1

Structural basis for polyuridine tract recognition by SARS-CoV-2 Nsp15

Authors

Fumiaki Ito[1,2,3], Hanjing Yang[1], Z. Hong Zhou[2,3], and Xiaojiang S. Chen[1,4,5,6*]

Affiliations

[1]Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA. [2]Department of Microbiology, Immunology and Molecular Genetics, and [3]California NanoSystems Institute, University of California, Los Angeles, CA90095, USA. [4]Genetic, Molecular and Cellular Biology Program, Keck School of Medicine, [5]Norris Comprehensive Cancer Center, and [6]Center of Excellence in NanoBiophysics, University of Southern California, Los Angeles, CA90089, USA.

[*]To whom correspondence should be addressed.

Tel: +1(213)740-5487; FAX: +1(213)740-4340;

Email: xiaojiac@usc.edu

Abstract

SARS-CoV-2 non-structural protein 15 (Nsp15) is critical for productive viral replication and evasion of host immunity. The uridine-specific endoribonuclease activity of Nsp15 mediates the cleavage of the polyuridine [poly(U)] tract of the negative-strand coronavirus genome to minimize the formation of dsRNA that activates the host antiviral interferon signaling. However, the molecular basis for the recognition and cleavage of the poly(U) tract by Nsp15 is incompletely understood. Here, we present cryogenic electron microscopy (cryoEM) structures of SARS-CoV-2 Nsp15 bound to viral replication intermediate dsRNA containing poly(U) tract at 2.7-3.3 Å resolution. The structures reveal one copy of dsRNA binds to the sidewall of an Nsp15 homohexamer, spanning three subunits in two distinct binding states. The target uracil is dislodged from the base-pairing of the dsRNA by amino acid residues W332 and M330 of Nsp15, and the dislodged base is entrapped at the endonuclease active site center. Up to 20 A/U base pairs are anchored on the Nsp15 hexamer, which explains the basis for a substantially shortened poly(U) sequence in the negative strand coronavirus genome compared to the long poly(A) tail in its positive strand. Our results provide mechanistic insights into the unique immune evasion strategy employed by coronavirus Nsp15.

Main

Coronaviruses have evolved a wide array of tactics to evade host antiviral immunity. When host cells detect invading foreign nucleic acids including viral genome and viral replication intermediates, they activate interferon (IFN) signaling via cytoplasmic pattern recognition receptors (PRRs) (1, 2). Remarkable stealth activities exhibited by coronaviruses are facilitated by a series of non-structural proteins (Nsps). For example, coronavirus Nsps possess unique machinery to perform 5′-end capping of their genomic and sub-genomic RNAs to mimic host mRNA, allowing them to escape from the antiviral action of host IFN while hijacking the host ribosome (3). During the initial stage of positive and negative strand RNA genome synthesis, viral RNA-dependent RNA polymerase (RdRp), which consists of Nsp7, Nsp8, and Nsp12, coordinates with viral capping enzymes and their associated cofactors such as Nsp9, Nsp10, Nsp13, Nsp14, and Nsp16. These Nsp proteins form a large replication-transcription complex (RTC) and add the $^{7Me}GpppA_{2'-OMe}$ cap at the 5′-end of the viral RNAs (4-7). This process enables coronaviruses to replicate outside the nucleus without the need to rely on host mRNA capping enzymes.

Another key disguising strategy during the coronavirus replication is carried out by Nsp15, which is conserved among known coronavirus lineages (8, 9). Nsp15 mediates the evasion of cytoplasmic dsRNA detection by host MDA5 and other PRRs in macrophages to suppress the antiviral IFN response (10, 11). Nsp15 is a uridine-specific endoribonuclease and is active on both ssRNA and dsRNA *in vitro* (12-14). Purified Nsp15 cleaves uridine-containing ssRNA with low specificity and dsRNA with high specificity (15, 16). Nsp15 is also a part of the coronavirus RTC assembly and its association with RTC is likely mediated by Nsp8, a known cofactor of RdRp (17-19).

Polyuridine [poly(U)] tract present in the coronavirus negative strand RNA, a transcription product of its positive strand RNA genome, has been identified as one of the physiological targets of Nsp15 (20). The study showed that an endonuclease-deficient mutant of Nsp15 caused the accumulation of dsRNA in the coronavirus-infected hepatocytes and that the negative strand genomic RNA is responsible for this dsRNA formation. They further demonstrated that Nsp15 limits the abundance and length of poly(U) within the negative strand RNA 5′-extension. The transfection of poly(U)-containing synthetic RNA into alpha mouse liver 12 cells caused the activation of MDA5-mediated IFN induction (20). During the coronavirus replication, a short poly(U) sequence in the negative strand is required to initiate the synthesis of a long poly(A) tail (100-130 bp) in the positive strand genome, which in turn is a prerequisite for the negative strand synthesis (21-23). Thus, Nsp15 plays a role in trimming down the initially synthesized poly(U) lead sequence to the optimal length that can suppress dsRNA formation but still serves as a template for poly(A) extension.

X-ray crystallography and cryogenic electron microscopy (cryoEM) studies conducted throughout the SARS-CoV-2 pandemic have advanced our understanding of the substrate binding mechanisms employed by Nsp15 for both ssRNA and dsRNA (24-26). However, the precise molecular mechanisms underlying the recognition of poly(U) by Nsp15 remain unknown. Here we reconstituted a complex of

70   SARS-CoV-2 Nsp15 with a viral replicative dsRNA intermediate containing 3′-end of the viral genome

71   followed by a 20-bp poly(A/U) extension. Cryogenic electron microscopy (cryoEM) and 3D classification

72   revealed Nsp15 hexamer structures at various functional states, including RNA-free and two RNA-bound

73   states. Comparison of these structures show that the poly(U) tract of the sequence is recognized by an Nsp15

74   hexamer via direct contacts with three subunits in two distinct states. The active site utilizes a base-flipping

75   mechanism to hold the target uracil base in the endonuclease catalytic center for cleavage. Overall, we

76   provide a snapshot of coronavirus Nsp15 bound to the poly(A/U) sequence of its genomic replicative

77   dsRNA intermediate for the evasion of the host antiviral response.

