

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

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

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

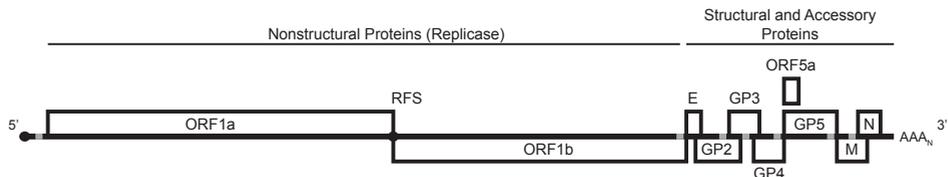
1 INTRODUCTION

2
3 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^{pro}) or main (M^{pro})
 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

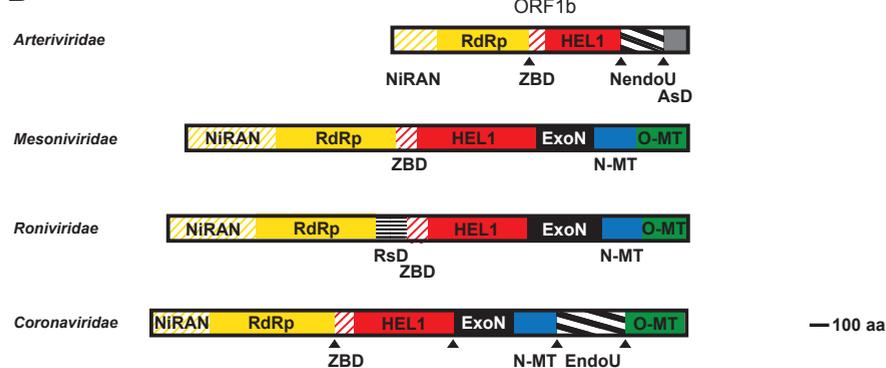


Figure 1. Genome organization and ORF1b-encoded enzymes and domains of nidoviruses. **(A)** Genome organization of *Equine arteritis virus* (EAV) including replicase open reading frames (ORFs) 1a and 1b, and 3' ORFs encoding structural proteins. Genomes of other nidoviruses employ similar organizations while they may vary in respect to size of different regions and number of 3' ORFs. RFS, ribosomal frameshift site. **(B)** ORF1b size and domain comparison between the four nidovirus families shown for EAV (*Arteriviridae*), *Nam Dinh virus* (*Mesoniviridae*), *Gill-associated virus* (*Roniviridae*), and *Severe acute respiratory syndrome coronavirus* (*Coronaviridae*). NiRAN, nidovirus RdRp-associated nucleotidyltransferase; RdRp, RNA-dependent RNA polymerase; ZBD, zinc-binding domain; Hel1, helicase superfamily 1 core domain; NendoU, nidovirus uridylate-specific endoribonuclease; ExoN, exoribonuclease; N-MT, N7-methyltransferase; O-MT, 2'-O-methyltransferase; AsD, arterivirus-specific domain; RsD, ronivirus-specific domain. Depicted is a simplified domain organization 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. Triangles, established cleavage sites by 3CL^{pro} in two virus families; ORF1b-encoded proteins of other viruses may be 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^{pro}, 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 α/β 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^{pro} 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, (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

