

Crystal Structures of Two Coronavirus ADP-Ribose-1"-Monophosphatases and Their Complexes with ADP-Ribose: a Systematic Structural Analysis of the Viral ADRP Domain[∇]

Yuanyuan Xu,¹ Le Cong,¹ Cheng Chen,¹ Lei Wei,¹ Qi Zhao,¹ Xiaoling Xu,¹
Yanlin Ma,² Mark Bartlam,³ and Zihe Rao^{1,2,3*}

Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China¹; National Laboratory of Biomacromolecules, Institute of Biophysics (IBP), Chinese Academy of Sciences, Beijing 100101, China²; and College of Life Sciences and Tianjin State Laboratory of Protein Science, Nankai University, Tianjin 300071, China³

Received 4 September 2008/Accepted 22 October 2008

The coronaviruses are a large family of plus-strand RNA viruses that cause a wide variety of diseases both in humans and in other organisms. The coronaviruses are composed of three main lineages and have a complex organization of nonstructural proteins (nsp's). In the coronavirus, nsp3 resides a domain with the macroH2A-like fold and ADP-ribose-1"-monophosphatase (ADRP) activity, which is proposed to play a regulatory role in the replication process. However, the significance of this domain for the coronaviruses is still poorly understood due to the lack of structural information from different lineages. We have determined the crystal structures of two viral ADRP domains, from the group I human coronavirus 229E and the group III avian infectious bronchitis virus, as well as their respective complexes with ADP-ribose. The structures were individually solved to elucidate the structural similarities and differences of the ADRP domains among various coronavirus species. The active-site residues responsible for mediating ADRP activity were found to be highly conserved in terms of both sequence alignment and structural superposition, whereas the substrate binding pocket exhibited variations in structure but not in sequence. Together with data from a previous analysis of the ADRP domain from the group II severe acute respiratory syndrome coronavirus and from other related functional studies of ADRP domains, a systematic structural analysis of the coronavirus ADRP domains was realized for the first time to provide a structural basis for the function of this domain in the coronavirus replication process.

The coronaviruses are positive-strand RNA viruses with the largest known genome sizes and the most complex replication mechanisms. After generations of evolution, the coronaviruses that have been characterized to date produce a striking number of virus-encoded nonstructural proteins (nsp's) which assemble into a large membrane-bound complex to perform the rapid viral replication process (23, 30, 35, 46). Current understanding of the coronavirus genome suggests that a single large replicase gene encodes all the proteins involved in the process. This gene contains two open reading frames (ORFs) (designated ORF1a and ORF1b) and is transcribed into two polyproteins, pp1a (from ORF1a) and pp1ab (from ORF1a and ORF1b) (46). The synthesis of the ORF1b-encoded part in the latter polyprotein requires a -1 ribosomal frameshift upon translation of the viral mRNA (8, 9). In order to produce functional nsp's, the two polyproteins are cleaved by two virus-encoded proteases, the main protease (M^{Pro} or $3CL^{Pro}$) and the papain-like protease (PL^{Pro}), to produce up to 16 nsp's (nsp1 to nsp16), the final product of this intricate process (46, 48). Among these nsp's, nsp3 is the largest and possesses a variety of putative domains that are conserved among coronaviruses. These domains have been shown to harbor diverse

enzymatic activities, including a domain with ADP-ribose-1"-monophosphatase (ADRP) activity (14, 37, 46, 47). As structural and functional evidence accumulates, it would appear that the enzymatic activities harbored by the viral nsp's are essential for the coronavirus to achieve its highly coordinated replication process (4, 5, 7, 17, 19, 20, 41, 42, 45).

The ADRP domain of nsp3 is proposed to belong to the macroH2A-like family, which is characterized by the possession of a structural module called the "macro domain" with high-affinity ADP-ribose (and, in some cases, poly-ADP-ribose [PAR]) binding (21). The macroH2A-like family is named after the nonhistone macro domain of the histone macroH2A, a prototype of this family (28). Noticeably, their recognition of ADP-ribose and its derivative in animal cells has been demonstrated to be associated with many key physiological processes including ADP ribosylation, an important posttranslational protein modification involved in DNA damage repair, transcription regulation, chromatin remodeling, and so on (1, 21, 33). The coronaviruses characterized to date all possess the ADRP domain as part of nsp3, yet very few other viruses are known to contain this module. Only rubella virus, alphaviruses, and hepatitis E virus have been shown to possess an ADRP domain to date (37). Given the ubiquity and functional significance of the macroH2A-like family of proteins, it would seem that viral ADRP domains may play an essential role in the replication of coronavirus or other viruses containing such a module. How this domain is involved in the complicated viral replication process or why it exists exclusively in such a limited

* Corresponding author. Mailing address: Laboratory of Structural Biology, Life Sciences Building, Tsinghua University, Beijing 100084, China. Phone: 86 10 62771493. Fax: 86 10 62773145. E-mail: raozh@xtal.tsinghua.edu.cn.

[∇] Published ahead of print on 5 November 2008.

range of virus families remains unclear. Until now, there has been no clear evidence to suggest any specific interactions between the viral ADRP domains and biological pathways in the host cells. Moreover, a reverse genetics study recently revealed that mutations in the active site of the viral ADRP domain resulted in no significant effects on virus replication when viral transcription levels were assayed in cell culture. Hence, it has been suggested that this domain may be involved in the regulation of viral replication rather than in the process itself (31).

In yeast (*Saccharomyces cerevisiae*) and plant cells, proteins with the macroH2A-like fold have been shown to involve in the tRNA splicing pathway by acting as an ADRP (22, 25, 36). Further studies from both structural and functional perspectives have confirmed that the ADRP domains in coronaviruses, including severe acute respiratory syndrome coronavirus (SARS-CoV), human coronavirus 229E (HCoV-229E), and transmissible gastroenteritis virus, also possess this enzymatic activity with high specificity. Although this may point toward a potential function of viral ADRP domains in regulating the metabolism of ADP-ribose derivatives, the poor turnover numbers in enzymatic assays (from 5 to 20 min⁻¹ for the three positive-strand RNA viruses reported) indicate an insufficiency in metabolite processing and argue against this hypothesis (12, 25, 31, 32, 34, 37). Another possibility is that viral ADRP domains could serve as PAR-recognizing modules and may interact with host proteins to regulate cellular responses to viral infection. Such processes may include a counteraction of apoptosis-signaling pathways induced by viral entry and the subsequent transcription of the viral RNA genome (16). In support of this hypothesis, a recent structural and functional study on the SARS-CoV ADRP domain demonstrated the mechanism of substrate binding and showed that viral ADRP domains have a high affinity for PAR (12). However, the question of how and why coronaviruses uniquely evolved this domain as part of their replication complex remains a mystery. Thus far, no studies have been conducted that could provide a comprehensive understanding of the significance of the conserved sequence of the ADRP domains among coronavirus and how this conservation is related to their three-dimensional structural features and corresponding functions in the viral replication process.

