Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Research Article

Biochemical Characterization of Emerging SARS-CoV-2 Nsp15 Endoribonuclease Variants

Isha M. Wilson [1], Meredith N. Frazier [1,2*], Jian-Liang Li [3], Thomas A. Randall [3] and Robin E. Stanley [1*]

*1 - **Signal Transduction Laboratory,** National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, 111 T. W. Alexander Drive, Research Triangle Park, NC 27709, USA*
*2 - **Department of Chemistry and Biochemistry,** College of Charleston, Charleston, SC, 29424, USA[†]*
*3 - **Integrative Bioinformatics Support Group,** National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, 111 T. W. Alexander Drive, Research Triangle Park, NC 27709, USA*

Correspondence to Meredith N. Frazier and Robin E. Stanley: Signal Transduction Laboratory, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, 111 T. W. Alexander Drive, Research Triangle Park, NC 27709, USA (M.N. Frazier). *robin.stanley@nih.gov (R.E. Stanley) @ishamyana (I.M. Wilson), @MNFrazier5 (M.N. Frazier), @RobinStanleyLab (R.E. Stanley)*
https://doi.org/10.1016/j.jmb.2022.167796
Edited by M.F. Summers

Abstract

Global sequencing efforts from the ongoing COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, continue to provide insight into the evolution of the viral genome. Coronaviruses encode 16 non-structural proteins, within the first two-thirds of their genome, that facilitate viral replication and transcription as well as evasion of the host immune response. However, many of these viral proteins remain understudied. Nsp15 is a uridine-specific endoribonuclease conserved across all coronaviruses. The nuclease activity of Nsp15 helps the virus evade triggering an innate immune response. Understanding how Nsp15 has changed over the course of the pandemic, and how mutations affect its RNA processing function, will provide insight into the evolution of an oligomerization-dependent endoribonuclease and inform drug design. In combination with previous structural data, bioinformatics analyses of 1.9 + million SARS-CoV-2 sequences revealed mutations across Nsp15′s three structured domains (N-terminal, Middle, EndoU). Selected Nsp15 variants were characterized biochemically and compared to wild type Nsp15. We found that mutations to important catalytic residues decreased cleavage activity but increased the hexamer/monomer ratio of the recombinant protein. Many of the highly prevalent variants we analyzed led to decreased nuclease activity as well as an increase in the inactive, monomeric form. Overall, our work establishes how Nsp15 variants seen in patient samples affect nuclease activity and oligomerization, providing insight into the effect of these variants *in vivo*.

Published by Elsevier Ltd.

Introduction

Coronaviruses are a family of large, positive stranded RNA viruses that have known zoonotic potential[1]. The Middle Eastern Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV-1) coronaviruses have caused several epidemics in the early 2000s. Currently a novel coronavirus, SARS-CoV-2, is responsible for a multi-year pandemic (COVID-19). Coronaviruses are also responsible for 10–30% of common colds[2]. RNA viruses, which replicate with an RNA-dependent-RNA polymerase are known to mutate frequently[3]. Mutations that

confer a benefit to the virus are useful, but many mutations are neutral or deleterious. Evolution of RNA viruses is generally accepted to be driven by both genetic drift and negative selection (removing deleterious mutations)[4]. Coronaviruses encode proof-reading machinery that reduces mutation rate, but it is still very high compared to the mutation rate of cellular replication of DNA. Having low replication fidelity (i.e. a high error rate) allows coronaviruses to sample the mutational space by supporting a wide range of sequences during an infection, which leads to fast adaptation in response to selective pressures[5]. Mutations in viral RNA can also arise through damage or the action of RNA editors[6]. The enormous sequencing effort throughout the COVID-19 pandemic has presented the opportunity to analyze patterns of mutations and investigate the functional effects of predominant mutants, lending insight into important structural and functional regions of viral proteins that may otherwise have remained uncharacterized[7–15].

The first two-thirds of the ~30 kb coronavirus genome encodes two open reading frames, pp1a and pp1ab, that encode 16 non-structural proteins (nsps) via a ribosomal frame-shifting mechanism[1,16]. The latter one-third encodes structural and accessory proteins, such as the spike protein. Many of the nsps function in the replication transcription complex (RTC) and are necessary for viral replication[17]. Studies have shown mutational frequency varies across the SARS-CoV-2 genome[18]; mutation frequency seems to be higher in regions encoding structural proteins, and lower in the region encoding nsps[18–19]. While structural proteins, especially the spike protein, have been extensively studied, nsps have had less focus[20–22]. However, given their conservation across coronaviruses and lower mutational rates compared to structural proteins, nsps make attractive anti-viral targets.

One such target is Nsp15, a uridine-specific endoribonuclease that processes viral RNA to prevent the activation of the dsRNA sensor MDA5[23–25]. *In vivo* studies in animal models infected with coronaviruses with inactivated Nsp15 show lower mortality and reduced pathogenicity[23,25–27]. Recent structural studies of Nsp15 have revealed important molecular details regarding how it processes both single and double stranded (ds) RNA[28–31]. SARS-CoV-2 Nsp15 is a hexamer formed from a dimer of trimers; monomeric Nsp15 is inactive[29–30,32–33]. Each protomer consists of an N-terminal (NTD), middle (MD), and catalytic endoribonuclease (EndoU) domain. Nsp15 is only active as a hexamer and the hexamer is critical for supporting dsRNA binding[28]. Beyond Nsp15, EndoU domains are found in all kingdoms of life and share a catalytic triad with RNase A[34]. Like RNase A, Nsp15 uses a two-step transesterification mechanism to cleave RNA 3′ of uridines[29]. While much progress has been made to understand how Nsp15 functions to processes viral RNA, many open questions remain. Moreover, in comparison with the EndoU domain the NTD and MD's from SARS-CoV-2 Nsp15 remain poorly characterized.

Increased understanding of the molecular details of Nsp15 structure and function provides a platform to better understand the effect of mutations in the Nsp15 coding sequence. Interestingly, in certain variants, Nsp15 mutations have been noted as clade defining markers[35–36]. For example, in an analysis of different Delta variant clades, Nsp15 H235Y is a marker for Delta subclade C and K260R is a marker for Delta clade E. Nsp15 K13N was identified as a marker for a subclade of the B.1.1.33 lineage in Brazil[37]. Additionally, Nsp15 T113I is a marker for Omicron[38] and was computationally predicted to have the greatest impact on fitness of Orf1b (Nsp12-16) mutations analyzed across 6.4 million SARS-CoV-2 genomes.[39] However, it remains unknown how any of these specific mutations impact Nsp15 activity.

Our goal was to use the Nsp15 mutational information from the first year of the COVID19 pandemic to characterize how mutations affect its endoribonuclease activity, either directly or through changes in its oligomerization state. We analyzed Nsp15 mutations found in sequences deposited in the GISAID database as of June 2021[40]. We selected mutations from each domain (NTD, MD and EndoU) based on frequency of mutation and reports of specific variants in the literature, as well as our knowledge of active site residues. We then carried out *in vitro* biochemical assays of the recombinantly purified mutants to evaluate their oligomeric state and test nuclease activity compared to WT. Through this approach we have established how these amino acids changes seen in SARS-CoV-2 patient samples impact Nsp15 function. Moreover, through this approach we have gained additional knowledge about important regions of Nsp15.

Results

Identifying Nsp15 mutants from SARS-CoV-2 genome sequencing

Approximately 1.9 million full length sequences of Nsp15 were extracted from the EpiCov section (allnuc0614) of the GISAID database, which is a worldwide repository of viral isolates. The sequence of Nsp15 of each viral isolate was compared to the original Wuhan isolate. In line with the relatively low replication fidelity of coronaviruses, single nucleotide variants were present across all three domains of the endoribonuclease. Of the full-length Nsp15 sequences (1038 bases), 1025 positions were observed to have nucleotide substitutions (Supplemental File 1), which were corresponding to changes in 341 out of 346 amino acids from

Nsp15. However, 87% of these variants were observed in less than 100 isolates (Supplemental File 1). Six amino acid variants were observed in over 10,000 isolates, where five of them (N74N, 43,964; D79D, 15,684; L214L, 40,846; L217L, 74,888; N278N, 10,718) are synonymous and one (D220Y, 10,323) is non-synonymous (Figure 1 (A)). For subsequent biochemical analysis we focused only on non-synonymous variants that led to a change in the protein sequence (Figure 1(B)). As expected, surface residues were more frequently mutated than core residues[19].