78

79   Results

80   CryoEM structures of apo- and RNA-bound Nsp15

81   To understand the mechanism of poly(U) targeting by SARS-CoV-2 Nsp15, we reconstituted a

82   ribonucleoprotein complex of Nsp15 and a 35-bp dsRNA substrate comprising the final 15-bp of 3′-end of

83   the SARS-CoV-2 genome and 20-bp poly(A/U) extension, which represents a coronavirus genome

84   replication intermediate. Nsp15 C-terminal domain (CTD) belongs to a family of endoribonuclease with

85   two conserved catalytic histidine residues (H234 and H249), which serve as general acid and general base,

86   respectively, to attack the 3′-phosphate of the target uridine (25, 27). To capture the RNA substrate bound

87   to Nsp15 and trap the target uracil base at the nuclease active site, catalytically inactive mutant H234A was

88   used for the reconstitution (Fig. 1A). SARS-CoV-2 Nsp8, a previously hypothesized cofactor of the Nsp15,

89   was also added to facilitate the Nsp15-RNA interaction (17, 18). CryoEM imaging and heterogeneous

90   refinement were performed to evaluate and optimize RNA binding in the reconstituted protein-RNA

91   complex (Fig. S1 and S2). An equimolar ratio of Nsp15:RNA resulted in mostly apo-Nsp15 (Data set 1 in

92   Fig. S2). We found that including an excess amount of RNA with the Nsp15:RNA ratio of 1:10 yielded a

93   complex with the available RNA binding sites within the functional oligomer of Nsp15 occupied by RNA

94   (Data sets 2 and 3 in Fig. S2).

95   2D classification of the reconstituted Nsp15-RNA complex particles in data sets 2 and 3 showed

96   that a subset of 2D class averages has extra helical densities stemming from the core of the 2D densities

97   (Fig. 1B). Multiclass *ab initio* 3D reconstruction showed that 25% of the particles belong to RNA-free apo-

98   Nsp15, and 39% belong to Nsp15 in complex with obvious dsRNA (Fig. S2). The apo-Nsp15 was

99   reconstructed as a homohexamer with D3 symmetry and refined to 2.3 Å resolution (Table S1). The

100   hexamer forms a barrel-like architecture with a central channel, which consists of a head-to-head stack of

101   two trimers (defined as top and bottom)(Figs. 1C, S3, and S4). The RNA-bound form was reconstructed as

102   a homohexamer with clear A-form like RNA duplex density and refined to 2.7 Å resolution (Table S1).

103   dsRNA density is diagonally attached on the outer periphery of the hexameric barrel and occupies a shallow

104   groove between two subunits of the top trimer (defined as subunit A and B)(Figs. 1D, S5, and S6). The

105   nuclease active site centers are located near the subunit interface between the neighboring subunits within
106   the top or bottom trimers and the bound dsRNA fully shields the active site of the subunit A. One end of
107   the dsRNA stretches along a groove between the CTD of subunit A and the middle domain (MD) of subunit
108   B, extending upwards beyond the top trimer. The other end of the dsRNA stretches towards the MD of the
109   subunit D within the bottom trimer. No extra densities were observed in the central channel of the hexamer.
110   Although we added an excess quantity of the substrate RNA (10-fold of Nsp15 in molarity), we only
111   observed Nsp15 hexamer with a single piece of dsRNA as a substrate-bound form. The active sites of the
112   other five subunits are unoccupied. Extensive 3D classification did not yield any 3D classes containing
113   more than one dsRNA piece per Nsp15 hexamer. These observations likely indicate that Nsp15 hexamer is
114   compatible with only one dsRNA substrate at a time.

115   We noticed that both ends of the bound dsRNA had relatively weak density (Fig. 1D). We therefore
116   hypothesized that there is conformational or compositional variability in the bound RNA structure and that
117   the observed RNA-bound form could be an average of multiple different states. Further heterogeneous 3D
118   refinement revealed that the dsRNA-bound form was subclassified into two states: A class with dsRNA
119   extending towards subunit B in the upper trimer (state 1) and the other class with dsRNA extending towards
120   subunit D in the bottom trimer (state 2). In the state 1 structure, the full 20 A/U pairs are readily visible,
121   including 11-bp preceding the target U at the active site (defined as U_0) and 9-bp following the U_0. The
122   state 2 structure also showed the full 20 A/U pairs, including 3-bp preceding the U_0 and 16-bp following
123   the U_0 (Figs. 2A and S7). Taken together, these observations indicate that a single piece of dsRNA can
124   engage in the Nsp15 hexamer during poly(U) RNA targeting and that a range of uridines within the poly(U)
125   tract can be targeted for cleavage.

126

127   Poly(U) uracil base flipping at active site center

128   The dsRNA is firmly held by the Nsp15 hexamer with a slight bent near the groove between the CTD of
129   subunit A and the NTD of subunit B (Fig. 2B). Within the central region of the dsRNA, we observed
130   unpaired bases: one base is flipped outside from the RNA duplex while its complementary counterpart
131   remains orienting inwards within the duplex (Fig. 3A). U/A pair was modeled at this location and the flipped
132   uracil is designated as U_0. The local resolution of the bound RNA ranges between 2.4 Å to 4.0 Å with the
133   highest resolution around the flipped U_0 base (Fig. S5). A total of 17 A/U base pairs were confidently built
134   for the consensus Nsp15-RNA structure including the 10-bp preceding the flipped U_0 and 6-bp following
135   the U_0. Outside this patch, the base densities are insufficiently featured to distinguish their identity.

136   At the endonuclease active site pocket, the pyrimidine ring of the flipped uracil base is sandwiched
137   between two hydrophobic residues Y342 and V291. The aliphatic chain of K344 additionally constitutes
138   this hydrophobic pocket to hold the uracil base. S293 plays a key role in conferring the selectivity for the
139   target uracil as a hydroxyl of the S293 side chain recognizes the N3 atom of the uracil base to form a
140   hydrogen bond. The main chain nitrogen of S293 forms another hydrogen bond with the O2 atom of the

141 uracil base. The catalytic H249 is located near the scissile 3′-phosphate of U_0 and its imidazole ring forms
142 a hydrogen bond with the 2′-OH of the ribose ring of U_0. As expected, the inactivated catalytic histidine
143 (H234A) is located near H249 and the scissile 3′-phosphate. Two polar residues Q244 and K289
144 additionally surround the 3′-phosphate, thereby stabilizing the position of the target uridine (Fig. 3B and
145 C).

146   The space that would have been occupied by the U_0 base within the RNA duplex is partially
147 occupied by W332 from the three-stranded anti-parallel β-sheet of CTD. W332 together with M330 and
148 Y342 from the same β-sheet create a hydrophobic surface and intercalates into the open major groove at
149 this location (Fig. 3D and E). Notably, W332 positions itself directly across the orphan A_0 base and
150 stabilizes the adjacent U_{+1} base by forming a stacking interaction. These three residues, M330, W332, and
151 Y342, responsible for dislodging the target uracil base are completely conserved across coronaviruses,
152 highlighting their importance (Fig. S8). Two polar residues K334 and E339 near the top edge of the β-sheet
153 interact with the U_{+2} and U_{+1} in the negative strand, respectively. A primary amine of K334 forms hydrogen
154 bonds with the O2 atom of the U_{+2} base, and 2′-OH of the ribose ring of U_{+2}. A side chain carboxyl of E339
155 forms hydrogen bonds with 2′-OH of the ribose ring of U_{+1} (Fig. 3D and E).