Here we report the crystal structures of two coronavirus nsp3 ADRP domains from avian infectious bronchitis virus (IBV) and HCoV-229E to 1.8-Å and 2.1-Å resolutions, respectively, along with those of their corresponding ADP-ribose complexes. These structures reveal a novel dimerization state in IBV, and, more significantly, observable variations in the structural organization of the substrate binding pocket, despite their conserved amino acid sequence. This is the first structure-based comparison of viral ADRP domains involving three distinct structures, from HCoV-229E, SARS-CoV, and IBV, which are related to each of the three main coronavirus lineages currently identified (38). Subsequent analysis of the structural and functional differences of viral ADRP domains found in the three coronavirus groups demonstrates a highly conserved active site among the coronavirus ADRP domains, from both sequence and structural perspectives. Thus, our work provides the first systematic study of how these highly conserved amino acid sequences translated into three-dimen-

sional structural features that direct the function of this domain in the coronavirus life cycle. Collectively, these results could provide insights into the potential role of the viral ADRP domain in the coronavirus replication process and host-virus interaction and in the evolution of coronavirus nsp3's. Additionally, our study may shed new light on the structurally based design of new antiviral drugs targeting the active site harbored in viral ADRP domains, an approach which has been demonstrated in previous reports concerning coronavirus main protease (42–44).

MATERIALS AND METHODS

Protein expression and purification. The sequences encoding the nsp3 ADRP domains from IBV (isolate M41, residues 1005 to 1178 of the polyprotein) and HCoV-229E (residues 1269 to 1436 of the polyprotein) were cloned from virus cDNA libraries by PCR. The two sequences were both inserted between the BamHI and XhoI sites of the pGEX-6p-1 plasmid (GE Healthcare). The forward and reverse PCR primers used for amplification were IBV-nsp3-ADRP-F (5'-CGGGATCCGTTAAACCAAGCTACATGTGA-3'), IBV-nsp3-ADRP-R (5'-CCGCTCGATTACTTACAAGTTGCATCGAAAT-3'), 229E-nsp3-ADRP-F (5'-CGCGGATCCAAAGAGAAGTTGACGCCT-3'), and 229E-nsp3-ADRP-R (5'-CCGCTCGAGTTACTACTAAACCAGACACAA-3'). The resulting plasmids with the two inserted sequences were transformed into *Escherichia coli* BL21(DE3) cells as glutathione *S*-transferase (GST) fusion proteins IBV-nsp3-ADRP-GST and 229E-nsp3-ADRP-GST and purified using glutathione affinity chromatography. The GST tag was removed by PreScission protease (GE Healthcare), leading to five additional residues (GPLGS) at the N terminus for both proteins. The proteins were further purified by cation-exchange chromatography using a Resource S column (GE Healthcare) with elution buffer containing 20 mM MES (morpholineethanesulfonic acid) (pH 6.0), 1 M NaCl and by size exclusion chromatography using a Superdex 75 column (GE Healthcare) in 20 mM MES (pH 6.0), 150 mM NaCl. The protein was finally concentrated to 25 mg · ml⁻¹ before crystallization.

Protein crystallization. The nsp3 ADRP domains from IBV and HCoV-229E were both crystallized by the hanging-drop vapor diffusion method at 291 K. A 1-μl drop of protein was mixed with 1 μl of reservoir solution, and the mixture was allowed to reach equilibrium over 400 μl of reservoir solution. For the IBV ADRP domain, optimum crystals with a cuboid shape were obtained using a reservoir solution containing 0.12 M magnesium chloride hexahydrate, 0.1 M HEPES, pH 7.5, and 22% (wt/vol) polyethylene glycol 3350. In the case of the HCoV-229E ADRP domain, the optimum conditions for the protein crystallization were obtained with a reservoir solution containing 0.1 M HEPES, pH 7.5, and 25% (wt/vol) polyethylene glycol 3350.

Diffraction data collection and processing. Prior to data collection, crystals were transferred to a solution containing 20% (wt/vol) polyethylene glycol 6000 and treated briefly for cryoprotection. A data set for the native nsp3 ADRP domain from IBV was collected in-house at 100 K using a Rigaku CuKα rotating-anode X-ray generator (MM-007) operating at 40 kV and 20 mA ($\lambda = 1.5418 \text{ \AA}$) with a Rigaku R-Axis IV⁺⁺ image plate detector. A data set from the ADRP domain:ADP-ribose complex was also collected in-house under the same conditions. The crystals belonged to space group P1 ($a = 41.1 \text{ \AA}$, $b = 43.2 \text{ \AA}$, $c = 48.9 \text{ \AA}$, $\alpha = 78.0^\circ$, $\beta = 80.1^\circ$, $\gamma = 73.6^\circ$). Each asymmetric unit in the crystal contains two molecules of the IBV nsp3 ADRP domain. Another data set of the native HCoV-229E nsp3 ADRP domain was collected following a similar procedure. In this case, the protein crystal belonged to space group P2₁2₁1 ($a = 47.8 \text{ \AA}$, $b = 50.9 \text{ \AA}$, $c = 68.3 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$). Only one molecule of the HCoV-229E ADRP domain is present in each asymmetric unit of the crystal. In order to solve the phase problem for the two proteins, crystals of the selenomethionyl (Se-Met) derivative for each were prepared. Data sets for the Se-Met derivatives of ADRP domains from IBV and HCoV-229E were collected at 100 K using an ADSC Quantum 315 detector on beam line BL-5 of the Photon Factory (Tsukuba, Japan). The Se-Met crystals from IBV and HCoV-229E diffracted to 1.8-Å and 2.1-Å resolutions, respectively. They have the same space group as and unit cell parameters similar to those of their respective native crystals. All data were processed, integrated, scaled, and merged using HKL-2000 (27). The data collection statistics are shown in Table 1.

Phasing, model building, and refinement. The structure of the IBV nsp3 ADRP domain and that of its complex with ADP-ribose was solved by the single-wavelength anomalous dispersion (SAD) method from a Se-Met derivative of the nsp3 ADRP domain and from a Se-Met-substituted crystal that had

TABLE 1. Data collection and refinement statistics

Parameter	IBV		HCoV-229E	
	ADRP domain	ADRP domain:ADP-ribose complex	ADRP domain	ADRP domain:ADP-ribose complex
Data collection statistics				
Space group	P1	P1	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell parameters				
a (Å)	41.139	41.364	47.820	47.776
b (Å)	43.201	43.985	50.852	51.024
c (Å)	48.940	49.266	68.278	68.077
α (°)	78.016	78.25	90.00	90.00
β (°)	80.057	79.45	90.00	90.00
γ (°)	73.574	73.39	90.00	90.00
Wavelength (Å)	0.9798	1.5418	0.9798	1.5418
Resolution range (Å) ^a	50.0–1.80 (1.85–1.80)	50.0–2.00 (2.05–2.00)	50.0–2.10 (2.15–2.10)	50.0–2.00 (2.06–2.00)
No. of all reflections	181,232	68,291	117,092	62,592
No. of unique reflections	25,343	22,223	9,732	11,578
Completeness (%)	90.0 (80.2)	85.2 (82.3)	99.6 (96.4)	99.3 (94.8)
R _{merge} ^b (%)	6.9 (41)	6.7 (29.9)	6.6 (23.4)	5.0 (20.4)
Redundancy	7.1 (5.6)	3.0 (2.5)	12.0 (6.3)	5.4 (3.8)
Mean I/sigma	10.0 (3.5)	19.5 (3.6)	11.1 (6.2)	18.1 (5.0)
Refinement statistics				
No. of reflections used	24,398	22,009	9,584	10,993
No. of reflections in testing site	1,298	1,203	1,024	550
R _{work} (%) ^c	17.1	22.4	20.4	20.8
R _{free} (%) ^c	23.8	26.3	28.2	26.3
Mean B factor (Å ²)	23.8	27.3	26.3	27.0
RMSD bond distance (Å)	0.015	0.017	0.019	0.021
RMSD bond angle (°)	1.544	1.881	1.886	2.176
Ramachandran plot (%) ^d	94.2/4.8	94.2/5.4	86.3/9.6	87.7/9.6

^a Values in parentheses refer to the highest-resolution shell.