A total of 83 variants were observed in over 1,000 isolates which were unlikely to be sequencing

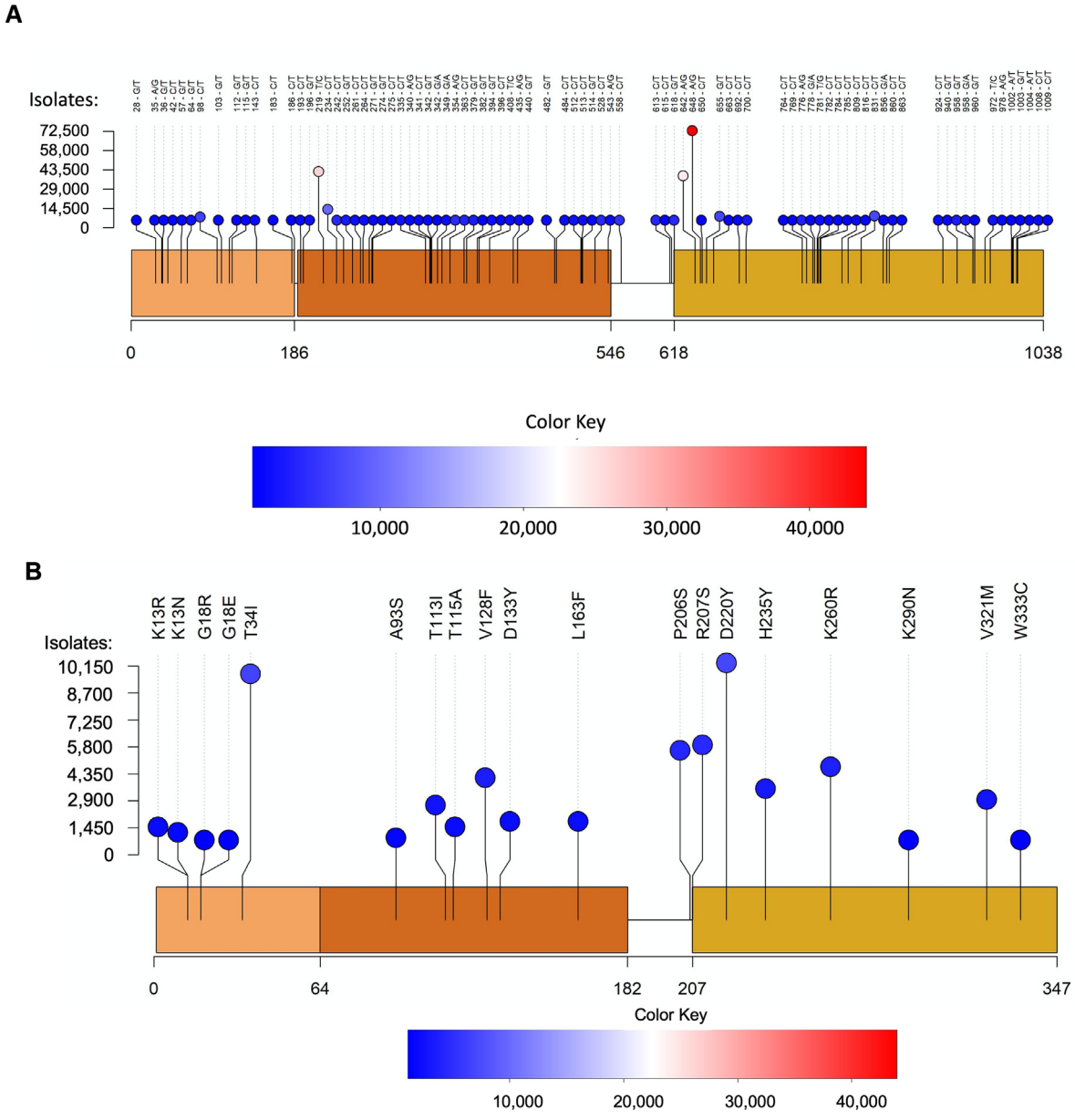

Figure 1. Variants identified in Nsp15 from the GISAID database. Approximately 1.9 million full-length NSP15 sequences (1038 bp) were downloaded from the GISAID database. The sequence of each Nsp15 isolate was compared to that of the original Wuhan isolate, and single nucleotide variants were then generated, which were presented in the lollipop-style plots. Each circle represents a variant observed at the indicated position. The height and the color of the circle indicate the frequency of the mutation event, ie. the number of isolates in which the variant was observed. **(A)** Variants are shown at the post-sequencing DNA level. Only variants found in at least 1000 isolates are shown. These single nucleotide variants were distributed widely throughout the three domains of Nsp15. **(B)** Selected variants are shown at the protein level. Variants with prominent numbers of isolates were selected to be biochemically characterized along with those found in residues previously shown to be important to the structure and function of Nsp15.

artifacts. Among them, 53 variants are non-synonymous and 30 are synonymous. We selected several prominent mutations to study from these 53 non-synonymous variants: K13N (1716), K13R (1339), T34I (9900), A93S (1081), T113I (2815), T115A (1810), V128F (4457), D133Y (1982), L163F (2072), P206S (5849), R207S (6174), D220Y (10323), H235Y (3630), K260R (4938), and V321M (3306). In addition to these variants, others were chosen based on previous biochemical data[29–30,41–42]. These variants were G18E (304), G18R (123), K290N (396), and W333C (304). Four of the selected residues (K13, T34, T115, and R207) had also been previously identified as major mutated residues of Nsp15[18,43]. The selected mutations covered all three domains of Nsp15 (Figure 2).

Four of the Nsp15 non-synonymous substitutions show multiple base substitutions, G18R, R207S, K290N, and W333C. In each case there is a definite skewing of the abundance of one of these mutations. For G18R, the G to A substitution is 152-fold more prevalent than the G to C substitution. A simple explanation for this could be the elevated transition mutations relative to transversions, which was noted long ago[44]. However, the G to A transition on the positive strand of SARS-CoV-2 is one of the least frequent transition

