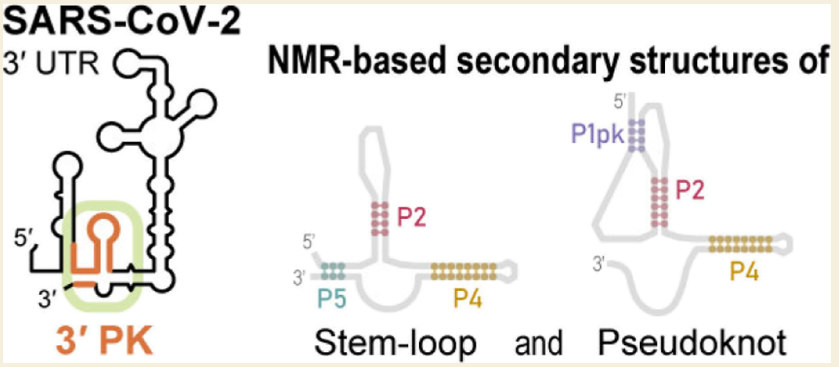
JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/jacsau
Article

NMR Studies of Genomic RNA in 3′ Untranslated Region Unveil Pseudoknot Structure that Initiates Viral RNA Replication in SARS-CoV-2

Takako Ohyama,* Takuo Osawa, Shun-ichi Sekine, and Yoshitaka Ishii*

Cite This: JACS Au 2024, 4, 1323−1333
Read Online



ABSTRACT: In the 3′ untranslated region of the SARS-CoV-2 virus RNA genome, genomic RNA replication is initiated in the highly conserved region called 3′PK, containing three stem structures (P1pk, P2, and P5). According to one proposed mechanism, P1pk and distal P2 stems switch their structure to a pseudoknot through base-pairing, thereby initiating transcription by recruiting RNA-dependent RNA polymerase complexed with nonstructural proteins (nsp)7 and nsp8. However, experimental evidence of pseudoknot formation or structural switching is unavailable. Using SARS-CoV-2 3′PK fragments, we show that 3′PK adopted stem-loop and pseudoknot forms in a mutually exclusive manner. When P1pk and P2 formed a pseudoknot, the P5 stem, which includes a sequence at the 3′ end, exited from the stem-loop structure and opened up. Interaction with the nsp7/nsp8 complex destabilized the stem-loop form but did not alter the pseudoknot form. These results suggest that the interaction between the pseudoknot and nsp7/nsp8 complex transformed the 3′ end of viral genomic RNA into single-stranded RNA ready for synthesis, presenting the unique pseudoknot structure as a potential pharmacological target.

KEYWORDS: *noncoding RNA, secondary structure, SARS-CoV-2, NMR*

The genome of the severe acute respiratory syndrome (SARS)-coronavirus 2 (CoV-2) virus, which is about 30,000 nucleotides in length, contains approximately 250 and 340 nucleotides of the untranslated region (UTR) located at the 5′ and 3′ ends of the genome, respectively[1] (Figure 1A). Several hypotheses have been proposed for the mechanism that transcribes the virus genome and subgenomic RNA;[2,3] in each case, genomic and subgenomic RNAs are transcribed from the 3′ end of positive-sense genomic RNAs of the SARS-CoV-2 virus. The 3′ UTR, which is noncoding RNA that plays an integral role in translation and transcription,[4,5] includes a stem-loop region called the bulged stem-loop (BSL), a large branched stem-loop that contains the hypervariable region (HVR), the octanucleotide motif (ONM), and the stem-loop II-like motif (S2M; Figure 1B). An additional region known as 3′PK contains part of the bulged stem-loop and includes a stem structure located between the bulged stem-loop and the hypervariable region (Figure 1B). The 3′ PK is a region found in coronavirus that is presumed to form pseudoknot structures. It was reported that deletions[6] of the putative pseudoknot region or mutations that disrupt putative pseudoknot structure[4,7,8] were lethal for virus replication. This region is widely believed to play a pivotal role in controlling replication[5,9,10] and is highly conserved among coronaviruses, including SARS-related coronaviruses (SARSr-CoV)[11,12] and other beta-coronaviruses, including bovine coronavirus (BCoV) and mouse hepatitis virus (MHV).[8] A pseudoknot structure is a structure motif of nucleic acids that has two (or

Received: October 20, 2023
Revised: February 20, 2024
Accepted: March 8, 2024
Published: March 20, 2024

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

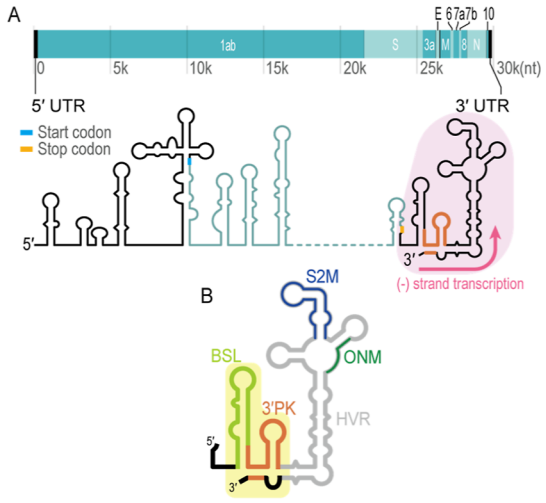


Figure 1. (A) Schematic drawing of genomic RNA and secondary structures of SARS-CoV-2. The 3′UTR shown in (B) is highlighted by light pink. (B) Schematic drawing of secondary structure in the 3′ UTR region, with the names of different motifs color-coded: yellow-green, bulged stem-loop; orange, 3′PK; gray, hypervariable region; green, octanucleotide motif; and blue, stem-loop II like motif. The region described in Figure 2 is highlighted by yellow in (B).

more) stem-loops; in a basic pseudoknot structure, one stem is nested into the loop region of the other stem-loop. Pseudoknot structures were found in structured RNAs involved in important biological processes, such as the frameshift element of viruses,[13,14] the telomerase RNA component,[15] and tRNA-like structures.[16]

Nonstructural proteins (nsp)7−16, which are coded in the *ORF1ab* gene at the 5′ terminus of the genome, play primary roles in replicating the virus genome. Genomic RNA is transcribed by a replication−transcription complex composed of RNA-dependent RNA polymerase (RdRp), also called nsp12, and the cofactor proteins nsp7 and nsp8.[17−20] In addition, nsp8 exhibits RNA polymerase activity,[21] and a nsp7−nsp8 complex acts as a primase.[22] Furthermore, RdRp is associated with other accessory proteins such as nsp9 and nsp13.[23,24] nsp9 forms a dimer and binds to nsp8 to interact with single-stranded RNA by nonspecific binding.[25−29] Recently, it was revealed that nsp9 in complex with nsp12 mediates the capping of RNA.[23] Nsp13 forms a dimer and acts as a helicase when binding to nsp8 and nsp12.[24] The widely adopted model of how negative-sense genomic RNA transcription is initiated in coronaviruses was proposed by Züst et al.[5] based on extensive studies of mutations and reverting mutations on 3′PK, nsp8, and nsp9 that modulated virus growth for closely related MHV; the model is illustrated in Figure 2. First, a primase complex composed of nsp7, nsp8, and nsp9 binds to the 3′PK stem-loop structure, causing the transcription of a small RNA primer. Next, the 3′PK structure switches to form a pseudoknot. It is also proposed that once the pseudoknot structure is formed, nsp12 binds to the RNA in the pseudoknot form through the primase complex with 3′PK and starts to transcribe genomic RNA. Mutations of nucleotides that disrupt the putative pseudoknot structure of the 3′PK region were lethal for viral replication.[4,7] In vivo and in virion whole or partial genome RNA analysis by SHAPE-MaPseq and DMS-MaPseq showed the secondary structure of 3′ UTR and suggested that the secondary structure of the 3′PK region is a stem-loop,[10,30−34] but evidence supporting the formation of a pseudoknot structure is scarce. Thermal unfolding profiles and a simplified sequencing of psoralen cross-linked, ligated, and selected hybrids (SPLASH) assay indirectly suggested that the formation of a pseudoknot structure in 3′PK is only marginally stable in mouse hepatitis virus[35] and SARS-CoV-2.[36] Furthermore, a recent NMR-based study of the secondary structure reported the presence of a stem-loop rather than a pseudoknot in the 3′PK region.[37] However, it is unsurprising that the pseudoknot structure has