^b $R_{\text{merge}} = \sum_i |I_i - \langle I \rangle| / \sum_i I_i$, where I_i is an individual intensity measurement and $\langle I \rangle$ is the average intensity for all the reflection i .

^c $R_{\text{work}} = \sum \|F_o - F_c\| / \sum F_o$, where F_o is the observed and F_c is the calculated structure factor amplitude. R_{free} is defined as R_{work} for a randomly selected subset containing 5% of reflections.

^d The percentages of residues located in the most favorable/additionally allowed regions of the Ramachandran plot are given.

been soaked for 2 h in 2 mM ADP-ribose prior to data collection, respectively. The same methods were also applied to the HCoV-229E ADRP domain and its ADP-ribose complex. Initial phases were calculated by the program SOLVE (40). Density modification (solvent flipping) and phase extension to 1.8 Å for IBV and 2.1 Å for HCoV-229E were performed using RESOLVE (39). The models of the two nsp3 ADRP domains were automatically traced using the program ARP/wARP (29) to approximately 90% completeness for the IBV ADRP domain and 70% completeness for the HCoV-229E ADRP domain. The structure was built further manually and refined using the programs *Coot* (13) and REFMAC (26). The IBV nsp3 ADRP domain crystal structure was refined at 1.8-Å resolution to a final R_{work} of 0.171 and R_{free} of 0.238, whereas its HCoV-229E counterpart was refined at 2.1-Å resolution to a final R_{work} of 0.204 and R_{free} of 0.282. The IBV and HCoV-229E ADRP domain:ADP-ribose complex structures were solved by molecular replacement method with CNS (10) using the native structure as a model and followed a similar refinement protocol. The validation of all final models was carried out with PROCHECK (24). Electrostatic surface charges were generated by APBS (6). All diagrams were prepared with PyMOL (<http://www.pymol.org/>). The final refinement statistics are summarized in Table 1.

Protein structure accession numbers. The coordinates for the coronavirus nsp3 ADRP domain crystal structures from IBV and HCoV-229E have been deposited in the RCSB Protein Data Bank (PDB) under accession numbers 3EWO (for the 1.8-Å IBV ADRP domain crystal structure), 3EWP (for the 2.0-Å IBV ADRP domain:ADP-ribose complex crystal structure), 3EWO (for the 2.1-Å HCoV-229E ADRP domain crystal structure), and 3EWR (for the 2.0-Å HCoV-229E ADRP domain:ADP-ribose complex crystal structure).

RESULTS AND DISCUSSION

Overall structure of the IBV and HCoV-229E nsp3 ADRP domains. The cDNA coding for the nsp3 ADRP domain from

IBV was amplified by PCR, and the coded protein contains amino acid residues 1005 to 1178 of pp1a, which are renumbered as 1 to 174 hereinafter for convenience. The crystal structure of the IBV ADRP domain was successfully determined using the SAD method from a Se-Met derivative diffracting to 1.8-Å resolution, as described in Materials and Methods. In the crystal, the IBV ADRP domain exists as a dimer with dimensions of approximately 40 by 40 by 70 Å³, which is unique among all ADRP structures solved to date (Fig. 1A). The two subunits in the asymmetric unit have very similar structures with pair-wise Cα root mean square deviations (RMSD) of less than 0.6 Å. After final refinement, electron density for a few residues at the N and C termini of one of the two monomers could not be observed. These include residues before Lys8 (including five leading residues left from the tag) and residue Lys174 in chain B. The final refinement statistics are listed in Table 1. The two monomeric units in the dimer are in a side-by-side arrangement with a rotation of approximately 90° between the two subunits.

The nsp3 ADRP domain from HCoV-229E was cloned and expressed in the same manner. The coded protein contains amino acid residues 1269 to 1436 of pp1a, which are renumbered 1 to 168 hereinafter for convenience. The crystal structure was determined using the same SAD method from a Se-Met derivative diffracting to 2.1-Å resolution, as described in Materials and Methods. In the HCoV-229E crystal, the nsp3

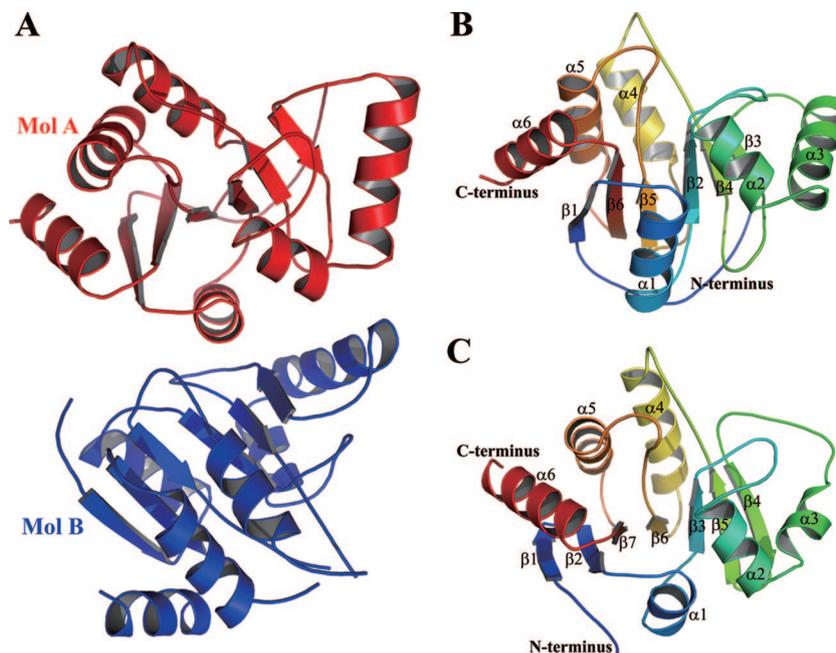


FIG. 1. Three-dimensional structures of the viral ADRP domains from IBV and HCoV-229E. (A) Overall structure of IBV ADRP domain in one asymmetric unit. Molecule A (Mol A; red) and Mol B (blue) form a homodimer. (B) Subunit of the IBV ADRP domain (Mol A). Secondary structures (helices, strands, and loops) are colored from blue (N terminus) to red (C terminus) in a rainbow fashion; α -helices are numbered from $\alpha 1$ to $\alpha 6$, and β -strands are numbered from $\beta 1$ to $\beta 6$. (C) Subunit of the HCoV-229E ADRP domain. Secondary-structure elements are colored in the same way as for IBV; α -helices are numbered from $\alpha 1$ to $\alpha 6$, and β -strands are numbered from $\beta 1$ to $\beta 7$.

ADRP domain exists as a single molecule in the asymmetric unit with dimensions of approximately 35 by 40 by 45 Å³. After final refinement, electron densities for the five leading residues left from the tag and Val168 at the C terminus were not observed. The final refinement statistics are also shown in Table 1.