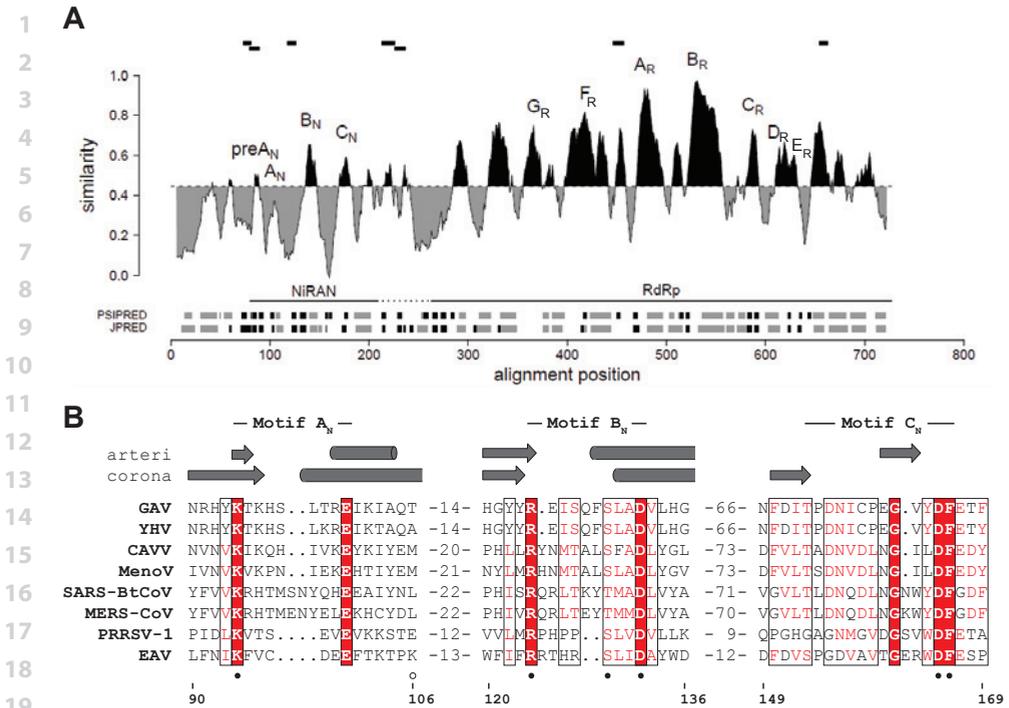


Figure 2: Delineation of the NiRAN domain in RdRp-containing proteins of nidoviruses. **(A)** Sequence variation, domain organization, and secondary structure of the RdRp-containing protein of arteriviruses, and location of peptides identified by mass spectrometry after FSBG-labeling of arterivirus nsp9. 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 α -helices, black for β -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^{pro}, 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.

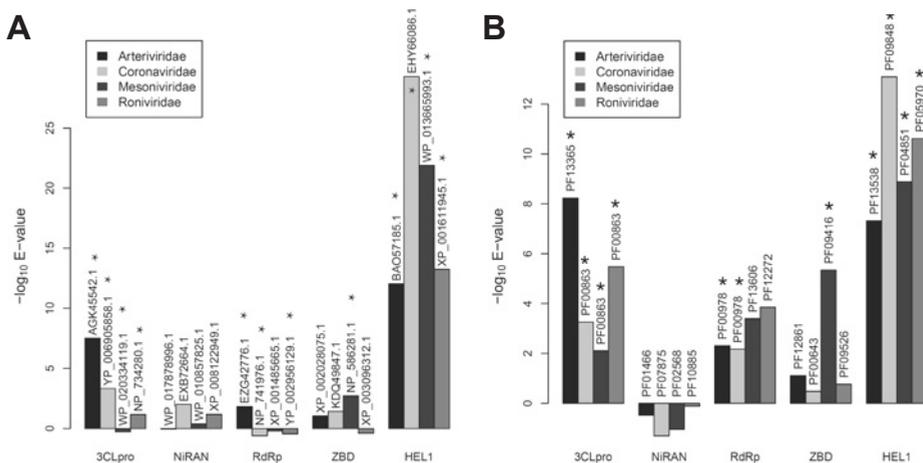


Figure 3: Comparison of nidovirus-wide conserved domains with sequence databases. Shown are histograms depicting E values of the best non-nidovirus hits obtained during HMMER-mediated profile-sequence (A) and HHsearch-mediated profile-profile (B) searches of the GenBank and PFAM A databases, respectively, 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 Mn²⁺-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 α/β 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 α -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}P -labeled at the α -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 γ -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)**.

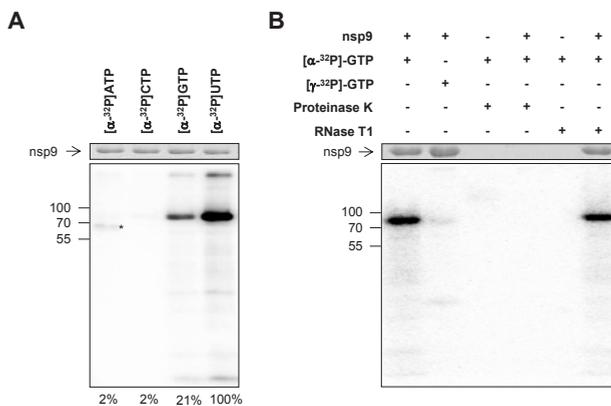
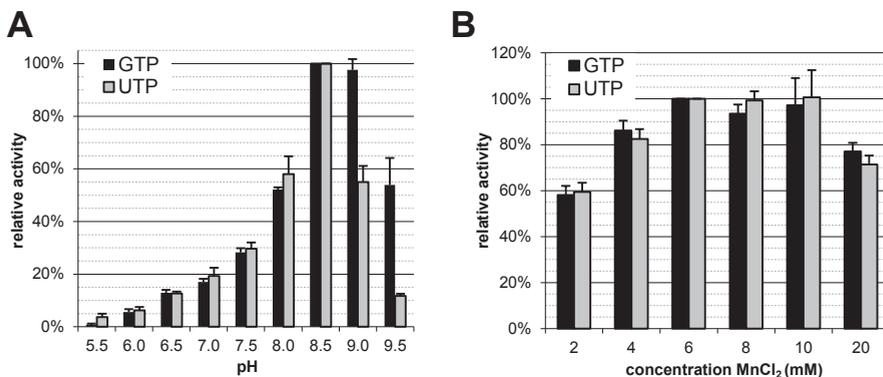