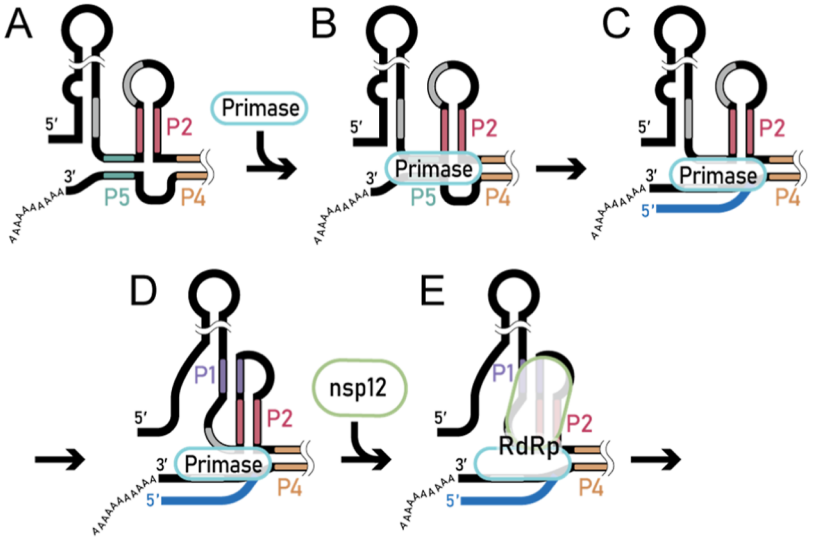


Figure 2. A hypothetical model of how genomic RNA replication is initiated in coronaviruses, based on a proposal by Züst et al.[5] with minor modifications. The region highlighted by yellow in Figure 1B is shown. (A) Stem-loop structure formed at 3′PK. (B) Primase, consisting of nonstructural proteins (nsp)7, 8, and 9, binds to the P5 stem. (C) Primase transcribes a primer for negative-sense genomic RNA (shown in blue) and (D) after the primer forms at the 3′ end of genomic RNA, the P5 stem is left unpaired, thus causing a pseudoknot structure. (E) Nsp12, acting as RNA-dependent RNA polymerase (RdRp), binds to the pseudoknot structure and an nsp7/nsp8/nsp9 complex and then elongates negative-sense genomic RNA.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

been difficult to detect, especially if it is formed only when RNA transcription is initiated, as the model by Züst et al. proposed. Our experiments, undertaken through NMR spectroscopy of SARS-CoV-2 3′PK fragments that were identified to stabilize one of the secondary structure forms, demonstrate that 3′PK can form a secondary structure that contains either a pseudoknot or a stem-loop. A comparison of the two forms of the secondary structure showed that the P5 stem, which includes the last 10 nucleotides at the 3′ end, was formed only in the stem-loop structure and not in the pseudoknot structure. In contrast, the P2 stem existed in both structures and exhibited only minor structural differences between the two forms. Thus, our results indicate that the conformation switch from the stem-loop structure to the pseudoknot structure allows the putative primase to transcribe the 3′ end of a positive, single-stranded sequence. We also examined interactions of the stem-loop and pseudoknot forms with the primase complexes of nsp7 and nsp8. Imino proton signal NMR analysis of the stem-loop structure showed that after the binding of nsp7/nsp8, the P5 stem was destabilized, whereas the pseudoknot RNA largely retained its secondary structure in the complex. Our results suggest that the binding mode of the nsp7/nsp8 complex to RNA is modulated by the RNA structure.

Our findings provide strong experimental evidence for the current hypothetical model by Züst et al.[5] of the initial events of genomic replication of the SARS-CoV-2 virus, which is widely accepted but has lacked structural evidence. In addition, our experimental data support the new possibility that the pseudoknot structure could be formed even when the primase complex does not synthesize an RNA primer, which may change some aspects of the existing mechanistic model of the initial events of RNA replication.

2. RESULTS

2.1. Two Secondary Structures Formed in the 3′PK Region Imply Their Role in a Structural Switch

First, to define the structured region and plausible structures of 3′PK, we prepared a series of RNA fragments that were excised from 3′PK (Table 1 and Supporting Information, and Materials and Methods). For the frame of these fragments, the secondary structure model determined by Rangan et al. was used as a ref 11. Note that the fragments excised from the original sequence do not involve any artificial modifications to stabilize particular structures. 2D ^{1}H−^{1}H NOESY spectra of the four designed RNA fragments (i.e., PK, SL, PKP4, and SLP4) are shown in Figure 3 as (A−C), (E−G), (I,K), respectively, together with their secondary structures elucidated from the NMR data (M−P). It should be noted that PK and SL formed pseudoknot and stem-loop structures, respectively, and these fragments were named after the structures. In addition, 2D ^{1}H−^{15}N HMQC spectra were collected for ^{13}C and ^{15}N-labeled samples of (D) PK, (H) SL, (J) PKP4, and (L) SLP4. To generate PKP4, we added the sequence for the loop between P4 and P5 to the PK fragment; in the SLP4 fragment, the sequence for the P4 stem was added to the SL; to generate PKP4, we added the sequence for the loop between P4 and P5 to the PK fragment.

The secondary structure of the entire SARS-CoV-2 genomic RNA has been reported by several groups,[30−34,36,37] and their analyses suggest that the 3′PK region forms a stem-loop structure.[30−34,37] In contrast, our NMR-based analysis

Table 1. RNA Sequences of Fragments

Name	Region, Length(nt)	Sequence
SL	29,622 - 29,657 and 29,857 - 29,867 49	GGAUUCUCGUAACUACAU AGCACAAGUAGAUGUAGU UACUUAGGAGAAU
PK	29,607 - 29,657, 54	GGGUCUUGUGCAGAAUGA AUUCUCGUAACUACAUAG CACAAGUAGAUGUAGUUA
SLL	29,622 - 29,657 and 29,850 - 29,867, 55	GGAUUCUCGUAACUACAU AGCACAAGUAGAUGUAGU UAUAGCUUCUUAGGAGAA U
SLP4	29,622 - 29,671 and 29,840 - 29,867[a], 85	GGGAUUCUCGUAACUACA UAGCACAAGUAGAUGUAG UUAACUUUAAUCUCACAG CAAUGUGAUUUUAAUAGC UUCUUAGGAGAAU
PKP4	29,608 - 29,671 and 29,840 - 29,867[a], 99	GGGCUUGUGCAGAAUGAA UUCUCGUAACUACAUAGC ACAAGUAGAUGUAGUUAA CUUUAAUCUCACAGCAAU GUGAUUUUAAUAGCUUCU UAGGAGAAU

[a]Connected by a GCAA linker. Structured regions were indicated in color, as shown in Figure 3. For clarity, the last three digits of the residue numbers are displayed in Figures 3, 5, and 6.

demonstrated that, in the PK and PKP4 fragments, the 3′PK region is capable of folding into a pseudoknot structure, which became apparent from the imino proton resonances. In general, imino proton resonances are not observed for RNA that does not form a base pair. In contrast, their resonances become detectable when imino protons form base pairs. This is because the rate of exchange of the imino proton with the solvent water proton, which is generally too fast for NMR detection, becomes sufficiently low when an imino proton is involved in hydrogen bonding. Assignments for imino proton signals of PK, SL, PKP4, and SLP4 were achieved from ^{1}H−^{1}H NOESY spectra using unlabeled RNA fragments and ^{1}H−^{15}N HMQC spectra using ^{13}C and ^{15}N-labeled RNA fragments (see Figure 3 and Table S1). Some NOEs between imino protons of the P5 stem were weak; thus, the P5 stem was assigned using imino−imino and A H2-imino NOEs (Figures 3E and 6A). The detailed assignments of PK, SL, PKP4, and SLP4 are shown in Figures S1−S4. The chemical shifts for each stem were mostly identical between RNA fragments containing the stem in both the HMQC and NOESY spectra.

