3464–3477 Nucleic Acids Research, 2014, Vol. 42, No. 5 Published online 24 December 2013
doi:10.1093/nar/gkt1310

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

Zengqin Deng[1], Kathleen C. Lehmann[2], Xiaorong Li[1], Chong Feng[1], Guoqiang Wang[1], Qi Zhang[1], Xiaoxuan Qi[1], Lin Yu[1], Xingliang Zhang[3], Wenhai Feng[1], Wei Wu[1], Peng Gong[4], Ye Tao[5], Clara C. Posthuma[2], Eric J. Snijder[2], Alexander E. Gorbalenya[2,6,*] and Zhongzhou Chen[1,*]

[1]State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China, [2]Department of Medical Microbiology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands, [3]Clinical Medicine Research Center, Affiliated Hospital of Guangdong Medical College, Guangdong 524001, China, [4]State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei 430071, China, [5]Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China and [6]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow 119991, Russia

Received October 14, 2013; Revised November 22, 2013; Accepted November 25, 2013

ABSTRACT

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

INTRODUCTION

Helicases and nucleic acid translocases are ATP-dependent motor proteins capable of moving along their nucleic acid substrates while either unwinding duplexed regions (helicases) or performing other functions (translocases), including protein displacement and the nucleation of larger RNA–protein complexes (1,2). These enzymes are known to be critical players in a wide variety of biological processes and are encoded by all organisms, as well as positive-stranded RNA (+RNA) viruses with genomes larger than about 7 kb [(3); for reviews, see (4–6)]. On the basis of sequence comparisons, helicases/translocases have been classified into six superfamilies (SF1 to SF6) (7,8), with +RNA viral helicases belonging to SF1, SF2 or SF3. Based on the direction of translocation, helicases of various superfamilies have been divided into (biochemical) classes A and B, which translocate along their nucleic acid substrates in the 3′-5′ or 5′-3′ direction, respectively (7). In the case of SF1 helicases (9,10), structurally characterized cellular enzymes of class B (SF1B) are further divided into the phylogenetically compact

*To whom correspondence should be addressed. Tel: +86 10 62734078; Fax: +86 10 62734078; Email: chenzhongzhou@cau.edu.cn
Correspondence may also be addressed to Alexander E. Gorbalenya. Tel: +31 71 526 1652; Fax: +31 71 526 6981; Email: a.e.gorbalenya@lumc.nl

Pif1-like (Pif1, RecD2), UvrD/Rep and Upf1-like (Upf1, Ighmbp2) groups, with the latter being able to unwind both DNA and RNA duplexes (11).

Helicase SF1 also includes a large number of (putative) helicases from a dozen +RNA virus families belonging to two diverse phylogenetic lineages, known as the alphavirus-like (or Sindbis virus-like) supergroup (12) and the order *Nidovirales* (13). More detailed studies on the SF1 helicases of two alphavirus-like viruses have recently been published. The helicase domain of the dendrolimus punctatus tetravirus (an insect virus form the *Alphatetraviridae* family) was found to have dsRNA-unwinding activity with 5′-3′ directionality (14). The helicase domain of the plant tomato mosaic virus (ToMV; family *Virgaviridae*) was not characterized enzymatically, but its crystal structure revealed the two canonical RecA-like α/β domains (1A and 2A) of the helicase core (15). Accessory domain insertions, an otherwise frequently observed phenomenon among cellular SF1 helicases, are lacking in the ToMV helicase. The SF1 helicases of nidoviruses (HEL1), one of which is the focus of this study, were characterized in some detail using bioinformatics, molecular genetics and biochemistry (see below), but structural information was lacking thus far.

Nidoviruses constitute an order of +RNA viruses composed of virus groups targeting a wide variety of mammalian, avian and invertebrate hosts. In mammals, nidovirus infection can be associated with severe respiratory disease, as in the case of porcine reproductive and respiratory syndrome (PRRS) (16), one of the leading swine diseases (caused by arteriviruses), and zoonotic coronavirus infections in humans, like severe acute respiratory syndrome (17) and Middle East respiratory syndrome (18). The continuing outbreak of the latter disease is currently attracting worldwide attention, in particular because of its ∼40% case fatality rate. Besides their pathogenic properties, nidoviruses have been studied for their extraordinarily large RNA genomes: even the shortest nidovirus genome (the 12.7-kb RNA of the arterivirus equine arteritis virus, EAV) outranks almost all other mammalian +RNA virus genomes, whereas coronavirus genomes (26.3–31.7 kb) are larger than those of any other RNA virus group. Their large genome size enabled nidoviruses to evolve substantial genetic complexity, which is evident from (among other properties) the acquisition of a variety of enzymatic activities and accessory proteins, many of which are lacking or rare in other +RNA viruses (19). These proteins appear to contribute to the regulation of the complex RNA synthesis of nidoviruses, which occurs exclusively in the cytoplasm of the infected cell, and to the elaborate array of virus–host interactions needed to support efficient virus replication (13,20). For example, nidoviruses with genomes >20 kb use a proofreading 3′-5′ RNA exonuclease that is proposed to promote the fidelity of viral RNA synthesis (19,21–27). However, it is completely unknown whether and how nidoviruses deal with translational quality control during the expression of their large multicistronic genomic RNAs, which also serve as mRNAs for the synthesis of the viral replicative enzymes.

Compared with other +RNA viruses, nidovirus replicase genes encode an exceptionally large number of non-structural proteins (nsps) (19,24,25,28). Nidovirus nsps are expressed from open reading frames (ORFs) 1a and 1b, which make up the 5′-proximal 65–75% of the genome RNA. ORF1a encodes polyprotein 1a (pp1a; size ranging from 1728 to 4550 aa) and following a -1 ribosomal frameshift pp1a can be extended with the ORF1b-encoded polyprotein to give pp1ab (3175–7183 aa) (29) (Supplementary Figure S1). Both polyproteins are subject to extensive proteolytic processing by multiple, internally encoded proteinases (19,30). The nidovirus replicase backbone consists of a conserved array of domains, arranged in a nidovirus-specific order and including the ORF1b-encoded RNA-dependent RNA polymerase (RdRp) and HEL1 domains, the core enzymes needed for genome RNA synthesis (replication) and subgenomic (sg) mRNA production (transcription). The latter process yields an extensive nested set of sg mRNAs, which are used to express up to a dozen structural and accessory proteins from smaller ORFs in the 3′-proximal part of the genome (31–33). In both corona- and arteriviruses, sg mRNAs contain a common leader sequence that is identical to the 5′ end of the genome. Their generation from sg negative-stranded templates involves a mechanism of discontinuous negative strand RNA synthesis (31,32).

Previous studies identified the nsp carrying RNA helicase activity (arterivirus nsp10 and coronavirus nsp13) as one of the two most evolutionarily conserved nidovirus proteins. Biochemical studies using recombinant arterivirus and coronavirus helicases revealed similar enzymatic properties, including nucleic acid-stimulated ATPase and 5′-3′ duplex unwinding activities on both RNA and DNA substrates containing 5′ single-stranded regions (34,35). A unique nidovirus helicase feature is the presence of an N-terminal (predicted) complex zinc-binding domain (ZBD) of 80–100 residues. ZBD includes 12 or 13 conserved Cys/His residues (36) and is a nidoviral genetic marker not found in any other RNA virus group (19). ZBD is separated from the downstream HEL1 domain by an uncharacterized domain that varies in size and sequence between arteri- and coronaviruses (37). For the arterivirus prototype EAV, the significance of the nsp10 ZBD was evaluated extensively using site-directed mutagenesis in combination with biochemical assays and reverse genetics. Amino acid substitutions in ZBD or the adjacent 'spacer' that connects it to the downstream domain can profoundly affect EAV helicase activity and RNA synthesis, with most replacements of conserved Cys or His residues yielding replication-negative virus phenotypes (36,37). Intriguingly, some mutations in the spacer region selectively inactivated transcription, while not affecting replication (36,38), strongly suggesting a specific role for nsp10 in the unique mechanism of discontinuous sg RNA synthesis.

Despite its importance as a key replicative enzyme and antiviral drug target (39), no 3D structural information has been reported for any nidovirus helicase. To understand the regulatory role of ZBD and the protein's interaction with nucleic acids, we characterized the structure of a helicase-competent derivative of EAV nsp10, alone and

in complex with poly(dT). The multi-domain nsp10 includes the canonical 1A and 2A core domains of a SF1 helicase, a flexible accessory domain that is sensitive to nucleic acid binding, and a complex ZBD displaying a novel structural organization. Strikingly, the protein was found to bear structural resemblance to the eukaryotic Upf1 helicases, which are multi-domain proteins involved in RNA quality control, including nonsense-mediated mRNA decay (40). Thus, our study not only highlights how nidovirus helicase activity depends on the extensive relay of interactions between the ZBD, accessory and HEL1 domains but also provides a framework to propose and explore a role for the enzyme in the post-transcriptional quality control of nidovirus RNAs.