156

157 **Nsp15-dsRNA interaction at multiple locations**

158 The dsRNA substrate contacts three subunits (subunits A, B, and D) extensively on the sidewall of the
159 Nsp15 hexameric barrel (Fig. 2B). Both the negative and positive strands make substantial contacts with
160 Nsp15 across approximately two and a half turns of the double-stranded helix. Aside from the base pair that
161 involves flipped U_0, base pairing is well-maintained throughout the duplex. Outside the endonuclease active
162 site pocket anchoring the flipped U_0 base, subunit A has two additional interface areas on either direction
163 of the active site (Fig. 4A and B). The first interface area includes the first strand of the three-stranded β-
164 sheet in the CTD. A patch of -^{314}VSKV^{317}- in β13 is in close contact with the orphan nucleotide A_0 and A_{+1}
165 in the positive strand RNA. S315 forms a hydrogen bond with the 2′-hydroxyl group of the ribose ring of
166 the A_0, while the backbone nitrogen of V317 forms another hydrogen bond with the 2′-hydroxyl of the
167 ribose ring of A_{+1}. The second interface in subunit A is near a patch of -^{242}HSQ^{244}- within a loop region,
168 which supports the backbone of the positive strand between positions A_{-8} to A_{-6}. Subunit B, on the other
169 hand, interacts with the RNA through its MD and NTD (Fig. 4C and D). The surface area comprised of
170 hydrophilic residues S128, E145, S147 and K173 in subunit B's MD contacts the major groove near U_{+8} to
171 U_{+10} in the negative strand and A_{-8} to A_{-6} in the positive strand (Fig. 4C). Another polar surface comprised
172 of residues K12, D16, Q18, and Q19 in subunit B's NTD contacts the backbone of the negative strand near
173 U_{-4} to U_{-1} (Fig. 4D). Lastly, Nsp15-RNA state 2 structure has an additional interface in subunit D from the
174 bottom trimer. A hydrophilic surface comprised of the polar residues K110, T112, E113, D132, N136, and
175 R135 in subunit D's MD contacts the minor groove area near U_{-14}, and U_{-13} in the negative strand and A_{+15},
176 and A_{+16} in the positive strand, further stabilizing the bound dsRNA (Fig. 4E).

177

Structural remodeling by dsRNA binding

179 Comparison of subunit A structures in its RNA-bound form and its apo-form showed an r.m.s.d. of 0.388
180 Å, indicating that the binding of RNA does not induce significant global conformational changes to Nsp15
181 protomer. Yet, the high-resolution structures of both apo- and RNA-bound forms of Nsp15 allowed us to
182 identify notable local structural changes. First, upon binding of RNA, there is a subtle linear shift of β-
183 strands 14 and 15, accompanied by their connecting β-turn ($-^{334}$KDGH^{337}-), towards the bound dsRNA.
184 This shift is likely caused by the presence of W332 on the β-strand 14 (Fig. S9A). Second, the C-terminal
185 tail of the subunit A including the terminal residues -^{344}KLQ^{346} underwent a structural remodeling upon
186 RNA binding. Superimposition of the apo-Nsp15 and RNA-bound subunit A structures showed that the C-
187 terminal end glutamine residue clashes with the flipped uracil base. In the RNA-bound structure, the
188 ^{344}KLQ^{346} patch swung away from the active site pocket, allowing the substrate uracil base to fit into the
189 pocket (Fig. S9B). The rest of the subunits (B to F) did not display any noticeable structural changes upon
190 RNA binding. Interestingly, the cryoEM density for residues W332 and M330 in subunit A, which play key
191 roles in the base-flipping of the target uracil, are more clearly defined than the same residues in the other
192 subunits, indicating that the side chains of these residues are stabilized by the RNA binding (Fig. S9C).

193

Discussion

195 Nsp15 is one of the indispensable components of the non-structural proteins that are initially expressed as
196 a large polyprotein and subsequently proteolytically cleaved to yield individual functional enzymes. Nsp15
197 plays a pivotal role in facilitating viral infection. Inactivation of its endonuclease activity severely
198 compromises the virus fitness, thereby making it a promising therapeutic target (10, 11). Our understanding
199 of Nsp15's physiological role remains limited due to a lack of information about its array of physiological
200 substrates, the regulatory mechanisms governing its endonuclease activity, and its association with other
201 viral or host proteins. Nsp15 targets the extended poly(U) lead sequence at the 5′-end of the coronavirus
202 negative strand genome. This activity curtails the accumulation of dsRNA, effectively suppressing the
203 activation of the antiviral response mediated by the cytoplasmic dsRNA sensor MDA5 (20). Our cryoEM
204 structures of SARS-CoV-2 Nsp15 in complex with its substrate RNA provide direct evidence of its
205 association with poly(U) in its double-stranded form, which represents a replication intermediate during the
206 initial synthesis of poly(U) tract using positive-strand poly(A) tail as a template. Our results also show that
207 the Nsp15-RNA complex exists at least in two distinct states, indicating that a range of uridines within the
208 poly(U) tract can be recognized for cleavage in the reconstituted system. The target uracil base is flipped
209 out from the duplex and is captured at the endonuclease active site pocket. The hydrophobic residues W332
210 and M330 are intercalated into the open major groove of the RNA duplex and are likely responsible for
211 dislodging the target uracil base.

212   Prior crystallographic and cryoEM studies of Nsp15 have provided insights into its binding to both
213   ssRNA and dsRNA (16, 24, 26). Overall dsRNA binding mode observed in our Nsp15-RNA structures
214   resembles that observed in the recently reported structure of SARS-CoV-2 Nsp15 bound to a 52-bp dsRNA
215   (26). A synthetic dsRNA duplex adopted from a substrate of the *Drosophila* Dicer-2 was used in earlier
216   studies (26, 28), in contrast to the poly(A/U)-containing physiological dsRNA substrate used in the current
217   study. The observed target uracil base recognition mode in our structure is consistent with both short
218   ssRNA-bound and dsRNA-bound structures (24, 26). Notably, the structure with 52-bp dsRNA shows the
219   adenine at the +1 position of the target strand. The structure presented in the current study shows that uridine
220   can also be readily accommodated at this location. Interestingly, the comparison of the structures around
221   the target U_0 base shows that the position of W332 is tuned to promote the interactions by stacking the
222   indole ring with the U_{+1} base while the flipped U_0 base remains precisely at the same position.