Figure 4. EAV nsp9 has nucleotidylation activity. Purified recombinant EAV nsp9 (78 kDa) was incubated with the indicated [^{32}P]NTP in the presence of MnCl_2 . Reaction products were visualized after denaturing SDS-PAGE by Coomassie brilliant blue staining (top panels) and phosphor imaging (bottom panels). Positions of molecular weight markers are depicted on the left in kDa. **(A)** Uridylylation and guanylylation activity as revealed by covalent binding of the respective radioactive nucleotide to nsp9. Note that the protein indicated with an asterisk likely is an *E. coli*-derived impurity reacting with ATP. Relative band intensities 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 MnCl₂ 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 MnCl_2 . This result made it unlikely that manganese oxidation and a concomitant decrease of available 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 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 (α -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 μM to 2 mM. Within this range a concentration of 100 μM 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 α -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 β -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 α -phosphate, this serine may in prin-

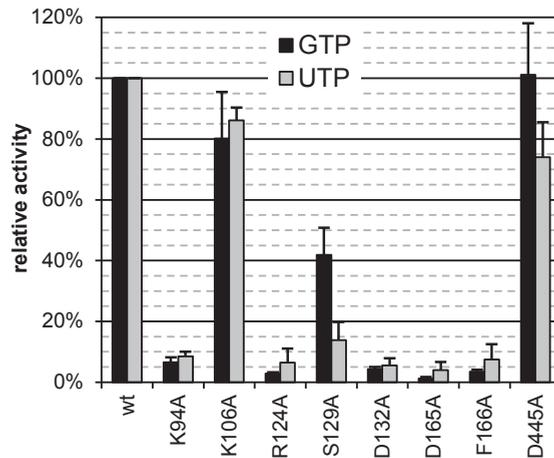
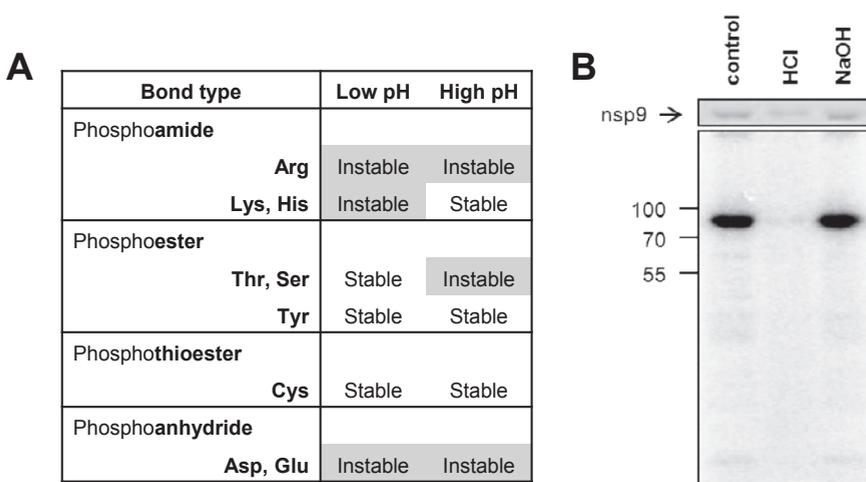


Figure 6. Alanine substitution of conserved NiRAN residues dramatically decreased the nucleotidylation activity of nsp9. In contrast, mutation of the non-conserved K106 in the NiRAN domain or the conserved D445 in the RdRp domain had only a mild effect on activity. Error bars represent the standard deviation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



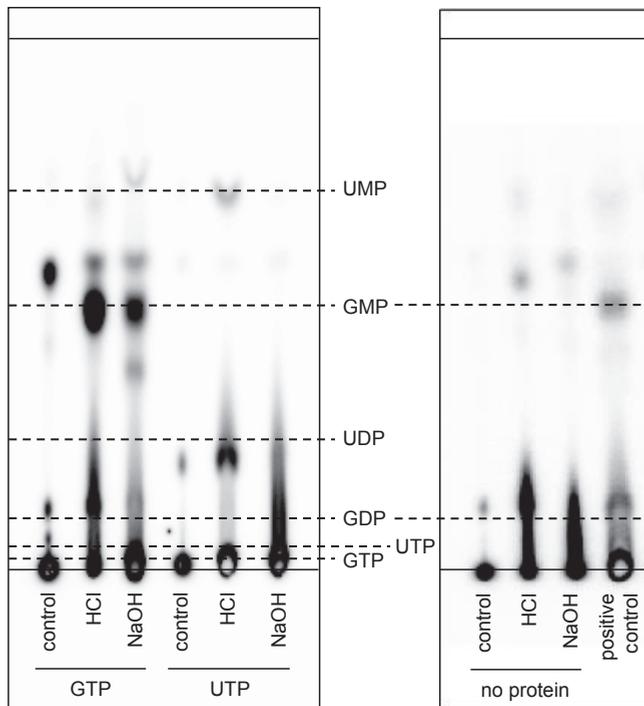
15 **Figure 7.** A phosphoamide bond is formed between nsp9 and the guanosine phosphate. **(A)** Chemical stability of different phosphoamino acid bonds. Adapted from (54). **(B)** The protein was labeled with [α - 32 P] GTP and subsequently incubated at pH 8.5 (control) or under acidic or alkaline conditions. 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 guanylylation and uridylylation 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 β - instead of the α -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