PK and PKP4 fragments exhibited imino proton signals in both the P1pk and P2 regions. In Figure 3B,I, the signals for three base pairs (U29,612−G29,611, U29,612−G29,613, and G29,639−G29,613; shown in purple) were attributed to the P1pk stem of PK and PKP4 on the basis of the NMR spectra of an RNA sequence that contains only the P1 single-stem structure (Figure S5A). U29,609 and U29,610 were observed only in the P1 single-stem structure and not in the PK or PKP4 fragments. Even when minor chemical shift changes were observed, the chemical shifts for the P1 single stem were consistent with those for P1pk being accommodated in a pseudoknot structure (PK and PKP4) (see Table S1 and Figure S5A). The peak positions for the P1pk region are also nearly identical in the HMQC spectra (see Figure 3D,J),

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

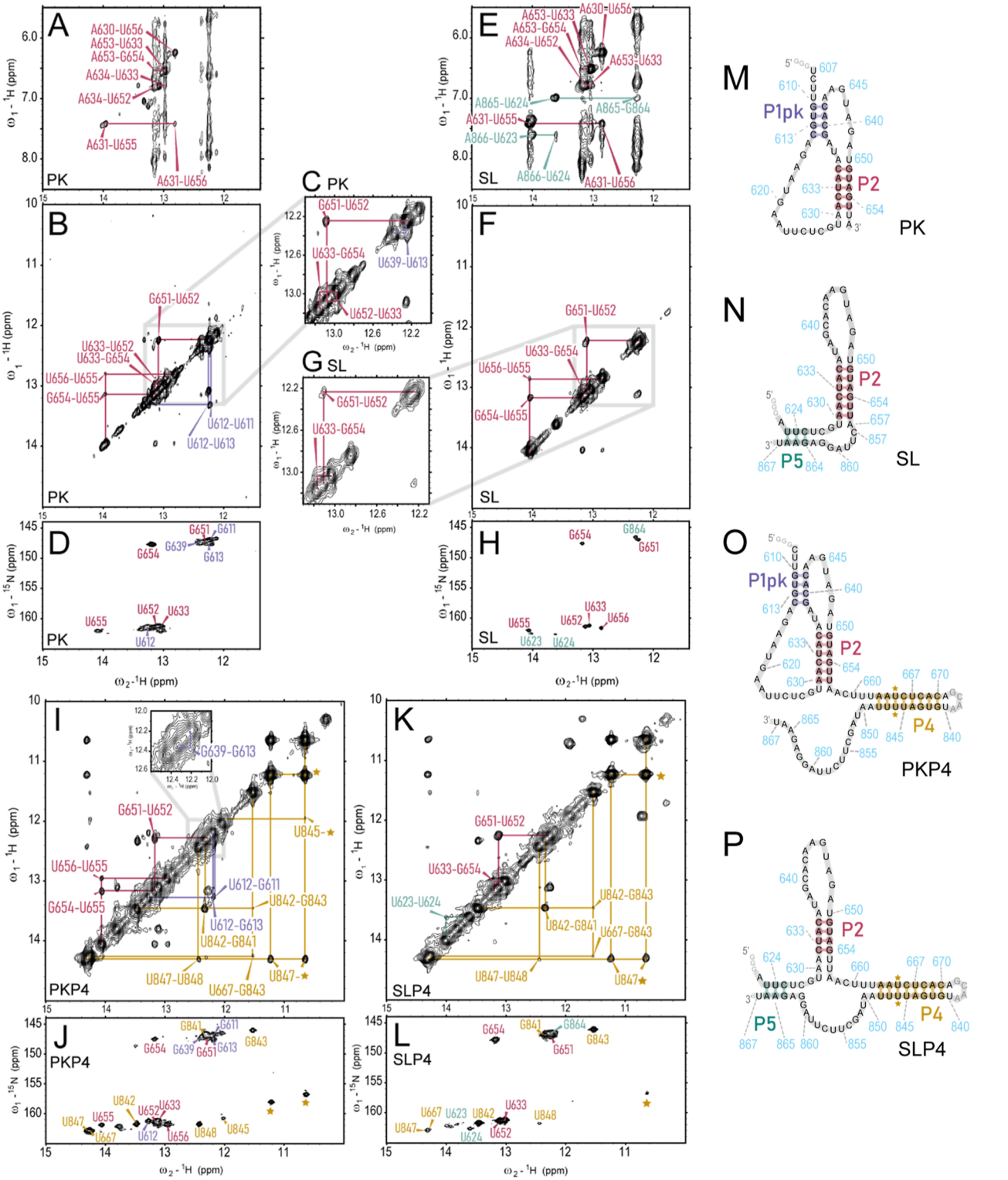


Figure 3. 2D ^{1}H−^{1}H NOESY spectra of (A) adenosine (A) H2-imino region, (B) imino−imino region of PK, and (C) 12.0−13.3 ppm region of (B) with (D) 2D ^{1}H−^{15}N HMQC spectrum of PK. 2D ^{1}H−^{1}H NOESY spectra of (E) H2-imino region, (F) imino−imino region of SL, and (G) 12.0−13.3 ppm region of (F) with (H) 2D ^{1}H−^{15}N HMQC spectrum of SL. 2D ^{1}H−^{1}H NOESY spectra of the imino region of (I) PKP4 and (K) SLP4 with 2D ^{1}H−^{15}N HMQC spectra of (J) PKP4 and (L) SLP4. NMR-based secondary structure of (M) PK, (N) SL, (O) PKP4, and (P) SLP4. (E,F) Were recorded at 298 K with a mixing time of 120 ms. The inset spectrum of (I) was recorded at 288 K with a mixing time of 120 ms. The other NOESY spectra were recorded at 288 K with a mixing time of 60 ms. All of the 2D ^{1}H−^{15}N HMQC spectra were recorded at 288 K. The assignment of each signal was color-coded, as described in the secondary structure of each fragment. The first two digits of the base number were omitted from the figures for simplicity. For example, U29,633 was indicated as U633. In (I−L), cross-peaks marked with asterisks are intrabase pair NOE between U665 and U846 imino protons or the ^{1}H−^{15}N imino cross-peak of U665/U846; those marked with U845/U847—* indicate correlations from U845/U847 to U665 or U846.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

strongly suggesting the consistent stem structure. Signals for the P2 stem in the PK and PKP4 fragments were identified from a comparison with those for SL and SLP4. P2 stem signals for the pseudoknot structure in PK and PKP4 fragments (shown in red in Figure 3B,I) were nearly identical with those for SL (Figure 3F). On the whole, the base-pairing patterns for PK and PKP4 were all consistent with the formation of the pseudoknot structure.

Interestingly, in the SLP4 fragment, imino proton signals of U29,655, U29,656, and U29,629 were not observed for P2 (Figure 3K,L). These three U−A pairs were also not observed in the 3_SL3base investigated by Wacker et al.,[37] which contains regions identical to those in SLP4. In contrast, we observed those signals for a fragment containing the P2 single stem (Figure S5B) and in SL (Figure 3F,H). SL contained a short bulge region (C29,857−G29,861) between P2 and P5 (Figure 3N) that corresponded to half of the full-length bulge-loop region between P4 and P5 (A29,850−G29,861) found in SLP4 (Figure 3P), whereas the P2 single stem had no bulge-loop region. Another designed fragment, called SLL, which has the P2 region and the P5 stem with a full-length bulge-loop, also showed very weak U29,655 and U29,656 signals, similar to those seen in SLP4 (see Figure S6). These data suggest that the bottom of the P2 region in SLP4 and SLL fragments adopts a noticeably different structure from that in other fragments (SL and P2 single stem) that lack the upstream section of the bulge-loop (A29,850−U29,856), possibly due to interaction between the bottom of P2 and this bulge-loop region. Indeed, the region between A29,850 and C29,854 is highly conserved among SARSr-CoV.[11] These results suggested that the upstream section of the bulge-loop region is likely to modulate the stem-loop structure and may be essential for recognition by primase or holo RdRp.

The P5 stem was formed only in fragments with a stem-loop structure (SL and SLP4; Figure 3E,H,K,L). Although PKP4 also contained this region, the P5 stem was not formed. Previous analysis of the whole genome RNA suggested the formation of both P1pk and P5 stems, and the data were presumed to be consistent with a mixture of two conformers, the pseudoknot, and the stem-loop;[36] our current study confirms, for the first time, that the pseudoknot structure can be formed by the 3′PK region.