223   Nsp15 is part of the large coronavirus RTC, which is responsible for RNA-dependent RNA
224   transcription and 5′-capping (17, 29). The RTC is formed by a series of Nsps including Nsp7, Nsp8, Nsp9,
225   Nsp10, Nsp12, Nsp13, Nsp14, and Nsp16 (5, 17, 19). The association of Nsp15 with RTC provides a timely
226   advantage to viral infection, as Nsp15 can trim the poly(U) tract immediately after its synthesis by RdRp
227   thus minimizing the duration of the presence of unprocessed long poly(U) in the infected cells. Within the
228   RTC, Nsp8 serves as a cofactor of Nsp15, which may jointly hold the substrate dsRNA (30). Despite our
229   attempt to assemble a complex of Nsp15 with Nsp8 in the presence and absence of the substrate RNA, we
230   were unable to observe complex formation between them under the conditions tested. Further study is
231   needed to understand the potential role of Nsp15 in the RTC and their native complex formation during
232   coronaviral replication.

233   In summary, our structures of the Nsp15-RNA complex in two states reveal direct interactions
234   between Nsp15 and poly(A/U) RNA, its only known physiological substrate. These structures offer
235   snapshots that inform how SARS-CoV-2 camouflages itself in infected cells to escape the host detection of
236   viral RNA. Given the high sequence conservation of endoribonuclease among known coronavirus lineages
237   and SARS-CoV-2 variants known to date, targeting Nsp15 activity may be a promising therapeutic strategy
238   against the current and future SARS-CoV-2 variants. Inhibitors of the Nsp15 activity would preserve the
239   natural innate immune responses against dsRNA derived from the coronavirus genome, giving rise to broad-
240   spectrum anti-viral drugs.

241

242   **Methods**

243   **Plasmids**

244   Nsp15 from SARS-CoV-2 isolate WA-CDC-02982586-001/2020 (GenBank: MN985325.1, residues 1-
245   346) with His_6-tag at N-terminus was cloned into Champion™ pET SUMO vector by excluding SUMO
246   fusion tag, and Nsp8 from the same SARS-CoV-2 isolate (residues 1-198) was cloned into pET28a vector

247 with His$_6$-tag at N-terminus. Cloning and mutagenesis were performed with In-Fusion cloning and
248 PrimeSTAR mutagenesis (Clontech) by following the manufacturer's instructions. The sequences of all the
249 constructs were verified by Sanger DNA sequencing (Azenta Life Sciences). The multiple sequence
250 alignments were generated with Linnaeo (https://github.com/beowulfey/linnaeo).

251

Protein expression and purification

253 His$_6$-Nsp15 catalytically inactive mutant H234A and wild-type His$_6$-Nsp8 expression vectors were
254 transformed into the *E. coli* strains BL21(DE3). The *E. coli* cells harboring the expression vectors were
255 grown in LB medium at 37°C until the OD$_{600}$ reaches 0.6. The recombinant proteins were induced by 0.2
256 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16°C for 18 hours.

257 For Nsp15, the cell pellets were resuspended with the buffer (25 mM HEPES-NaOH (pH 7.5), 500
258 mM NaCl, and 0.5 mM TCEP) containing RNase A (0.1 mg/ml, Qiagen), lysed by sonication, and cellular
259 debris was removed by centrifugation. The supernatant containing the His$_6$-Nsp15 was loaded onto the Ni-
260 NTA agarose column (Qiagen). The nickel column was extensively washed with wash buffer (25 mM
261 HEPES-NaOH (pH 7.5), 500 mM NaCl, 50 mM imidazole, and 0.5 mM TCEP) and the protein was eluted
262 with elution buffer (25 mM HEPES-NaOH (pH 7.5), 500 mM NaCl, 500 mM imidazole, and 0.5 mM
263 TCEP). The eluted proteins were concentrated and subjected to Superdex 200 Increase 10/300 GL column
264 (Cytiva) equilibrated with the buffer (25 mM HEPES-NaOH (pH 7.5), 150 mM NaCl, and 0.5 mM TCEP).
265 The peak fractions corresponding to the hexamer form were collected and concentrated for cryoEM study.

266 For Nsp8, the cell pellets were resuspended with the buffer (20 mM Tris-HCl (pH 8.0), 500 mM
267 NaCl, and 0.5 mM TCEP) containing RNase A (0.1 mg/ml, Qiagen), lysed by sonication, and cellular debris
268 was removed by centrifugation. The supernatant containing the His$_6$-Nsp8 was loaded onto the Ni-NTA
269 agarose column (Qiagen). The nickel column was extensively washed with wash buffer (20 mM Tris-HCl
270 (pH 8.0), 500 mM NaCl, 20 mM imidazole, and 0.5 mM TCEP) and the protein was eluted with elution
271 buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 300 mM imidazole, and 0.5 mM TCEP). The eluted
272 proteins were concentrated and subjected to Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated
273 with the buffer (20 mM Tris-HCl (pH 8.0), 250 mM NaCl, and 0.5 mM TCEP). The peak fractions were
274 collected and concentrated for cryoEM study. Protein purity was assessed by SDS-PAGE at each
275 purification step.

276

Negative-stain EM

278 5 µl of 0.02 mg/ml purified Nsp15 sample was applied onto glow-discharged ultrathin formvar/carbon
279 supported copper 400-mesh grids (Electron Microscopy Sciences), blotted and stained with 2.0% uranyl

280 acetate. Negative-stained grids were imaged on a Talos F200C transmission electron microscope (Thermo
281 Fisher Scientific) operated at 200 kV.

282

CryoEM data acquisition

284 Three data sets were collected in separate TEM sessions. For reconstitution of Nsp15-RNA complex, pre-
285 annealed dsRNA (chain 1: 5′-
286 rUrCrUrUrArGrGrArGrArArUrGrArCrArArArArArArArArArArArArArArArArArArArArA-3′, chain 2:
287 5′-rUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrGrUrCrArUrUrCrUrCrCrUrArArGrA-3′)
288 substrate was synthesized (Integrated DNA Technologies). For data set 1, Nsp15, Nsp8, and dsRNA were
289 mixed by 1:0.5:1 molar ratio (7.5 μM Nsp15, 3.75 μM Nsp8, and 7.5 μM dsRNA) in a buffer (25 mM
290 HEPES-NaOH, 150 mM NaCl, pH 7.5). For data sets 2 and 3, Nsp15, Nsp8, and dsRNA were mixed by
291 1:1:10 molar ratio (7.5 μM Nsp15, 7.5 μM Nsp8, and 75 μM dsRNA) in a buffer with lower salt (25 mM
292 HEPES-NaOH, 100 mM NaCl, pH 7.5). The mixture was incubated on ice for 30-60 min before plunge-
293 freezing. 4 ul aliquots of the mixture were applied to UltrAu foil R1.2/1.3 gold 300-mesh grids (Electron
294 Microscopy Sciences). Grids were then blotted and vitrified in liquid ethane using Vitrobot Mark IV
295 (Thermo Fisher Scientific). CryoEM data was collected in a Glacios (Thermo Fisher Scientific) equipped
296 with Falcon-4 direct electron detector operated at 200 kV in electron counting mode. Movies were collected
297 at a nominal magnification of 150,000× and a pixel size of 0.92 Å in EER format. A total dose of 52 e$^-$/Å^2
298 per movie was used with a dose rate of 5-6 e$^-$/Å^2/sec. 7,268, 10,001, and 5,150 movies were recorded for
299 the data set 1, 2, and 3, respectively, by automated data acquisition with EPU.

