SARS-CoV-2 Genome from the Khyber Pakhtunkhwa Province of Pakistan

Muhammad Tahir Khan, Sajid Ali, Anwar Sheed Khan, Noor Muhammad, Faiza Khalil, Muhammad Ishfaq, Muhammad Irfan, Abdullah G. Al-Sehemi, Shabbir Muhammad, Arif Malik,* Taj Ali Khan,* and Dong Qing Wei*

ABSTRACT: Among viral outbreaks, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is one of the deadliest ones, and it has triggered the global COVID-19 pandemic. In Pakistan, until 5th September 2020, a total of 6342 deaths have been reported, of which 1255 were from the Khyber Pakhtunkhwa (KPK) province. To understand the disease progression and control and also to produce vaccines and therapeutic efforts, whole genome sequence analysis is important. In the current investigation, we sequenced a single sample of SARS-CoV-2 genomes (accession no. MT879619) from a male suspect from Peshawar, the KPK capital city, during the first wave of infection. The local SARS-CoV-2 strain shows some unique characteristics compared to neighboring Iranian and Chinese isolates in phylogenetic tree and mutations. The circulating strains of SARS-CoV-2 represent an intermediate evolution from China and Iran. Furthermore, eight complete whole genome sequences, including the current Pakistani isolates which have been submitted to Global Initiative on Sharing All Influenza Data (GSAID), were also investigated for specific mutations and characters. Some novel mutations [NSP2 (D268del), NSP5 (N228K), and NS3 (F105S)] and specific characters have been detected in the coding regions, which may affect viral transmission, epidemiology, and disease severity. The computational modeling revealed that a majority of these mutations may have a stabilizing effect on the viral protein structure. In conclusion, the genome sequencing of local strains is important for better understanding the pathogenicity, immunogenicity, and epidemiology of causative agents.

INTRODUCTION

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a positive RNA virus (+ssRNA) with a single-stranded genome of the betacoronavirus family, which also includes MERS-CoV and SARS-CoV. Among the RNA viruses, coronaviruses possess the largest genome (30 kb) that contains structural and accessory genes, ample replicas, and other nonstructural proteins (NSPs).1–4 The pandemic appears to be spreading worldwide through human-to-human transmission.5 According to the World Health Organization report of 3rd December 2020, globally there have been 63,965,092 confirmed cases and 1,488,120 deaths. In a recent study6 the SARS-CoV-2 trajectory prediction was carried out, and it was concluded that during the beginning of the pandemic, Wuhan performed a more active and effective nonpharmaceutical intervention.

The ORF1a/b, the largest region (two-third) of the SARS-CoV-2 genome, is translated into two large polyproteins, pp1a and pp1ab (NSP1-NSP16).7 The matrix (M), nucleocapsid phosphoprotein (N), envelope (E), and spike (S) are structural proteins, functioning along with RNA and NSP1-16, to facilitate the replication of the virus within the host cell. The complete virion includes the structural proteins in combination with the RNA genome.

The processing of pp1a and pp1ab resulted in 16 viral NSPs,8 which were assembled into replication and transcription complexes. These complexes are involved in multiple functions, ranging from replication to the processing of polypeptides.9–17 The SARS-CoV-2 genome continuously undergoes significant variation in NSPs particularly, S protein, NSP3, and RNA-dependent RNA polymerase (RdRp).18,19 The S protein is the key factor of evolution, pathogenicity, and transmission that might be important in vaccine
development. Moreover, RdRp might be an important target for designing antivirals against SARS-CoV-2.23 In a recent study, 12,706 mutations were detected within 46,723 data sets of SARS-CoV-2 genomic isolates from worldwide patients, among which 398 were strongly supported recurrent mutations.20

A virus’s sequence mutation rate is driven by multiple factors, including selective pressure, and processes such as 3’-exonuclease activity, replicative repairing, and so on1,22 play an important role in evolution. During a pandemic, the whole genome sequence analysis is an important strategy to monitor the disease progression, control, and therapeutic efforts.23 Many important proteins harbor mutations that could possibly make them poor drug targets. Mutations also play a key role in the transmission and pathogenicity of the virus. Therefore, before the development of drugs and diagnostic tools, constant molecular scrutiny should be performed for the better management of COVID-19.