The monomer fold. In the crystal of the full-length IBV nsp3 ADRP domain, each subunit is comprised of six α -helices and six β -strands (Fig. 1B). As typically observed for the macroH2A-like fold, the six β -strands assume an almost parallel three-dimensional arrangement in the order of $\beta 1$ - $\beta 6$ - $\beta 5$ - $\beta 2$ - $\beta 4$ - $\beta 3$ to form a central six-stranded β -sheet (21). The last strand on one side of the sheet, namely, the $\beta 3$ strand, is uniquely antiparallel to the rest. The surrounding six α -helices have a sandwich-like topology and form a three-layered $\alpha/\beta/\alpha$ motif with the central β -sheet, with three on one side of the sheet, namely, $\alpha 1$, $\alpha 2$, and $\alpha 3$, and the other three on the other side. In the HCoV-229E nsp3 ADRP domain crystal, despite the same $\alpha/\beta/\alpha$ three-layer overall arrangement, the monomer has an additional β -strand at the N terminus compared with its counterpart from IBV (Fig. 1C). This β -strand and the other six β -strands constitute the central β -sheet in the order $\beta 1$ - $\beta 2$ - $\beta 7$ - $\beta 6$ - $\beta 3$ - $\beta 5$ - $\beta 4$. The first and last strands are antiparallel to the rest. The overall topology of the HCoV-229E nsp3 ADRP domain is thus similar to that of the equivalent domain from SARS-CoV, which has been demonstrated in previous reports (34).

In order to further analyze the structural features of the viral ADRP domain, a Dali (18) search was applied using one of the chains of IBV nsp3 ADRP domain as a model. A comparison with other known structures in the PDB revealed the presence

of several structural homologs. Among them the most noteworthy are a putative phosphatase from *Escherichia coli*, ER58 (PDB code, 1SPV; Z-score of 20.2; RMSD of 1.9 Å for 154 superimposed C α atoms); the SARS ADRP domain (PDB code, 2FAV; Z-score of 18.8; RMSD of 2.0 Å for 151 superimposed C α atoms); and a hypothetical protein from *Archaeoglobus fulgidus*, AF1521 (PDB code, 1HJZ; Z-score of 18.6; RMSD of 2.5 Å for 156 superimposed C α atoms). These structures are typical of the “macro domain-like” fold, with the same three-layered $\alpha/\beta/\alpha$ topological arrangement (2). Another close match from the Dali search was the core histone macroH2A.1 (PDB code, 1YD9; Z-score of 17.8; RMSD of 2.1 Å for 155 superimposed C α atoms), which confirms the close relationship between the coronavirus ADRP domain and the macroH2A-like domain. A similar Dali search using HCoV-229E ADRP domain as a model yields similar results, with a Z-score of 23.1 for SARS ADRP domain (RMSD of 1.8 Å for 162 superimposed C α atoms), a Z-score of 20.4 for AF1521 (RMSD of 2.1 Å for 160 superimposed C α atoms), and a Z-score of 20.0 for ER58 (RMSD of 2.1 Å for 153 superimposed C α atoms). Thus, these results from the structure-based comparison, in combination with previous reports on the SARS-CoV nsp3 ADRP domain, unambiguously demonstrate that the viral nsp3 ADRP domain in all three main lineages of coronavirus belongs to the canonical macroH2A-like fold family (34).

Dimeric association of IBV nsp3 ADRP domain. The IBV nsp3 ADRP domain protein forms a crystallographic dimer via a twofold axis (Fig. 2A). The interface area between the two subunits is approximately 2,600 Å² and is formed by a majority of nonpolar residues (55%). Residues in $\alpha 1$ of monomer A,

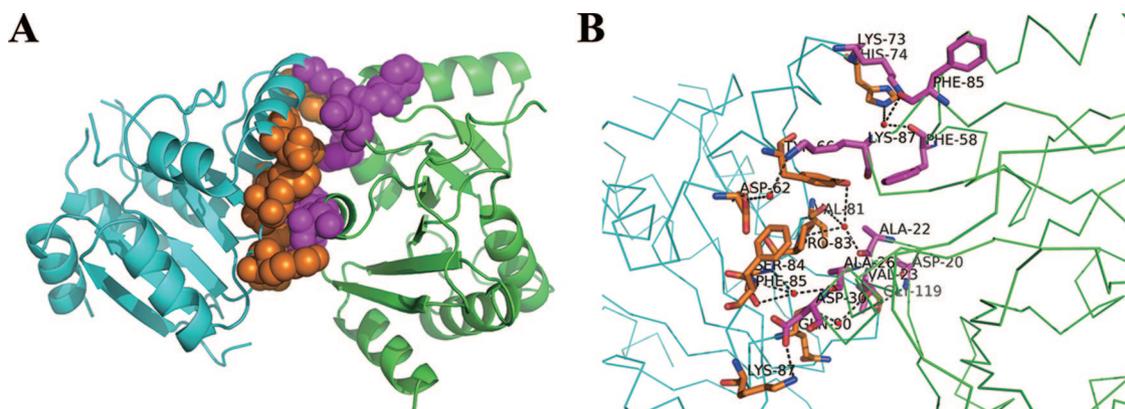


FIG. 2. Dimeric association of the ADRP domain from IBV. (A) The two monomers are shown in green (molecule A) and cyan (molecule B). Residues located in the dimerization interface are shown in sphere representation, and colored separately for each molecule (magenta for molecule A and orange for molecule B). (B) Detailed mechanism of dimer association. The molecules and residues are colored the same as in panel A. The residues involved in the dimer association are shown in a stick model and are labeled. Water molecules involved in the hydrogen bonding are colored red. The dashed lines show the polar contacts between the residues and water molecules.

namely, Asp20, Val23, and Ala26, are involved in the interfacial contacts with a long loop connecting strands $\beta 3$ and $\beta 4$ of monomer B, including Val81, Pro83, and Ser84. The interactions are mediated mainly by hydrogen bonding via water molecules in this region. Additionally, residue Asp30 on $\alpha 1$ of monomer A is negatively charged and interacts with the corresponding positively charged residue, Lys87, in the long loop connecting strands $\beta 3$ and $\beta 4$ of monomer B to form a salt bridge. Besides this electrostatic interaction, hydrogen bonding between side chains of the residues on the contacting surface also contributes to the stability of the dimer. These residues are located mainly in the two loop regions in monomer A: the short loop spanning helices $\alpha 2$ and $\alpha 3$, and the long loop connecting strands $\beta 3$ and $\beta 4$. These residues form hydrogen bonds with residues on helix $\alpha 3$ of monomer B. Five water molecules buried in the dimerization interface are also involved in the interchain hydrogen-bonding network (Fig. 2B).

Systematic structural analysis for ADP-ribose binding. Previous reports on viral ADRP domains demonstrated that they are capable of hydrolyzing ADP-ribose-1"-monophosphate (ADPR-1"-P) to ADP-ribose with high specificity, thus giving rise to the name ADRP domain. And the corresponding structure of the ADRP:ADP-ribose complex from SARS-CoV (group II) has been solved to explain the mechanism of this activity (12, 31, 32, 34, 37). Nevertheless, there have been no investigations to date on the differences between nsp3 ADRP domains from the three main lineages of coronavirus from a structural perspective. This lack of information hinders efforts to explain how coronaviruses evolved this domain with such a highly specific enzymatic activity and to what extent it is conserved or modified among the three coronavirus lineages. In order to provide a systematic understanding of the viral ADRP domain, we solved the structures of the ADRP domains from HCoV-229E (a group I coronavirus) and IBV (a group III coronavirus) in complex with ADP-ribose.