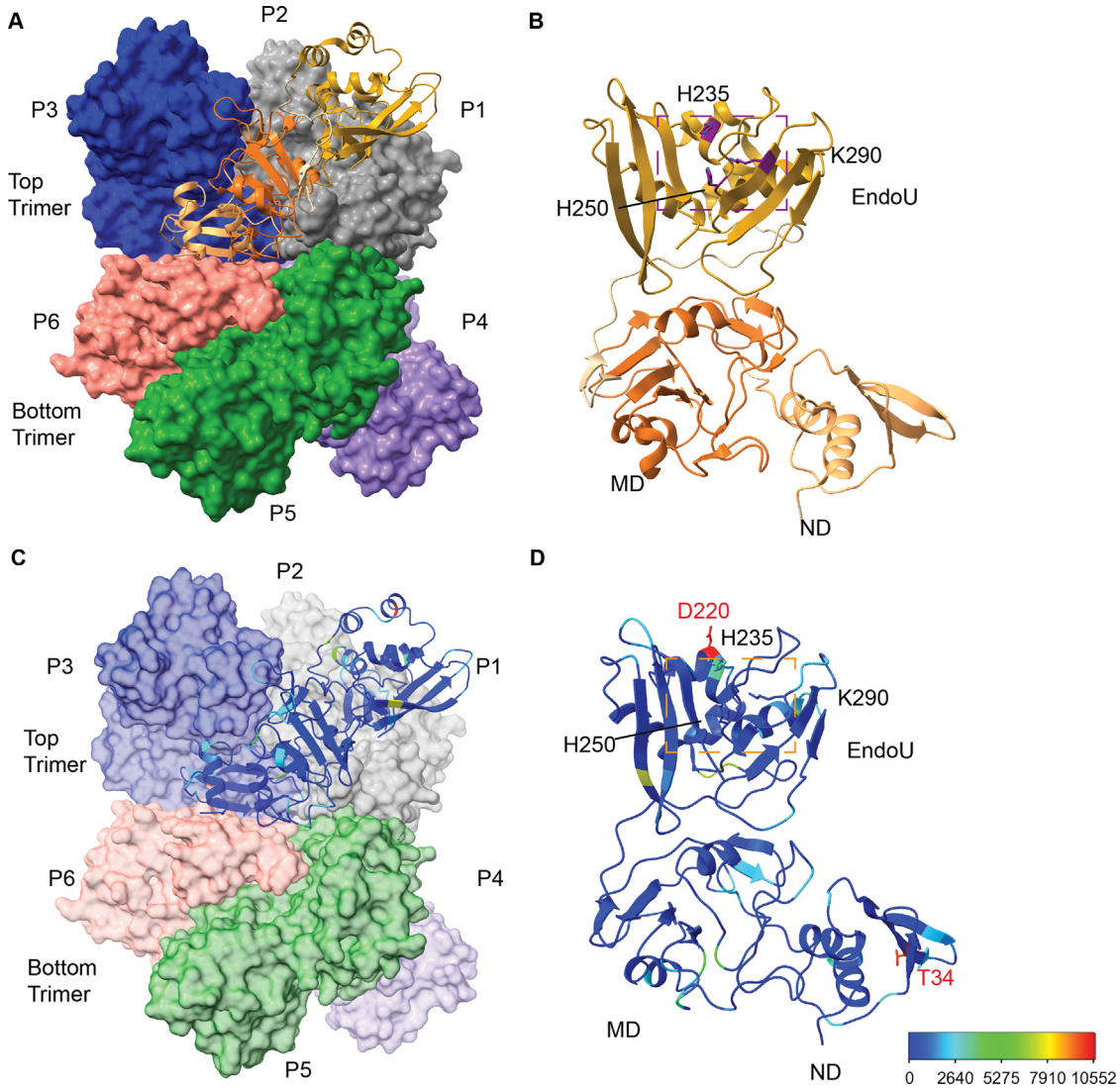

Figure 2. Overview of SARS-CoV-2 Nsp15 protein structure. (A) The Nsp15 hexamer forms from a dimer of back to back trimers. P1 is shown in ribbon diagram, while P2-6 are shown in surface representation (PDB ID: 7 N06). **(B)** Zoom in of an Nsp15 protomer, colored as in Figure 1, by domain (ND, MD, EndoU). The catalytic triad is shown in stick representation, colored purple and labeled. **(C-D)** The number of mutations at each residue were colored using a rainbow palette (see scale bar at bottom left). **(C)** One colored protomer is docked into the hexamer. **(D)** Zoom in of the mutation mapped protomer. In addition to the catalytic triad, the two residues with the highest number of mutations are shown (T34 and D220).

substitutions observed in SARS-CoV-2[45]. The other three notable transversions are all either G to T or G to C transversions. The G to T transversion is 152- to 3087-fold more prevalent in these cases. This is substantially different from what is seen in either SARS-CoV-2 or human genomes[46] where G to C transversions are more common. The over-representation of the G to T events could also be due to a jackpot effect. While these observations were made with the GISAID allnuc0614 dataset (Supplemental File 1), these differences persist in newer versions (allnuc0215; Supplemental File 2). There has been significant recent analysis of genome-wide mutational spectra, especially as it relates to human cancer. Of particular interest has been the distribution of trinucleotide-centered mutational motifs and the mechanism(s) by which they occur. The only statistically significant trinucleotide motif found mutated in SARS-CoV-2 genomes was uCn[45]. None of the codons in the substitutions of interest within Nsp15 match that motif.

While the initial analysis of SARS-CoV-2 sequences in June 2021 provided the rationale for selecting Nsp15 mutations to characterize biochemically, we have continued to monitor SARS-CoV-2 variant lineages to understand how the Nsp15 mutations we selected appear in Delta and Omicron (Supplemental File 3). Delta, the predominant SARS-CoV-2 VOC through most of 2021, was replaced by the Omicron VOC beginning in late 2021. Omicron is not a direct descendant of delta; its origin is completely independent[47]. One can see this in our analysis of the mutational frequencies of the non-synonymous amino acid mutations within Nsp15 (Supplemental Table 2, Supplemental File 4). If Delta was the progenitor of Omicron, some of the mutations present in Delta would be present in nearly all of the GISAID Omicron genomes yet we see that for all non-synonymous substitutions only a fraction of the Omicron genomes carry any one substitution seen in the Delta genomes. Subsets of both Delta and Omicron genomes carry nearly all of these substitution mutations suggesting convergent evolution of these mutations in different SARS-CoV-2 lineages. Their persistence in both lineages suggests that any functional alteration to Nsp15 is well tolerated, perhaps due to compensation by other non-structural proteins.

N-terminal domain variants

We assessed the impact of mutations on K13, G18, and T34I (Figure 3(A)) from the NTD of Nsp15, which has been shown to be critical for oligomerization[30,43,48]. In particular, K13 sits at the interface between neighboring protomers, and participates in water-mediated interactions with RNA[30] Supplemental Figure 2). G18 is also positioned near the dsRNA binding platform within the Nsp15 hexamer[28]. Given the importance of the NTD in oligomerization, we compared the amount of hexamer and monomer from each purified variant. WT Nsp15 has a hexamer/monomer ratio around 1 (Table 1, Figure 3(B)). K13R had a slightly greater amount of hexamer, while T34I had a ratio similar to WT. The other variants all had increased amounts of monomer. This provides additional evidence that disturbances in the NTD affect hexamer stability and suggests that these variants would result in less active Nsp15 in virus infected cells since monomeric Nsp15 is inactive.

Using an established FRET nuclease assay[29,42] the cleavage activity of the hexameric forms of the K13N, K13R, G18E, G18R, and T34I variants was measured (Figure 3(B)). Nsp15 K13N showed the largest decrease in cleavage activity, while K13R, G18E, and T34I also had less activity compared to WT Nsp15 (Figure 3(B)). Nsp15 G18R nuclease activity was unchanged compared to WT Nsp15 (Figure 3(B)). Given the high number of isolates with T34I, its decrease in activity contradicts the notion that variants with high amounts of isolates are more advantageous for the endoribonuclease. T34 is a core residue in the NTD, surrounded by non-polar residues (Supplementary Figure 2). As a core residue, it is likely important for maintaining the correct fold and stability of the NTD. Analysis of T34 variants in the SARS-Cov-2 Molecular Dynamics database[58] suggests that T34I would have a stabilizing effect on Nsp15 while mutation to alanine, arginine, or lysine at the same position would have a destabilizing effect. Thus, we assume that introduction of a non-polar isoleucine at this position would likely not disturb the fold of the NTD. The stabilizing effect of T34I could impact nuclease activity in other ways such as altering conformation dynamics within the hexamer.

Overall, this FRET data supports the hypothesis that the NTD provides charge-mediated stability to RNA substrates[30]. The addition of a positive charge (G18R) did not negatively impact nuclease activity, while the loss of a positive residue (K13N) or the introduction of a negative residue (G18E) decreased activity. We also tested nuclease activity on a longer, more physiologically relevant RNA substrate corresponding to the Transcriptional Regulatory Sequence (TRS) for the SARS-CoV-2N protein using a gel-based cleavage assay. The trends seen in the FRET assay were confirmed by gel-based cleavage (Table 1, Supplemental Figure 3).

Middle domain variants

We analyzed middle domain variants from two distinct faces of Nsp15's MD (Figure 4(A)). T113, V128, D133, and L163 lie on one face of this domain, while A93 and T115 sit on the opposite face. All the MD variants had increased amounts of monomer, highlighting that the MD also plays an important role in stabilizing the hexamer. Previous work on MERS Nsp15 characterized several residues in the MD involved in oligomerization[41]. When we mapped these variants