14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35 **Figure 8.** GMP is released from labeled EAV nsp9 under acidic conditions. (A) nsp9 was labeled with $[\alpha\text{-}^{32}\text{P}]$
 36 GTP or $[\alpha\text{-}^{32}\text{P}]$ UTP and was incubated at pH 8.5 (control) or under acidic or alkaline conditions after removal
 37 of non-incorporated nucleotides. Resulting products were separated with PEI-cellulose TLC. Solid lines rep-
 38 resent the position where samples have been spotted (bottom) and the running front (top). Dashed lines
 39 represent the respective mobilities of the indicated nucleotides. (B) $[\alpha\text{-}^{32}\text{P}]$ GTP was incubated under the
 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 β -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

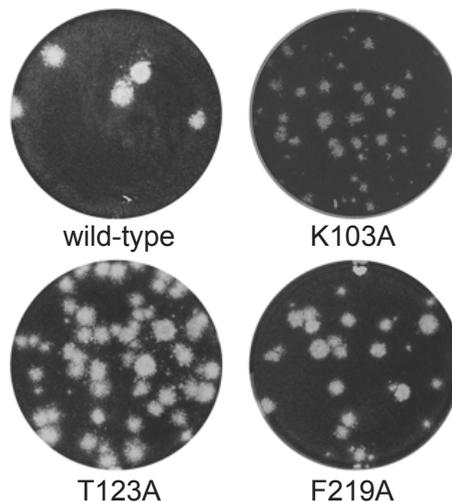
12 To establish the importance of the NiRAN domain for nidoviral replication, reverse ge-
 13 netics was used to engineer both EAV and SARS-CoV mutants in which conserved NiRAN
 14 residues were substituted with alanine. Following transfection of *in vitro*-transcribed full-
 15

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 ^a
EAV	-	wt	-	$1 \cdot 10^7$, $2 \cdot 10^8$	n.d.
	A _N	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 _N	R124A	CGU <u>GCU</u>	<20, <20	Reversion
	B _N	S129A	UCG <u>GCG</u>	$1 \cdot 10^4$, $5 \cdot 10^3$	Reversion
	B _N	D132A	GAU <u>GCU</u>	$3 \cdot 10^4$, $6 \cdot 10^3$	Reversion
	C _N	D165A	GAU <u>GCU</u>	$3 \cdot 10^3$, $1 \cdot 10^4$	Reversion
	C _N	F166A	UUU <u>GCU</u>	<20, <20	n.a.
	A _R	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 _N	K73A	AAG <u>GCC</u>	<20, <20	n.a.
	Non-conserved	K103A	AAG <u>GCA</u>	<20, <20	GCA
	B _N	R116A	CGU <u>GCU</u>	<20, <20	n.a.
	B _N	T123A	ACA <u>GCU</u>	$1 \cdot 10^5$, $4 \cdot 10^5$	GCU
	B _N	D126A	GAU <u>GCG</u>	<20, <20	n.a.
	C _N	D218A	GAU <u>GCU</u>	<20, <20	n.a.
	C _N	F219A	UUC <u>GCG</u>	$2 \cdot 10^4$, $8 \cdot 10^2$	GCG
	A _R	D618A	GAU <u>GCG</u>	<20, <20	n.a.

37 ^aVirus-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.



15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
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.

33
34
35
36
37
38
39

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. 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.

5 6 **Multiple sequence alignments**

7
8 MSAs of five nidovirus-wide conserved protein domains: 3C-like protease (3CL^{Pro}),
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).

32 33 **Homology detection**

34
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 OD₆₀₀ >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 Co²⁺ 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⁻¹cm⁻¹ 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 μ l containing, unless
4 specified otherwise, 50 mM Tris, pH 8.5, 6 mM MnCl_2 , 5 mM DTT, up to 2.5 μ M nsp9,
5 and 0.17 μ M [α - 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 μ 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 μ 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 μ 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 μ g trypsin in 100 μ 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 [α -³²P]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 5×10^6 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

2

3 Supplementary Material and Methods

4

5 Synthesis of 5'-(4-fluorosulfonylbenzoyl)guanosine (FSBG)

6

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).