In the current study, a single genome of the SARS-CoV-2 from a male suspect was sequenced to analyze the common mutations along with seven other whole genome sequences from other locations in Pakistan, which were retrieved from GISAID.24 The genomes harbor some novel mutations and characters that could be useful in the analysis of the COVID-19 epidemiology and severity.

### RESULTS AND DISCUSSION

The sequence of SARS-CoV-2 (SARS-CoV-2-KPK-KUST-SJTU/2020) has been submitted to GenBank under accession number MT879619. The statistics of the sequence are shown in Table 1. The sequencing coverage was found to be 99.97% with 367,235 reads.

| statistics type       | number       |
|-----------------------|--------------|
| aligned bases         | 45,417,209   |
| aligned reads number  | 367,235      |
| coverage %            | 99.97        |
| duplication rate %    | 31.4093      |
| indel rate %          | 0.0055       |
| mean read length      | 125.6        |
| mismatch rate %       | 0.24         |
| sequencing depth      | 1438.63      |
| variants number       | 15           |

A total of fifteen mutations were detected, among which eight were nonsynonymous. Four of these were detected in the orf1ab, one in each of S, ORF3a, N, and orf10, respectively (Table 2). The mutation L3606F in orf1ab was also reported in Japanese isolate (accession no. LC528232), while 2702Q > H and 5561A > T seem novel in Pakistani isolates. Pro4715Leu has been detected in NSP12/RdRP (orf1ab).

The variant D614G in S protein detected in the current study is more commonly present in European isolates, such as those from Spain, Belgium, France, Italy, Switzerland, and the Netherlands, and appears more severe and fatal, accounting for a huge death toll (https://www.worldometers.info/coronavirus/);25 Germany, Kuwait, and Pakistan have the wild-type 614D at S in a majority of strains, with a lower death toll. The scenario in Germany, compared to other European countries, remains uncertain. Therefore, continuous follow-up will be vital to assess the SARS-CoV-2 genomic variations and severity. The mutation 327S > L is present in the C-terminal domain of N proteins. Recently, it was observed that 327S > L may increase the stability in N proteins but decrease the molecular flexibility.26

The nucleocapsid (N), orf3a, and orf10 harbored only one mutation, and these proteins have many conserved regions as variants, which have not been detected in the majority of cases.27 Among the major targets, the nucleocapsid (N) is an important protein involved in RNA binding and is essential for RNA activities such as replication. The N protein primarily promotes the binding and packing of the RNA ribonucleoprotein complex (nucleocapsid).28−31 A single mutation 241C > T in the 5′ UTR region was detected in the current study, which was also reported recently in a whole genome sequence from Gilgit (accession no. MT240479), Pakistan. Mutation Q57H in our genome (Table 4) seems very common in Indian isolates.32 A total of 128 genomes from Indian patients were analyzed, among which all the SARS-CoV2 genomes had mutations at Q57H of the protein ORF3a, except a single genome (accession no. MTS09503) from Junagadh, apart from other mutations (Table 3).

The genome size of the SARS-CoV-2 is about 29.8 kb to 29.9 kb. The 3′ genes are E, S, and N proteins and M (Figure 1). Six proteins, encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes, are also known as accessory proteins.33−35

The phylogenetic tree from 943 isolates of our two severely affected neighbors Iran (25) and China (917), along with the current isolate, downloaded from GISAID,24 was constructed using Archaeopteryx and Phylo.io36,37 ([https://mafft.cbrc.jp/alignment/server/](https://mafft.cbrc.jp/alignment/server/)) from MAFT server (Figure 2). Phylo.io with distinctive features is the scalability to large trees, rooting, and leaf order identification with the best matching, high usability, and standard HTML5 implementation. The tree based on nucleotide sequences showed the incidence of many clades and clusters of the SARS-CoV-2. The current isolate seems close to the Iranian isolate (Figure 2A) and more divergent from that collected from the National Institute of Health (NIH), Islamabad, Pakistan.