300

CryoEM data processing

302 The movies from three data sets were imported into cryoSPARC software package (31) and subjected to
303 patch motion correction and CTF estimation in cryoSPARC. For data set 1, reference-free manual particle
304 picking in a small subset of data was performed to generate 2D templates for auto-picking. A total of
305 2,478,629 particles were picked initially, extracted, and down-sampled by a factor of 4, on which 2D
306 classification was performed. 1,961,839 particles from 2D class averages were selected and re-extracted
307 with full-resolution. 3D *ab initio* reconstruction was then performed to generate three initial volumes. A
308 single dominant class containing 68% of the particles showed a feature of hexamer form Nsp15. Further
309 classification did not yield any 3D classes containing RNA or Nsp8 density. For data sets 2 and 3, 2,983,555
310 and 1,730,135 particles were picked initially by using the templates generated from the data set 1, extracted,
311 and down-sampled by a factor of 4, on which 2D classification was performed. Additional RNA densities
312 were present in a subset of 2D classes, which are not present in the data set 1. 1,330,310 particles (data set
313 2) and 1,119,650 particles (data set 3) from 2D class averages were selected and re-extracted with full-
314 resolution. 3D *ab initio* reconstruction was then performed to generate three initial volumes. A class

315   containing 39% of the particles in each data set showed a clear feature of dsRNA density attached to the
316   Nsp15 hexameric barrel. A class containing 37-38% of the particles in each data set showed a feature of
317   hexamer form, which is similar to the class observed in the data set 1. The 3D classes representing hexamer
318   form with no obvious RNA densities from the three data sets were combined and non-uniform refinement
319   (32) was performed with D3 symmetry to yield the final 2.3 Å resolution map. The 3D classes representing
320   RNA-bound form from the data sets 2 and 3 were combined and non-uniform refinement was performed
321   with C1 symmetry to yield the final 2.7 Å resolution map. We noticed that RNA density in the 3D map was
322   anisotropic, and the map could be an average of different conformational states. To further classify into
323   possible different classes, heterogeneous refinement was performed to yield four classes. A class containing
324   24% of the particles showed strong RNA density along the top trimer (state 1) while two classes containing
325   53% of the particles showed strong RNA density along the bottom trimer (state 2). Each state was subjected
326   to non-uniform refinement to yield the final 3.3 Å (state 1) and 3.1 Å (state 2) resolution maps, respectively.
327   Any additional classification did not yield 3D classes with Nsp8, or Nsp15-RNA complex with more than
328   one dsRNA bound to the Nsp15 hexamer. All resolution evaluation was performed based on the gold-
329   standard criterion of Fourier shell correction (FSC) coefficient at 0.143 (33).

330

Model building and refinement

332   An atomic model derived from crystal structure of SARS-CoV-2 Nsp15 (PDB ID: 6VWW)(27) was docked
333   into the cryoEM map of apo-Nsp15 using UCSF Chimera (34). The apo-Nsp15 model was refined with the
334   phenix.real_space_refine module in Phenix, with secondary structure restraints and geometry restraints (35,
335   36). The atomic models went through iterative cycles of manual adjustment in COOT (37) and real-space
336   refinement in Phenix (38). For Nsp15-RNA complex consensus form, standard A-form double-stranded
337   RNA was generated and docked together with apo-Nsp15 model into the cryoEM map using UCSF
338   Chimera. The RNA model was manually adjusted while keeping proper RNA geometry using COOT.
339   Nsp15-RNA complex states 1 and 2 models were built based on the consensus form by extending and
340   refining the RNA strands. The final atomic models were validated using the comprehensive cryoEM
341   validation tool implemented in Phenix (Table S1) (39). All structural figures were generated with UCSF
342   ChimeraX (40).

343

References

345   1.   Kato H, *et al.* (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*
346   441(7089):101-105.
347   2.   Kang DC, *et al.* (2002) mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-
348   dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A*
349   99(2):637-642.