MATERIALS AND METHODS

Cloning, expression and purification of soluble EAV nsp10

Nsp10 of the EAV-Bucyrus isolate (NCBI Reference Sequence NC_002532) is composed of amino acids 2371–2837 of replicase pp1ab, which will throughout this study be referred to as nsp10 residues 1–467. The full-length nsp10 sequence or a C-terminally truncated version comprising residues 1–402 (nsp10Δ) were cloned into a modified pET28a vector with a tobacco etch virus (TEV) protease cleavage site. Mutations were generated using the QuikChange protocol and confirmed by DNA sequencing. The proteins were overexpressed at 37°C in *Escherichia coli* strain BL21 (DE3) grown to an OD_{600} of ~0.8 in Luria–Bertani medium in the presence of 50 μg/ml kanamycin. Protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside for 12 h at 16°C. Cell pellets were resuspended in lysis buffer (20 mM HEPES, pH 7.0, for nsp10Δ or pH 8.0 for full-length nsp10, 500 mM NaCl and 30 mM imidazole), supplemented with protease inhibitor cocktail (Roche) and disrupted by sonication. Lysates were clarified at 20 000*g* for 30 min and the soluble fraction was applied to a Ni^{2+} chelating column. After sample loading, the column was washed (20 mM HEPES, pH 7.0 or 8.0, 500 mM NaCl and 60 mM imidazole) and the protein was eluted (20 mM HEPES, pH 7.0 or 8.0, 500 mM NaCl and 400 mM imidazole). Proteins intended for ATPase or helicase assays were dialysed against storage buffer (20 mM HEPES, pH 7.0 or 8.0, 100 mM NaCl, 50% glycerol) and stored at −20°C. Truncated protein for crystallization studies was digested with 10% (w/w) TEV protease to remove the His-tag. Further purification was performed by size-exclusion chromatography using a Superdex 200 column (GE Healthcare) with GF buffer (20 mM HEPES, pH 7.0, 500 mM NaCl). The peak fraction was collected and analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis.

Crystallization and data collection

Purified nsp10Δ was concentrated to 10 mg/ml and initial crystallization trials were performed at 16°C using the sitting-drop vapour-diffusion method by mixing 1 μl of protein solution with 1 μl of reservoir solution. The conditions were then optimized and high-quality crystals were obtained in 1.6 M $(NH_4)_2SO_4$, 0.1 M HEPES, pH 7.1, 25 mM KCl and 20% ethylene glycol. To obtain crystals of the protein–DNA complex, purified protein and partially double-stranded DNA with a 5′ single-stranded poly-thymidine overhang (the two partially complementary sequences were 5′-TTTTTTTTTTGCAGTGCT CG-3′ and 5′-CGCGAGCACTGC-3′) were mixed in a 1:1.5 molar ratio and incubated at 4°C overnight. The complex was further purified by size-exclusion chromatography (Superdex 200, GE Healthcare) and concentrated to 5 mg/ml. The condition for obtaining crystals was 14% PEG 3350, 0.1 M HEPES, pH 7.0, and 0.2 M calcium acetate. For data collection, crystals were cryoprotected in mother liquor containing 25% (v/v) ethylene glycol and flash cooled to −173°C.

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

Structure determination

The structure of nsp10Δ was determined by the MAD method. Initial phases were calculated by SOLVE, and phases were subsequently improved using RESOLVE (42). The figure of merit from the MAD phasing was 0.36 and the Z score was 15.7. Several segments of the protein could be automatically modelled into the electron-density map by RESOLVE, although in part only as poly-alanine chains. Manual rebuilding was performed in COOT (43), and refinement was performed with REFAMC5 (44). Further rounds of refinement were done with Translation/Libration/Screw (TLS) refinement (45). The structure was refined to 2.0 Å with an R_{work} of 19.5% and an R_{free} of 22.4%.

Using the structure of free nsp10Δ without domain 1B as input model, the structure of nsp10Δ in complex with DNA was successfully solved by molecular replacement. The initial model was obtained by MOLREP from the CCP4 program suite (46). A good match for domains ZBD, 1A and 2A with electron density was found. Domain 1B was manually added with the aid of $2F_o–F_c$ and $F_o–F_c$ maps using COOT (43). DNA molecules were included in the final stages of refinement. Difference Fourier maps clearly showed electron densities for seven bound deoxyribonucleotides. The final model was refined to 2.65 Å with an R_{work} of 23.2% and an R_{free} of 25.7%. All figures in this article displaying molecular structure were made using PYMOL (47).

RESULTS

C-terminally truncated EAV nsp10 retains ATPase and helicase activity

Full-length EAV nsp10 and a series of truncated variants were overexpressed in and purified from *E. coli*. After extensive crystallization trials, diffracting crystals could only

Table 1. Data collection and refinement statistics of nsp10Δ and the nsp10Δ–DNA complex[a]

Data collection	Zn-Peak	Zn-edge	Zn-remote	Nsp10Δ	Nsp10Δ-DNA
Wavelength	1.2827	1.2831	1.0000	1.0000	1.0000
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	P1
Cell dimensions					
a,*b*,*c*(Å)	89.1,90.6,56.9	89.1,90.6,56.9	89.1,90.6,56.9	89.8,91.0,57.7	56.6,88.8,128.8
α,β,γ (°)	90,90,90	90,90,90	90,90,90	90,90,90	81.7,90.0,71.4
Resolution (Å)[b]	50–2.83 (2.88–2.83)	50–2.80 (2.85–2.80)	50–2.75 (2.80–2.75)	50–2.0 (2.03–2.0)	50–2.65 (2.7–2.65)
*R*merge (%)	8.7 (38.9)	9.0 (39.9)	9.3 (42.8)	7.4 (67.0)	10.0 (73.2)
I/σ	15.2 (2.2)	15.5 (1.9)	17 (2.8)	42.0 (2)	25.6 (3.4)
Completeness (%)	97.6 (87.2)	98.8 (84.5)	99.1 (89)	97.8 (86.1)	96.1 (95.3)
Redundancy	6.9 (5.3)	6.8 (4.3)	10.1 (6.7)	12.5 (8.9)	4.9 (4.8)
Refinement					
Resolution (Å)				50–2.0 (2.05–2.0)	50–2.65 (2.72–2.65)
Number of reflections				30451 (1874)	62140 (4421)
R_{work}/R_{free} (%)				19.5/22.4	23.2/25.7
Number of atoms					
Protein				2986	11594
DNA/ion				29	567
Water				252	822
B-factors					
Protein				42.7	49.6
DNA/ion				44.3	62.0
Water				50.4	42.0
RMSD					
Bond lengths (Å)				0.008	0.007
Bond angles (°)				1.17	1.22
Ramachandran plot (%)[c]				93.9/6.1/0/0	83.3/15.9/0.8/0

[a]Three crystal experiments for each structure.
[b]Statistics for highest resolution shell.
[c]Residues in most favoured, additional allowed, generously allowed and disallowed regions of the Ramachandran plot.

be obtained for a truncated form of nsp10 (aa 1–402) lacking the 65 C-terminal residues. For simplicity, we will hereafter refer to this protein as nsp10Δ, which was used throughout this study unless otherwise specified. To verify that nsp10Δ, which contained all characteristic SF1 helicase sequences (motifs), is enzymatically active, we performed *in vitro* enzyme assays to compare full-length and truncated nsp10. In agreement with previously published results (35), full-length nsp10 displayed only weak ATPase activity in the absence of nucleic acid, but was strongly stimulated by the addition of poly-uridine (polyU). In the absence of polyU nsp10Δ showed a 5-fold higher ATPase activity than the full-length protein (Figure 1A), yet this increased ATP turnover did apparently not translate into increased helicase activity. Unwinding of a partially double-stranded DNA substrate by nsp10Δ was incomplete, but went to completion when using full-length nsp10 (Figure 1B). As expected, replacement of the conserved lysine of the Walker A motif, which is essential for ATP hydrolysis (35), with glutamine (mutant K164Q) completely abolished ATPase and consequentially also helicase activity. This confirmed that the observed activities could be completely attributed to the recombinant EAV proteins used, rather than to potential trace amounts of contaminating bacterial enzymes.

The observed enzymatic differences between nsp10 and nsp10Δ may be caused by the latter's truncation and could, in principle, be explained by one or multiple defects, like decreased unwinding velocity and/or processivity, loss of affinity towards the substrate or uncoupling of ATPase from helicase activity. The results of the ATPase assay lead us to propose that the observed reduction of duplex unwinding may be due to unproductive ATP hydrolysis, originating from the fact that the ATPase reaction is independent of nucleic acid substrate binding. Accordingly, the input ATP in the nsp10Δ assay may have been depleted before complete unwinding was achieved. Regardless of which interpretation is correct, the C-terminal 65 amino acids clearly are dispensable for the helicase activity of EAV nsp10. This result is in good agreement with the fact that the truncated protein retained all HEL1 key domains (Figure 2A) previously shown to be evolutionary conserved and essential in both *in vitro* enzyme assays and *in vivo* studies with virus mutants.

The crystal structure of EAV nsp10Δ reveals a multi-domain organization of the arterivirus replicative helicase

Because 3D structures of orthologous proteins were not available, we took advantage of the zinc-binding properties of nsp10 and used the zinc multiple-wavelength anomalous dispersion (MAD) method (42) to solve the EAV nsp10Δ structure. The presence and position of three zinc atoms were established with anomalous data collected from the zinc absorption edge (Table 1). The final model included EAV nsp10 residues 1–401, three zinc ions in the N-terminal ZBD, five sulphate ions and 267 water molecules.