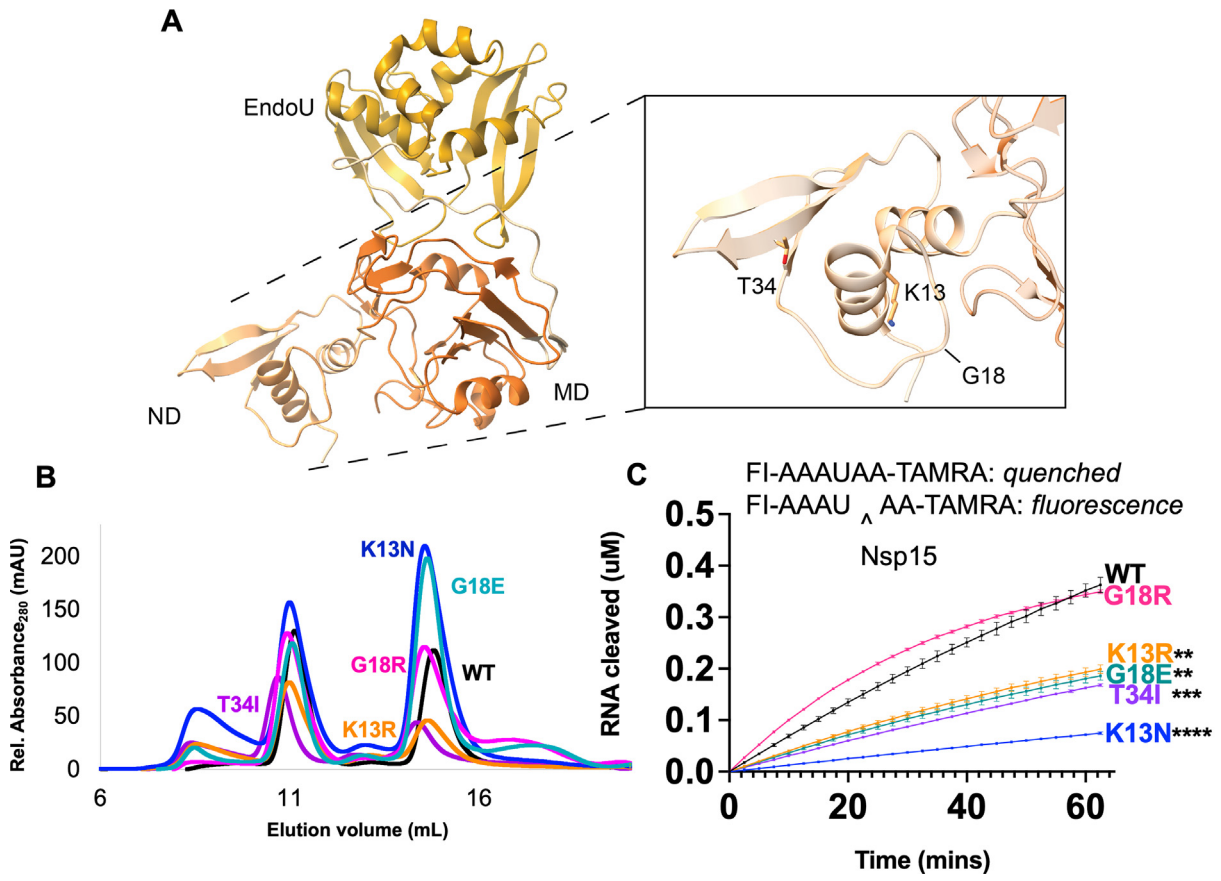

Figure 3. Characterization of Nsp15 N terminal domain variants. (A) Displayed on the structure are the residues K13, G18, and T34. **(B)** S200 elution profiles of NTD variants. Hexameric Nsp15 elutes at 11 mL, monomeric Nsp15 at 15 mL. **(C)** An embedded visual of the Fluorescence resonance energy transfer cleavage assay to accompany the FRET time course data for Nsp15 N terminal domain mutants. Nsp15 WT and variants (2.5 nM) were incubated with RNA (0.8 μM) at room temperature and fluorescence was monitored every 2.5 min for an hour. The average of a representative technical triplicate is plotted with standard deviation error bars. At least two biological replicates were performed for each mutant. Each mutant is represented by a different color: WT Nsp15 (black), Nsp15 K13N (blue), Nsp15 K13R (orange), Nsp15 G18E (teal), Nsp15 G18R (pink), and Nsp15 T34I (purple). **p < 0.01, ***p < 0.001, ****p < 0.0001.

onto our SARS-CoV-2 Nsp15 structure, we see T115 and L163 facilitate interactions with neighboring protomers. T113 occupies the same loop as T115 but faces the opposite direction in our model and is further removed from the hexamer interface. Interestingly, T113I is the only MD variant in our analysis which showed increased hexamer formation. A recently determined structure of Nsp15 bound to dsRNA revealed that T113 forms part of a large platform spanning three protomers from the hexamer that supports dsRNA engagement[28]. T115 lies near the interface with a protomer from the bottom trimer, while L163 sits at the interface with a neighboring protomer in the same (top) trimer (Supplemental Figure 2). L163F dramatically reduced the amount of purified hexamer, suggesting this residue is particularly important for oligomerization (Table 1, Figure 4(B)).

Our FRET activity assay revealed a large, significant increase in activity with the T113I and V128F mutants, and a significant decrease in activity for L163F (Figure 4(C)). The remaining variants A93S, T115A, and D133Y showed no significant changes in cleavage activity. The activity trends seen by FRET were again confirmed using a longer RNA substrate in a gel-based cleavage assay (Table 1, Supplemental Figure 3-4).

Located in the middle domain, V128 resides in a surface exposed loop away from protomer interfaces; therefore, structural analysis does not provide a clear reason for this residue affecting cleavage significantly—Nsp15 V128F is twofold more active than WT Nsp15. However, the change from a small hydrophobic residue to a large hydrophobic residue could possibly lead to changes in the domain conformation to avoid surface exposure. Similarly, T113I also lies in a surface exposed loop that would not be predicted to lead to an increase in activity in the cleavage of

Table 1 **Summary of Nsp15 variants analyzed in this study.** A summary table of the variants categorized by domain of Nsp15 (light orange, ND; medium orange, MD; dark orange, EndoU). The rate refers to mutation rate as a percentage. The impact score *q* value describes its position in the distribution of all variants where the range is 0–1 and 0.5 is the mean q score[58]. The hexamer/monomer ratio is displayed along with the activity of the variant compared to WT as a percentage (based on FRET endpoint data; NSC, no significant change). Residues with asterisks (*) denote variants below the 1,000 threshold in GISAID that were selected based on previous biochemical data.

Residue	Rate (%)	Impact score (q)	Hex/Mono Ratio	Activity (%)
K13N	0.070	0.83	0.63	19
K13R	0.090	0.83	1.18	54
G18E*	0.006	0.24	0.56	50
G18R*	0.016	0.24	0.88	NSC
T34I	0.520	0.69	1.01	44
A93S	0.057	0.30	0.49	NSC
T113I	0.148	0.10	1.79	250
T115A	0.095	0.14	0.55	NSC
V128F	0.234	0.36	0.50	275
D133Y	0.104	0.73	0.48	NSC
L163F	0.109	0.33	0.23	24
P206S	0.307	0.19	2.49	NSC
R207S	0.324	0.77	0.14	22
D220Y	0.542	0.40	0.63	NSC
H235Y	0.191	0.90	1.61	5.8
K260R	0.260	0.48	1.80	42
K290N*	0.021	0.61	2.12	0.2
V321M	0.173	0.45	0.65	194
W333C*	0.016	0.73	0.77	17

single stranded RNA. This is not seen when T115, two residues down on the same loop, is mutated. These two variants also show opposite trends in oligomerization; T113I yields increased hexamer, while T115A leads to increased monomer. These data suggest surface exposed loops can alter structural conformations to affect cleavage. In virus, increased nuclease activity could potentially tip the balance from being protective (cleaving viral RNA to avoid dsRNA sensors) to harmful (preventing viral replication due to increased cleavage). Thus, our *in vitro* nuclease assays suggest these mutations could be detrimental to viral proliferation, yet they were found in numerous isolates. In fact, T113I was computationally predicted to increase viral fitness[39]. Finally, in a recently determined structure of Nsp15 bound to a dsRNA substrate, T113 forms part of a large platform for dsRNA binding, thus T113 likely facilitates the binding of longer RNA substrates (Supplemental Figure 5)[28]. In the future to better understand the impact of the T113I mutation and the other variants, mutational studies in virus would be beneficial.

Along with a significant reduction in hexamer, the remaining L163F hexamer was less active than WT. The combined oligomerization and nuclease activity results for this variant suggest this mutant would produce very little active enzyme in virus. This residue is located near residues N164 in SARS-CoV-2 (equivalent to N157 in MERS)[41]. Mutation of MERS N157 decreased RNA binding affinity, reduced nuclease activity, decreased oligomerization, and decreased the thermal stability of the endoribonuclease[41]. Our analysis of L163F in SARS-CoV-2 supports earlier work that this interface is critical for nuclease activity and protein stability.

EndoU domain variants

Finally, we analyzed specific variants from the EndoU domain, which contains the uridine binding pocket and catalytic core (Figure 5(A)). P206 (which exists at the end of the long linker between the MD and EndoU domains), R207, K260 and V321 occupy the face of the EndoU domain opposite of the active site. H235, K290, and W333 lie in the active site on the other side of the EndoU domain and have been well-characterized biochemically[29–30,41–42]. H235 and K290, along with H250, form the catalytic triad and interact with the scissile phosphate. It is therefore somewhat surprising that H235Y would be found in so many isolates (3630)—in fact, recent work characterized this mutation as a lineage marker for a clade of the Delta variant, Delta D. W333 forms important π-stacking interactions with the base 3′ of the cleaved uridine. D220 occupies the same face as the active site but is farther removed from the RNA binding sites. It is surface exposed, and the side chain does not form any interactions with other residues (defined by a distance of <5 Å). In contrast to the other domains, several EndoU variants showed an increased hexamer/monomer ratio: P206S, H235Y, K260R, and K290N. R207S,