350   3.   Daffis S, *et al.* (2010) 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members.
351   *Nature* 468(7322):452-456.
352   4.   Yan L, *et al.* (2021) Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex
353   Reveals an Intermediate State in Cap Synthesis. *Cell* 184(1):184-193 e110.
354   5.   Yan L, *et al.* (2021) Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2
355   polymerase reveals mechanisms for capping and proofreading. *Cell* 184(13):3474-3485 e3411.
356   6.   Park GJ, *et al.* (2022) The mechanism of RNA capping by SARS-CoV-2. *Nature* 609(7928):793-800.
357   7.   Yan L, *et al.* (2022) A mechanism for SARS-CoV-2 RNA capping and its inhibition by nucleotide analog
358   inhibitors. *Cell* 185(23):4347-4360 e4317.
359   8.   Deng X & Baker SC (2018) An "Old" protein with a new story: Coronavirus endoribonuclease is important
360   for evading host antiviral defenses. *Virology* 517:157-163.
361   9.   Deng X, *et al.* (2019) Coronavirus Endoribonuclease Activity in Porcine Epidemic Diarrhea Virus Suppresses
362   Type I and Type III Interferon Responses. *J Virol* 93(8).
363   10. Kindler E, *et al.* (2017) Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus
364   replication. *PLoS Pathog* 13(2):e1006195.
365   11. Deng X, *et al.* (2017) Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits
366   apoptosis in macrophages. *Proc Natl Acad Sci U S A* 114(21):E4251-E4260.
367   12. Bhardwaj K, Guarino L, & Kao CC (2004) The severe acute respiratory syndrome coronavirus Nsp15 protein
368   is an endoribonuclease that prefers manganese as a cofactor. *J Virol* 78(22):12218-12224.
369   13. Bhardwaj K, Sun J, Holzenburg A, Guarino LA, & Kao CC (2006) RNA recognition and cleavage by the SARS
370   coronavirus endoribonuclease. *J Mol Biol* 361(2):243-256.
371   14. Bhardwaj K, *et al.* (2008) Structural and functional analyses of the severe acute respiratory syndrome
372   coronavirus endoribonuclease Nsp15. *J Biol Chem* 283(6):3655-3664.
373   15. Ivanov KA, *et al.* (2004) Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc
374   Natl Acad Sci U S A* 101(34):12694-12699.
375   16. Frazier MN, *et al.* (2021) Characterization of SARS2 Nsp15 nuclease activity reveals it's mad about U. *Nucleic
376   Acids Res* 49(17):10136-10149.
377   17. Athmer J, *et al.* (2017) In Situ Tagged nsp15 Reveals Interactions with Coronavirus Replication/Transcription
378   Complex-Associated Proteins. *mBio* 8(1).
379   18. Zhang L, *et al.* (2018) Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by
380   Middle East Respiratory Syndrome Coronavirus. *J Virol* 92(22).
381   19. Xu W, *et al.* (2021) Compartmentalization-aided interaction screening reveals extensive high-order
382   complexes within the SARS-CoV-2 proteome. *Cell Rep* 36(5):109482.
383   20. Hackbart M, Deng X, & Baker SC (2020) Coronavirus endoribonuclease targets viral polyuridine sequences
384   to evade activating host sensors. *Proc Natl Acad Sci U S A* 117(14):8094-8103.
385   21. Hofmann MA & Brian DA (1991) The 5' end of coronavirus minus-strand RNAs contains a short poly(U) tract.
386   *J Virol* 65(11):6331-6333.
387   22. Wu HY, Ke TY, Liao WY, & Chang NY (2013) Regulation of coronaviral poly(A) tail length during infection.
388   *PLoS One* 8(7):e70548.
389   23. Kim D, *et al.* (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell* 181(4):914-921 e910.
390   24. Kim Y, *et al.* (2021) Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-
391   CoV-2. *Commun Biol* 4(1):193.
392   25. Pillon MC, *et al.* (2021) Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into
393   nuclease specificity and dynamics. *Nat Commun* 12(1):636.
394   26. Frazier MN, *et al.* (2022) Flipped over U: structural basis for dsRNA cleavage by the SARS-CoV-2
395   endoribonuclease. *Nucleic Acids Res* 50(14):8290-8301.
396   27. Kim Y, *et al.* (2020) Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci*
397   29(7):1596-1605.

398 28. Sinha NK, Iwasa J, Shen PS, & Bass BL (2018) Dicer uses distinct modules for recognizing dsRNA termini.
399   *Science* 359(6373):329-334.
400 29. Shi ST*, et al.* (1999) Colocalization and membrane association of murine hepatitis virus gene 1 products and
401   De novo-synthesized viral RNA in infected cells. *J Virol* 73(7):5957-5969.
402 30. Perry JK*, et al.* (2021) An atomistic model of the coronavirus replication-transcription complex as a hexamer
403   assembled around nsp15. *J Biol Chem* 297(4):101218.
404 31. Punjani A, Rubinstein JL, Fleet DJ, & Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised
405   cryo-EM structure determination. *Nat Methods* 14(3):290-296.
406 32. Punjani A, Zhang H, & Fleet DJ (2020) Non-uniform refinement: adaptive regularization improves single-
407   particle cryo-EM reconstruction. *Nat Methods* 17(12):1214-1221.
408 33. Chen S*, et al.* (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D
409   structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135:24-35.
410 34. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J
411   Comput Chem* 25(13):1605-1612.
412 35. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure
413   solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
414 36. Afonine PV*, et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta
415   Crystallogr D Biol Crystallogr* 68(Pt 4):352-367.
416 37. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D
417   Biol Crystallogr* 66(Pt 4):486-501.
418 38. Afonine PV*, et al.* (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr
419   D Struct Biol* 74(Pt 6):531-544.
420 39. Afonine PV*, et al.* (2018) New tools for the analysis and validation of cryo-EM maps and atomic models.
421   *Acta Crystallogr D Struct Biol* 74(Pt 9):814-840.
422 40. Goddard TD*, et al.* (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein
423   Sci* 27(1):14-25.

424

Acknowledgments

Author contributions

Z.H.Z. and X.S.C. supervised the project and acquired the funding. F.I. conceived the project and designed the experiments. F.I. and H.Y. generated clones and purified the proteins. F.I. performed the cryoEM grid screening, data collection, image processing, and atomic model building. F.I. wrote the manuscript with inputs from all authors.