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

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	
Coordinates on target (aa residues)	Target length	Target species		Target description		
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	
Coordinates on target (aa residues)	Target length	Target species		Target description		
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	
Coordinates on target (aa residues)	Target length	Target species		Target description		
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
Coordinates on target (aa residues)	Target length	Target species		Target description	
4-246	246	Saccharomyces cerevisiae		ATP-dependent molecular chaperone HSP82	
query: Arteriviridae RdRp (alignment of nsp9, columns 224-727, 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)	73,964	2,00E-06	CERT, CERT, NA, MEDIUM	263-727
Coordinates on target (aa residues)	Target length	Target species		Target description	
1-486	487	Sapporo virus			
query: Coronaviridae RdRp (alignment of nsp12, columns 311-1012, 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)	73,91	2,00E-06	CERT, CERT, LOW, MEDIUM	483-965
Coordinates on target (aa residues)	Target length	Target species		Target description	
1-478	478	Murine norovirus 1			

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

query: Mesoniviridae RdRp (alignment of nsp12t, columns 239-1103, 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)	
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			
query: Roniviridae RdRp (alignment of nsp12t, columns 253-1033, 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)	
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	
Coordinates on target (aa residues)	Target length	Target species		Target description		
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	
Coordinates on target (aa residues)	Target length	Target species		Target description		
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	
Coordinates on target (aa residues)	Target length	Target species		Target description		
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		