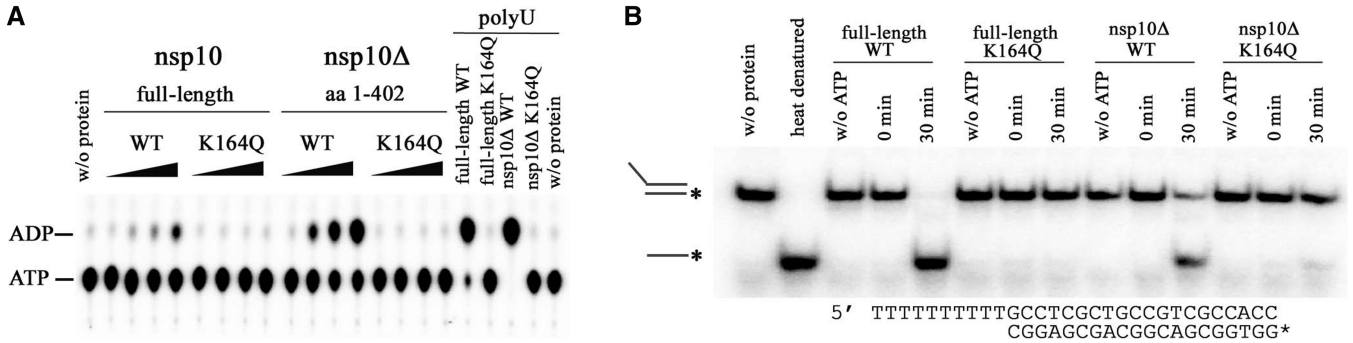

Figure 1. EAV nsp10 *in vitro* enzymatic activity assays. (**A**) ATPase activity of full-length EAV nsp10, the C-terminally truncated nsp10Δ (aa 1–402) and respective active site mutants carrying a Lys-164 to Gln substitution in their Walker A box were analysed as described in Supplementary Experimental Procedures. In the absence of nucleic acid, ATPase activity was measured by incubation at 20°C for 0, 5, 15 and 30 min, respectively. ATPase activity was strongly stimulated by the presence of polyU, as measured by incubation at 20°C for 5 min. (**B**) Helicase activity of full-length nsp10 and nsp10Δ, and their respective K164Q mutants. Activity was determined with the indicated DNA substrate (the asterisk marks the position of the radioactive label). Samples were incubated for 0 or 30 min at 30°C. Control samples without protein or ATP were incubated for 30 min.

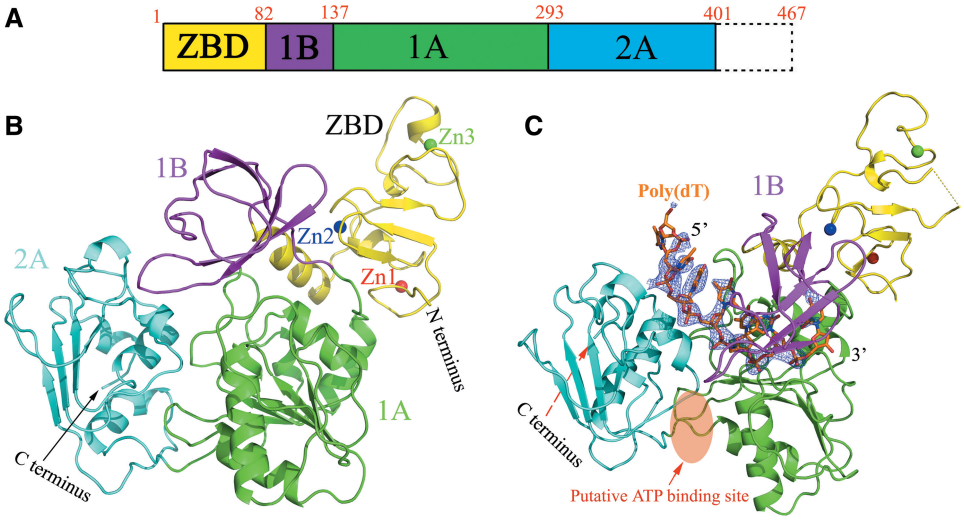

Figure 2. Overall structures of EAV nsp10Δ and the nsp10Δ–DNA binary complex. (**A**) Domain organization of EAV nsp10 depicting the N-terminal ZBD (yellow), the two RecA-like domains 1A (green) and 2A (cyan) of HEL1 and an additional regulatory domain 1B (magenta). Structure of (**B**) free and (**C**) nucleic acid-bound nsp10Δ. Also the F_o-F_c differential electron density map of the bound single-stranded part of a partially double-stranded DNA substrate at 2.5 σ is presented. The putative ATP binding site is shown as a brown oval.

Two RecA-like α/β domains (1A and 2A) form the structure's C-terminal part (Figure 2B; cyan and green) and constitute the helicase core (HEL1). Domain 1A contains a parallel five-stranded β-sheet that is sandwiched by three α-helices on one side and two α-helices on the other. Domain 2A contains a parallel four-stranded β-sheet with five α-helices on the side facing domain 1A. Upstream of domain 1A, we identified an additional domain with a characteristic β-barrel fold (Figure 2A and B; magenta). It consists of five β-strands arranged as two tightly packed anti-parallel β–sheets and is juxtaposed to domain 1A (Figure 2B). The location of this domain in the protein sequence and its orientation relative to the HEL1 domain resemble those of domain 1B in helicases of the SF1B Upf1-like subfamily (Figure 3B and C), and it was therefore named accordingly in our nsp10Δ structure. The domain has no counterpart in the only other solved structure of a viral SF1 helicase, that from ToMV (Figure 3A) (15), whereas its counterpart in helicases of the Pif1-like subfamily is inserted in domain 2A (Figure 3D) (48).

Our structure further revealed that the N-terminal ZBD (Figure 2; yellow) has a compact fold containing three structural zinc atoms. Based on secondary structure analysis with DIAL (49), we could partition ZBD into three elements (Figure 4). Two adjacent and structurally different zinc fingers, an N-terminal RING-like module (residues 1–40, pink) and a treble-clef zinc finger (residues 41–65, red) constitute the main body of ZBD. The third element is a C-terminal linker region (Linker1) that includes the long loop L7, which crosses the entire domain, and helix α4 (residues 66–82, yellow), which connects the two zinc fingers with domain 1B (Figure 4A). This classification is further supported by the observation that the connecting residues between the RING module and treble-clef zinc finger are disordered

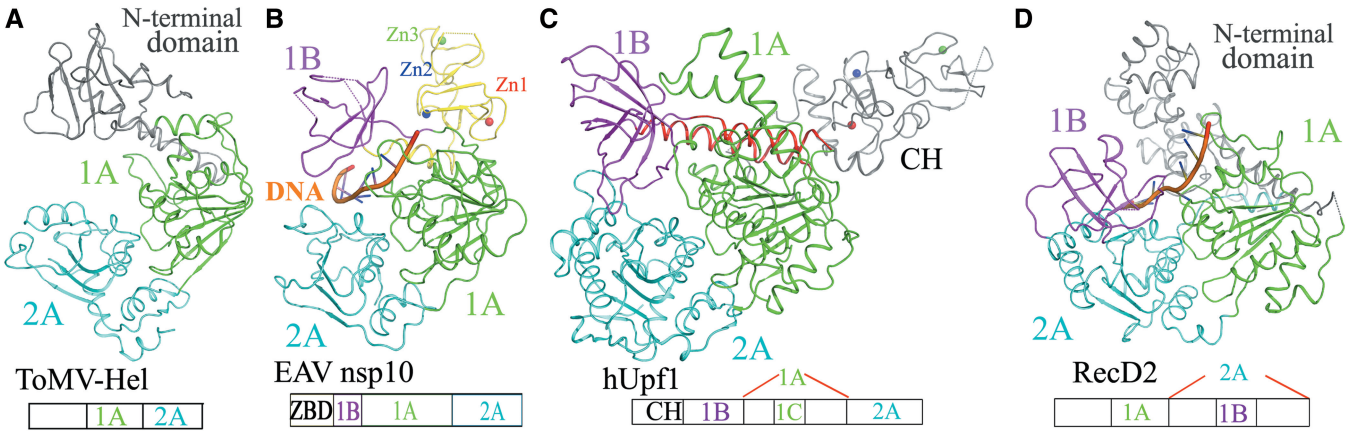

Figure 3. Structural comparison of EAV nsp10Δ with selected SF1 helicases. (**A**) ToMV-HEL (pdb code: 3vkw), (**B**) EAV nsp10Δ, (**C**) hUpf1 (pdb code: 2wjy) and (**D**) RecD2 (pdb code:3gp8). Domain colours are the same as in Figure 2.

(Supplementary Figure S2 and Figure 4D). Only 12 out of the 13 Cys/His residues are involved in zinc binding, rather than all 13 residues as proposed previously [(36); Figure 4B and C]. Not involved is His34, which is not conserved in other arteri- and coronaviruses (Supplementary Figure S3B).