2.2. Binding Assay Showed Both Stem-Loop and Pseudoknot RNA Bind to the nsp7/nsp8 Mixture with Comparably High Affinities

Next, by the use of an electrophoretic mobility shift assay, we examined the interactions of the four fragments that form either a stem-loop (SL and SLP4) or a pseudoknot (PK and PKP4) structure with nsp7 and/or nsp8. The RNA fragments mixed with nsp7 did not show any bands for the RNA/nsp7 complex. The fragments mixed with nsp8 or a mixture of nsp7 and nsp8 showed new shifted bands for RNA/protein complexes (Figure 4). These results were consistent with previous reports that nsp7 does not bind directly to duplex[38] or single-stranded[39] RNA. Our results also showed comparably high apparent affinities to the nsp7/nsp8 complex for PKP4 (K_d = 2.6 μM; Figure S7) and SLP4 (K_d = 2.4 μM; Figure S7). The results are consistent with a current hypothesis that the stem-loop RNA structure in 3′PK recruits the nsp7/nsp8 complex and changes its conformation to the pseudoknot structure while being bound to nsp7/nsp8.[5] The results also indicate a new possibility that the nsp7/nsp8 complex can bind

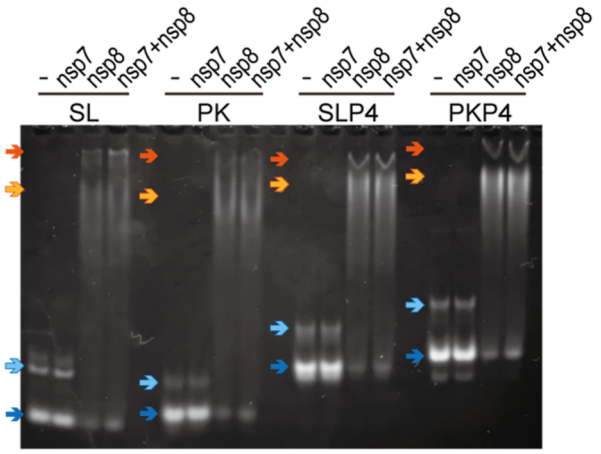


Figure 4. Gel image of the four RNA fragments (SL, PK, SLP4, and PKP4) tested for binding to nonstructural proteins (nsp)7 and nsp8. Bands indicated by blue and pale blue arrows correspond to folded and single-stranded RNA, respectively; those indicated by orange and yellow arrows correspond to protein-bound RNA and another conformer of protein-bound RNA, respectively. Bands in each RNA fragment shifted when nsp8 or the nsp7/nsp8 mixture was added. Reaction buffer solutions of 0 or 4.95 μM nsp7 and 0 or 4.95 μM nsp8 were mixed with 0.33 μM RNA.

directly to the pseudoknot form of 3′PK in equilibrium with the stem-loop form. The apparent affinities to nsp7/nsp8 were notably lower for PK (K_d = 5.3 μM) and SL(K_d = 5.7 μM; Figure S7). The results suggest that the P4 stem or the loop between P4 and P5 may be involved in protein binding. The apparent affinity of random sequence RNA rd85 (K_d = 3.7 μM; Figure S7) was comparable to the previously reported K_d for SARS-CoV nsp8 to dsRNA binding (~3.3 μM).[40] Rd85 has the same length as SLP4, but its K_d to nsp7/nsp8 was higher than those to SLP4 and PKP4, suggesting that the stem-loop or pseudoknot structure in 3′PK was mildly more favorable for nsp7/nsp8 binding than the normal duplex RNA structure. Our NMR analysis also demonstrated that PKP4 can stably form a complex when the RNA/nsp7/nsp8 ratio was 1:2:2, while SLP4 did not form a stable complex under the same condition. Considering that two nsp8s and one nsp7 are incorporated into the structure of the RNA-bound RdRp of SARS-CoV2[41] and nsp7 and nsp8 are reported to form a 2:2 complex,[42] it is possible that binding of the second nsp8s to the PKP4 site initiates the primase activity. As this work is mainly focused on the two major RNA structures for the 3′PK region, further studies are needed to evaluate this hypothesis. In each RNA fragment, the addition of nsp8 or a mixture of nsp7 and nsp8 gave slower mobility bands due to protein binding (orange and yellow arrows in Figure 4). The Hill coefficient derived from the Hill−Langmuir equation for each RNA fragment was greater than 1 (see Figure S7B). In particular, the Hill coefficient values for the PKP4 and SLP4 fragments were ~2.7; this means that multiple nsp7/nsp8 complexes were likely to interact with these RNA fragments. It has been reported that a mixture of nsp7 and nsp8 shows several multimer conformers with binding ratios such as nsp7/nsp8 = 2:2 and nsp7/nsp8 = 8:8;[38,40,43] therefore, these bands can be considered to reflect differences in protein multimeric conformation bound to the RNA fragments.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

2.3. NMR Data Suggest Structural Changes of RNA by Binding with nsp7/nsp8 Complex

Because the gel-shift assay showed that PKP4 and SLP4 RNA fragments have a higher affinity than PK and SL to the nsp7/nsp8 mixture, we decided to use PKP4 and SLP4 to obtain more detailed information on the interaction between RNA and the proteins. NMR signals of the nsp7 and nsp8 mixture without any RNA fragments showed notable line-broadening, suggesting nonuniform and/or a faster complexation rate than the NMR time scale (Figure 5C). Indeed, it was previously

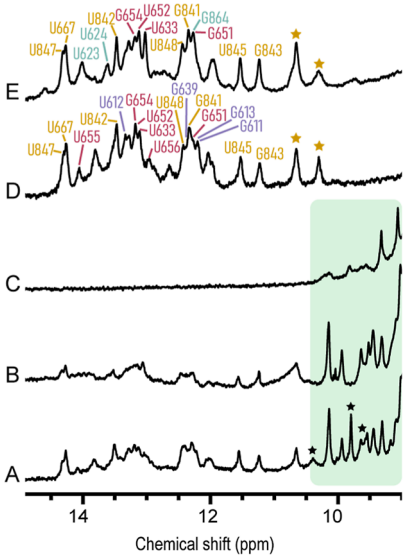


Figure 5. Structural changes caused by complexation in RNA with an nsp7/nsp8 mixture. The imino proton region of 1D 1H NMR spectra of (A) 1:2:2 mixture of PKP4/nsp7/nsp8, (B) 1:1:1 mixture of SLP4/nsp7/nsp8, (C) 1:1 mixture of nsp7/nsp8, (D) PKP4, and (E) SLP4. Asterisks mark signals that were observed only from the PKP4/nsp7/nsp8 mixture. Assignments for the free RNA signal are shown in color, as shown in Figure 3. The region exhibiting protein amide proton signals was highlighted in light green.

reported that a mixture of nsp7 and nsp8 exhibits a wide variety of multimer formation patterns, with the major component being the nsp7/nsp8 = 2:2 complex.[43,44] Adding PKP4 to the nsp7 and nsp8 proteins in a 1:2:2 ratio dramatically altered NMR spectral features for the proteins (Figure 5A), compared with those for an nsp7/nsp8 mixture without RNA (Figure 5C). In particular, several new sharp signals were observed at 9−10 ppm by adding RNA (Figure 5A). These new signals, which were attributed to amide proton signals of the proteins (see Figure S8), indicated that structural changes of the nsp7/nsp8 complex were induced upon the RNA binding. When SLP4 was added to the nsp7 and nsp8 proteins in a 1:1:1 ratio, similar prominent changes and signal sharpening occurred (Figure 5B) compared to the spectrum of the nsp7 and nsp8 proteins without RNA (Figure 5C). Note that we reduced the compositions of nsp7 and nsp8 because the 1:2:2 mixture of the samples yielded substantial line-broadening in the imino proton NMR signals for SLP4 (Figure S6), making it difficult to observe NOE signals. This may be due to the increase in molecular weight resulting from complex formation and/or due to exchange broadening for unstable complex formation. In contrast, the PKP4/nsp7/nsp8 = 1:1:1 mixture showed almost identical chemical shifts with broader spectral features compared with the 1:2:2 mixture (see Figure S9). The broadening may be attributed to the exchange between two different RNA-protein binding modes (or between proteins with and without RNA). The appearance of the sharp NMR signals shown in Figure 5A,B suggested the stabilization of a particular multimer of nsp7/nsp8 that was bound to the RNA fragments. Also, NOEs observed for the nsp7/nsp8 mixture were not observed for the SLP4/nsp7/nsp8 or PKP4/nsp7/nsp8 mixtures (see Figure S8). This finding suggests that most nsp7 and nsp8 proteins would form complexes in the mixtures with the RNA fragments PKP4 and SLP4. Figure 5A shows some signals due to protein (marked with asterisks) that were observed only for the PKP4/nsp7/nsp8 mixture. Thus, the modes of interaction of the proteins with the RNA fragments are likely different between PKP4 and SLP4.