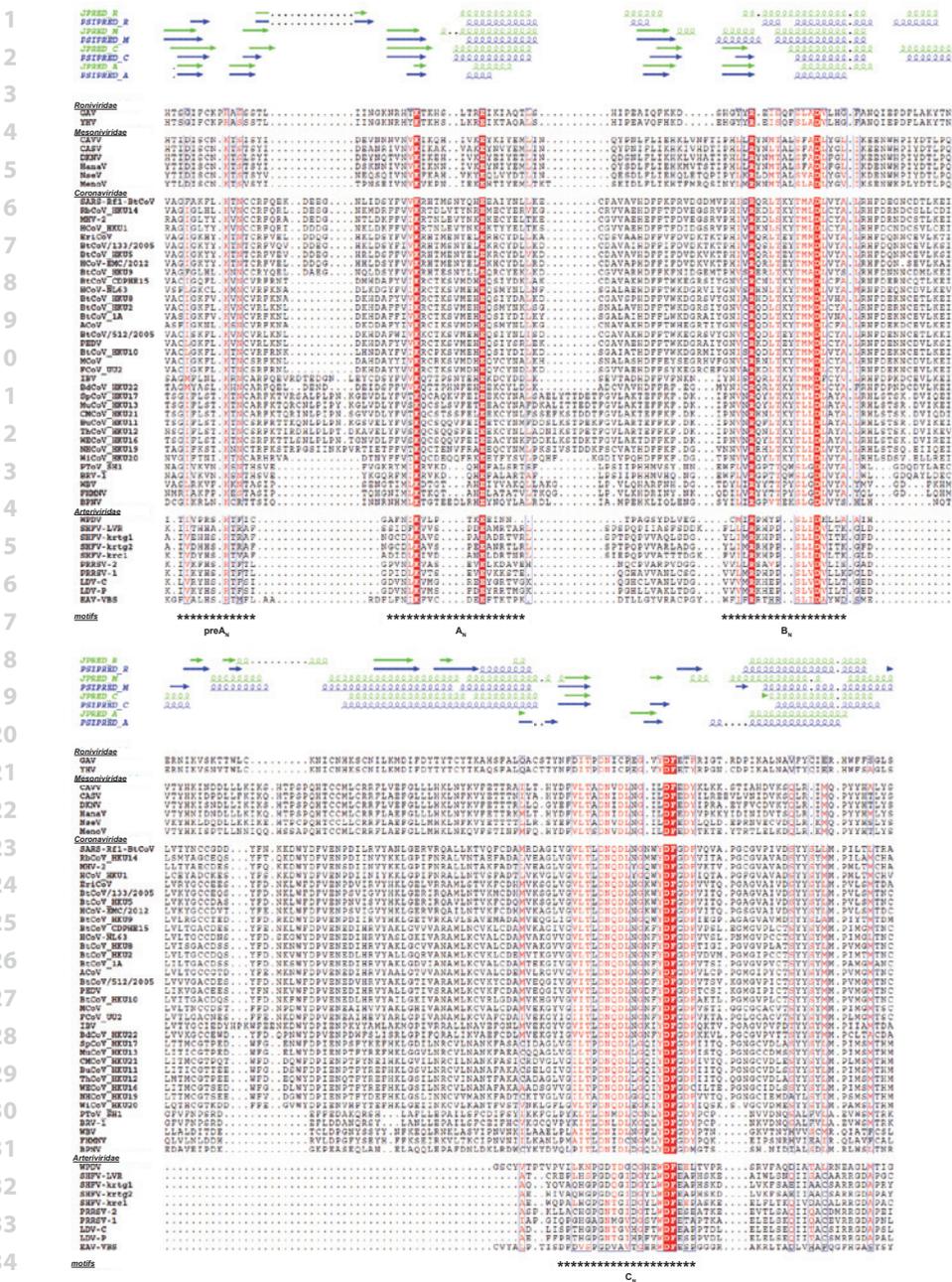


Figure S1: Core part of the nidovirus-wide NiRAN MSA encompassing conserved motifs. Virus names and accession numbers are listed in Table S1. Fully and partially conserved residues are depicted in red