The N-terminal RING-like module has a notable binuclear structure with a cross-brace topology involving 6 Cys and 2 His residues that coordinate two zinc ions (Figure 4A). A three-stranded antiparallel β-sheet (β1–β3) sits in the centre and packs against helix α1 following β2 (Figure 4B). The first zinc ion (Zn1) is coordinated by four cysteine residues (Cys4, Cys7, Cys22 and Cys25) within a treble-clef zinc finger-like motif. Residues Cys4 and Cys7 are provided by the zinc knuckle within loop L1, whereas Cys22 is positioned at the C-terminus of β2 and Cys25 comes from the N-terminus of helix α1. The second zinc ion (Zn2) is coordinated by residues Cys17, Cys33, His29 and His32, which are arranged in an αββ zinc finger-like motif. The second pair of the zinc-coordinating residues of both zinc-binding motifs of the RING module may include both His and Cys residues in other arteri- and coronaviruses. Overall, the RING module of these viruses can be described by a characteristic conserved $Cys2_A$-Cys_B-Cys[His/Cys]$_A$-[His/Cys]3_B pattern (where applicable, A and B refer to residues chelating the first and second zinc ion, respectively; brackets indicate positions at which His and Cys can alternate).

The C-terminal zinc finger of ZBD adopts a treble-clef fold distinct from that of the RING module (see above; Figure 4C). Two one-turn helices α2 and α3 are stabilized by a zinc atom (Zn3) that is chelated by residues Cys42 and His44 of a Zn-knuckle within loop L5, while Cys53 and Cys56 originate from L6 and α3, respectively. An extensive array of hydrogen bonds is observed between the main chains of residues in loop L7 and Thr54 in α3 (Figure 4D). These multiple hydrogen-bonding interactions play a major role in the formation of a compact zinc finger. Arteri- and coronaviruses appear to tolerate replacements (Cys for His, or vice versa) at the second and fourth residues of this finger (36,37), which can be described by the characteristic, conserved C[H/C]C[C/H] pattern. Finally, Linker 1 includes only one structured element (α4), but it plays a central role in the interaction between the main body of ZBD and HEL1, as detailed below.

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

Previously, ZBD mutagenesis demonstrated the *in vitro* and *in vivo* importance of this domain for nsp10 enzyme activities, genome replication and transcription, and arterivirus viability. The solved structure now provides us with a structural basis for these observations. ZBD packs against the HEL1 domains through extensive hydrophobic and hydrophilic interactions (Figure 5A and B). Specifically, residues Leu138, Val141, Val143, Leu147, Pro247, Val248, Leu280 and Trp281 in domain 1A together with residues Ile71, Leu72, Leu75, Leu76 and Ile79 from α4 in ZBD create an extensive hydrophobic surface. The total interface area between ZBD and the HEL1 is 1019 $Å^2$, as determined by Protein Interfaces, Surfaces and Assemblies (PISA) server (50). A major part of this interface involves the α4 helix, which is located in a groove formed by two helices and a loop of domain 1A, while making extensive contacts to the main body of ZBD and, to lesser extent, domain 1B (Figure 5). The interface areas between α4 and domain 1A, on the one hand, and the ZBD fingers (including zinc ions) on the other hand, are 558.1 and 402.4 $Å^2$, respectively. In addition, four hydrogen bonds between ZBD and the HEL1 enhance the interaction (Figure 5B), and a salt bridge is observed between His78 in ZBD and Asp136 in domain 1B (Figure 5B). The large size of these interface surfaces and the large number of interactions suggest the existence of a signalling network through which ZBD could affect both the fold and activity of HEL1.

The proposed signalling network can now be used to rationalize, in a structural context, the previously reported phenotypes of EAV ZBD mutants carrying replacements of residues not directly involved in Zn-binding. For instance, a replication-negative phenotype was described for mutant D45A (36). It is now clear that Asp45 forms

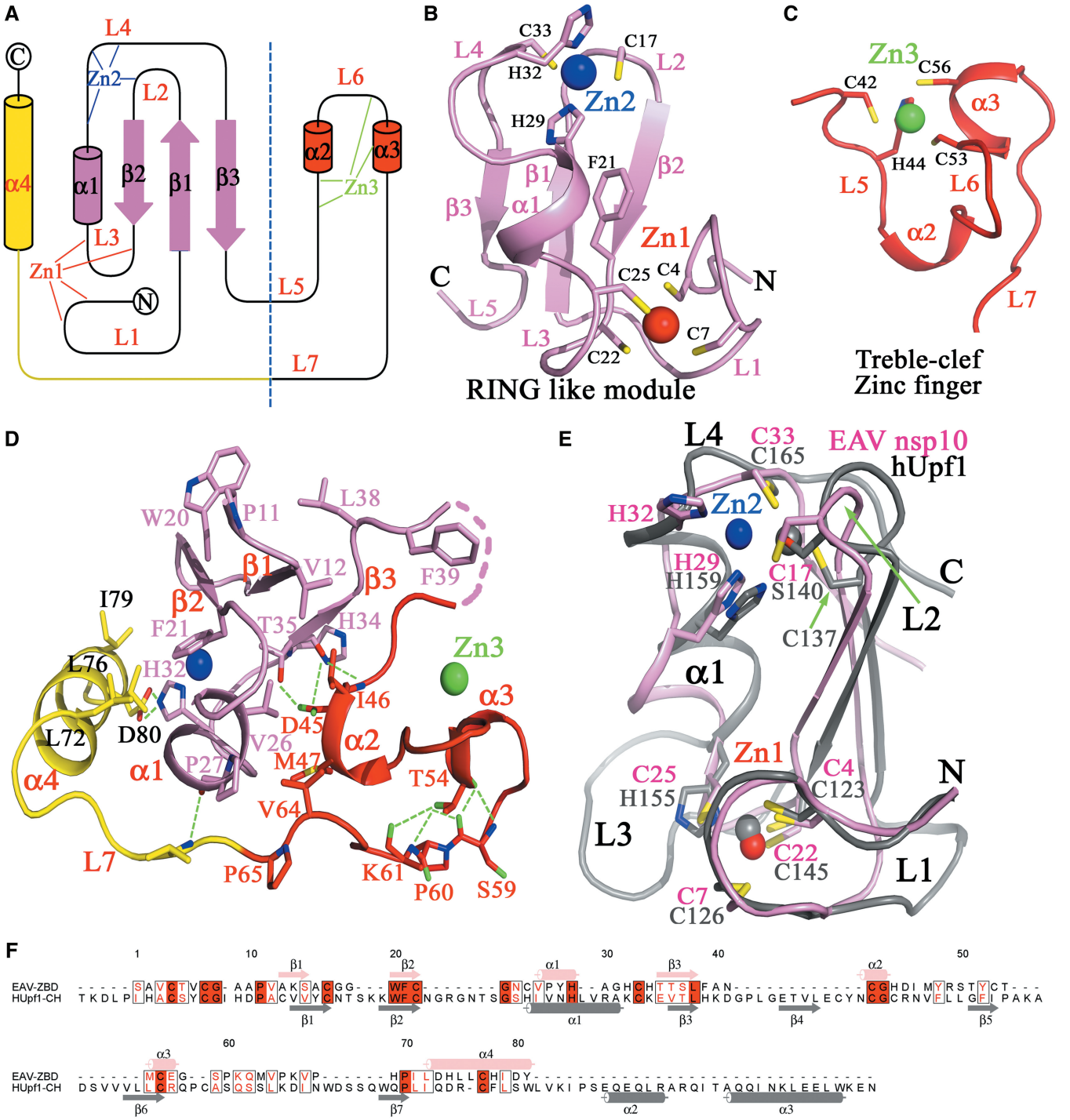

Figure 4. Structural characterization of the EAV nsp10Δ ZBD. (**A**) Topology of the ZBD with its RING-like module (pink), treble-clef zinc finger (red) and Linker 1 (yellow) indicated. (**B**) Structure of the RING-like module and (**C**) treble-clef zinc finger. The residues coordinating the Zn^{2+} ions are shown as sticks. (**D**) Interactions between the RING-like module and the treble-clef zinc finger. (**E**) Superposition of the RING-like modules of EAV nsp10 (pink) and hUpf1 (pdb code:2wjy; grey). (**F**) Sequence alignment of ZBD with the CH domain of hUpf1.

two hydrogen bonds with the main and side chain of Thr35 and electrostatically interacts with the side chain of His34, which both belong to the RING-like zinc finger (Figure 4D). Replacement of Asp45 may thus greatly reduce these interactions and disrupt ZBD integrity, potentially affecting the structural integrity of the HEL1. Another residue, Ser59, was probed extensively by mutagenesis after the finding that a virus mutant (EAV030F) carrying a S59P mutation replicates its genomic RNA with wild-type efficiency, while being completely defective in sg mRNA synthesis (38). This transcription-negative phenotype was attributed to the severe