Phylogenetically, the current genome (SARS-CoV-2-KPK-KUST-SJTU/2020) shows a unique position in the tree (Figure 2). The most common mutations detected in the Pakistani isolates were NSP6_L37F (5), spike_D614G, NSP12_P323L (4), and NS3_Q57H (4). In infected individuals, G614 has been associated with high viral loads in the upper respiratory tract, but it is not involved with disease severity;36 however, it has a stabilizing effect on the protein structure (Figure 3). NS3_Q57H mutation is a clade determinant (GH) mutation. All the GH clades harbor Q57H mutation in NS3 proteins.39

**Mutation in NSP6.** Mutations (L3606F) (Table 2), which link with the position L37F of NSP6 (Table 3), are present in isolates from the USA, China, Hong Kong, France, Singapore, and Italy. In the current study, NSP6 harbors two mutations—L37F (5 isolates) and M86I (2 isolates). The SARS-CoV-2 NSP6 is associated with NSP3 and NSP4, which are transmembrane proteins, forming double-membrane-like vesicles.31 Mutation L37F is present outside the transmembrane and as a part coil segment. The residue Val at position 37 is conserved among all the sarbecoviruses’ NSP6 protein, except in SARS-CoV-2 (L37). A substitution of aliphatic Leu with an aromatic Phe amino acid might have
Table 2. Mutations Detected in SARS-CoV-2 Whole Genome of the KPK Isolate

| s.no. | position | Ref* | Alt* | gene | variant type | protein position | codon position |
|-------|----------|------|------|------|--------------|-----------------|----------------|
| 1     | 241      | C    | T    | orf1ab | upstream  | QHD34315.1     | gene-orf1ab    |
| 2     | 2416     | C    | T    | orf1ab | synonymous | 717Y > Y        | 2151TAC > TAT  |
| 3     | 3037     | C    | T    | orf1ab | synonymous | 924F > F        | 2772 TTC > TTT |
| 4     | 8371     | G    | T    | orf1ab | missense   | 2702Q > H         | 8106 CAG > CAT |
| 5     | 9208     | T    | C    | orf1ab | synonymous | 2981S > S      | 8943 TCT > TCC |
| 6     | 10741    | C    | T    | orf1ab | synonymous | 3492D > D    | 10476 GAC > GAT |
| 7     | 11083    | G    | T    | orf1ab | missense   | 3606L > F (L37F on NSP6) | 10818 TTG > TTT |
| 8     | 12565    | G    | A    | orf1ab | synonymous | 4100Q > Q   | 12300 CAG > CAA |
| 9     | 14408    | G    | T    | orf1ab | missense   | 471SP > L   | 14144 CCT > CTT |
| 10    | 16945    | G    | A    | orf1ab | missense   | 5561A > T | 16681 GCA > ACA |
| 11    | 22477    | C    | T    | S     | synonymous | 305S > S    | 915 TCC > TCT |
| 12    | 23403    | A    | G    | S     | missense   | 614D > G   | 1841 GAT > GGT |
| 13    | 25563    | G    | T    | orf3a | missense   | 57Q > H    | 1714 CAT > CAT |
| 14    | 29253    | C    | T    | N     | missense   | 327S > L  | 980 TCG > TTG |
| 15    | 29645    | G    | T    | orf10 | missense   | 30V > L   | 88 GTA > TTA |

*Ref: Reference, Alt: Alteration. *Novel.

Upregulation of the host-immune activity, where the Phe residue performs cation–π interactions, affecting protein interactions in the L37F MT. The effect of this mutation on structural dynamics could not be unveiled due to the lack of experimental data in the Protein Data Bank for NSP6 homology modeling using templates. Mutation L37F in NSP6 leads to a weak SARS-CoV-2 subtype which may help in SARS-CoV-2 transmission and evolution across various regions over time during the pandemic.42 The effect of mutation NSP3_E92K could not be evaluated due to the unavailability of the complete crystal structure and suitable template for the homology model.

**Mutation in NSP12 (RdRp).** Mutation P323L in NSP12 (RdRp) in Pakistani isolates is the second-most common variant (Table 2). Due to the unique property of proline, it cyclizes back from the side chain onto the backbone, contributing to the development of the secondary structure because of the immense pyrrolidine ring. Mutant P323L shifts the structure integrity and might have functional consequences.43 Here, in MT, we detected a decrease in RdRp flexibility but a more stabilizing effect (Figure 4), which might be stable while interacting with RNA during infection. Mutation P323L lies in residue A250−R365, which is known as the interface domain of the RdRp. This domain is a putative docking site,43 and mutation in this region may affect the interactions with some antivirals such as filbuvir, simprevir, tegobuvir, and so on.