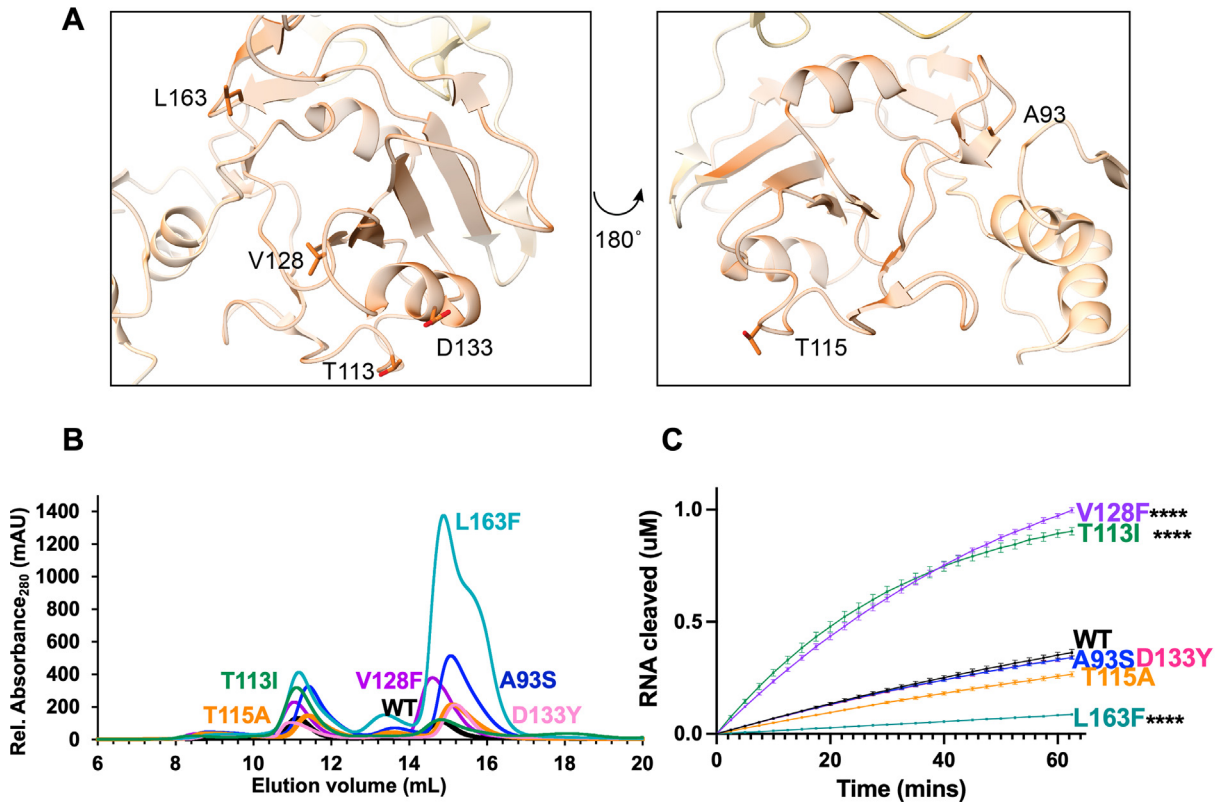

Figure 4. Characterization of Nsp15 Middle domain variants. **(A)** Displayed on the structure are the residues A93, T115, V128, D133, and L163. **(B)** S200 elution profiles of MD variants. Hexameric Nsp15 elutes at 11 mL, monomeric Nsp15 at 15 mL. **(C)** FRET time course data for Nsp15 Middle domain mutants. Nsp15 WT and variants (2.5 nM) were incubated with RNA (0.8 μM) at room temperature and fluorescence was monitored every 2.5 min for an hour. The average of a representative technical triplicate is plotted with standard deviation error bars. At least two biological replicates were performed for each mutant. Each mutant is represented by a different color: WT Nsp15 (black), Nsp15 A93S (blue), Nsp15 T113I (green), Nsp15 T115A (orange), Nsp15 V128F (purple), Nsp15 D133Y (pink), and Nsp15 L163F (teal). ****p < 0.0001.

D220Y, V321M and W333C showed a decreased hexamer/monomer ratio indicative of increased monomer (Figure 5, Table 1).

As expected, our FRET endoribonuclease assay confirmed that the catalytic residue variants H235Y and K290N were inactive (Figure 5(C)). The other active site residue W333C also showed significantly reduced cleavage activity, confirming its importance in pi-stacking with the RNA substrate. R207S and K260R also showed significantly reduced activity, while P206S and D220Y were not significantly different and V321M had increased activity. Gel-based cleavage assays with a longer RNA substrate showed similar trends to the FRET assay data (Table 1, Supplemental Figure 6).

V321M appears to be involved in the structural integrity of the EndoU fold, as part of a β-sheet packed against an α-helix on the side opposite the active site. Our cleavage data suggests a larger sidechain at that position may change the shape of the EndoU domain to promote more efficient cleavage. This is interesting because V321M was also identified in the same computational fitness analysis as T113I, making it another great candidate for future studies in virus.

P206S and D220Y having similar activity as WT aligns with our initial hypothesis that mutants with high numbers of isolates would not significantly change nuclease function. Too much viral RNA cleavage and too little viral RNA cleavage are both likely to be detrimental to the virus, thus we would expect the more common Nsp15 variants to retain relatively the same nuclease activity.

Discussion

In this work we biochemically characterized Nsp15 variants extracted from the GISAID database of SARS-CoV-2 sequences acquired during the pandemic. Nsp15 variants are unlikely to drive viral evolution and become lineage defining markers because the protein is not involved directly in viral entry or replication. Several studies have identified Nsp15 mutations

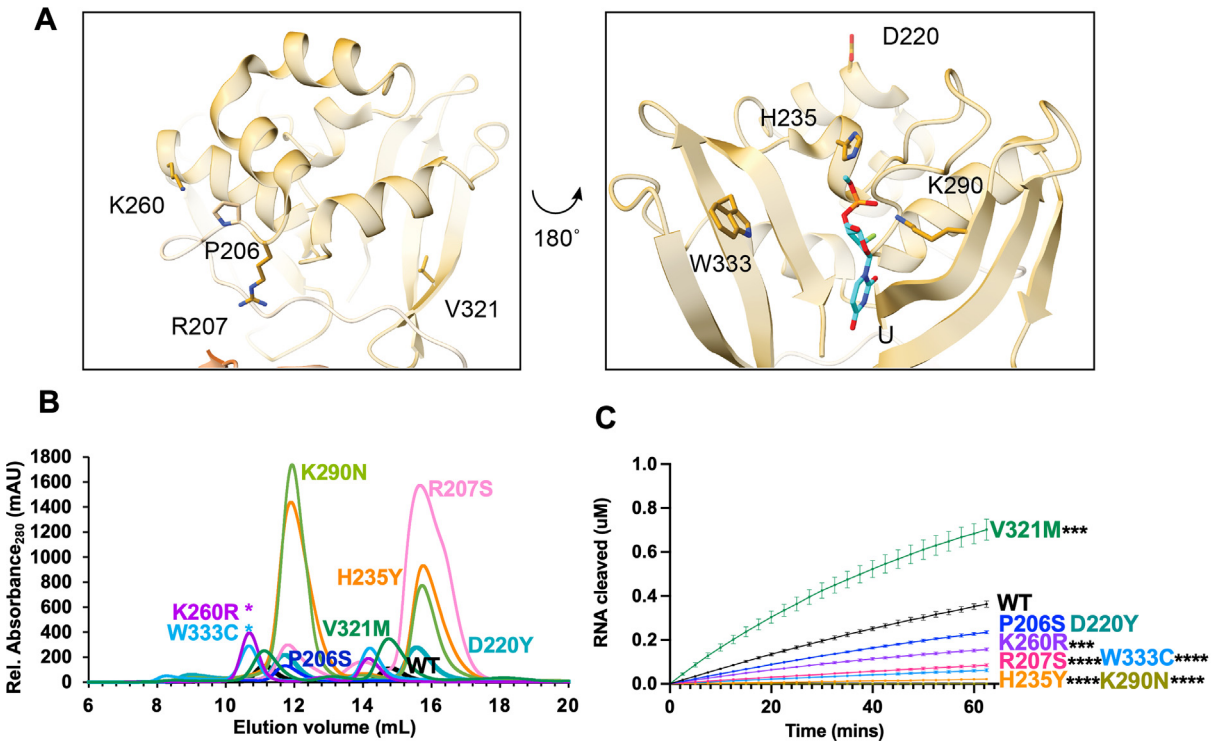

Figure 5. Characterization of Nsp15 EndoU domain variants. (A) Displayed on the structure are the residues P206, R207, D220, H235, K260, K290, V321, and W333. H235 and K290 interact with the 3′-PO_4 of uridine in the pre-cleavage RNA structure (PDB: 7 N33) **(B)** S200 elution profiles of EndoU variants. Hexameric Nsp15 elutes at 11 mL, monomeric Nsp15 at 15 mL. **(C)** FRET time course data for Nsp15 EndoU domain mutants. Nsp15 WT and variants (2.5 nM) were incubated with RNA (0.8 μM) at room temperature and fluorescence was monitored every 2.5 min for an hour. The average of a representative technical triplicate is plotted with standard deviation error bars. At least two biological replicates were performed for each mutant. Each mutant is represented by a different color: WT Nsp15 (black), Nsp15 P206S (blue), Nsp15 R207S (pink), Nsp15 D220Y (teal), Nsp15 H235Y (orange), Nsp15 K260R (purple), Nsp15 K290N (green), Nsp15 V321A (green) and Nsp15 W333C (light blue). Asterisk (*) signifies mutants were separated using a different AKTA system which affected elution volume. ***$p < 0.001$, ****$p < 0.0001$.

that may have increased fitness and clinical significance, likely by suppressing interferon I response[39,49]. Thus, characterizing variants seen in patient samples provides new knowledge about protein function beyond characterizing the active site residues directly involved in catalysis. Identifying new regions of Nsp15 that affect oligomerization and nuclease activity builds a base for mapping regulatory regions and additional drug design targets.