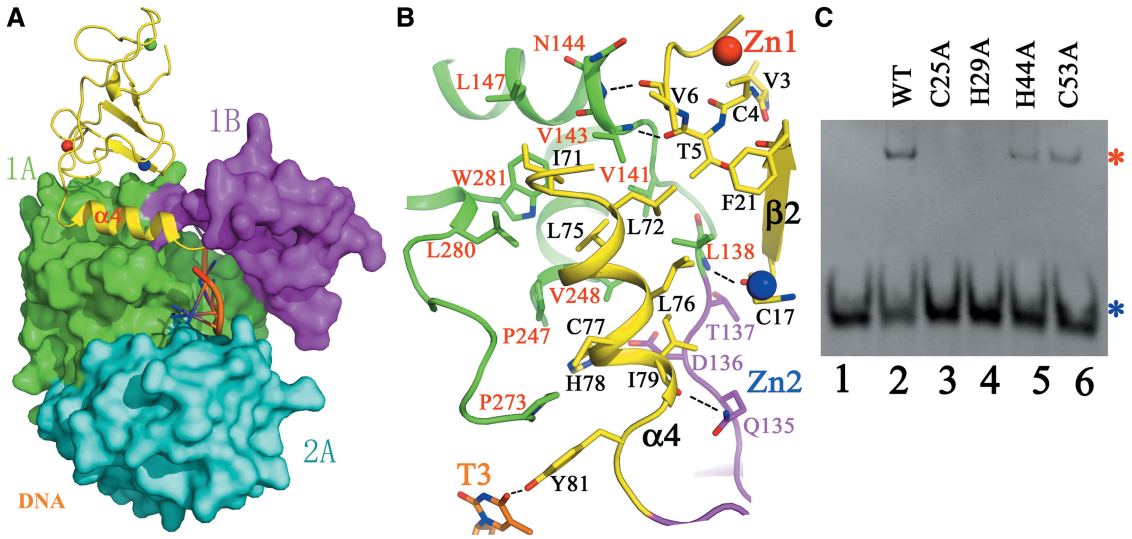

Figure 5. Inter-domain interactions of ZBD and HEL1 domains 1A and 1B. (**A**) Overview of the spatial orientation of the essential interaction helix α4 of ZBD. (**B**) Close-up view of the domain interface. Residues engaged in interactions are shown as sticks. Domain colours are the same as in Figure 2A. (**C**) DNA-binding assay of EAV nsp10Δ mutants with reduced Zn^{2+}-binding capabilities. Position of free DNA and protein–DNA complexes are indicated by blue and red asterisk, respectively.

structural constraints exerted by Pro residues on the local conformation of the proposed hinge region, as various substitutions of Ser59 alone (to Ala, Cys, Gly, His, Leu or Thr) yielded virus mutants with a wild-type phenotype, while combining the neutral S59G mutation with a P60G substitution reproduced the specific defect in sg mRNA synthesis (36). This interpretation is now further supported by the nsp10Δ structure in which Ser59 and Pro60 are located in the hinge connecting the treble-clef zinc finger and α4 of ZBD. The main chain of Ser59 forms three hydrogen bonds with the treble-clef Thr54, which is also connected to the Pro60 side chain and Lys61 main chain (Figure 4D). Owing to the unique properties of the Pro residue, the Ser59-to-Thr54 bonds are likely disrupted by the S59P mutation, but are not affected by the alternative replacements tested. Consequently, also owing to the main chain rigidity associated with the introduction of a Pro residue, the orientation of α4 relative to 1A and/or the main body of ZBD is likely affected in mutant S59P, which carries adjacent Pro residues at positions 59 and 60. Likewise, the introduction of two Gly residues at these positions [double mutant S59G/P60G; (36)] probably gives rise to excessive flexibility of the hinge region, thus compromising nsp10 function in a similar manner.

To further explore the role of ZBD, we tested the effect of four mutations (C25A and H29A in the RING-like module; H44A and C53A in the treble-clef zinc finger) expected to affect the ability to bind Zn1, Zn2 or Zn3, respectively. In agreement with the proposed structural role of these zinc ions, soluble His-tagged proteins containing these mutations could not be obtained and only low yields of GST-nsp10 fusion proteins carrying the same mutations could be recovered. For mutants C25A and H29A, band shift analysis revealed a complete loss of binding to a partially double-stranded DNA substrate containing a 5′ single-stranded poly(dT) overhang (substrate 5′DNA-T10; Figure 5C, lane 3–4). These results complement previous findings, showing a complete loss of both ATPase and helicase activity for these mutants (37). In contrast, the level of nucleic acid binding by mutants H44A and C53A was comparable with that of the wild-type protein (Figure 5C, lanes 5–6), consistent with nsp10-H44A retaining a limited level of ATPase and helicase activity (37). On further testing, we observed that the addition of 40 mM EDTA altered the overall conformation of nsp10Δ, as detected by changes in circular dichroism (Supplementary Figure S4A), and reduced its binding to 5′-DNA-A10 (Supplementary Figure S4B). In summary, these results reveal that ZBD interacts extensively with the HEL1 domain and that its integrity is an essential determinant of nsp10Δ properties in *in vitro* assays.

Structural resemblance between EAV nsp10Δ and mRNA decay factor Upf1

Next we analysed the existence of structural similarity between EAV nsp10Δ and other proteins by scanning a protein data bank using the DALI server (51). The structure of the nsp10Δ HEL1 domain was found to be most similar [*Z* score, 20.9; root-mean-square deviation (RMSD), 3.5 Å] to the helicase core of nonsense-mediated mRNA decay factor Upf1 and its homolog Ighmbp2 (*Z* score, 19.9; RMSD, 3.0 Å), which both belong to the Upf1-like helicase subfamily (11). Further comparisons revealed that this resemblance extends into the respective N-terminal ZBDs: the binuclear RING-like module of nsp10Δ ZBD was found to be most similar to RING-like module 1 in the CH-domain of Upf1 (Figure 4E). This similarity was rather limited (*Z*-score of 1.9 and RMSD of 2.2 Å) because only six out of the eight Zn-chelating residues in the two domains could be juxtaposed

(Figure 4F) and because loops L1, L3 and helix α1 in nsp10Δ are shorter than the corresponding elements in Upf1. We did not detect significant similarity of the treble-clef zinc finger with other proteins, although we note that the Upf1 CH-domain also has a zinc finger (but of a different fold) downstream of the RING1 module. Thus, EAV nsp10 ZBD prototypes a novel and complex multi-domain zinc finger with distinct structural properties. On the other hand, EAV nsp10 and Upf1 share a similar domain organization, including structurally similar RING and helicase domains. These similarities are further enhanced by the 5′-3′ directionality of duplex unwinding shared by both these helicases and likely extends to other nidovirus helicases in view of the observed sequence conservation (Supplementary Figure S3B).

Structure of EAV nsp10Δ in complex with a nucleic acid substrate

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

Nucleic acid binding induces profound conformational changes outside the HEL1 domain of nsp10Δ

The Cα atoms of domains 1A and 2A of free nsp10Δ and the nsp10Δ–DNA complex can be superimposed with an RMSD of 0.6 Å, indicating that the relative orientations of these core domains are barely affected by DNA binding (Figure 6A). However, outside these domains, the effect of DNA binding was considerable, with the RMSD between the Cα atoms of the two forms of nsp10Δ increasing to 1.8 Å. Particularly large conformational changes were observed in domain 1B, which rotates ∼28° towards ZBD in the nsp10Δ–DNA complex (Figure 6A). The RMSD between the Cα atoms of the two forms of domain 1B is 1.8 Å, with loop residues being affected most profoundly (Supplementary Figure S6A). Both width and height of the polynucleotide substrate channel formed by domains 1A and 1B (originally ∼5 and 11 Å, respectively) are increased by 2 Å on this rotation. This reorganization makes this channel large enough to accept single-stranded nucleic acids, although it remains too narrow for a nucleic acid duplex (Figure 6B). Consequently, double-stranded nucleic acids must be unwound at the entrance of the substrate channel to let a single-stranded chain enter. Besides this large conformational change, temperature factor calculations suggest that the regions at the surface of domain 1B not directly involved in DNA binding may become flexible (Supplementary Figure S2). For example, domain 1B residues Arg95, Gly125 and Ala131 become disordered after DNA binding (Figures 2C and 3B and Supplementary Figure S2).

On DNA binding, a structural change was also observed in the treble-clef zinc finger of ZBD, as reflected by its relatively high temperature factor (compared with that of domains 1A and 2A) in the nsp10Δ–DNA complex as opposed to nsp10Δ alone (Supplementary Figure S2).

Substrate recognition by EAV nsp10Δ is sequence-independent

As outlined above, the single-stranded part of the DNA substrate is bound to a nucleic acid-binding channel formed by domains 1A, 1B and 2A (Figure 2C). The backbone phosphates of the poly(dT) are located on top of domains 1A and 2A, with the thymine bases exposed to the solvent (Supplementary Figure S7A). The majority of contacts with the bound DNA are made via the phosphodiester backbone and non-specific protein–base interactions as depicted in Figure 7. Consistent with this observation, the base orientation varies in the four EAV nsp10–poly(dT) complexes of the asymmetric unit, while the position of the DNA backbone is rather rigid (Supplementary Figure S7B and C). Several key residues from domains 1A and 2A contact the DNA backbone in the channel of the protein (Figure 7A and B). Base T1, the most 5′ one, is exposed to the solvent and protrudes outwards, causing a bend in the DNA backbone between T1 and T2. The bases T2 and T3 as well as T5 and T6 stack with each other at an average distance of 3.7 Å. In contrast, base T4 is almost perpendicular to T3, with its edge exposed to protein side chains that make specific contacts. Val271 in domain 1A forms van der Waals contacts with the base and the sugar ring of T4 and thus stabilizes the DNA conformation. Moreover, the binding is stabilized by several hydrogen bonds between His186, His339, Thr348, Ser351 and the backbone of the DNA, and by van der Waals contacts between Thr185, Leu227, Val230, Tyr338 and the phosphate groups of the DNA. While the interactions described above do not involve specific bases, six further interactions specific for thymine were found. For example, the backbone NH of

