling a nonsense-mediated mRNA decay helicase; *Nucleic*  
24 *Acids Res.* 2014 Mar; 42(5):3464-77

25

26 Dong Han, **Kathleen C. Lehmann**, Gerhard Krauss. SSO1450 – a CAS1 protein from  
27 *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA; *FEBS Lett.* 2009 Jun 18;  
28 583(12):1928-32

29

30

31

32

33

34

35

36

37

38

39