We specifically analyzed changes in oligomerization state and nuclease activity. The majority of the variants studied resulted in decreased Nsp15 activity. This work highlights the importance of combining bioinformatics and structural data to predict mutation effects and then testing those hypotheses *in vitro*. For example, based on Nsp15 structures alone, we would not have predicted a large impact of the T34I mutation, as it is a core residue that only interacts with other residues in the N-terminal domain. Indeed, the oligomeric state is not perturbed; however, our nuclease assays reveal a significant decrease in activity. Similarly, V128F would also not be predicted to have a great impact based solely on structural information, since it is not directly involved in any protomer interfaces and far from any EndoU active sites. However, our assays showed it decreases the formation of hexamer, but the hexamer formed has a substantial increase in nuclease activity. This suggests that structural conformations in that region could be manipulated to regulate nuclease activity, perhaps making it a surface where protein–protein interactions take place inside infected cells. Biochemical characterization is not high throughput, therefore identification and selection of variants to test is important. Mutations with a high number of isolates are an easy selection criterion; our selection of K13, T34, T115, and R207 agree with bioinformatics analysis of main mutated residues across the SARS-CoV-2 proteome published in early 2021[18]. Knowledge of molecular mechanisms is also important; this led us to select N-terminal residue and active site residue mutations to further test the established roles of these amino acids[30,41–42,48]. We recently deter-

mined a structure of Nsp15 bound to dsRNA[28]; of the Nsp15 mutants we characterized here, the only non-active site residues in the area of the dsRNA are T113 (~4 Å), D133 (~5 Å) and V128 (~9 Å) (Supplemental Figure 5).

Coronavirus genomes have a bias for uridines[50–51] so the evolution of a uridine-specific nuclease is interesting. Clearly, regulation of viral RNA through cleavage of uridines both in the polyU sequence of the negative strand (complementary to the polyA tail)[23]and throughout the positive strand[52] is important for the coronavirus life cycle. Except for the Nsp15 H235Y mutation, which serves as a marker of a Delta viral clade[35], none of the active site residues accumulated mutations above the 1000 isolate threshold we used. The lack of nuclease dead isolates signifies the importance of nuclease activity for the virus. Thus, Nsp15 H235Y stands out as a candidate to further study in viruses.

A recent computational modeling study suggested Nsp15 serves as a scaffold for the assembly of the Replication Transcription Complex (RTC)[53]. Thus, some of the variants may disrupt RTC protein–protein interactions. In the MD, Nsp15 T115 appears to interact with Nsp8, while D133 is positioned near Nsp10 (Supplemental Figure 7). Nsp15 EndoU residues D220 and V3231 neighbor the Nsp16/10 complex in the RTC, with V321 at the end of Nsp10 not interfacing with Nsp16 and D220 close to Nsp16 (Supplemental Figure 7). In this function of scaffolding the RTC, nuclease activity may not be necessary. Thus, the H235Y and K290N mutations which kill catalytic activity but increase hexamer formation (Table 1), could support the RTC model. The Delta clade with Nsp15 H235Y has adaptations in other virally-encoded immune evasion proteins that may fill the gap of a nuclease dead Nsp15. However, a recent preprint analyzed transmission dynamics of SARS-CoV-2 variants and found the H235Y mutation along with several other mutations in that Delta clade led to transmission suppression compared to other lineages[54], evidence that a lack of Nsp15 activity may contribute to poor viral transmission.

Work to develop virus-like particles to study SARS-CoV-2 variants found that an important RNA packaging signal overlaps with the Nsp15 coding sequence[55]; therefore, variants that do not affect oligomerization or activity nonetheless could affect RNA packaging. Similarly, synonymous amino acid mutations may also result in changes to RNA packaging. More research is needed to understand the determinants of that *cis*-acting element[55]. Additionally, a recent preprint proposes an RTC model that positions Nsp15 to discriminate cleavage sites based on RNA structure as opposed to sequence[56]. This may be another way that variants impact function at the RNA level as opposed to the protein level.

Biochemical and *in vivo* studies of SARS-CoV-2 protein variants beyond the spike protein will contribute to a fuller picture of how the coronavirus proteome functions and changes over time. Recent work analyzing RdRp variants and the effect on remdesivir inhibition serve as one example[57]. Earlier this year a SARS-CoV-2 molecular dynamics database was created that takes the wealth of bioinformatics data and structural information we have on SARS-CoV-2 and applied molecular dynamics simulations to estimate the impact of variants[58]. This high-throughput source of information is a great starting point to better understand viral protein evolution. The addition of *in vitro* data points, like the analyses presented in this paper, as well as *in vivo* analyses, to repositories like this database could facilitate a holistic understanding of the molecular impacts of viral mutations.

Materials and Methods

Nsp15 variants analysis

A total of 51,536,473 entries of the SARS-CoV-2 nucleotide sequences were downloaded from GISAID (version: allnuc0614). The downloaded data contains the coding sequences for genes including Nsp15 based on the hCoV-19/Wuhan/WIV04/2019 reference. Only full-length Nsp15 sequences (1038 bp) were retained for the downstream analysis, so 1,905,411 entries were included in the mutational analysis. The sequences of Nsp15 isolates were then aligned to that of the original Wuhan isolate (GenBank NC_045512.2) using the nucmer command from MUMmer 4.0 package[59] with default parameters. Variant data from the nucmer alignments were generated using show-snps command and the output was parsed directly using a custom perl script to convert MUMmer snps file to a tab delimited vcf-like table. Protein annotation was included in the final table (Supplemental File 1). Variants selected for testing were primarily at the amino acid level. Mutations that did not result in a specific amino acid change were discarded. The snps caused the same amino acid change (i.e. H235Y) were then summed to provide the total number of isolates per amino acid mutation. Visualizations of the variants on the Nsp15 DNA and protein sequences were generated via the lollipop function in trackViewer R package[60].

Protein expression of Nsp15 variants

WT Nsp15 was previously synthesized by Genscript (Piscataway, NJ), and contains an N-terminal His tag with thrombin and TEV cleavage sites in pET14b[29]. For this study, Genscript mutated the WT sequence to our variants of interest (Supplemental Table 1). WT Nsp15 and Nsp15 variants were overexpressed in *E. Coli* C41 (DE3) competent cells cultured in Terrific Broth supplemented with 100 μg/mL ampicillin. Transformed cell cultures

were grown to an optical density (600 nm) of 0.6–1.0 at 37 °C prior to induction with 0.2 mM IPTG and overnight incubation at 16 °C. Harvested cells were stored at −80 °C until needed.

Protein purification

Protein purification was carried out as previously described[29]. Cells were resuspended in lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 5% glycerol, 5 mM β-ME, 5 mM imidazole) and then supplemented with complete EDTA-free protease inhibitor tablets (Roche). They were disrupted by sonication and the lysate was clarified by centrifugation at 15,000 rpm for 50 min at 4 °C, followed by incubation for 45 minutes with a TALON metal affinity resin (Clontech). His-Thrombin-TEV-Nsp15 variants were eluted with 250 mM imidazole and incubated with thrombin (Sigma) at room temperature in Thrombin Cleavage Buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5% glycerol, 2 mM β-ME, 2 mM $CaCl_2$) for a 4 h time period. Thrombin cleavage was quenched by the addition of 1 mM PMSF (phenylmethylsulfonyl fluoride). The cleavage reactions were incubated with TALON metal affinity resin, and tagless protein was eluted in batch and resolved by gel filtration using a Superdex-200 column equilibrated with SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MnCl_2$, 5 mM β-ME). The peak fraction corresponding to the hexamer was used in subsequent assays. SDS-PAGE was used to assess protein purity (Supplemental Figure 1).