**Mutations in NS8 (ORF8).** Among the three mutations (NS8_L84S, NS8_E92K, and NS8_W45L) in NS8 (ORF8) of SARS-CoV-2, the appearance of E92K has not been reported in earlier studies.44,45 Mutation L84S in the current study has been detected in two isolates, which is a strain determining mutation of clade S (Table 3). ORF8 is remarkably divergent and consists of a predicted Ig-like fold next to the N-terminal signal sequence.46 ORF8 from SARS-CoV and SARS-CoV-2 owns a signal sequence for endoplasmic reticulum import. Here, in the lumen, SARS-CoV-2 ORF8 interacts with ER-associated degradation factors and a variety of other host proteins.47 It is secreted because ORF8 antibodies are among the major markers in SARS-CoV-2 patients.48 ORF8 disrupts IFN-I signaling and downregulates MHC-I in cells.49 Mutation L84S in ORF8 has been associated with decreased stability of ORF8.44 L84S destabilizes the folding, which may cause upregulation of the host-immune activity.50 (Figure 5). We predicted the stability effect of MTs NS8_E92K and NS8_W45L through DynaMut online server.51,52 The result shows a destabilizing effect, which might be important for the upregulation of the immune activity and, thus, the successful eradication of the infection.

**Mutations in Nucleocapsid (N) Proteins.** In the current study, two mutations, N_S202N and N_S327L, were detected (Figures 1, 6 and Table 3). S202N is present in the serine–arginine-rich region (residues 184−2024), close to the N-terminal domain (N-NTD). Due to the unavailability of the crystal structure of this region, its impact on the dynamics of N proteins could not be explored. A recent study53,54 reported that the interactions between Nsp3 and N are mediated by residues 1 to 194 (N1a−N1b) and 195 to 257 (N2a). Mutation_N_S202N was detected in N2a; however, its impacts on the interaction with Nsp3 need to be evaluated. Mutation S327L has been detected in the C-terminal domain (247−364) (Figure 1), and its impact on the N-CTD dynamics is shown in Figure 6. Compared to wild type, the MT L327 seems more stabilized and has molecular flexibility. The N protein is 419 amino acid long and present in the nucleocapsid of the SARS-CoV-2. The N protein comprises the N-terminal domain (NTD) (46−176 aa), also called the RNA-binding domain, that binds to the 3′-end of the viral RNA, a linker of the serine–arginine-rich region (182−247 aa) which interacts directly with RNA and plays a part in cell signaling.55,56 and residues 247 to 364 of the C-terminal domain (CTD).53,54 All the three domains electrostatically interact with the SARS-CoV-2 genomic RNA and modulate unwinding. Screening mutation in the local isolate might help identify new targets and design a vaccine for the better management of COVID-19.

**Mutation in NSP13 (Helicase).** Helicases play a key role in viral RNA replication and a vital step in viral propagation and pathogenesis. Therefore, they are ideal targets for antiviral drugs. In the current study, a single substitution A237T was detected in the NSP13, which shows a more stabilizing effect on the helicase structure activity (Figure 7). Furthermore, molecular flexibility decreases in the MT residue T237, forming more interactions. Before designing a potential drug, mutations should be screened and characterized for the better management of inhibitors.
Table 3. Mutations Detected in Whole Genome Sequences of Pakistani Isolates

| Clade | Length (nt) | Length (aa) | Mutations | Mutation Description |
|-------|-------------|-------------|-----------|----------------------|
| hCoV-19/Pakistan/Gilgit1 | 29,836 | 9710 | 4 | NSP2_V198I, NSP2_R27C, NSP4_P202L, NSP6_L37F |
| hCoV-19/Pakistan/KHI1 | 29,819 | 9709 | 1 | NSP2_D268del |
| hCoV-19/Pakistan/NIH-44905 | 29,876 | 9710 | 5 | NSP6_M86I, Spike_D830A, NS8_E92K, NS8_L84S, N_S202N |
| hCoV-19/Pakistan/NIH-45143 | 29,877 | 9710 | 7 | NS3_F105S |
| hCoV-19/Pakistan/NIH-45090 | 29,880 | 9710 | 6 | NSP3_Q1884H, NSP6_L37F, NSP12_P323L, Spike_D614G, NS3_Q57H |
| hCoV-19/Pakistan/NIH-HAS001 | 29,881 | 9710 | 5 | NSP6_M86I, Spike_D830A, NS8_E92K |
| hCoV-19/Pakistan/KPK-KUST-SJTU | 29,897 | 9710 | 7 | NSP3_Q1884H, NSP6_L37F, NSP12_P323L, NSP13_A237T, N_S327L |