boxes or red font, respectively. Sequence motifs are indicated by stars. Secondary structure predictions 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).

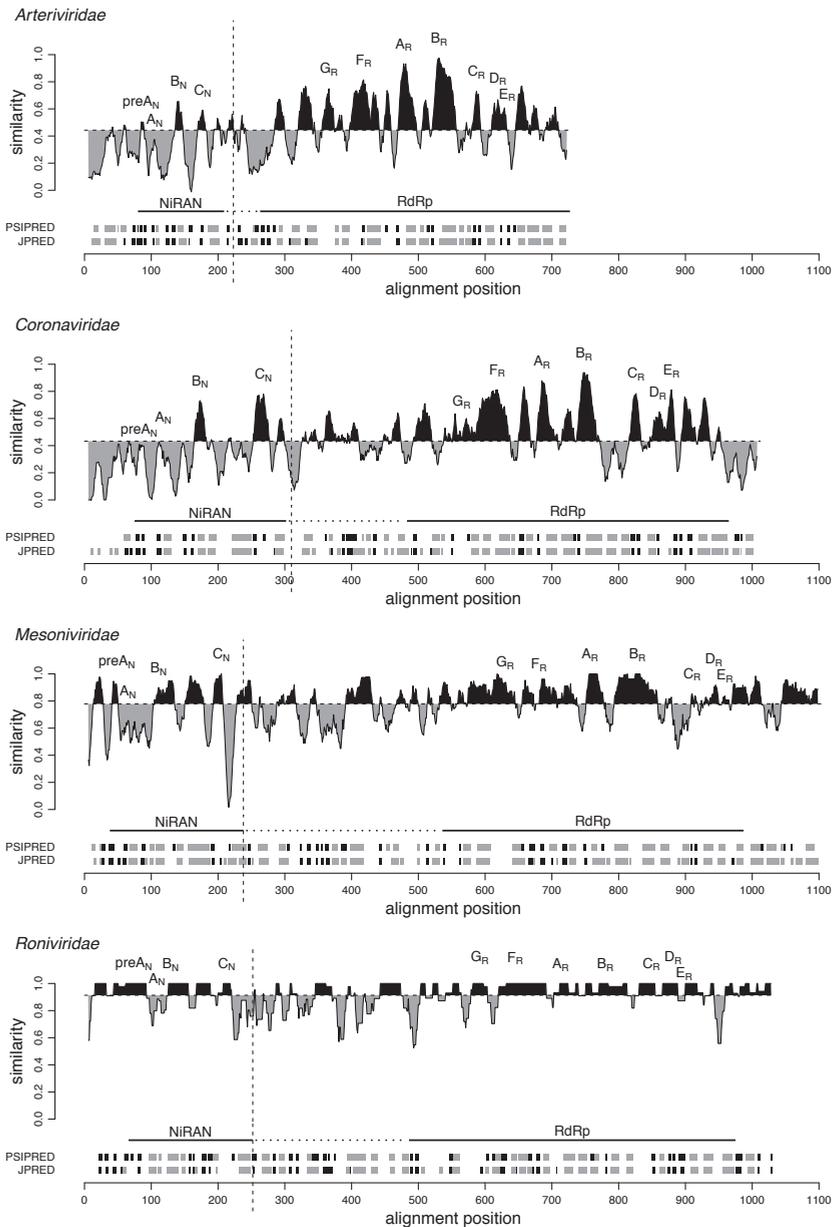
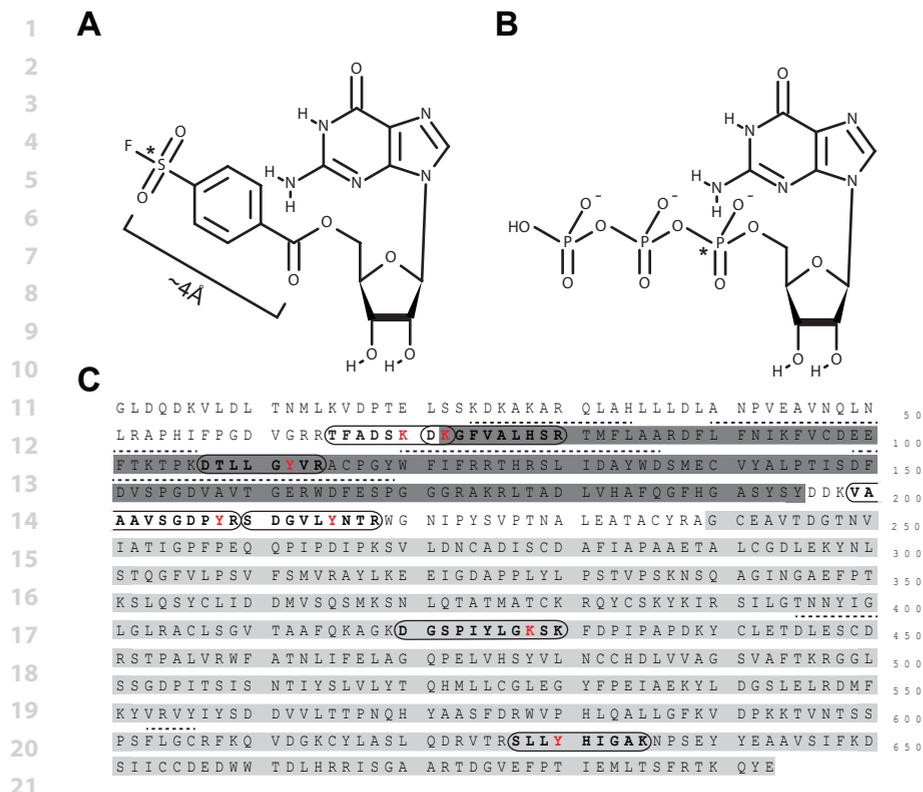