Nsp15 endoribonuclease FRET assay

Real-time Nsp15 RNA cleavage was monitored as previously described[29–30] with minor modifications to the protocol. The 5′-fluorescein (FI) label on the RNA substrate is quenched by its 3′-TAMRA label (5′-FI-AAAUAA-TAMRA-3′). A negative control substrate (5′-FI-AAAAAA-TAMRA-3′ was also used (no cleavage observed; data not shown). The FRET RNA substrate (0.8 μM) was then incubated with a constant amount of Nsp15 variant (2.5 nM) in RNA cleavage buffer (20 mM HEPES pH 7.5, 75 mM NaCl, 5 mM $MnCl_2$, 5 mM β-ME) at 25 °C for a 60 min time period. RNA cleavage was measured as an increase in fluorescein fluorescence. The fluorescence was measured every 2.5 min using a POLARstar Omega plate reader (BMG Labtech) set to excitation and emission wavelengths of 485 ± 12 nm and 520 nm, respectively. Three technical replicates were performed to calculate the mean, standard deviation, and pairwise comparison test. (Dunnett's T3 multiple corrections test, Prism/Graphpad).

Gel-based endoribonuclease assay

Gel-based cleavage assays were performed as described previously[30]. Double labeled RNA substrates (5′-FI and 3′-Cy5, 500 nM) were incubated with Nsp15 (50 nM) in RNA cleavage buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MnCl_2$, 5 mM DTT, 1 u/μl RNasin ribonuclease inhibitor) at room temperature for 30 min with the reaction samples collected at 0, 1, 5, 10 and 30 min. The reactions were quenched with 2x urea loading buffer (8 M urea, 20 mM Tris pH 8.0, 1 mM EDTA). Loading buffer without dye was used due to the expected size of cleavage products and the size of bromophenol blue. To monitor the gel front, control lanes of protein only with bromophenol blue were run. To generate a ladder, alkaline hydrolysis of the RNA was carried out for 15 min at 90 °C using 1 μM RNA in alkaline hydrolysis buffer (50 mM sodium carbonate pH 9.2, 1 mM EDTA) and quenched with 2x urea loading buffer. The cleavage reactions were separated using 15% TBE-urea PAGE gels and visualized with a Typhoon RGB imager (Amersham) using Cy2 (λ_{ex} = 488 nm, λ_{em} = 515–535 nm) and Cy5 (λ_{ex} = 635 nm, λ_{em} = 655–685 nm) channels.

CRediT authorship contribution statement

Isha M. Wilson: Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Visualization, Methodology. **Meredith N. Frazier:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Visualization, Methodology, Supervision, Project administration. **Jian-Liang Li:** Investigation, Writing – original draft, Visualization, Data curation, Formal analysis. **Thomas A. Randall:** Investigation, Writing – original draft, Visualization, Data curation, Formal analysis. **Robin E. Stanley:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Acknowledgements

We would like to thank Drs. Andrea Kaminski and Zachary Kockler for their critical review of this manuscript, and Dr. Juno Krahn for writing a Perl script converting number of mutations to PDB B-factors. This work was supported by the US National Institutes of Health (NIH) Intramural Research Program; US National Institute of Environmental Health Sciences (NIEHS, NIEHS/NIH ZIA ES103247 (RES) from the Division of Intramural Research of the NIH/NIEHS. This work was also supported by the NIH Intramural Targeted Anti-COVID-19 (ITAC) Program funded by the National Institute of Allergy and Infectious Diseases (NIAID, NIAID/NIH 1ZIAES103340 (RES)). The GISAID database was used to access SARS-CoV-2 sequence information; see

Supplemental File 5 for table acknowledging the contributions of submitting and originating laboratories.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Conflicts of Interests

The authors declare no conflicts of interest.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2022.167796.

Received 10 May 2022;
Accepted 15 August 2022;
Available online 19 August 2022

Keywords:
Nsp15;
SARS-CoV-2;
EndoU;
endoribonuclease;
coronavirus

† Present address.

Abbreviations:
COVID-19, Coronavirus Disease 2019; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; Nsp15, Non-structural protein 15; RNA, ribonucleic acid; MERS-CoV, Middle Eastern Respiratory Syndrome Coronavirus; SARS-CoV-1, Severe Acute Respiratory Syndrome Coronavirus 1; pp1a, polyprotein 1a; pp1ab, polyprotein 1ab; Non-structural proteins, nsps; Replication Transcription Complex, RTC; dsRNA, double stranded RNA; MDA5, melanoma differentiation-associated protein 5; NTD, N-terminal domain; MD, Middle domain; EndoU, endonuclease domain; GISAID, Global initiative on sharing all influenza data; WT, wild-type; IPTG, Isopropyl β- d-1-thiogalactopyranoside; β-ME, 2-mercaptoethanol; EDTA, Ethylenediamine tetraacetic acid; TEV, Tobacco Etch Virus cleavage site; PMSF, phenylmethylsulfonyl fluoride; SEC, size exclusion chromatography; FRET, fluorescence resonance energy transfer; FI, fluorescein; TAMRA, Carboxytetramethylrhodamine; PAGE, polyacrylamide gel electrophoresis

References

1. V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H., Thiel, V., (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170.
2. Paules, C.I., Marston, H.D., Fauci, A.S., (2020). Coronavirus Infections-More Than Just the Common Cold. *JAMA* **323**, 707–708.
3. Dolan, P.T., Whitfield, Z.J., Andino, R., (2018). Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. *Annu. Rev. Virol.* **5**, 69–92.
4. Hughes, A.L., Hughes, M.A., (2007). More effective purifying selection on RNA viruses than in DNA viruses. *Gene* **404**, 117–125.
5. Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., (2020). Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol. Cell* **79**, 710–727.
6. Christofi, T., Zaravinos, A., (2019). RNA editing in the forefront of epitranscriptomics and human health. *J. Transl. Med.* **17**, 319.
7. Saville, J.W., Mannar, D., Zhu, X., Srivastava, S.S., Berezuk, A.M., Demers, J.P., (2022). Structural and biochemical rationale for enhanced spike protein fitness in delta and kappa SARS-CoV-2 variants. *Nat. Commun.* **13**, 742.
8. Han, P., Li, L., Liu, S., Wang, Q., Zhang, D., Xu, Z., (2022). Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell* **185** 630-40 e10.
9. Cui, Z., Liu, P., Wang, N., Wang, L., Fan, K., Zhu, Q., (2022). Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. *Cell* **185** 860-71 e13.
10. Cerutti, G., Guo, Y., Liu, L., Liu, L., Zhang, Z., Luo, Y., (2022). Cryo-EM structure of the SARS-CoV-2 Omicron spike. *Cell Rep.* **38**, 110428.
11. Zhang, J., Xiao, T., Cai, Y., Lavine, C.L., Peng, H., Zhu, H., (2021). Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science* **374**, 1353–1360.
12. Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S.M., (2021). Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* **372**, 525–530.
13. Yang, T.J., Yu, P.Y., Chang, Y.C., Liang, K.H., Tso, H.C., Ho, M.R., (2021). Effect of SARS-CoV-2 B.1.1.7 mutations on spike protein structure and function. *Nat. Struct. Mol. Biol.* **28**, 731–739.
14. Wang, Y., Xu, C., Wang, Y., Hong, Q., Zhang, C., Li, Z., (2021). Conformational dynamics of the Beta and Kappa SARS-CoV-2 spike proteins and their complexes with ACE2 receptor revealed by cryo-EM. *Nat. Commun.* **12**, 7345.
15. Gobeil, S.M., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Stalls, V., (2021). Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* **373**
16. Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., Glaunsinger, B.A., (2020). The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934.
17. Malone, B., Urakova, N., Snijder, E.J., Campbell, E.A., (2022). Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.* **23**, 21–39.
18. Vilar, S., Isom, D.G., (2021). One Year of SARS-CoV-2: How Much Has the Virus Changed? *Biology (Basel)* **10**
19. Jaroszewski, L., Iyer, M., Alisoltani, A., Sedova, M., Godzik, A., (2021). The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins. *PLoS Comput. Biol.* **17**, e1009147.
20. Magazine, N., Zhang, T., Wu, Y., McGee, M.C., Veggiani, G., Huang, W., (2022). Mutations and Evolution of the SARS-CoV-2 Spike Protein. *Viruses.* **14**