By soaking a native IBV nsp3 ADRP domain crystal in 2 mM ADP-ribose for 2 h, we successfully determined the structure of the ADRP:ADP-ribose complex by use of the native IBV nsp3 ADRP domain structure as a search model (Fig.

3A). After final refinement, residues 1 to 174 (including two additional residues left by the N-terminal tag) in monomer A and residues 7 to 174 in monomer B were built, and two ADP-ribose molecules could be clearly identified from the electric density map. There is one ADP-ribose molecule in each of the two monomers in the asymmetric unit of the crystal. In this case, the ADP-ribose binding site in the ADRP domain was not buried in the dimerization interface, and thus ADP-ribose could diffuse into both monomers, explaining the presence of two ADP-ribose molecules in the dimer structure. In each monomer, the ADP-ribose molecule is located in a binding pocket formed mainly by the N-terminal residues of $\alpha 1$, the long loop connecting strand $\beta 2$ and helix $\alpha 2$, the long loop connecting strand $\beta 5$ and helix $\alpha 5$, and the short loop spanning strand $\beta 6$ and helix $\alpha 6$. Through the same approach employed for the IBV nsp3 ADRP domain, we obtained the structure of the HCoV-229E ADPR:ADP-ribose complex. In this case, ADP-ribose is also tightly bound to the binding pocket formed in the corresponding topological region (Fig. 3B). However, the numbers of the strands that form the pocket are different due to the extra strand at the N terminus in HCoV-229E ADRP domain, as described earlier.

The ADP-ribose binding site is shown to be an open and solvent-accessible cavity from the surface representation of the ADRP domain (Fig. 3C). By calculating the solvent-accessible surface potential, the binding site was revealed to be a mainly positively charged floor, correlating to its capacity for nucleoside diphosphate binding. Upon binding of the ADP-ribose, the most significant conformational change could be observed for the two long loops that form the binding pocket, namely, the long loop connecting strand $\beta 2$ and helix $\alpha 2$ and the long loop connecting strand $\beta 5$ and helix $\alpha 5$ in IBV, along with the long loop connecting strand $\beta 3$ and helix $\alpha 2$ and the long loop connecting strand $\beta 6$ and helix $\alpha 5$ in HCoV-229E, respectively.

In both cases, the ADP-ribose adopts a curved shape as it binds into the pocket. The adenine moiety fits into the hydrophobic cavity formed by residues Leu21, Ala40, Val51, Pro127, Ile133, and Phe159 of the IBV ADRP domain and by residues Val20, Leu46, Pro120, Ile126, Phe150, and Tyr152 of the

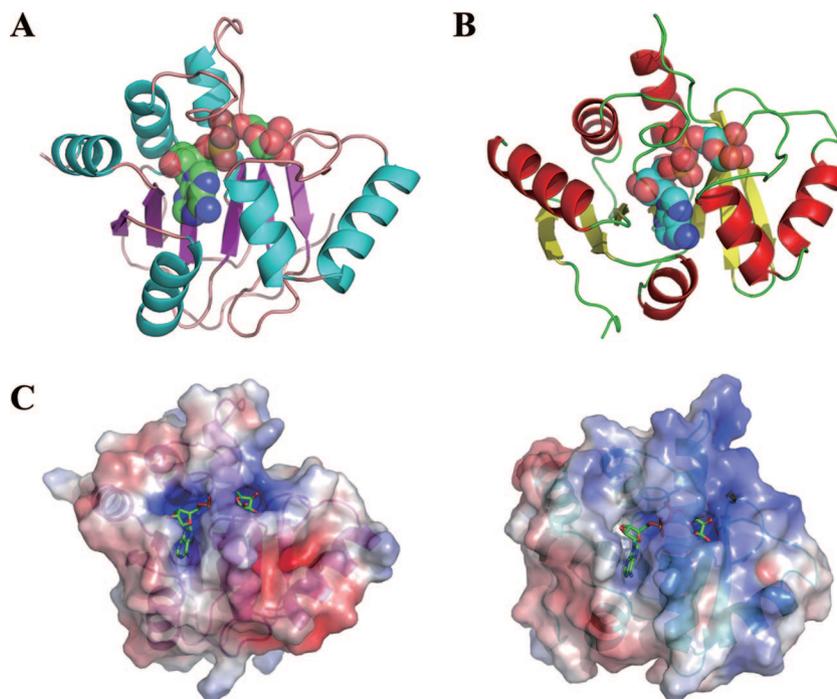


FIG. 3. ADP-ribose binding model of the ADRP domains from IBV and HCoV-229E. (A) The IBV ADRP domain:ADP-ribose complex structure. The ADRP domain is colored by secondary-structure elements (cyan, α -helices; magenta, β -strands; pink, loops). The bound ADP-ribose is shown as a sphere model and is colored by element. (B) The HCoV-229E ADRP domain:ADP-ribose complex structure. The ADRP domain is colored by secondary-structure features (red, α -helices; yellow, β -strands; green, loops). The bound ADP-ribose is represented by spheres and colored by element. (C) Surface model of ADRP domains from IBV and HCoV-229E shown covered by an electrostatic surface potential. Positively charged residues are colored blue; negatively charged residues are colored red. The bound ADP-ribose is shown in a stick representation and colored according to element.

HCoV-229E ADRP domain. A series of hydrogen bonds are also involved in the binding of ADP-ribose. The N6 atom of the adenine ring makes three hydrogen bonds with surrounding water molecules, through which it interacts with Asp20 in IBV or with the equivalent Asp19 in HCoV-229E (Fig. 4A). The equivalent residue in the SARS-CoV ADRP domain is Asp23, which has also been demonstrated to be involved in hydrogen bonding with the adenosine moiety from previous

structural reports (12). This residue has been revealed to be critical for the binding specificity of the ADRP domain by a study on AF1521, a macro domain from *Archaeoglobus fulgidus* (2). Structure-based sequence alignment of the viral ADRP domain also shows that this residue is highly conserved among the three main coronavirus lineages (Fig. 5). Collectively, these facts indicate that Asp20 in the IBV ADRP domain is indeed conserved in terms of both amino acid sequence and structural

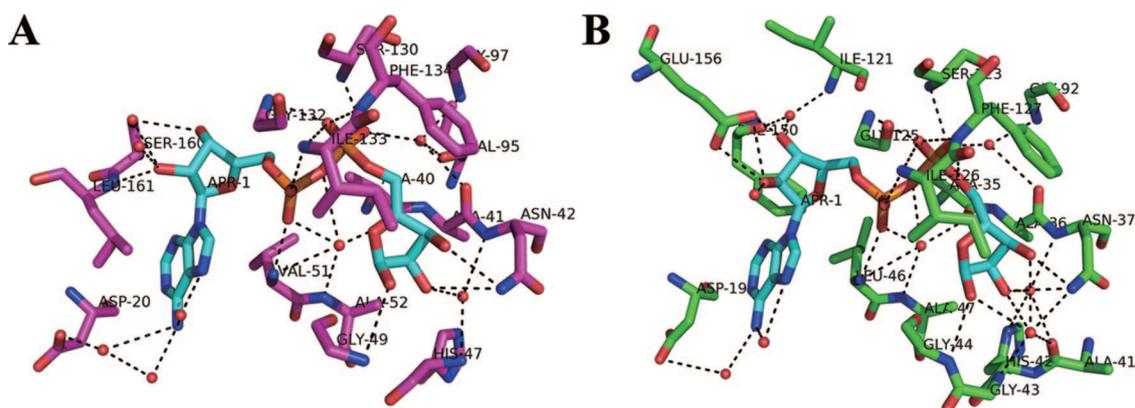


FIG. 4. Close-up view of the interactions in the ADRP domain:ADP-ribose complex from IBV and HCoV-229E. (A) Interactions between the IBV ADRP domain and bound ADP-ribose. Protein residues and ADP-ribose are shown in a stick model and colored magenta and cyan, respectively. Oxygen, nitrogen, and phosphorus atoms are shown in red, blue, and orange, respectively. The dashed lines indicate hydrogen bonds. Water molecules involved in hydrogen bonding are shown in red. (B) Interactions between the HCoV-229E ADRP domain and bound ADP-ribose. Protein residues and ADP-ribose are shown in green and cyan, respectively. The other representations are the same as in panel A.

