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 Rao1,2,3*

Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China;1 National Laboratory of Biomacromolecules, Institute of Biophysics (IBP), Chinese Academy of Sciences, Beijing 100101, China;2 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 essential ORF1a and ORF1b) and is transcribed into two

* 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 be involved 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 coronavirus 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 nsp5s. 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 that 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 1209 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'-CCGGTACCTTTACACGG CATCAGTA-3'), IBV-nsp3-ADRP-R (5'-CCGCGTCA GTTACTTCAAGGTGTCAGCAAAAT-3'), 229E-nsp3-ADRP-F (5'-CCGCGGAT CCAAGAGAAGTGACACGCCT-3'), and 229E-nsp3-ADRP-R (5'-CCGCTCG AGTTACCAAACTCAGACACAA-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 (GPGLS) 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 (ω = 1.5418 Å) 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 = 47.8 Å, b = 50.9 Å, c = 68.3 Å, α = β = γ = 90°). 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
### 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 pairwise 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 monomer and a few residues at the N- and C-terminal ends of the other could not be observed.

The 3D coordinates of the IBV ADRP domain crystal structures from IBV and HCoV-229E have been deposited in the RCSB Protein Data Bank (PDB) under accession numbers 3EXW (for the 3.0-Å IBV ADRP domain crystal structure), 3EWO (for the 2.1-Å HCoV-229E ADRP domain crystal structure), and 3EWR (for the 2.0-Å HCoV-220E ADRP domain crystal structure).

### Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>IBV ADRP domain</th>
<th>IBV ADRP domain:ADP-ribose complex</th>
<th>HCoV-229E ADRP domain</th>
<th>HCoV-229E ADRP domain:ADP-ribose complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
<td>P2;3/2;1</td>
<td>P2;3/2;1</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>41.139</td>
<td>41.364</td>
<td>47.820</td>
<td>47.776</td>
</tr>
<tr>
<td>b (Å)</td>
<td>43.201</td>
<td>43.985</td>
<td>50.852</td>
<td>51.024</td>
</tr>
<tr>
<td>c (Å)</td>
<td>48.940</td>
<td>49.266</td>
<td>68.278</td>
<td>68.077</td>
</tr>
<tr>
<td>α (°)</td>
<td>78.016</td>
<td>78.25</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>β (°)</td>
<td>80.057</td>
<td>79.45</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td>γ (°)</td>
<td>73.574</td>
<td>73.39</td>
<td>90.00</td>
<td>90.00</td>
</tr>
<tr>
<td><strong>Refinement statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of reflections used</td>
<td>24,398</td>
<td>22,009</td>
<td>9,584</td>
<td>10,993</td>
</tr>
<tr>
<td>No. of reflections in testing site</td>
<td>1,298</td>
<td>1,203</td>
<td>1,024</td>
<td>550</td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>17.1</td>
<td>22.4</td>
<td>20.4</td>
<td>20.8</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>23.8</td>
<td>26.3</td>
<td>26.3</td>
<td>26.3</td>
</tr>
<tr>
<td>Mean B factor (Å²)</td>
<td>23.8</td>
<td>27.3</td>
<td>26.3</td>
<td>27.0</td>
</tr>
<tr>
<td>RMSD bond distance (Å)</td>
<td>0.115</td>
<td>0.107</td>
<td>0.019</td>
<td>0.021</td>
</tr>
<tr>
<td>RMSD bond angle (°)</td>
<td>1.544</td>
<td>1.881</td>
<td>1.886</td>
<td>2.176</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td>94.2/4.8</td>
<td>94.2/5.4</td>
<td>86.3/9.6</td>
<td>87.7/9.6</td>
</tr>
</tbody>
</table>

* Values in parentheses refer to the highest-resolution shell.

b $R_{merge} = \Sigma_{i,j} F_{i,j} - \langle F \rangle$ where $F_{i,j}$ is an individual intensity measurement and $\langle F \rangle$ is the average intensity for all the reflection $i$.

$R_{work} = \Sigma_{obs} |F_o| - |F_c| / \Sigma_{obs} |F_o|$, 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.
ADRP domain exists as a single molecule in the asymmetric unit with dimensions of approximately 35 by 40 by 45 Å³. After 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 β1-β6-β5-β2-β4-β3 to form a central six-stranded β-sheet (21). The last strand on one side of the sheet, namely, the β3 strand, is uniquely antiparallel to the rest. The surrounding six α-helices have a sandwich-like topology and form a three-layered α/β/α motif with the central β-sheet, with three on one side of the sheet, namely, α1, α2, and α3, and the other three on the other side. In the HCoV-229E nsp3 ADRP domain crystal, despite the same α/β/α 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 β1-β2-β7-β6-β3-β5-β4. 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 Ca atoms); the SARS ADRP domain (PDB code, 2FAV; Z-score of 18.8; RMSD of 2.0 Å for 151 superimposed Ca 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 Ca atoms). These structures are typical of the “macro domain-like” fold, with the same three-layered α/β/α 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 Ca 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 1.5 Å for 162 superimposed Ca atoms), a Z-score of 20.4 for AF1521 (RMSD of 2.1 Å for 160 superimposed Ca atoms), and a Z-score of 20.0 for ER58 (RMSD of 2.1 Å for 153 superimposed Ca 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 α1 of monomer A,
namely, Asp20, Val23, and Ala26, are involved in the interfacial contacts with a long loop connecting strands β3 and β4 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 α1 of monomer A is negatively charged and interacts with the corresponding positively charged residue, Lys87, in the long loop connecting strands β3 and β4 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 α2 and α3, and the long loop connecting strands β3 and β4. These residues form hydrogen bonds with residues on helix α3 of monomer B. Five water molecules buried in the dimerization interface also involve 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 α1, the long loop connecting strand β2 and helix α2, the long loop connecting strand β5 and helix α5, and the short loop connecting strand β6 and helix α6. Through the same approach employed for the IBV nsp3 ADRP domain, we obtained the structure of the HCoV-229E ADRP: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 β2 and helix α2 and the long loop connecting strand β5 and helix α5 in IBV, along with the long loop connecting strand β3 and helix α2 and the long loop connecting strand β6 and helix α6 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...
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. 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.

---

**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.
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/H9251 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 catalytic function.
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 $\mathrm{C}_\alpha$ 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 complexes from IBV, SARS-CoV, and HCoV-229E 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 $\mathrm{C}_\alpha$ 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