Regarding NMR signals of RNA fragments, the RNA/protein mixture samples showed line-broadening at 10.5−14.5 ppm (Figure 5A,B) upon binding. It is unsurprising because the molar masses of the RNA/protein complexes are greater than those of the RNA fragments alone for both PKP4 and SLP4. Some imino proton signals for PKP4 and SLP4 mixed with the nsp7/nsp8 mixture (Figure 5A,B) exhibited chemical shift changes from those for the RNA fragments alone (Figure 5D,E). Figure 6 shows 2D $^{1}H/^{1}H$ NOESY spectra of (A) SLP4, (B) PKP4, (C) SLP4/nsp7/nsp8 mixture (1:1:1), and (D) PKP4/nsp7/nsp8 mixture (1:2:2). Both SLP4 and PKP4 showed few imino−imino NOEs when the nsp7/nsp8 mixture was added (see Figure S10). Therefore, intra-base pair NOEs between imino protons G H1 and U H3 at 11.5−14.5 ppm and C H41 and A H2 at 5.5−8.5 ppm, which were observed only for hydrogen-bonded base pairs, were used for further analysis. The spectral patterns for PKP4 bound to nsp7/nsp8 complexes (Figure 6D) are largely unchanged from those of free PKP4 (Figure 6B). After nsp7/nsp8 binding, PKP4 retained most of the cross-peaks due to inter-residue NOE (Figure 6D) for P1pk (purple arrows), P2 stem (red arrows), and P4 stem (orange arrows). The observed chemical shift changes were within 0.06 ppm for P1pk and P2 (Figure 6F). The results suggested that PKP4 binds to nsp7/nsp8 complexes while maintaining its pseudoknot structure (Figure 6B,D,F). In contrast, the spectral patterns are drastically different between free SLP4 (Figure 6A) and SLP4 bound to the nsp7/nsp8 complex (Figure 6C). The cross-peaks due to inter-residue NOE for the P2 stem (arrowed in red) and P4 (arrowed in orange) were observed in both spectra. Although line-broadening meant that some weak NOE signals between neighboring bases in the P2 stem were missing for SLP4 bound to nsp7/nsp8, all the strong intra-base pair NOE signals were retained. Interestingly, intra-base pair NOEs in the P5 region were not observed for SLP4 complexed with nsp7/nsp8 (Figure 6C), whereas unbound SLP4 showed clear NOEs for the P5 region (Figure 6A, yellow lines and arrows). Thus, the results suggested that binding by nsp7/nsp8 complexes destabilized the P5 stem. However, note that PKP4 has no P5 stem; thus, these results suggested that the transition from SLP4 to PKP4 may start from the destabilization of P5 upon nsp7/nsp8 binding. While there is some possibility that the signals for the P5 stem are subject to exchange broadening, the selective disappearance of these signals indicates that the transient binding of the nsp7/nsp8 complexes is likely to alter or destabilize the hydrogen-binding structure of the P5 stem substantially.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

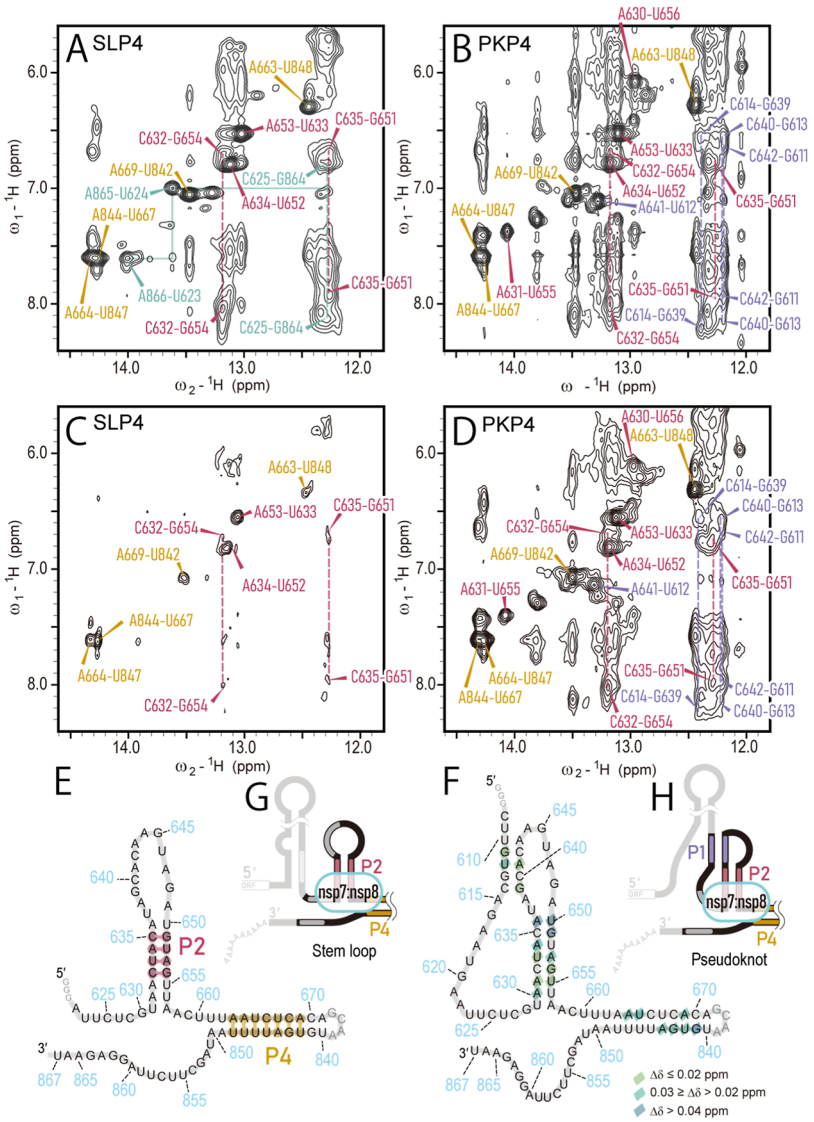


Figure 6. Structural changes in SLP4 and PKP4 were caused by binding with the nsp7/nsp8 mixture. Imino-base proton region of $^{1}H-^{1}H$ NOESY spectra of (A) SLP4, (B) PKP4, (C) a 1:1:1 mixture of SLP4/nsp7/nsp8, and (D) a 1:2:2 mixture of PKP4/nsp7/nsp8. (E) Schematic representation of the secondary structure for SLP4 bound to the nsp7/nsp8 complex. Destabilization of the P5 stem region (yellow) that existed in SLP4 was confirmed by the lack of intra-base pair NOE signals. (F) Schematic representation of the secondary structure for PKP4 bound to nsp7/nsp8 complexes, with the chemical shift, changes $\Delta\delta$ shown by color (see key). ^{1}H positions (A H2, C H41 or H42, U H3, and G H1) were omitted in figures (A–D). Schematic representation of (G) SLP4/nsp7/nsp8 and (H) PKP4/nsp7/nsp8 complexes. By binding of nsp7/nsp8, the P5 stem was single-stranded in both cases.

Both the PKP4 and SLP4 RNA fragments exhibited small but noticeable chemical shift changes in the P4 stem when a nsp7/nsp8 mixture was added, indicating that the P4 stem is likely to be involved in protein interaction (Figure 6F and Table S2). This is consistent with the fact that PKP4 and SLP4 have a higher apparent binding affinity than PK and SL, which lack the P4 region. Our NMR results supported the view that the interaction between P4 and the nsp7/nsp8 complex increased the affinity for PKP4 and SLP4.

3. DISCUSSION

Despite recent intensive studies on SARSr-CoV, including SARS-CoV-2, understanding the mechanism by which RNA transcription is initiated has remained elusive. Transcription of negative-strand RNA from the positive strand is the first and essential step in transcribing positive RNA and subgenomic RNA for the betacoronaviruses, including SARSr-CoV.[2,3] Genomic RNA is transcribed by virus-derived RdRp that consists of nsp12, nsp7, and nsp8.[18,19] Cryo-EM analysis showed the RdRp consisted of nsp12/nsp7/nsp8 = 1:1:2, with the template RNA lying in a hole at the interface between nsp12 and two nsp8.[18] In contrast, the binding ratio of the putative primase complexes, nsp7 and nsp8, was 2:2 (or 8:8).[38,40,43] Interestingly, the ratios of nsp7 to nsp8 of RdRp and putative primase were different. According to the hypothetical model by Züst et al.,[5] the transcription is initiated by nsp7/nsp8 complex binding to the stem-loop structure in 3′PK. Then, this binding prompts primer extension at the 3′ end of RNA and subsequently induces a structural change of 3′PK to a pseudoknot. However, the existence of a pseudoknot structure in 3′PK had not been experimentally determined. Here, we demonstrated that the 3′PK of SARS-CoV-2 could fold into not only a stem-loop but also a pseudoknot structure.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

Previous studies using a computational 3D structure prediction program[45] predicted that the pseudoknot structure included P1pk, P2, and P5 stems; however, our NMR analysis (Figure 3I) clearly suggested that the P5 stem was not included in the pseudoknot structure.