Note: Mut's: mutants. L: reference clade, Char: Characters. Freq: Frequency of each mutations, NSP6_L37F = 5, NSP6_M86I = 2, NSP3_Q1884H = 4, NS3_Q57H = 4, NSP12_P323L = 4, N_S202N = 3, N_S327L = 1, Spike_D614G = 4, Spike_D830A = 2, NS8_L84S = 2, NS8_E92K = 2, NS8_W45L = 1, NSP2_L270F = 1, NSP13_A237T = 1, and NSP14_T250I = 1.
Mutation S327L in CTD of N increases the flexibility, while it stabilizes the structure. Point mutations (L84S, E93K, and W45L) in NS8 (ORF8) have a destabilizing effect on the structure. However, E93K and W45L exhibited increased flexibility compared to L84S. A single novel mutation N228K detected in M\(^{63}\), located in domain III, may have catalytic consequences, which need to be validated through experimental approaches. For the better management of SARS-CoV-2 infections, whole genome sequence analysis may offer useful information behind the transmission, pathogenicity, and severity of SARS-CoV-2 isolates in specific geographic regions.

### MATERIALS AND METHODS

**Area of Sample Collection.** A single sample of nasopharyngeal swab was collected from a suspected SARS-CoV-2 male patient of district Peshawar, Khyber Pakhtunkhwa (KPK), during the first wave of infection on June 15th, 2020. The person has no idea regarding from whom he contracted the infection. Four persons in his family developed similar symptoms. However, we collected the sample from a single 54-year-old person who had developed severe symptoms including, fever, fatigue, dry cough, bone pain, shortness of breath, and loss of smell, taste, and sensation. The patient also complained about the loss of sexual desire for more than a month. The duration of smell and taste loss was not confirmed by the patient. The infection was not fatal and the patient felt better 22 days after the appearance of first symptoms of dry cough and fever. The samples (SARS-CoV-2-KPK-KUST-SJTU/2020) were submitted to GenBank under accession number MT879619 (https://www.ncbi.nlm.nih.gov/nuccore/MT879619).

**Sample Processing and Confirmation.** The sample was collected according to the complete protocol of the biosafety interim guideline. To confirm the sample was COVID-19 positive, the nasopharyngeal swab specimen was taken and placed in 3 mL of the normal saline medium. It was mixed by inverting it a minimum of five times. The diluted specimen was then transferred to a SARS-CoV-2 Xpert cartridge through a sterile dropper and loaded into the GeneXpert System platform (Rapidmicrobiology Xpert Xpress SARS-CoV-2 Point-of-Care Test).\(^{62,63}\) The GeneXpert Dx System (Cepheid, Sunnyvale, CA) is an automated sample-processing and real-time PCR component with a completely closed cartridge, containing sample-processing and lyophilized form of the real-time RT-PCR reagent. The machine consists of modules, each for separate testing purposes, to avoid cross-contamination. It has built-in auto-sample preparation, extraction, amplification, and detection for target sequence detection. The Xpert SARS-CoV-2 test targets the E and N2 genes.\(^{64}\) The system consists of a software-based auto-interpretation of results,\(^{65}\) which are automatically compared and analyzed through auto built-in, pre-established software.

**Viral RNA Extraction and Sequencing.** The sample was processed for viral RNA extraction using a QIAamp kit (Qiagen, Germany), and amplification was carried out according to the protocol of ARTIC nCoV-2019.\(^{66}\) RNA Quality Control was checked using a Qubit RNA BR assay kit (Invitrogen), and cDNA was synthesized using Revert Aid First Strand cDNA synthesis kit. (Thermo Scientific) PCR was carried out using Phusion Flash High-Fidelity PCR Master Mix, and (Thermo Scientific) library preparation was performed through NextEra XT DNA library preparation kit, Illumina, San Diego, CA. Sequencing was performed using Illumina MiSeq at Rehman Medical Institute, Peshawar, Pakistan.