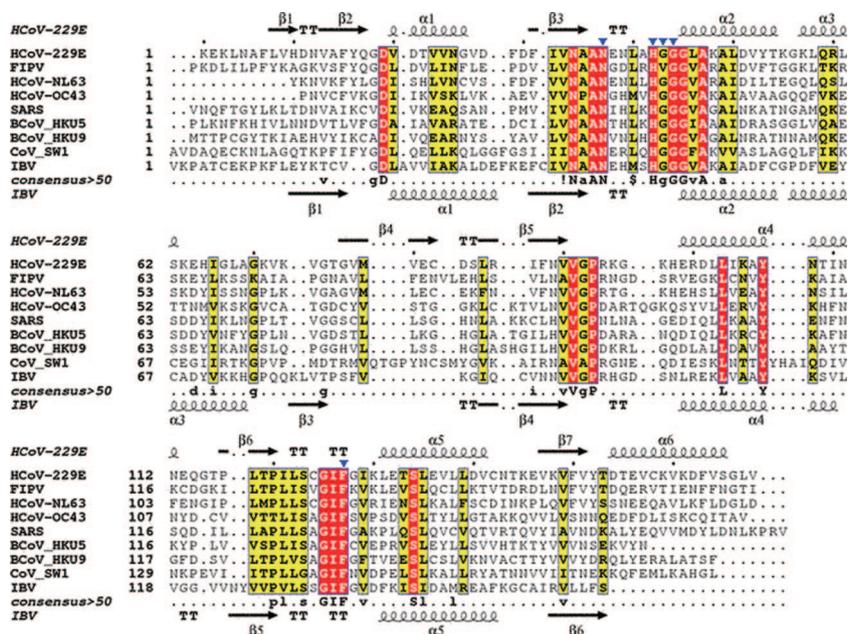


FIG. 5. Structure-based sequence alignment of the viral ADRP domains from all three main coronavirus lineages. Shown are the following: HCoV-229E (group Ib, DDBJ/EMBL/GenBank accession number P0C6U2); feline infectious peritonitis virus (FIPV; group Ia, DDBJ/EMBL/GenBank accession number Q98VG9); HCoV-NL63 (group Ib, DDBJ/EMBL/GenBank accession number P0C6X5); HCoV-OC43 (group IIa, DDBJ/EMBL/GenBank accession number P0C6X6); SARS-CoV (group IIb, DDBJ/EMBL/GenBank accession number P0C6X7); bat coronavirus HKU5 (BCoV_HKU5; group IIc, DDBJ/EMBL/GenBank accession number P0C6W4); bat coronavirus HKU9 (BCoV_HKU9; group II, DDBJ/EMBL/GenBank accession number P0C6W5); coronavirus SW1 (CoV_SW1; group III, DDBJ/EMBL/GenBank accession number P0C6V5); and IBV (group III, DDBJ/EMBL/GenBank accession number P0C6V5). Secondary structures of the HCoV-229E ADRP domain (above) and the IBV ADRP domain (below) are indicated in the aligned sequence. Residue numbers of ADRP domain from HCoV-229E are indicated by black dots above the HCoV-229E sequence (one dot corresponding to 10 residues). The residues located in the active site of the ADRP domain, namely, Asn37, His42, Gly43, Gly44, and Phe127 (numbering from HCoV-229E), are labeled by blue arrows. The sequence alignment was generated using MUSCLE (11) and presented using ESPrnt (15).

interactions, confirming its role in conveying the substrate specificity to the viral ADRP domain. The first ribose moiety and the two phosphate groups make strong hydrogen bonds with the main chain of surrounding residues. This complicated set of residues includes Gly49, Val51, Ala52, Ser130, Gly132, Ile133, and Phe134 in the IBV ADRP domain and Gly44, Leu46, Ala47, Ser123, Gly125, Ile126, and Phe127 in the HCoV-229E ADRP domain. Surprisingly, although these residues are involved only in the binding of the ADP moiety, all of them are highly conserved in sequence among different coronavirus species (Fig. 5).

The terminal ribose, which harbors the site of cleavage in the catalytic hydrolysis reaction, interacts with Asn42, His47, Gly49, and Phe134 in the IBV ADRP domain through a complex hydrogen-bonding network (Fig. 4A). Noticeably, a water molecule serves as an intermediate bridge between the cleavage site on the terminal ribose and the catalytically significant residues, i.e., Asn42 and His47. This indicates that Asn42 and His47 may be responsible for the catalytic activity of the ADRP domain through which ADPR-1''-P is converted into ADP-ribose. This result is consistent with previous structural data obtained from the yeast ADRP domain, in which it was shown to employ similar residues to achieve its catalytic activity (22). Additional biochemical studies on the viral ADRP domain also demonstrated that when the residues in the SARS-CoV ADRP domain corresponding to Asn42, His47, Gly49, and Phe134 in

IBV are mutated, the ADRP domain will lose most of its catalytic activity (12).

Similar structural organization is also observed for the HCoV-229E ADRP domain. In this case, residues Asn37, His42, Gly43, and Gly44 make hydrogen bonds with the terminal ribose with the aid of surrounding water molecules (Fig. 4B). Previous site-directed mutagenesis studies showed that residues Asn1302, Asn1305, His1310, Gly1312, and Gly1313 in the HCoV-229E ADRP domain (corresponding to Asn34, Asn37, His42, Gly43, and Gly44, respectively, herein) form part of the active site of the enzyme (31). Our structure provides direct evidence for the location of the ADRP active site. In the ADRP domain:ADP-ribose complex structure, all residues with the exception of Asn34 indeed participate in the hydrogen bonding between the ADRP domain and the ADP-ribose. However, Asn34, which was proposed to be located at the active site in the previous study, has no observable interaction with the ADP-ribose in the crystal structure; the distance between its C α and the RC1* of the terminal ribose is 8.7 Å. Since the substrate for the ADRP activity is ADPR-1''-P, which has an additional terminal phosphate compared to ADP-ribose, it is possible that this residue may contribute to the catalytic activity by interacting with the terminal phosphate through water-mediated hydrogen bonding, or it may serve as part of the scaffold supporting the residues at the active site so that they may adopt the optimal conformation to perform their