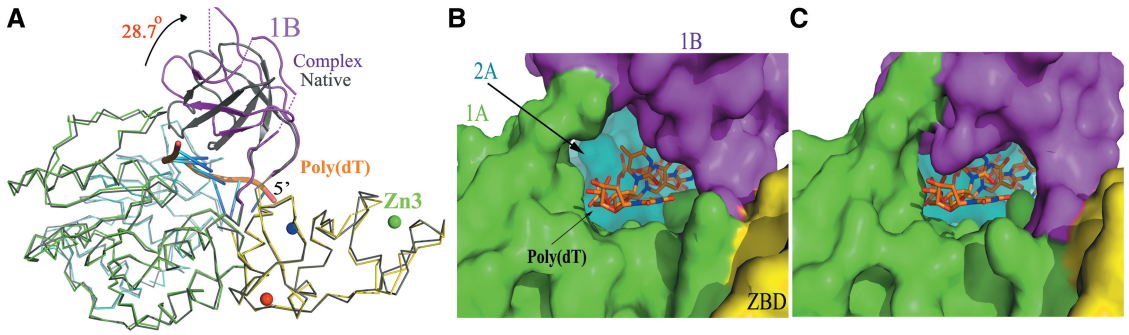

Figure 6. Conformational changes of EAV nsp10Δ on nucleic acid binding. (**A**) Comparison of the free and DNA-bound states of nsp10Δ. The arrow indicates the movement of domain 1B (cartoon) in the DNA-bound state compared with the free state. (**B**) Surface model of the channel formed by domains 1A and 1B in the DNA-bound state and (**C**) the DNA-free state. Domain colours are the same as used in Figure 2A. Note that the DNA in Figure 5C was extracted from the complex structure of DNA-bound state.

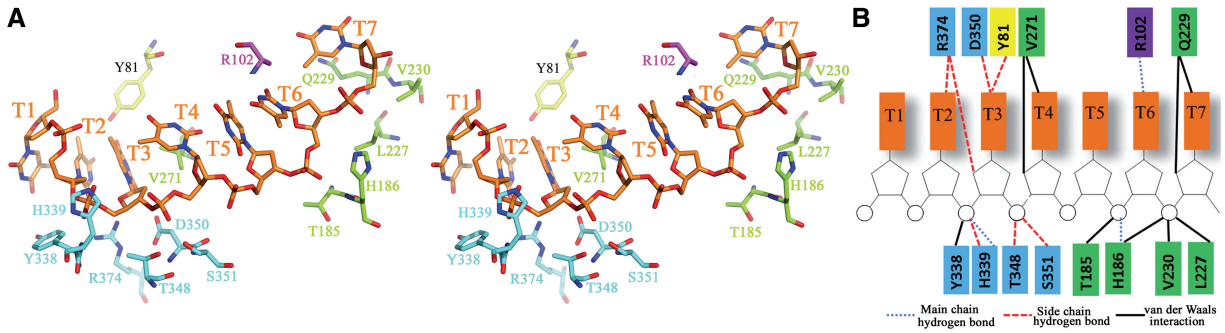

Figure 7. Interactions between EAV nsp10Δ and a DNA substrate. (**A**) Stereo view of the nucleic acid-binding pocket of nsp10Δ. Bound single-stranded DNA and the interacting residues are shown as sticks. Nucleotides are numbered (T1 to T7) in the 5′-3′ direction and are shown in orange. Residues are coloured according to their domain origin as indicated in Figure 2A. (**B**) Schematic representation of the contacts between nsp10Δ residues and DNA.

Arg102 forms a hydrogen bond with the O4 atom of T6. The O2 and O4 atoms of base T3 form hydrogen bonds with the side chains of Asp350 and Tyr81. Also, several residues, such as Arg374 and Gln229, interact with both the base and the sugar ring. However, no interaction was observed between nsp10Δ and position C2′ of the ribose ring of the DNA substrate. This observation may explain why EAV nsp10 has the ability to unwind both DNA and RNA, in agreement with the substrate specificity observed for other helicases (52,53) possessing or lacking the ability to interact with the 2′ OH moiety of the RNA backbone.

Disscusion

Among +RNA viruses, whose RdRps generally have a high error rate, nidoviruses stand out for their large to very large genome size (13–32 kb). Consequently, the replication fidelity of nidoviruses, in particular coronaviruses, has been the subject of intense study. Most recently, the identification of a unique 3′-to-5′ exoribonuclease (ExoN) activity has provided the basis for the hypothesis that a primitive proofreading mechanism operates to promote the fidelity of RNA-dependent RNA synthesis in nidoviruses with >20 kb genomes (21–27).

Despite this recent progress, the two central subunits of the nidovirus replicase, the RdRp and the unique ZBD-containing RNA helicase, have remained poorly characterized, also due to the lack of structural information. Remarkably, our present analysis of the arterivirus helicase structure revealed a number of important similarities with Upf1 helicases, eukaryotic enzymes involved in quality control of RNAs through multiple pathways, including nonsense-mediated mRNA decay (54–56). In contrast to the ExoN-driven control of replication fidelity (see above), the possibility of post-transcriptional quality control of nidovirus mRNAs has not been considered thus far. Yet, replicase ORF1ab is extremely large (from 3175 to >7000 codons) and its correct expression by translation of the viral genome is a critical first step in the production of the enzymes directing genome replication and expression. Therefore, our study not only provides the first insights into the structural basis for nidovirus RNA helicase function, but also creates a basis to propose a role for this protein in the post-transcriptional quality control of viral mRNAs. This role may be common to all nidoviruses, regardless of their genomes size, which would distinguish it from the ExoN-based proofreading mechanism that appears to be restricted to nidoviruses with a >20 kb genome. On the time scale of nidovirus evolution, the acquisition of

ZBD-HEL1 may have been a critical event to facilitate the genome expansion of ancestral small-sized nidoviruses, thus setting the stage for the subsequent ExoN-driven expansion towards even larger nidovirus genomes (19,57).

EAV nsp10 represents a multi-domain helicase conserved in nidoviruses

Previously, using bioinformatics, biochemistry and molecular genetics, it was established that nsp10 of arteriviruses and its orthologs in other nidoviruses are multi-domain proteins. Of its domains, ZBD and the HEL1 domains are critical for the enzyme's ATPase and helicase activities *in vitro* and for the regulation of viral replication and transcription in infected cells. Our structural and biochemical studies extended the characterization of known domains and delineated two hitherto uncharacterized domains: one (domain 1B) flanked by ZBD and HEL1, and the other (C-terminal domain) located downstream of the HEL1, with its structure remaining to be solved. Our data show that, along with ZBD, these two non-enzymatic domains may regulate HEL1 function. Given that nsp10/nsp13 is one of only three proteins whose nidovirus-wide conservation can be detected at the sequence level (19,24,25,28), the nsp10Δ structure should be applicable to other nidovirus helicases, including those of PRRS viruses and coronaviruses. However, considerable size differences exist between arteri- and coronaviruses in the most conserved ZBD and HEL1 domains, whereas the 1B and C-terminal domains lack appreciable sequence conservation. Thus, helicase structures from other small- and large-genome nidoviruses will be required to fully understand the enzyme's function.

The nsp10 C-terminal domain: coupling ATPase and helicase activities?

While attempting to solve the EAV nsp10 structure, we were confronted with the low stability of the full-length recombinant protein expressed in *E. coli*. We solved this problem by characterizing the C-terminally truncated nsp10Δ, which lacks the 65 residues (C-terminal domain) downstream of the known HEL1 motifs. This protein was found to bind partially double-stranded DNA and display the previously reported *in vitro* ATPase and helicase activities. Because, compared with full-length nsp10, nsp10Δ appeared to be somewhat more active as an ATPase but somewhat less active as a helicase, the C-terminal truncation may have affected the coupling of these two enzymatic activities. This suggests that the C-terminal domain may have evolved to (co)regulate nsp10 helicase-mediated functions *in vivo*, implying that it must be able to communicate with the nsp10 active site. This could be achieved either directly, by interacting with the nucleic acid- or ATP-binding site (the nsp10Δ C-terminus is ~22.5 Å apart of the active centre; Figure 2C), or indirectly, through a protein signal transduction network. Importantly, the C-terminal domain is poorly conserved among arteri- and coronaviruses in terms of both sequence and size (Supplementary Figure S3A, and data not shown), arguing that such a putative regulatory function could be executed in a virus- and, possibly, host-specific manner.