Previous chemical probing analyses for the whole genome[30−34] reported base pairings of A29,636−U29,650, U29,637−A29,649, G29,639−U29,646, and C29,640−G29,645 in 3′PK. In contrast, in our NMR study, these base pairs were not observed in SL and SLP4 fragments (Figure 3E,K). Except for A29,636−U29,650, these base pairs were also not observed in the 3_SL3base RNA construct that was previously reported.[37] These bases formed the P1pk stem or were placed at the junction of the P1pk and P2 regions when the pseudoknot structure was formed. Having experimentally determined the stem-loop and pseudoknot secondary structures for 3′PK, we can deduce that the unreactive bases located in the loop of the P2 stem in the chemical probing analysis may be attributed to the small amount of RNA forming the pseudoknot structure.

The gel-shift assay and NMR analysis of the RNA fragments with the nsp7/nsp8 mixture showed that the protein mixture can interact with not only a single stem-loop with a dangling end but also a branched stem-loop (SLP4) and a pseudoknot (PKP4). Both PKP4 and SLP4 fragments exhibited mild chemical shift changes in the ^{1}H NMR signals for the P4 stem when a nsp7/nsp8 mixture was added (Table S2) and had slightly higher apparent affinity than SL and PK. These results suggested that P4 is likely to be involved in nsp7/nsp8 binding. Our 2D NMR data also demonstrated that the P5 stem in SLP4 was destabilized upon nsp7/nsp8 binding. We also found that the P5 region of PKP4 did not form base pairs either when bound by nsp7/nsp8 complexes or when unbound. The results provide the first experimental support that the 3′ end of genomic RNA can be single-stranded after genomic RNA binds to nsp7/nsp8 in the 3′PK region (Figures 6G,H and S8).

Interestingly, NMR signals for the nsp7/nsp8 mixture showed significant chemical shift changes upon the addition of PKP4 or SLPK (highlighted in light green in Figure 5A,B). These signals with sharp spectral features were different from those of free nsp7,[46] free nsp8 (BMRB ID 51325) or the nsp7/nsp8 mixture without the RNA fragments. In addition, broad signals due to multimerization or fast exchange of proteins in the nsp7/nsp8 mixture were diminished by adding PKP4 or SLP4. Previously reported X-ray crystallography structures suggested that the binding ratio of the nsp7/nsp8 complex without RNA is a mixture of multiple conformers such as 2:2 and 8:8.[38,40,43] Our results suggested that binding to the RNA fragments was likely to induce a conformational change of nsp7/nsp8 complexes in varied multimerized states into a unique complex with a specific nsp7/nsp8 multimerized state. ^{1}H NMR signals at 9−10 ppm, which were attributed to the nsp7/nsp8 mixture, exhibited different chemical shifts in protein complexes with SLP4 from those in protein complexes with PKP4 (Figure 5A,B, asterisked signals). The results suggested that nsp7/nsp8 complexes bound with SLP4 or PKP4 have some structural differences, depending on the RNA structure. We do not yet know why the nsp7/nsp8 complex for the SLP4/nsp7/nsp8 mixture showed sharp signals for the 1:1:1 mixture but exhibited line-broadening for the 1:2:2 mixture. The difference might imply that the SLP4/nsp7/nsp8 complex was destabilized by loading more nsp7/nsp8, as suggested in Figure 2. Further studies are needed to examine the structural changes in the nsp7/nsp8 complex that are formed with SLP4 and PKP4, but this task lies outside the scope of this study, which focuses on structural changes in the 3′PK region upon nsp7/nsp8 binding.

In summary, our results provide the first experimental evidence of the pseudoknot structure, which is a key concept in a hypothetical model of the initiation of genomic RNA replication for SARS-CoV-2 and other SARSr-CoV complexes (Figure 2). Because this structure is rather rare and is critical for genomic RNA replication, our results indicate that PKP4 and related fragments can be a therapeutic target for developing an anticoronavirus drug, including against SARSr-CoV. The 3′PK regions are highly conserved among betacoronaviruses.[47] In addition, mutation analysis of SARS-CoV-2 and related viruses showed that rarely occurring mutations at the 3′PK region kept the stem-loop structure.[48] These results suggest that the binding agents specifically involved in forming the structure may potentially suppress virus proliferation. In addition, our results provide, for the first time, site-specific details of the RNA structures for the pseudoknot and stem-loop structures of 3′PK upon interaction with the nsp7/nsp8 complex. Furthermore, our findings show that the P5 stem of 3′PK is likely to change its structure to single-stranded (Figure S11B,C) once the primase complex is bound to 3′PK in its stem-loop form. Atomic details of how the primase interacts with 3′PK will be presented in a forthcoming work.

4. MATERIALS AND METHODS

4.1. Preparation of RNA

All RNA fragments were synthesized by T7 RNA polymerase and purified by electrophoresis. Further details of the RNA preparation and sequences of template DNA were described in the Supporting Information.

4.2. Preparation of nsp7 and nsp8

The SARS-CoV-2 nsp7 and nsp8 genes were cloned into the pCR2.1-TOPO vector (Invitrogen, Waltham, Massachusetts, USA). Each gene has an N-terminal Ni-binding tag, followed by a cleavage site for the HRV 3C protease. *Escherichia coli* KRX cells (Promega, Madison, Wisconsin, USA) were transformed by the resultant plasmid and were cultured for protein expression, which was induced by adding 0.1% rhamnose. The harvested cells were resuspended with buffer A [20 mM Tris−HCl pH 7.9 or pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM PMSF, and cOmplete protease inhibitors (Hoffmann-La Roche, Basel, Switzerland)] and were disrupted by sonication. After centrifugation of the cell lysates, the supernatants were mixed with Ni Sepharose 6 Fast Flow resin (Cytiva, Marlborough, Massachusetts, USA) and then were incubated at 4 °C for 1−1.5 h. The resin was transferred to a column and washed. To remove the N-terminal tag, the eluate from the resin was mixed with the HRV 3C protease solution and was dialyzed against buffer A without PMSF or cOmplete protease inhibitors. The dialyzed protein solution was further purified using Ni Sepharose 6 Fast Flow resin and a HiTrap Q HP column (Cytiva). Finally, the protein solution was subjected to gel filtration on Superdex 200 Increase 10/300 (Cytiva) equilibrated with buffer B (20 mM HEPES at pH 8.0, 150 mM NaCl, and 1 mM DTT). Fractions containing the protein of interest were combined and concentrated.

4.3. Electrophoretic Mobility-Shift Assay

Storage and reaction buffer components for RNA fragments and proteins were 20 mM HEPES (pH 7.4, adjusted by NaOH), 140 mM KCl, and 0.8 mM $MgCl_2$. Reaction-buffer solutions of 0 or 4.95 μM nsp7 and 0 or 4.95 μM nsp8 were mixed with 0.33 μM RNA. The mixed solutions were incubated at 20 °C for 10 min and then loaded onto a 6% polyacrylamide (19:1 = acrylamide/bis) gel containing 35

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

mM HEPES, 43 mM imidazole, and 10 mM KCl buffer and run at 100 V for 60 min at 4 °C. Buffer components for electrophoresis were selected to obtain pH 7.4.[49]

Bands were stained with GelStar Nucleic Acid Gel Stain (Lonza, Basel, Switzerland) and visualized with a 470 nm blue LED light. An apparent dissociation constant (K_d) was obtained from the Hill–Langmuir equation below

$$\theta = \frac{[L]^n}{K_d + [L]^n}$$

θ was obtained from {1 − (free RNA/total RNA)}, $[L]$ was [nsp8]. Free RNA/total RNA was determined from the ratio of RNA not mixed with protein to free RNA in each RNA–protein mixture.