438

Competing financial interests

440   The authors declare no competing financial interests.

441
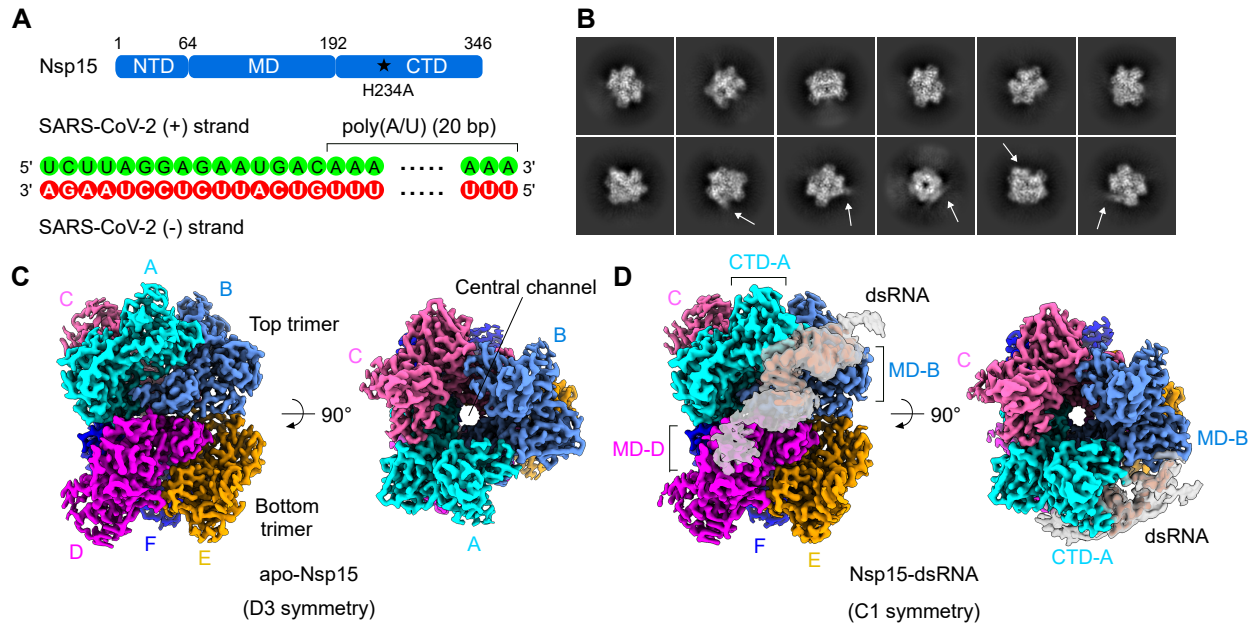
Fig. 1. CryoEM reconstructions of apo- and RNA-bound SARS-CoV-2 Nsp15. (A) Schematic representation of the domain organization and construct design of Nsp15 (top). dsRNA substrate sequence used for the reconstitution of Nsp15-RNA complex (bottom). (B) 2D class averages of the reconstituted Nsp15-RNA complex. A subset of the 2D classes shows extra density stemming from the core of the 2D density (indicated by arrows). (C) Two orthogonal views of the cryoEM reconstruction of apo-Nsp15 at 2.3 Å resolution. (D) Two orthogonal views of the cryoEM reconstruction of the Nsp15-RNA complex at 2.7 Å resolution. RNA density is shown in two different isosurface threshold levels (grey and red surfaces) to show both high-resolution and low-resolution features.
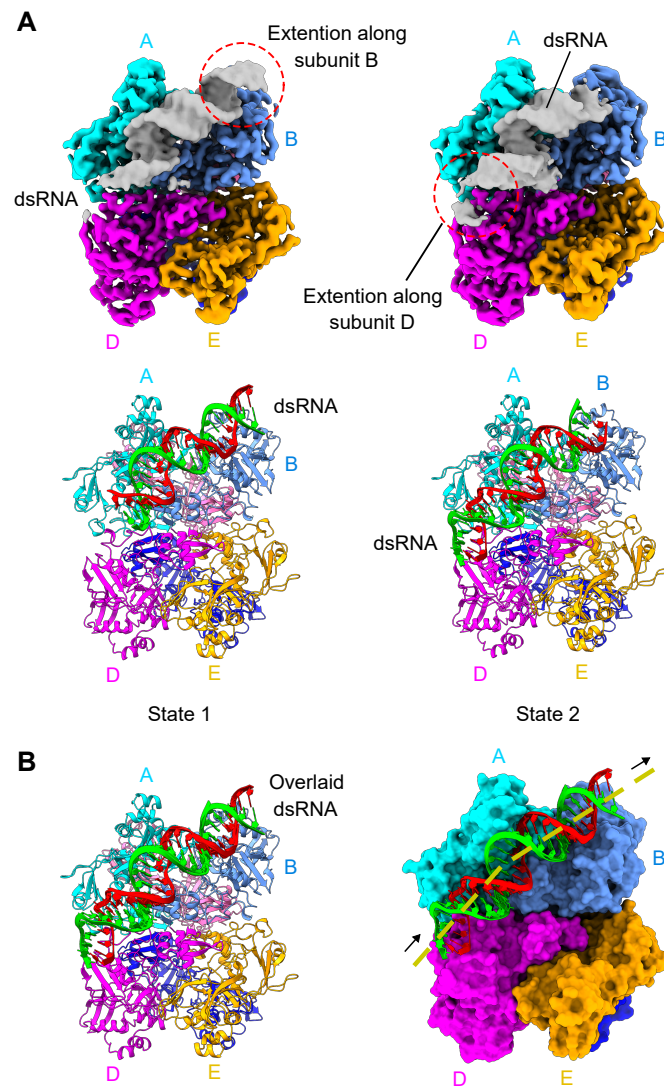
Fig. 2. Two distinct states of Nsp15-RNA complex. (A) CryoEM density maps (top) and the corresponding atomic models (bottom) of the Nsp15-RNA state 1 (left) and state 2 (right) structures. The RNA chains in green and red correspond to positive and negative strands of the coronavirus genome, respectively. (B) Overlay of the RNA models from the two states on the consensus Nsp15 structure in the ribbon model (left) and surface model (right). The dotted line indicates the trajectory of the bound dsRNA.
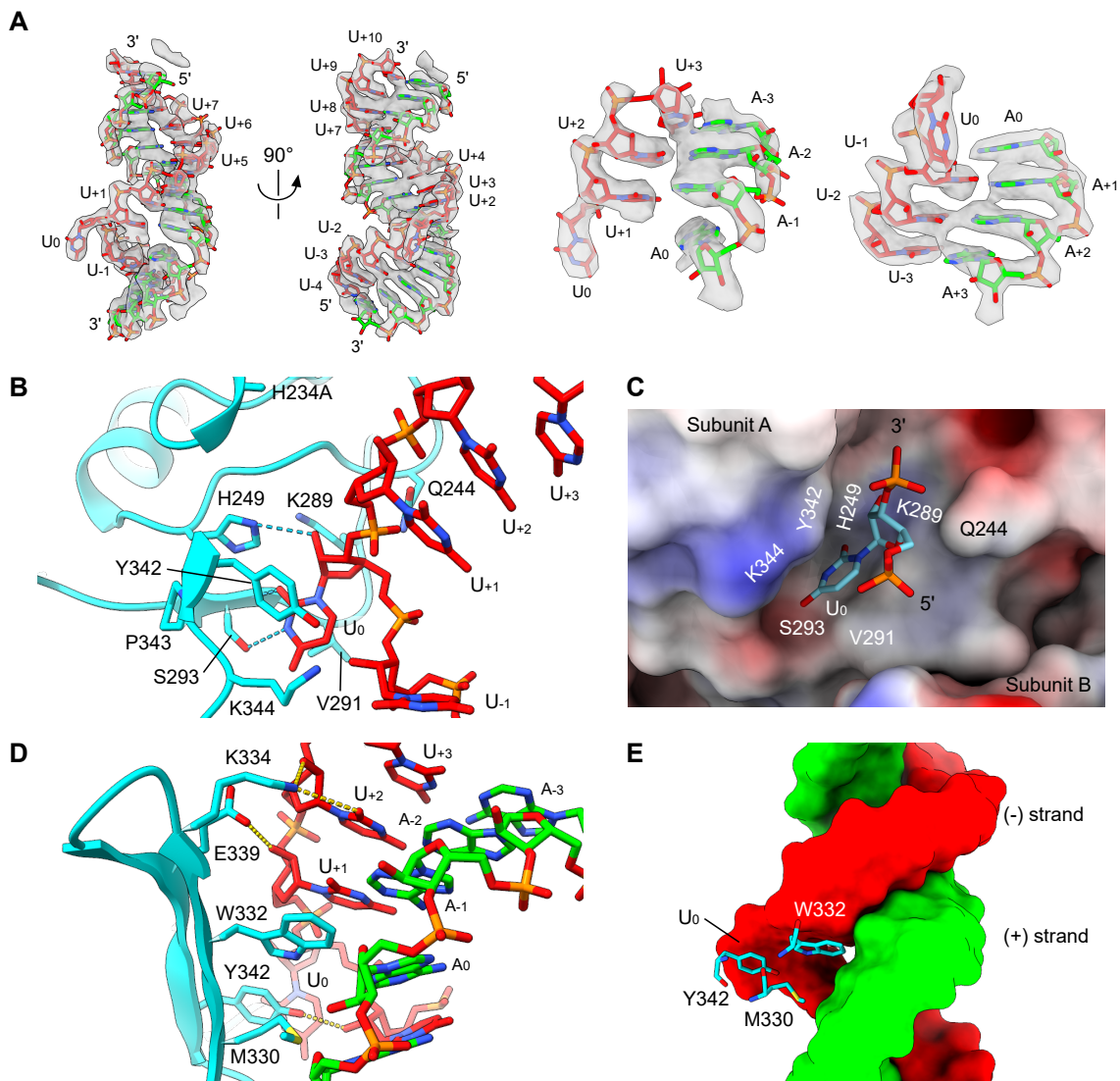
Fig. 3. Poly(U) uracil base flipping at nuclease active site center. (A) CryoEM density and the corresponding atomic model of the RNA bound to Nsp15. Two orthogonal views of the bound dsRNA density and its model (left). Density fitting of the RNA model of the U_0-U_{+3}:A_0-A_{+3} (middle). Density