The nsp10 structure: defining a complex ZBD

Our characterization of the EAV nsp10 structure verified and revised a model of the N-terminal ZBD based on prior studies (36,37,58). It resolved the uncertainty about the number of zinc ions bound (now established to be three) and the fold of this domain (a unique structure combining a RING-like module fused with a treble-clef zinc finger). Furthermore, it redefined the C-terminal border of ZBD and placed it 13 residues downstream to include a third hitherto unrecognized structural element (helix α4). Previously, we analysed a variety of EAV nsp10 ZBD mutants in which putative zinc-binding residues were replaced in a manner (Cys→His or His→Cys) that could preserve zinc binding (36,37). From the solved structure, it is now apparent that the replication-negative phenotypes of these virus mutants can likely be attributed to the detrimental impact of the respective mutations on ZBD integrity and, through the extensive interaction network, HEL1 domains. It presently remains unclear why the replacement of His44 by Cys in the treble-clef zinc finger was partially tolerated. On the other hand, structural superposition of the RING-like modules of nsp10 and hUpf1 (Figure 4E) reveals how the only other similarly tolerated replacement (36,37), that of the Zn1-coordinating Cys25 by His (found in the equivalent position in hUpf1), could be accommodated by nsp10. The RING-like module 1 of Upf1 also shares structural similarity with RING-box domains of E3 ubiquitin ligases (59) and the involvement of this module in self-ubiquitination of Upf1 was indeed demonstrated (60). It would be interesting to see whether these results are relevant for nsp10 and its ZBD. Recently, arterivirus papain-like protease 2 was found to have deubiquitinase activity, which suppresses the innate immune response in infected host cells (61,62).

The nsp10-nucleic acid complex: towards the dsRNA unwinding mechanism

To understand how nsp10 unwinds its natural dsRNA substrates, we analysed a complex of nsp10Δ with a partially double-stranded DNA substrate. Only seven thymidine residues could be confirmed in the structure of that complex (Figure 2C). The DNA-bound nsp10Δ structure revealed two possible RNA-binding clefts at the surface of nsp10, which are formed by domains 1B and ZBD (named putative exit site 1), and 1A and ZBD (putative exit site 2), respectively (Supplementary Figure S8). Both have continuous positively charged surfaces, with the latter (Supplementary Figure S8, right panel) being sufficiently large to bind a ssRNA >10 bp, which could be especially suited for unwinding complex secondary structures. This organization suggests that, after unwinding, one of the separated RNA chains would be guided through the narrow nucleic acid substrate tunnel formed by domains 1A, 1B and 2A, while the path of the other strand remains to be defined. No matter which cleft is actually used for

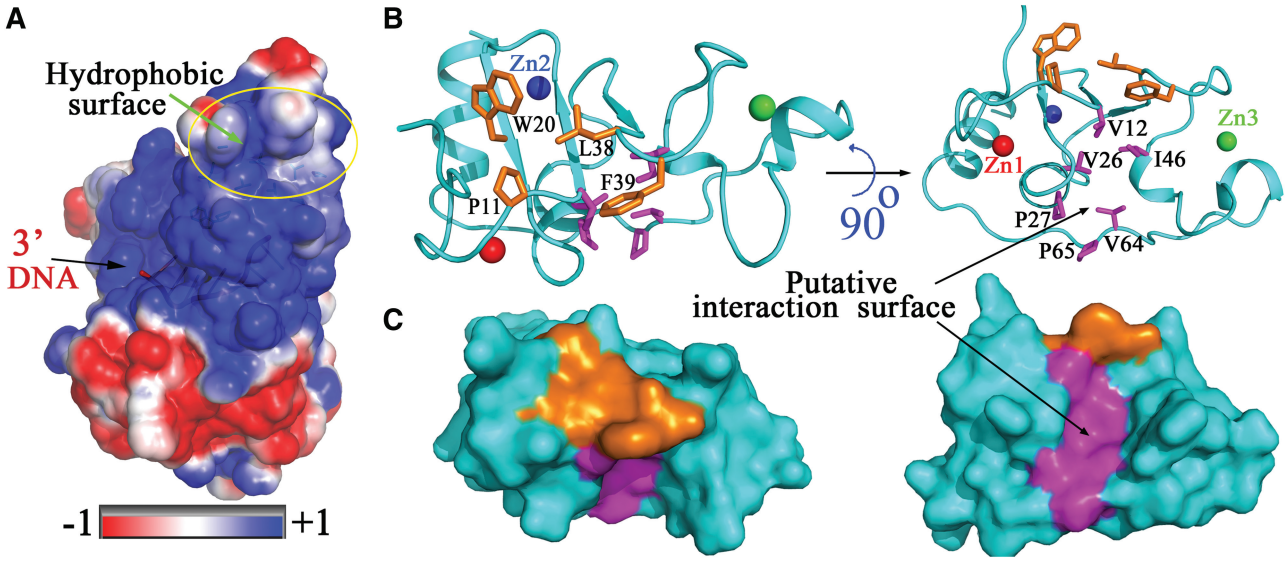

Figure 8. Putative protein interaction surfaces of the EAV nsp10Δ ZBD. (**A**) Overview of surface charges of nsp10Δ. The putative protein interaction surfaces are indicated by a yellow circle. (**B**) Close-up view of zone 1 (orange) and zone 2 (magenta). Hydrophobic residues are shown as sticks. (**C**) Surface representation of the two putative interaction zones. The orientation is the same as in panel B.

RNA binding, the positively charged ZBD, and especially its RING-like module, would be involved.

Like the protein-binding surface of the Upf1 C/H domain (54), ZBD has a putative protein interaction surface composed of two major hydrophobic zones that are almost perpendicular to each other (Figure 8). Nucleic acid binding induced a conformational change (Supplementary Figure S6B) of these two zones. In addition, the temperature factor of the treble-clef zinc finger was higher and several residues are disordered in the structure of the nsp10Δ-DNA complex (Supplementary Figure S2). Together, these findings imply that these two zones are readily accessible for interactions with other proteins, which may further influence nucleic acid binding.

Substrate binding by nsp10 is accompanied by structural changes in domain 1B and the treble-clef zinc finger, which may be recognized by yet-to-be identified interaction partners modulating nsp10 function. The treble-clef finger is fairly distant from the bound substrate, suggesting long-distance signal transduction within nsp10, possibly involving helix α4, which interacts with 1A, 1B and nucleic acid and is directly connected to the treble-clef zinc finger. The flexibility of the hinge region connecting the treble-clef zinc finger and helix α4 is likely compromised by the previously described S59P and S59G/P60G mutations that, importantly, were found to impair viral sg mRNA synthesis but not genome replication (36). Consequently, the described inter-domain communication channel may be used by nsp10 and its partners for switching from a role in genome replication to directing viral transcription, a hypothesis that will be the subject of future studies.

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

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

If the nidovirus helicase possesses some of the properties of Upf1, this could explain the exclusive conservation of ZBD in nidoviruses, which stand out for their large to very large single-stranded RNA genomes. For instance, by providing post-transcriptional quality control of genomic RNA, i.e. detection of nonsense and/or other mutations and elimination of defective molecules, the nidovirus helicase could alleviate the consequences of the generally low fidelity of RNA virus genome replication. Such a role of ZBD-HEL1 may have protected an ancestral nidovirus from the mutational meltdown of its expanding genome, similar to the proposed fixation of the proofreading ExoN domain at a later stage of nidovirus evolution (19,24,25,28). Subsequently, the enzyme would have facilitated expansion to the genome size observed in contemporary arteriviruses, and remained a critical factor in the further ExoN-driven genome expansion to evolve middle- and large-sized nidoviruses. Thus, the proposed Upf1-like role of the nidovirus helicase can be accommodated in a meaningful evolutionary scenario incorporating several of the structural and functional observations made in this study. The structural similarity between nsp10 and Upf1 establishes a new connection between research on viral and cellular helicases, which

could be mutually insightful for understanding the evolution and function of this group of vitally important enzymes.

ACCESSION NUMBERS

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

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

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

FUNDING

National Basic Research Program of China [973 Program, 2011CB965304 and 2009CB825501]; National Natural Science Foundation of China [31370720, 31222032, 90919043 and 31070664]; Chinese Universities Scientific Fund [2013QJ027]; Specialized Research Fund for the Doctoral Program of Higher Education [20100008 110009]; Open Research Fund Program of the State Key Laboratory of Virology of China [2013IOV003]; National Laboratory of Medical Molecular Biology (PUMC) to Zhongzhou Chen; European Union Seventh Framework program through the EUVIRNA project [264286]; the Netherlands Organization for Scientific Research [NWO; TOP-GO 700.10.352]; the Collaborative Agreement on Bioinformatics between Leiden University Medical Center and Moscow State University (MoBiLe); Leiden University Fund. Funding for open access charge: [2011CB965304] and Leiden University Medical Center.

Conflict of interest statement. None declared.