4.4. NMR Measurements

All the RNA samples were dissolved in 20 mM MES (pH 6.0 adjusted by NaOH), 140 mM KCl, and 0.8 mM $MgCl_2$, containing 8 or 100% D_2O, with the solution then adjusted to 56–334 μM. RNA and protein mixture samples were dissolved in 20 mM MES (pH 6.0, adjusted by NaOH), 140 mM KCl, 0.8 mM $MgCl_2$, 1 mM DTT, and 40 units RiboGuard RNase inhibitor (LGC Biosearch Technologies), containing 8% D_2O with the solution then adjusted to 50 μM SLP4, 50 μM nsp7, and 50 μM nsp8 for the SLP4/nsp7/nsp8 mixture and 50 μM SLP4, 100 μM nsp7, and 100 μM nsp8 for the PKP4/nsp7/nsp8 mixture. 4 mm diameter Shigemi tubes (Shigemi, Tokyo, Japan) were used for all the NMR samples. Monovalent and divalent cation concentrations of NMR buffer were selected based on the ionic strength of the cytosol.[50]

The NMR spectra were recorded on Bruker Avance 700, 800, and 900 MHz spectrometers with cryogenic probes (Bruker Biospin, Billerica, Massachusetts, USA). 1D ^{1}H NMR spectra were collected with data points of 8k and a spectral width of 25 ppm at 288 K. The spectral widths (the number of total data points) of each ^{1}H–^{1}H 2D NOESY spectra were 25 ppm (2048 or 4096) for the direct dimension and 25 ppm (512 or 600) for the indirect dimension, with a mixing time of 60 or 120 ms and a recycle delay of 1.5 s at 288 or 298 K. The spectral widths (number of total data points) of each ^{1}H–^{15}N HMQC spectrum was 25 ppm (4096) for the direct dimension and 32 ppm (512 or 600) for the indirect dimension with a J_{HN} = 120 Hz at 288 or 298 K. Water suppression was achieved with a Watergate 3–9–19 pulse sequence. Spectra were analyzed by using NMRFAM-SPARKY NMR assignment and integration software[51] (https://nmrfam.wisc.edu/nmrfam-sparky-distribution/) and TopSpin 3.6.2 NMR processing software (Bruker Biospin).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.3c00641.

Supporting Information and methods and additional NMR and gel shift data, and chemical shift table for the P1, P2, PK, SL, PKP4, and SLP4 RNA fragments (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Takako Ohyama − *Laboratory for Advanced NMR Application and Development, Center for Biosystems Dynamics Research, RIKEN, Yokohama 230-0045 Kanagawa, Japan; School of Life Science and Technology, Tokyo Institute of Technology, Yokohama 226-8503 Kanagawa, Japan;* orcid.org/0000-0002-9269-6732; Email: takako.ohyama@riken.jp

Yoshitaka Ishii − *Laboratory for Advanced NMR Application and Development, Center for Biosystems Dynamics Research, RIKEN, Yokohama 230-0045 Kanagawa, Japan; School of Life Science and Technology, Tokyo Institute of Technology, Yokohama 226-8503 Kanagawa, Japan;* orcid.org/0000-0002-7724-6469; Email: ishii@bio.titech.ac.jp

Authors

Takuo Osawa − *Laboratory for Transcription Structural Biology, Center for Biosystems Dynamics Research, RIKEN, Yokohama 230-0045 Kanagawa, Japan*

Shun-ichi Sekine − *Laboratory for Transcription Structural Biology, Center for Biosystems Dynamics Research, RIKEN, Yokohama 230-0045 Kanagawa, Japan;* orcid.org/0000-0001-8174-8704

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.3c00641

Author Contributions

T. Oh and Y.I. conceived the project, designed the experiments, and interpreted the data. T. Oh performed and analyzed most of the experiments, and prepared the RNA samples. T. Os and S.S. expressed and purified the protein samples for the gel shift assay and NMR experiments. T. Oh, T. Os, S.S., and Y.I. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Mikako Shirouzu and Yoshinori Yanagisawa (Laboratory for Protein Functional and Structural Biology, RIKEN Center for Biosystems Dynamics Research) for their helpful comments and discussions. We also thank Mie Goto and Mari Aoki (Laboratory for Transcription Structural Biology, RIKEN Center for Biosystems Dynamics Research) for their help in protein preparation. This work was supported by the JST-Mirai Program (grant no. JPMJMI17A2, Japan) to Y. I and President's Discretionary Funds of RIKEN.

■ ABBREVIATIONS

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; nsp7, nonstructural protein 7; nsp8, nonstructural protein 8; RdRp, RNA dependent RNA polymerase; NMR, nuclear magnetic resonance; NOE, nuclear overhauser effect; NOESY, nuclear overhauser effect spectroscopy; HMQC, heteronuclear multiple quantum correlation

■ REFERENCES

(1) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579* (7798), 265–269.

(2) Mizutani, T.; Repass, J. F.; Makino, S. Nascent synthesis of leader sequence-containing subgenomic mRNAs in coronavirus genome-length replicative intermediate RNA. *Virology* **2000**, *275* (2), 238–243.

(3) Sethna, P. B.; Hung, S. L.; Brian, D. A. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (14), 5626–5630.

(4) Williams, G. D.; Chang, R.-Y.; Brian, D. A. A Phylogenetically Conserved Hairpin-Type 3′ Untranslated Region Pseudoknot Functions in Coronavirus RNA Replication. *J. Virol.* **1999**, *73* (10), 8349–8355.

(5) Züst, R.; Miller, T. B.; Goebel, S. J.; Thiel, V.; Masters, P. S. Genetic interactions between an essential 3′ cis-acting RNA pseudoknot, replicase gene products, and the extreme 3′ end of the mouse coronavirus genome. *J. Virol.* **2008**, *82* (3), 1214–1228.