Figure S2: Sequence variation, domain organization, and secondary structure of NiRAN-RdRp-containing proteins of nidovirus families. For each family, the similarity density plot obtained for the MSA of proteins including the NiRAN and RdRp domains is shown. To highlight the regional deviation of conservation from that of the MSA average, areas above and below the mean similarity are shaded in black and gray, respectively. Sequence motifs of NiRAN and RdRp are labelled. Uncertainty in respect to the domain boundary between NiRAN and RdRp is indicated by dashed horizontal lines. Domain boundaries used for all bioinformatics analyses are indicated by dashed vertical lines. Below each similarity density plot predicted secondary structure elements are presented in gray for α -helices and black for β -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 carrying
26 the sulfonylbenzoyl modification are colored in red. Sequence or structural motifs are indicated by
27 dashed lines above the sequence in the order preA_N, A_N, B_N, C_N, A_R, and E_R. See also Figure 2A.

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

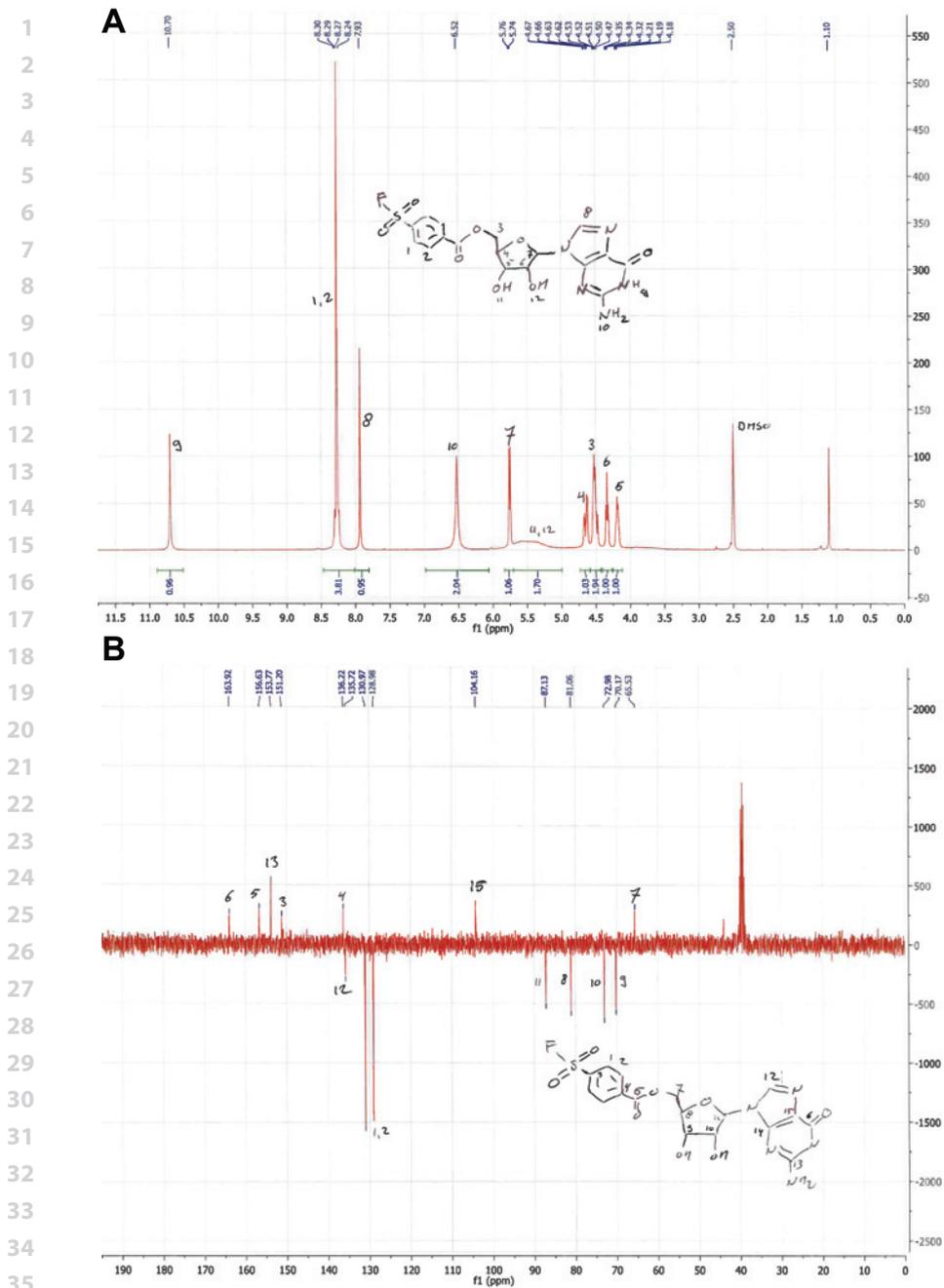


Figure S5: NMR analysis of 5'-(4-fluorosulfonylbenzoyl)guanosine. **(A)** ^1H NMR (300 MHz, DMSO-d_6) δ 10.70 (s, 1H), 8.38 – 8.12 (m, 4H), 7.93 (s, 1H), 6.52 (broad s, 2H), 5.75 (d, $J = 4.8$ Hz, 1H), 5.75 (broad s, 2H), 4.65 (dd, $J = 11.9, 3.6$ Hz, 1H), 4.59 – 4.42 (m, 2H), 4.34 (t, $J = 5.1$ Hz, 1H), 4.25 – 4.12 (m, 1H). **(B)** ^{13}C NMR (75 MHz, DMSO-d_6) δ 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
- 9 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.
- 10
- 11
- 12 4. Coleman CM, Frieman MB. Coronaviruses: important emerging human pathogens. *J.Virol.* 2014; 88(10):5209-5212.
- 13
- 14 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.
- 15
- 16 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.
- 17
- 18 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.
- 19
- 20
- 21 8. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J.Gen.Virol.* 2006;87(Pt 6):1403-1421.
- 22
- 23
- 24 9. Gorbalenya AE, Enjuanes L, Ziebuhr J, *et al.* *Nidovirales*: evolving the largest RNA virus genome. *Virus Res.* 2006;117(1):17-37.
- 25
- 26
- 27 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.
- 28
- 29 11. Lehmann KC, Snijder EJ, Posthuma CC, *et al.* What we know but do not understand about nidovirus helicases. *Virus Res.* 2014;(in press)
- 30
- 31 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.
- 32
- 33
- 34 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.
- 35
- 36
- 37 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.
- 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. *Nucleic Acids Res.* 1989;17(12):4847-4861.
- 27 71. Paul AV, Rieder E, Kim DW, *et al.* Identification of an RNA hairpin in poliovirus RNA that serves as
28 the primary template in the *in vitro* uridylylation of VPg. *J.Virol.* 2000;74(22):10359-10370.
- 29 72. Ambros V, Baltimore D. Protein is linked to the 5' end of poliovirus RNA by a phosphodiester
30 linkage to tyrosine. *J.Biol.Chem.* 1978;253(15):5263-5266.
- 31 73. Pan J, Lin L, Tao YJ. Self-guanylylation of birnavirus VP1 does not require an intact polymerase
32 activity site. *Virology* 2009;395(1):87-96.
- 33
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