REFERENCES

1. Linder,P. and Jankowsky,E. (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.*, **12**, 505–516.
2. Abdelhaleem,M. (2010) Helicases: an overview. *Methods Mol. Biol.*, **587**, 1–12.
3. Gorbalenya,A.E. and Koonin,E.V. (1989) Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.*, **17**, 8413–8440.
4. Kadare,G. and Haenni,A.L. (1997) Virus-encoded RNA helicases. *J. Virol.*, **71**, 2583–2590.
5. Steimer,L. and Klostermeier,D. (2012) RNA helicases in infection and disease. *RNA Biol.*, **9**, 751–771.
6. Rajagopal,V. and Patel,S.S. (2009) Viral helicases. In: Raney,K.D., Gotte,M. and Cameron,C.E. (eds), *Viral Genome Replication*. Springer, NY, pp. 429–466.
7. Singleton,M.R., Dillingham,M.S. and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
8. Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, **3**, 419–429.
9. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1988) A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett.*, **235**, 16–24.
10. Hodgman,T.C. (1988) A new superfamily of replicative proteins. *Nature*, **333**, 22–23.
11. Fairman-Williams,M.E., Guenther,U.P. and Jankowsky,E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
12. Strauss,J.H. and Strauss,E.G. (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.*, **58**, 491–562.
13. de Groot,R.J., Cowley,J.A., Enjuanes,L., Faaberg,K.S., Perlman,S., Rottier,P.J.M., Snijder,E.J., Ziebuhr,J. and Gorbalenya,A.E. (2012) Order Nidovirales. In: King,A.M.Q., Adams,M.J., Carstens,E.B. and Lefkowitz,E.J. (eds), *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, Amsterdam, pp. 785–795.
14. Wang,Q., Han,Y., Qiu,Y., Zhang,S., Tang,F., Wang,Y., Zhang,J., Hu,Y. and Zhou,X. (2012) Identification and characterization of RNA duplex unwinding and ATPase activities of an alphatetravirus superfamily 1 helicase. *Virology*, **433**, 440–448.
15. Nishikiori,M., Sugiyama,S., Xiang,H., Niiyama,M., Ishibashi,K., Inoue,T., Ishikawa,M., Matsumura,H. and Katoh,E. (2012) Crystal structure of the superfamily 1 helicase from Tomato mosaic virus. *J. Virol.*, **86**, 7565–7576.
16. Chand,R.J., Trible,B.R. and Rowland,R.R. (2012) Pathogenesis of porcine reproductive and respiratory syndrome virus. *Curr. Opin. Virol.*, **2**, 256–263.
17. Peiris,J.S., Guan,Y. and Yuen,K.Y. (2004) Severe acute respiratory syndrome. *Nat. Med.*, **10**, S88–S97.
18. de Groot,R.J., Baker,S.C., Baric,R.S., Brown,C.S., Drosten,C., Enjuanes,L., Fouchier,R.A., Galiano,M., Gorbalenya,A.E., Memish,Z.A. *et al.* (2013) Middle east respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus study group. *J. Virol.*, **87**, 7790–7792.
19. Gorbalenya,A.E., Enjuanes,L., Ziebuhr,J. and Snijder,E.J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res.*, **117**, 17–37.
20. Perlman,S., Gallagher,T. and Snijder,E.J. (2008) *Nidoviruses*. ASM Press, Washington, DC.
21. Denison,M.R., Graham,R.L., Donaldson,E.F., Eckerle,L.D. and Baric,R.S. (2011) Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.*, **8**, 270–279.
22. Minskaia,E., Hertzig,T., Gorbalenya,A.E., Campanacci,V., Cambillau,C., Canard,B. and Ziebuhr,J. (2006) Discovery of an RNA virus 3′→5′ exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl Acad. Sci. USA*, **103**, 5108–5113.
23. Bouvet,M., Imbert,I., Subissi,L., Gluais,L., Canard,B. and Decroly,E. (2012) RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl Acad. Sci. USA*, **109**, 9372–9377.
24. Snijder,E.J., Bredenbeek,P.J., Dobbe,J.C., Thiel,V., Ziebuhr,J., Poon,L.L., Guan,Y., Rozanov,M., Spaan,W.J. and Gorbalenya,A.E. (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.*, **331**, 991–1004.
25. Nga,P.T., Parquet Mdel,C., Lauber,C., Parida,M., Nabeshima,T., Yu,F., Thuy,N.T., Inoue,S., Ito,T., Okamoto,K. *et al.* (2011) Discovery of the first insect nidovirus, a missing evolutionary link

in the emergence of the largest RNA virus genomes. *PLoS Pathog.*, **7**, e1002215.
26. Eckerle,L.D., Becker,M.M., Halpin,R.A., Li,K., Venter,E., Lu,X., Scherbakova,S., Graham,R.L., Baric,R.S., Stockwell,T.B. *et al.* (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.*, **6**, e1000896.
27. Smith,E.C., Blanc,H., Vignuzzi,M. and Denison,M.R. (2013) Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.*, **9**, e1003565.
28. Ziebuhr,J. (2005) The coronavirus replicase. *Curr. Top. Microbiol. Immunol.*, **287**, 57–94.
29. Brierley,I., Digard,P. and Inglis,S.C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell*, **57**, 537–547.
30. Ziebuhr,J., Snijder,E.J. and Gorbalenya,A.E. (2000) Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.*, **81**, 853–879.
31. Pasternak,A.O., Spaan,W.J.M. and Snijder,E.J. (2006) Nidovirus transcription: how to make sense...? *J. Gen. Virol.*, **87**, 1403–1421.
32. Sawicki,S.G., Sawicki,D.L. and Siddell,S.G. (2007) A contemporary view of coronavirus transcription. *J. Virol.*, **81**, 20–29.
33. Sola,I., Mateos-Gomez,P.A., Almazan,F., Zuniga,S. and Enjuanes,L. (2011) RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol.*, **8**, 237–248.
34. Seybert,A., Hegyi,A., Siddell,S.G. and Ziebuhr,J. (2000) The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5′-to-3′ polarity. *RNA*, **6**, 1056–1068.
35. Seybert,A., van Dinten,L.C., Snijder,E.J. and Ziebuhr,J. (2000) Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. *J. Virol.*, **74**, 9586–9593.
36. van Dinten,L.C., van Tol,H., Gorbalenya,A.E. and Snijder,E.J. (2000) The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis. *J. Virol.*, **74**, 5213–5223.
37. Seybert,A., Posthuma,C.C., van Dinten,L.C., Snijder,E.J., Gorbalenya,A.E. and Ziebuhr,J. (2005) A complex zinc finger controls the enzymatic activities of nidovirus helicases. *J. Virol.*, **79**, 696–704.
38. van Dinten,L.C., den Boon,J.A., Wassenaar,A.L., Spaan,W.J. and Snijder,E.J. (1997) An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc. Natl Acad. Sci. USA*, **94**, 991–996.
39. Keum,Y.S. and Jeong,Y.J. (2012) Development of chemical inhibitors of the SARS coronavirus: viral helicase as a potential target. *Biochem. Pharmacol.*, **84**, 1351–1358.
40. Isken,O. and Maquat,L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.*, **21**, 1833–1856.
41. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol.*, **276**, 307–326.
42. Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 849–861.
43. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
44. Murshudov,G.N., Skubak,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 355–367.
45. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 439–450.
46. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 22–25.
47. Schrödinger,L. (2010) *The PyMOL Molecular Graphics System, Version 1.3r1.*
48. Saikrishnan,K., Powell,B., Cook,N.J., Webb,M.R. and Wigley,D.B. (2009) Mechanistic basis of 5′-3′ translocation in SF1B helicases. *Cell*, **137**, 849–859.
49. Pugalenthi,G., Archunan,G. and Sowdhamini,R. (2005) DIAL: a web-based server for the automatic identification of structural domains in proteins. *Nucleic Acids Res.*, **33**, W130–W132.
50. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
51. Holm,L. and Rosenstrom,P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
52. Appleby,T.C., Anderson,R., Fedorova,O., Pyle,A.M., Wang,R., Liu,X., Brendza,K.M. and Somoza,J.R. (2011) Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV. *J. Mol. Biol.*, **405**, 1139–1153.
53. Lim,S.C., Bowler,M.W., Lai,T.F. and Song,H. (2012) The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1. *Nucleic Acids Res.*, **40**, 11009–11022.
54. Clerici,M., Mourao,A., Gutsche,I., Gehring,N.H., Hentze,M.W., Kulozik,A., Kadlec,J., Sattler,M. and Cusack,S. (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.*, **28**, 2293–2306.
55. Lim,S.C., Bowler,M.W., Lai,T.F. and Song,H. (2012) The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1. *Nucleic Acids Res.*, **40**, 11009–11022.
56. Cheng,Z., Muhlrad,D., Lim,M.K., Parker,R. and Song,H. (2007) Structural and functional insights into the human Upf1 helicase core. *EMBO J.*, **26**, 253–264.
57. Lauber,C., Goeman,J.J., Parquet,M.D.C., Nga,P.T., Snijder,E.J., Morita,K. and Gorbalenya,A.E. (2013) The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathog.*, **9**, e1003500.
58. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1989) Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.*, **17**, 4847–4861.
59. Kadlec,J., Guilligay,D., Ravelli,R.B. and Cusack,S. (2006) Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA*, **12**, 1817–1824.
60. Takahashi,S., Araki,Y., Ohya,Y., Sakuno,T., Hoshino,S.I., Kontani,K., Nishina,H. and Katada,T. (2008) Upf1 potentially serves as a RING-related E3 ubiquitin ligase via its association with Upf3 in yeast. *RNA*, **14**, 1950–1958.
61. van Kasteren,P.B., Bailey-Elkin,B.A., James,T.W., Ninaber,D.K., Beugeling,C., Khajehpour,M., Snijder,E.J., Mark,B.L. and Kikkert,M. (2013) Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells. *Proc. Natl Acad. Sci. USA*, **110**, E838–E847.
62. Frias-Staheli,N., Giannakopoulos,N.V., Kikkert,M., Taylor,S.L., Bridgen,A., Paragas,J., Richt,J.A., Rowland,R.R., Schmaljohn,C.S., Lenschow,D.J. *et al.* (2007) Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe*, **2**, 404–416.
63. Chakrabarti,S., Jayachandran,U., Bonneau,F., Fiorini,F., Basquin,C., Domcke,S., Le Hir,H. and Conti,E. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell*, **41**, 693–703.