21. Pereson, M.J., Flichman, D.M., Martinez, A.P., Bare, P., Garcia, G.H., Di Lello, F.A., (2021). Evolutionary analysis of SARS-CoV-2 spike protein for its different clades. *J. Med. Virol.* **93**, 3000–3006.
22. Li, F., (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol.* **3**, 237–261.
23. Hackbart, M., Deng, X., Baker, S.C., (2020). Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc. Natl. Acad. Sci. U S A.* **117**, 8094–8103.
24. Deng, X., Baker, S.C., (2018). An "Old" protein with a new story: Coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology* **517**, 157–163.
25. Deng, X., Hackbart, M., Mettelman, R.C., O'Brien, A., Mielech, A.M., Yi, G., (2017). Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. U S A.* **114**, E4251–E4260.
26. Deng, X., Buckley, A.C., Pillatzki, A., Lager, K.M., Faaberg, K.S., Baker, S.C., (2020). Inactivating Three Interferon Antagonists Attenuates Pathogenesis of an Enteric Coronavirus. *J. Virol.* **94**
27. Liu, X., Fang, P., Fang, L., Hong, Y., Zhu, X., Wang, D., (2019). Porcine deltacoronavirus nsp15 antagonizes interferon-beta production independently of its endoribonuclease activity. *Mol. Immunol.* **114**, 100–107.
28. Frazier, M.N., Wilson, I.M., Krahn, J.M., Butay, K.J., Dillard, L.B., Borgnia, M.J., (2022). Flipped over U: structural basis for dsRNA cleavage by the SARS-CoV-2 endoribonuclease. *Nucleic Acids Res.*.
29. Pillon, M.C., Frazier, M.N., Dillard, L.B., Williams, J.G., Kocaman, S., Krahn, J.M., (2021). Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat. Commun.* **12**, 636.
30. Frazier, M.N., Dillard, L.B., Krahn, J.M., Perera, L., Williams, J.G., Wilson, I.M., (2021). Characterization of SARS2 Nsp15 nuclease activity reveals it's mad about U. *Nucleic Acids Res.* **49**, 10136–10149.
31. Youngchang Kim, J.W., Maltseva, Natalia, Chang, Changsoo, Jedrzejczak, Robert, Wilamowski, Mateusz, Kang, Soowon, Nicolaescu, Vlad, Randall, Glenn, et al., (2020). Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. *bioRxiv*. https://doi.org/10.1101/2020.06.26.173872.
32. Kim, Y., Wower, J., Maltseva, N., Chang, C., Jedrzejczak, R., Wilamowski, M., (2021). Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Commun. Biol.* **4**, 193.
33. Kim, Y., Jedrzejczak, R., Maltseva, N.I., Wilamowski, M., Endres, M., Godzik, A., (2020). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* **29**, 1596–1605.
34. Mushegian, A., Sorokina, I., Eroshkin, A., Dlakic, M., (2020). An ancient evolutionary connection between Ribonuclease A and EndoU families. *RNA* **26**, 803–813.
35. Adi Stern, S.F., Kustin, Talia, Dotan, Edo, Mandelboim, Michal, Erster, Oran, (2021). Israel Consortium of SARS-CoV-2 sequencing, Ella Mendelson, Orna Mor, Neta S. Zuckerman. The unique evolutionary dynamics of the SARS-CoV-2 Delta variant. *medRxiv*.
36. Anindita Banerjee, A.M., Roy, Jayita, Majumdar, Agniva, Chatterjee, Ananya, Biswas, Nidhan K., Sarkar, Mamta Chawla, Maitra, Arindam, Dutta, Shanta, et al., (2022). Evolution of Delta variant by non-Spike signature co-appearing mutations: trailblazer of COVID-19 disease outcome. *bioRxiv*.
37. Resende, P.C., Graf, T., Paixao, A.C.D., Appolinario, L., Lopes, R.S., Mendonca, A., (2021). A Potential SARS-CoV-2 Variant of Interest (VOI) Harboring Mutation E484K in the Spike Protein Was Identified within Lineage B.1.1.33 Circulating in Brazil. *Viruses.* **13**
38. Qussai Abbas, A.K., Sharrouf, K., Jyakhwo, S., Komissarov, A.S., (2022). Follow-up investigation and detailed mutational characterization of the SARS-CoV-2 Omicron variant lineages (BA.1, BA.2, BA.3 and BA.1.1). *bioRxiv*.
39. Obermeyer, F., Jankowiak, M., Barkas, N., Schaffner, S.F., Pyle, J.D., Yurkovetskiy, L., (2022). Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. *Science* **376**, 1327–1332.
40. Shu, Y., McCauley, J., (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro. Surveill.* **22**
41. Zhang, L., Li, L., Yan, L., Ming, Z., Jia, Z., Lou, Z., (2018). Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. *J. Virol.* **92**
42. Bhardwaj, K., Palaninathan, S., Alcantara, J.M.O., Li Yi, L., Guarino, L., Sacchettini, J.C., (2008). Structural and functional analyses of the severe acute respiratory syndrome coronavirus endoribonuclease Nsp15. *J. Biol. Chem.* **283**, 3655–3664.
43. Zheng, A., Shi, Y., Shen, Z., Wang, G., Shi, J., Xiong, Q., (2018). Insight into the evolution of nidovirus endoribonuclease based on the finding that nsp15 from porcine Deltacoronavirus functions as a dimer. *J. Biol. Chem.* **293**, 12054–12067.
44. Fitch, W.M., (1967). Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. *J. Mol. Biol.* **26**, 499–507.
45. Klimczak, L.J., Randall, T.A., Saini, N., Li, J.L., Gordenin, D.A., (2020). Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. *PLoS ONE* **15**, e0237689.
46. Lynch, M., (2010). Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U S A.* **107**, 961–968.
47. Viana, R., Moyo, S., Amoako, D.G., Tegally, H., Scheepers, C., Althaus, C.L., (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **603**, 679–686.
48. Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., (2007). Crystal structure of a monomeric form of severe acute respiratory syndrome coronavirus endonuclease nsp15 suggests a role for hexamerization as an allosteric switch. *J. Virol.* **81**, 6700–6708.
49. Ichikawa, T., Torii, S., Suzuki, H., Takada, A., Suzuki, S., Nakajima, M., (2022). Mutations in the nonstructural proteins of SARS-CoV-2 may contribute to adverse clinical outcome in patients with COVID-19. *Int. J. Infect. Dis.* **122**, 123–129.
50. Berkhout, B., van Hemert, F., (2015). On the biased nucleotide composition of the human coronavirus RNA genome. *Virus Res.* **202**, 41–47.
51. Kustin, T., Stern, A., (2021). Biased Mutation and Selection in RNA Viruses. *Mol. Biol. Evol.* **38**, 575–588.

52. Ancar, R., Li, Y., Kindler, E., Cooper, D.A., Ransom, M., Thiel, V., (2020). Physiologic RNA targets and refined sequence specificity of coronavirus EndoU. *RNA* **26**, 1976–1999.
53. Perry, J.K., Appleby, T.C., Bilello, J.P., Feng, J.Y., Schmitz, U., Campbell, E.A., (2021). An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15. *J. Biol. Chem.* **297**, 101218.
54. Hsin-Chou Yang, J.-H.W., Yang, Chih-Ting, Lin, Yin-Chun, Hsieh, VHan-Ni, Chen, Po-Wen, Liao, Hsiao-Chi, Chen, Chun-Houh, Liao, James C., (2022). Subtyping of major SARS-CoV-2 variants reveals different transmission dynamics. *bioRxiv*.
55. Syed, A.M., Taha, T.Y., Tabata, T., Chen, I.P., Ciling, A., Khalid, M.M., (2021). Rapid assessment of SARS-CoV-2-evolved variants using virus-like particles. *Science* **374**, 1626–1632.
56. Jianguang Liang, J.S., Chen, Shunmei, Duan, Guangyou, Yang, Fan, Cheng, Zhi, Li, Xin, Ruan, Jishou, Mi, Dong, Gao, (2022). How the replication and transcription complex functions in jumping transcription of SARS-CoV-2. *bioRxiv*.
57. Stevens, L.J., Pruijssers, A.J., Lee, H.W., Gordon, C.J., Tchesnokov, E.P., Gribble, J., (2022). Mutations in the SARS-CoV-2 RNA dependent RNA polymerase confer resistance to remdesivir by distinct mechanisms. *Sci. Transl. Med.*, eabo0718.
58. Torrens-Fontanals, M., Peralta-Garcia, A., Talarico, C., Guixa-Gonzalez, R., Giorgino, T., Selent, J., (2022). SCoV2-MD: a database for the dynamics of the SARS-CoV-2 proteome and variant impact predictions. *Nucleic Acids Res.* **50**, D858–D866.
59. Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., Zimin, A., (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944.
60. Ou, J., Zhu, L.J., (2019). trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. *Nat. Methods* **16**, 453–454.