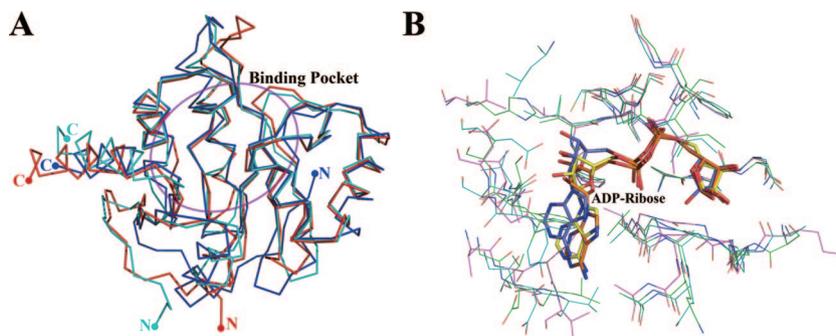


FIG. 6. Structural superposition of viral ADRP domains from the three main coronavirus lineages (HCoV-229E from group I, SARS-CoV from group II, and IBV from group III). (A) Superposition of the three ADRP domain structures from HCoV-229E (cyan), SARS-CoV (red), and IBV (blue). The structures of the three different lineages are quite similar except in the N and C termini and certain loop regions. The highly conserved binding pocket of the viral ADRP domain is shown by the magenta circle and labeled. (B) Close-up view of the active site in the superposed ADRP domain:ADP-ribose complexes from IBV, SARS-CoV, and HCoV-229E. Protein residues are shown in lines and colored in cyan for HCoV-229E, magenta for SARS-CoV, and green for IBV, respectively. The bound ADP-ribose molecules are shown in a stick model and colored orange for HCoV-229E, light blue for SARS-CoV, and yellow for IBV. Oxygen, nitrogen, and phosphorus atoms are shown in red, blue, and orange, respectively.

catalytic function, thus explaining the loss of enzymatic activity after mutation of this residue. Overall, the active-site residues are highly conserved in all three available structures of coronavirus nsp3 ADRP domains.

Systematic structure comparison among coronavirus species. In order to gain further insights into the similarities and differences of the viral ADRP domains among the three main coronavirus lineages, a superposition of the overall structure of the three available coronavirus ADRP domains from HCoV-229E (group I), SARS-CoV (group II), and IBV (group III) was performed to compare their structural features (12). The major characteristics of the macroH2A-like fold are well conserved, with appreciable variations only in the N- and C-terminal ends and some residues in the two loop regions: the short loop spanning helix $\alpha 3$ and strand $\beta 3$, along with the long loop connecting strand $\beta 4$ and helix $\alpha 4$ (secondary-structure numbering follows that of IBV), in the coronavirus ADRP domains (Fig. 6A). This observation was further confirmed by the Dali search results as previously described, which showed that the calculated RMSD for all superimposed C α atoms is less than 2.0 Å between any pair formed from the three available coronavirus ADRP domain crystal structures.

For a better understanding of the exact organization through which the conserved amino acid sequences are interpreted into three-dimensional protein structures to perform physiological functions, it is necessary to study the active sites of the ADRP domains in more detail. To do this, the residues surrounding the ADP-ribose binding site in the ADRP domain:ADP-ribose complex structures from the three representatives of coronavirus were superposed (Fig. 6B). A number of hydrophobic residues in the binding pocket are highly conserved among coronavirus species in terms of both sequence alignment and structural superposition. For example, Pro127 in IBV, Pro120 in HCoV-229E, and Pro126 in SARS-CoV are almost perfectly superposed in the same three-dimensional position. However, the superposition demonstrates that the majority of them have structural variations rather than being strictly conserved. Most noticeably, the residues conveying the substrate specificity, namely, Asp19 in HCoV-229E, Asp23 in SARS-CoV, and

Asp20 in IBV, are located in different positions and assume distinctive conformations in the three coronavirus species, with an average distance of 2.1 Å between C α atoms for the three residues. The mechanisms through which they form hydrogen bonds are also considerably different. In SARS-CoV, this residue interacts directly with the N6 atom of the adenosine ring, while in the other two cases the hydrogen bonding is mediated by surrounding water molecules. In addition, residues that flank the ADP moiety to stabilize it in the binding cavity also vary significantly, as shown in the superposition result. Thus, even though the majority of residues interacting with the ADP moiety are highly conserved in sequence, the structural superposition clearly indicates that this region is quite flexible, especially those parts that bind the adenosine ring and the first ribose moiety (Fig. 5 and Fig. 6B). Despite the maintenance of sequence homology, most residues that are responsible for substrate specificity and binding capacity in the ADRP domain binding pockets for ADP-ribose are structurally related but not rigorously conserved among the three different coronavirus lineages (12, 31).

The residues constituting the catalytic site of the ADRP domains are, on the other hand, strictly conserved among the three main coronavirus lineages. Noticeably, Asn42, His47, Gly49, and Phe134 (residue numbers are from IBV), the four residues identified by site-directed mutagenesis studies of the SARS-CoV and HCoV-229E ADRP domains, are all located at almost exactly the same position around the RC1* of the terminal ribose and exhibit strong interactions. This demonstrates the significant conservation of these catalytically important residues from a structural perspective, confirming previous reports that these residues have indeed evolved to perform a unifying biochemical function in viral ADRP domains. Even though the low turnover numbers in enzymatic assays and reverse genetics suggest that this catalytic activity is more likely to play a regulatory rather than essential role in viral replication, this conservation in sequence and structural analysis indicates that it is necessary to perform further studies of this ADRP activity in a host-virus interaction context to elucidate its physiological significance (12, 32, 34). Recent studies have

also shown that another possible explanation for the function of viral ADRP domains may be its ability to bind PAR (3, 12). As representative structures are now available for ADRP domains from all three main coronavirus lineages, further studies will be able to use these results as a basis to support PAR binding models, if the mechanisms through which this viral PAR binding ability interacts with host cell pathways are elucidated.

ACKNOWLEDGMENTS

We thank Ming Liao and K. Y. Yuen for providing cDNAs of IBV strain M41 and HCoV-229E, respectively.

This work was supported by Project 973 of the Ministry of Science and Technology of China (grant nos. 2006CB806503 and 2007CB914301), the National Natural Science Foundation of China (NSFC) (grant no. 30221003 and 30730022), the Chinese Academy of Sciences Knowledge Innovation Project (grant no. KSCX1-YW-R-05), and the Ministry of Education (grant no. 307002).