fitting of the RNA model of U_{-3}-U_0:A_{+3}-A_0 (right). (B) Structure of the endonuclease active site center and recognition mechanism of flipped scissile U_0 base. (C) Surface electrostatic potential of the Nsp15 around the endonuclease active site center and the interactions with the target U_0 with both 3′- and 5′-phosphates (depicted in sticks). The surface area is colored according to the calculated electrostatic potential from -10.0 kT/e (red) to +10.0 kT/e (blue). (D) Recognition of the open major groove of the dsRNA substrate by Nsp15. W332 is intercalated into the space that would have been occupied by the flipped U_0 base within the RNA duplex. M330 and Y342 additionally create the hydrophobic surface to facilitate the major groove interaction. Two polar residues K334 and E339 interact with U_{+2} and U_{+1}, respectively. (E) The key hydrophobic residues responsible for base-flipping. The side chain of W332 is deeply intercalated into the major groove of the RNA duplex. M330 and Y342 additionally participate in the major groove interaction.
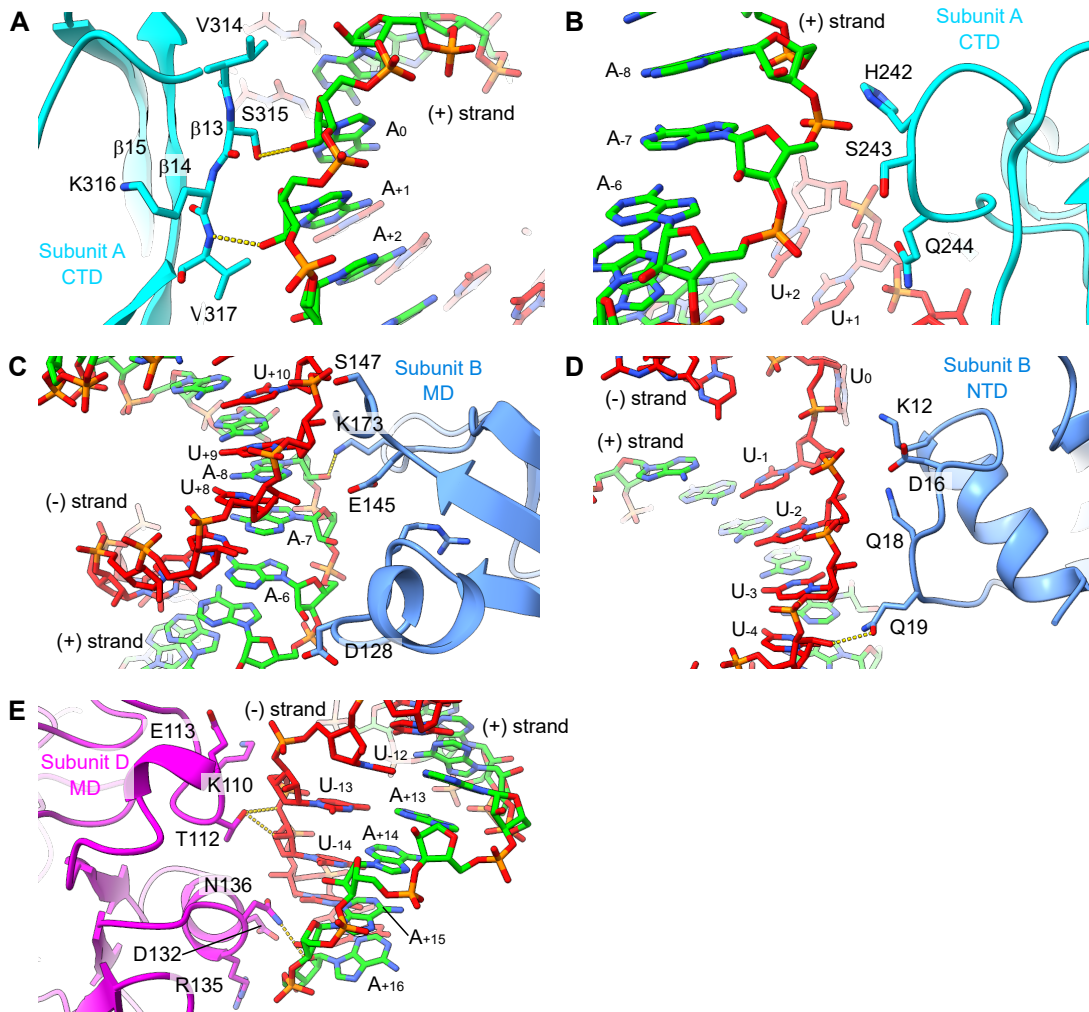
Fig. 4. Bound RNA forms extensive interactions with Nsp15 hexamer across three subunits. (A) Interface between a patch of -^{314}VSKV^{317}- of subunit A's CTD and A_0 and A_{+1} of the positive strand. (B) Interface between a patch of -^{242}HSQ^{244}- of subunit A's CTD and A_{-8} and A_{-7} of the positive strand and U_{+1} of the negative strand. (C) Interface between subunit B's MD and U_{+10} of the negative strand and A_{-8}, A_{-6} of the positive strand. (D) Interface between subunit B's NTD and the U_{-4}-U_{-1} of the negative strand. (E) Interface between subunit D's MD and U_{-14}-U_{-12} of the negative strand and A_{+16} of the positive strand.