(6) Lin, Y. J.; Zhang, X.; Wu, R. C.; Lai, M. M. The 3′ untranslated region of coronavirus RNA is required for subgenomic mRNA transcription from a defective interfering RNA. *J. Virol.* **1996**, *70* (10), 7236−7240.
(7) Goebel, S. J.; Hsue, B.; Dombrowski, T. F.; Masters, P. S. Characterization of the RNA Components of a Putative Molecular Switch in the 3′ Untranslated Region of the Murine Coronavirus Genome. *J. Virol.* **2004**, *78* (2), 669−682.
(8) Goebel, S. J.; Taylor, J.; Masters, P. S. The 3′ cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J. Virol.* **2004**, *78* (14), 7846−7851.
(9) Yang, D.; Leibowitz, J. L. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus Res.* **2015**, *206*, 120−133.
(10) Zhao, J.; Qiu, J.; Aryal, S.; Hackett, J. L.; Wang, J. The RNA Architecture of the SARS-CoV-2 3′-Untranslated Region. *Viruses* **2020**, *12* (12), 1473.
(11) Rangan, R.; Zheludev, I. N.; Hagey, R. J.; Pham, E. A.; Wayment-Steele, H. K.; Glenn, J. S.; Das, R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA* **2020**, *26* (8), 937−959.
(12) Jungreis, I.; Sealfon, R.; Kellis, M. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. *Nat. Commun.* **2021**, *12* (1), 2642.
(13) Brierley, I.; Dos Ramos, F. J. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.* **2006**, *119* (1), 29−42.
(14) Yan, S.; Zhu, Q.; Hohl, J.; Dong, A.; Schlick, T. Evolution of coronavirus frameshifting elements: Competing stem networks explain conservation and variability. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (20), No. e2221324120.
(15) Feng, J.; Funk, W. D.; Wang, S.-S.; Weinrich, S. L.; Avilion, A. A.; Chiu, C.-P.; Adams, R. R.; Chang, E.; Allsopp, R. C.; Yu, J.; et al. The RNA Component of Human Telomerase. *Science* **1995**, *269* (5228), 1236−1241.
(16) Dreher, T. W. Role of tRNA-like structures in controlling plant virus replication. *Virus Res.* **2009**, *139* (2), 217−229.
(17) Subissi, L.; Posthuma, C. C.; Collet, A.; Zevenhoven-Dobbe, J. C.; Gorbalenya, A. E.; Decroly, E.; Snijder, E. J.; Canard, B.; Imbert, I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (37), No. E3900.
(18) Hillen, H. S.; Kokic, G.; Farnung, L.; Dienemann, C.; Tegunov, D.; Cramer, P. Structure of replicating SARS-CoV-2 polymerase. *Nature* **2020**, *584* (7819), 154−156.
(19) Kirchdoerfer, R. N.; Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **2019**, *10* (1), 2342.
(20) te Velthuis, A. J.; Arnold, J. J.; Cameron, C. E.; van den Worm, S. H.; Snijder, E. J. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res.* **2010**, *38* (1), 203−214.
(21) Imbert, I.; Guillemot, J.-C.; Bourhis, J.-M.; Bussetta, C.; Coutard, B.; Egloff, M.-P.; Ferron, F.; Gorbalenya, A. E.; Canard, B. A second, non-canonical RNA-dependent RNA polymerase in SARS Coronavirus. *EMBO J.* **2006**, *25* (20), 4933−4942.
(22) Imbert, I.; Guillemot, J. C.; Bourhis, J. M.; Bussetta, C.; Coutard, B.; Egloff, M. P.; Ferron, F.; Gorbalenya, A. E.; Canard, B. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.* **2006**, *25* (20), 4933−4942.
(23) Park, G. J.; Osinski, A.; Hernandez, G.; Eitson, J. L.; Majumdar, A.; Tonelli, M.; Henzler-Wildman, K.; Pawłowski, K.; Chen, Z.; Li, Y.; et al. The mechanism of RNA capping by SARS-CoV-2. *Nature* **2022**, *609* (7928), 793−800.
(24) Yan, L.; Zhang, Y.; Ge, J.; Zheng, L.; Gao, Y.; Wang, T.; Jia, Z.; Wang, H.; Huang, Y.; Li, M.; et al. Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat. Commun.* **2020**, *11* (1), 5874.
(25) Egloff, M.-P.; Ferron, F.; Campanacci, V.; Longhi, S.; Rancurel, C.; Dutartre, H.; Snijder, E. J.; Gorbalenya, A. E.; Cambillau, C.; Canard, B. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (11), 3792−3796.
(26) de O Araújo, J.; Pinheiro, S.; Zamora, W. J.; Alves, C. N.; Lameira, J.; Lima, A. H. Structural, energetic and lipophilic analysis of SARS-CoV-2 non-structural protein 9 (NSP9). *Sci. Rep.* **2021**, *11* (1), 23003.
(27) Littler, D. R.; Gully, B. S.; Colson, R. N.; Rossjohn, J. Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. *iScience* **2020**, *23* (7), 101258.
(28) Sutton, G.; Fry, E.; Carter, L.; Sainsbury, S.; Walter, T.; Nettleship, J.; Berrow, N.; Owens, R.; Gilbert, R.; Davidson, A.; et al. The nsp9 Replicase Protein of SARS-Coronavirus, Structure and Functional Insights. *Structure* **2004**, *12* (2), 341−353.
(29) El-Kamand, S.; Du Plessis, M.-D.; Breen, N.; Johnson, L.; Beard, S.; Kwan, A. H.; Richard, D. J.; Cubeddu, L.; Gamsjaeger, R. A distinct ssDNA/RNA binding interface in the Nsp9 protein from SARS-CoV-2. *Proteins: Struct., Funct., Bioinf.* **2022**, *90* (1), 176−185.
(30) Sun, L.; Li, P.; Ju, X.; Rao, J.; Huang, W.; Ren, L.; Zhang, S.; Xiong, T.; Xu, K.; Zhou, X.; et al. In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell* **2021**, *184* (7), 1865−1883.e20.
(31) Manfredonia, I.; Nithin, C.; Ponce-Salvatierra, A.; Ghosh, P.; Wirecki, T. K.; Marinus, T.; Ogando, N. S.; Snijder, E. J.; van Hemert, M. J.; Bujnicki, J. M.; et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **2020**, *48* (22), 12436−12452.
(32) Huston, N. C.; Wan, H.; Strine, M. S.; de Cesaris Araujo Tavares, R.; Wilen, C. B.; Pyle, A. M. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* **2021**, *81* (3), 584−598.e5.
(33) Lan, T. C. T.; Allan, M. F.; Malsick, L. E.; Woo, J. Z.; Zhu, C.; Zhang, F.; Khandwala, S.; Nyeo, S. S. Y.; Sun, Y.; Guo, J. U.; et al. Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.* **2022**, *13* (1), 1128.
(34) Cao, C.; Cai, Z.; Xiao, X.; Rao, J.; Chen, J.; Hu, N.; Yang, M.; Xing, X.; Wang, Y.; Li, M.; et al. The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.* **2021**, *12* (1), 3917.
(35) Stammler, S. N.; Cao, S.; Chen, S.-J.; Giedroc, D. P. A conserved RNA pseudoknot in a putative molecular switch domain of the 3′-untranslated region of coronaviruses is only marginally stable. *RNA* **2011**, *17* (9), 1747−1759.
(36) Zhang, Y.; Huang, K.; Xie, D.; Lau, J. Y.; Shen, W.; Li, P.; Wang, D.; Zou, Z.; Shi, S.; Ren, H.; et al. In vivo structure and dynamics of the SARS-CoV-2 RNA genome. *Nat. Commun.* **2021**, *12* (1), 5695.
(37) Wacker, A.; Weigand, J. E.; Akabayov, S. R.; Altincekic, N.; Bains, J. K.; Banijamali, E.; Binas, O.; Castillo-Martinez, J.; Cetiner, E.; Ceylan, B.; et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **2020**, *48* (22), 12415−12435.
(38) Zhai, Y.; Sun, F.; Li, X.; Pang, H.; Xu, X.; Bartlam, M.; Rao, Z. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **2005**, *12* (11), 980−986.
(39) Zhang, C.; Li, L.; He, J.; Chen, C.; Su, D. Nonstructural protein 7 and 8 complexes of SARS-CoV-2. *Protein Sci.* **2021**, *30* (4), 873−881.
(40) te Velthuis, A. J.; van den Worm, S. H.; Snijder, E. J. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Res.* **2012**, *40* (4), 1737−1747.
(41) Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **2020**, *368* (6492), 779−782.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333

(42) Konkolova, E.; Klima, M.; Nencka, R.; Boura, E. Structural analysis of the putative SARS-CoV-2 primase complex. *J. Struct. Biol.* **2020**, *211* (2), 107548.
(43) Biswal, M.; Diggs, S.; Xu, D.; Khudaverdyan, N.; Lu, J.; Fang, J.; Blaha, G.; Hai, R.; Song, J. Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Res.* **2021**, *49* (10), 5956−5966.
(44) Krichel, B.; Bylapudi, G.; Schmidt, C.; Blanchet, C.; Schubert, R.; Brings, L.; Koehler, M.; Zenobi, R.; Svergun, D.; Lorenzen, K.; et al. Hallmarks of Alpha- and Betacoronavirus non-structural protein 7 + 8 complexes. *Sci. Adv.* **2021**, *7* (10), No. eabf1004.
(45) Rangan, R.; Watkins, A. M.; Chacon, J.; Kretsch, R.; Kladwang, W.; Zheludev, I. N.; Townley, J.; Rynge, M.; Thain, G.; Das, R. De novo 3D models of SARS-CoV-2 RNA elements from consensus experimental secondary structures. *Nucleic Acids Res.* **2021**, *49* (6), 3092−3108.
(46) Tonelli, M.; Rienstra, C.; Anderson, T. K.; Kirchdoerfer, R.; Henzler-Wildman, K. (1)H, (13)C, and (15)N backbone and side chain chemical shift assignments of the SARS-CoV-2 non-structural protein 7. *Biomol. NMR Assignments* **2021**, *15* (1), 73−77.
(47) Park, J. H.; Moon, J. Conserved 3′ UTR of Severe Acute Respiratory Syndrome Coronavirus 2: Potential Therapeutic Targets. *Front. Genet.* **2022**, *13*, 893141.
(48) Chan, A. P.; Choi, Y.; Schork, N. J. Conserved Genomic Terminals of SARS-CoV-2 as Coevolving Functional Elements and Potential Therapeutic Targets. *mSphere* **2020**, *5* (6), No. e00754.
(49) McLellan, T. Electrophoresis buffers for polyacrylamide gels at various pH. *Anal. Biochem.* **1982**, *126* (1), 94−99.
(50) Thier, S. O. Potassium physiology. *Am. J. Med.* **1986**, *80*, 3−7.
(51) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325−1327.

https://doi.org/10.1021/jacsau.3c00641
JACS Au 2024, 4, 1323−1333