REFERENCES

1. Aguiar, R. C. T., K. Takeyama, C. Y. He, K. Kreinbrink, and M. A. Shipp. 2005. B-aggressive lymphoma family proteins have unique domains that modulate transcription and exhibit poly(ADP-ribose) polymerase activity. *J. Biol. Chem.* **280**:33756–33765.
2. Allen, M. D., A. M. Buckle, S. C. Cordell, J. Lowe, and M. Bycroft. 2003. The crystal structure of AF1521 a protein from *Archaeoglobus fulgidus* with homology to the non-histone domain of macroH2A. *J. Mol. Biol.* **330**:503–511.
3. Ame, J. C., C. Spenlehauer, and G. de Murcia. 2004. The PARP superfamily. *Bioessays* **26**:882–893.
4. Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld. 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**:3213–3224.
5. Anand, K., J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld. 2003. Coronavirus main proteinase (3CL(pro)) structure: basis for design of anti-SARS drugs. *Science* **300**:1763–1767.
6. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**:10037–10041.
7. Bartlam, M., H. T. Yang, and Z. H. Rao. 2005. Structural insights into SARS coronavirus proteins. *Curr. Opin. Struct. Biol.* **15**:664–672.
8. Brierley, I., M. E. Boursnell, M. M. Binns, B. Bilimoria, V. C. Blok, T. D. Brown, and S. C. Inglis. 1987. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6**:3779–3785.
9. Brierley, I., P. Digard, and S. C. Inglis. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**:537–547.
10. Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**:905–921.
11. Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797.
12. Egloff, M. P., H. Malet, A. Putics, M. Heinonen, H. Dutartre, A. Frangeul, A. Gruetz, V. Campanacci, C. Cambillau, J. Ziebuhr, T. Ahola, and B. Canard. 2006. Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. *J. Virol.* **80**:8493–8502.
13. Emsley, P., and K. Cowtan. 2004. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**:2126–2132.
14. Gorbalenya, A. E. 2001. Big nidovirus genome—when count and order of domains matter. *Adv. Exp. Med. Biol.* **494**:1–17.
15. Gouet, P., E. Courcelle, D. I. Stuart, and F. Metz. 1999. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**:305–308.
16. Hay, S., and G. Kannourakis. 2002. A time to kill: viral manipulation of the cell death program. *J. Gen. Virol.* **83**:1547–1564.
17. Hilgenfeld, R., K. Anand, J. R. Mesters, Z. H. Rao, X. Shen, H. L. Jiang, J. Z. Tan, and K. H. G. Verschueren. 2006. Structure and dynamics of SARS coronavirus main proteinase (M-pro). *Adv. Exp. Med. Biol.* **581**:585–591.
18. Holm, L., and C. Sander. 1996. Mapping the protein universe. *Science* **273**:595–602.
19. Ivanov, K. A., T. Hertzog, M. Rozanov, S. Bayer, V. Thiel, A. E. Gorbalenya, and J. Ziebuhr. 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. USA* **101**:12694–12699.
20. Ivanov, K. A., V. Thiel, J. C. Dobbe, Y. van der Meer, E. J. Snijder, and J. Ziebuhr. 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.* **78**:5619–5632.
21. Karras, G. I., G. Kustatscher, H. R. Buhecha, M. D. Allen, C. Pugieux, F. Sait, M. Bycroft, and A. G. Ladurner. 2005. The macro domain is an ADP-ribose binding module. *EMBO J.* **24**:1911–1920.
22. Kumaran, D., S. Eswaremoorthy, F. W. Studier, and S. Swaminathan. 2005. Structure and mechanism of ADP-ribose-1^o-monophosphatase (Appr-1^o-pase), a ubiquitous cellular processing enzyme. *Protein Sci.* **14**:719–726.
23. Lai, M. M. C., and K. V. Holmes. 2001. *Coronaviridae*: the viruses and their replication, p. 1163–1185. *In* D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), *Fields virology*, 4th ed., vol. 1. Lippincott Williams & Wilkins, Philadelphia, PA.
24. Laskowski, R. A., J. A. C. Rullmann, M. W. MacArthur, R. Kaptein, and J. M. Thornton. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**:477–486.
25. Martzen, M. R., S. M. McCraith, S. L. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, and E. M. Phizicky. 1999. A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**:1153–1155.
26. Murshudov, G. N., A. A. Vagin, and E. J. Dodson. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**:240–255.
27. Otwinowski, Z., and W. Minor. 1997. Processing of X-ray diffraction data collected in oscillation mode, p. 307–326. *In* C. W. Carter, Jr., and R. M. Sweet (ed.), *Macromolecular crystallography*, part A, vol. 276. Academic Press, San Diego, CA.
28. Pehrson, J. R., and V. A. Fried. 1992. MacroH2A, a core histone containing a large nonhistone region. *Science* **257**:1398–1400.
29. Perrakis, A., R. Morris, and V. S. Lamzin. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**:458–463.
30. Prentice, E., J. McAuliffe, X. T. Lu, K. Subbarao, and M. R. Denison. 2004. Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J. Virol.* **78**:9977–9986.
31. Putics, A., W. Filipowicz, J. Hall, A. E. Gorbalenya, and J. Ziebuhr. 2005. ADP-ribose-1^o-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture. *J. Virol.* **79**:12721–12731.
32. Putics, A., A. E. Gorbalenya, and J. Ziebuhr. 2006. Identification of protease and ADP-ribose 1^o-monophosphatase activities associated with transmissible gastroenteritis virus non-structural protein 3. *J. Gen. Virol.* **87**:651–656.
33. Rouleau, M., R. A. Aubin, and G. G. Poirier. 2004. Poly(ADP-ribosyl)ated chromatin domains: access granted. *J. Cell Sci.* **117**:815–825.
34. Saikatendu, K. S., J. S. Joseph, V. Subramanian, T. Clayton, M. Griffith, K. Moy, J. Velasquez, B. W. Neuman, M. J. Buchmeier, R. C. Stevens, and P. Kuhn. 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1^o-phosphate dephosphorylation by a conserved domain of nsP3. *Structure* **13**:1665–1675.
35. Schelle, B., N. Karl, B. Ludewig, S. G. Siddell, and V. Thiel. 2005. Selective replication of coronavirus genomes that express nucleocapsid protein. *J. Virol.* **79**:6620–6630.
36. Shull, N. P., S. L. Spinelli, and E. M. Phizicky. 2005. A highly specific phosphatase that acts on ADP-ribose 1^o-phosphate, a metabolite of tRNA splicing in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **33**:650–660.
37. Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. M. Poon, Y. Guan, M. Rozanov, W. J. M. Spaan, and A. E. Gorbalenya. 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.
38. Spaan, W. J. M., and D. Cavanagh. 2004. Virus taxonomy, p. 945–962. VIIIth report of the ICTV. Elsevier/Academic Press, London, United Kingdom.
39. Terwilliger, T. C. 2000. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**:965–972.
40. Terwilliger, T. C., and J. Berendzen. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**:849–861.
41. Thiel, V., K. A. Ivanov, A. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr. 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**:2305–2315.
42. Xue, X. Y., H. W. Yu, H. T. Yang, F. Xue, Z. X. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X. J. C. Zhang, M. Liao, M. Bartlam, and Z. H. Rao. 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* **82**:2515–2527.
43. Yang, H. T., M. Bartlam, and Z. H. Rao. 2006. Drug design targeting the main protease, the Achilles' heel of coronaviruses. *Curr. Pharm. Des.* **12**:4573–4590.
44. Yang, H. T., W. Q. Xie, X. Y. Xue, K. L. Yang, J. Ma, W. X. Liang, Q. Zhao, Z. Zhou, D. Q. Pei, J. Ziebuhr, R. Hilgenfeld, K. Y. Yuen, L. Wong, G. X. Gao, S. J. Chen, Z. Chen, D. W. Ma, M. Bartlam, and Z. H. Rao. 2005.

- Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**:2044.
45. **Yang, H. T., M. J. Yang, Y. Ding, Y. W. Liu, Z. Y. Lou, Z. Zhou, L. Sun, L. J. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. H. Rao.** 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. USA* **100**:13190–13195.
 46. **Ziebuhr, J.** 2005. The coronavirus replicase. *Curr. Top. Microbiol. Immunol.* **287**:57–94.
 47. **Ziebuhr, J.** 2004. Molecular biology of severe acute respiratory syndrome coronavirus. *Curr. Opin. Microbiol.* **7**:412–419.
 48. **Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya.** 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**:853–879.