**Data Analysis.** The read quality of fastq files was checked using the FastQC tool (v0.11.8). The Trimmomatic tool (v0.39) was used to remove low-quality base calls (Q < 30) and index adapter sequences from both ends of the sequenced reads. The filtered reads were aligned with the Wuhan reference genome (accession no. NC 045512) using the default settings for the Burrows Wheeler Aligner (BWA, v0.6). Using Picard Tools (v2.21.6), the PCR duplicates were removed from the reads. To solve the mapping problems resulting from the existence of small Indels, mapped reads were analyzed using the Genome Analysis Toolkit and command-line tools “RealignerTargetCreator” and “IndelRealigner” (GATK v. 3.3.0). GATK tool “HaplotypeCaller” was used to call SNPs and Indels for variant calling, through local de-novo assembly of haplotypes in the regions showing deviation. Using default settings, GATK “Variant Filtration” was used to exclude potential false variants from the raw call set of variants. The annotation was performed using publicly available tools, at China’s National Genomics Data Center (NGDC). Variants were detected through Genome-to-Variants [https://bigd.big.ac.cn/ncov/online/tool/variation].
The identified variants were cross-validated manually by loading the sequences in BioEdit and checking the reported variants one-by-one. Seven other whole genome complete sequences originating in Pakistan were also downloaded, and the variants were analyzed through GISAID (Table 4).

**Mutation Effect on Viral Protein.** The mutations in the current study were studied to explore their effect on protein structures and dynamics using the DynaMut server. The server implements two distinct, normal-mode methods, which can be used to analyze and assess the mutations’ effect on protein stability and dynamics, resulting from changes in vibrational entropy. The impact of a mutation is predicted on protein stability through the integration of normal-mode dynamics along with graph-based signatures. This approach outperforms to predict the mutations’ effect on protein stability and flexibility (P-value < 0.001). The results are displayed in good resolution in different tabs to observe the analyses available for mutations’ effect on protein dynamics and stability.

**AUTHOR INFORMATION**

**Corresponding Authors**

Arif Malik — Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore, Lahore 58810, Pakistan; Phone: +92 (0)42 111865865; Email: arif.malik@imbb.uol.edu.pk

---

Figure 2. Phylogenetic analysis of SARS-CoV-2-KPK-KUST-SJTU/2020 (accession no. MT879619). (A) SARS-CoV-2-KPK-KUST-SJTU/2020 (red arrow), Iranian isolate (number 25), and Chinese isolates (total number 917). (B) SARS-CoV-2-KPK-KUST-SJTU/2020 (red arrow) and seven other isolates. The long name at the end of each node represents the serial number among the country isolates followed by country name, specific name given to each isolate, GISAID accession ID, and date of collection.
Figure 3. Effect of point mutation (D614G) on spike protein dynamics. ΔΔG; Free energy difference. ΔΔS_{Vib} ENCoM; vibrational entropy energy. This effect has been predicted through DynaMut online server. (A) Increase in molecular flexibility (red region) due to D614G point mutation. The total energy calculated for mutants (MT) shows a stabilizing effect on the protein structure. (B) Interactions of amino acids in wild type (WT) with surrounding residues. (C) Interactions of amino acid G614 (MT) with surrounding residues.

Figure 4. Effect of point mutation (P323L) on NSP12 (RdRp) dynamics. This effect has been predicted through DynaMut online server. (A) Decrease in RdRp flexibility (blue region). The effect of P323L seems stabilizing on the protein structure. Interactions of amino acids in WT and MT with surrounding residues have been encircled. MT has more interactions than WT.
Figure 5. Effect of point mutations (L84S, E93K, and W45L) on NS8 (ORF8) structure and dynamics. L84S, E93K, and W45L have a destabilizing effect. E93K and W45L exhibited an increase in flexibility while L84S shows decrease.

Figure 6. Effect of S327L mutation on N protein structure and dynamics (PDB ID 6yun). Flexibility seems increased due to substitution of leucine in place of serine at position 327. This mutation has a stabilizing effect as shown in blue.

Dong Qing Wei — State Key Laboratory of Microbial Metabolism, Shanghai—Islamabad—Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P. R. China; Peng Cheng Laboratory, Shenzhen, Guangdong 518055, P. R. China; orcid.org/0000-0003-4200-7502; Email: dqweig@sjtu.edu.cn

Muhammad Tahir Khan — Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore, Lahore 58810, Pakistan; State Key Laboratory of Microbial Metabolism, Shanghai—Islamabad—Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P. R. China; Peng Cheng Laboratory, Shenzhen,
Figure 7. Mutation in NSP13 (Helicase) at position A237T and its dynamic effect. The structure was downloaded from Swiss-Model server (PRO_0000449630). The MT exhibits lower flexibility than WT. This mutation has a stabilizing effect as shown in blue. The MT and WT residue has been shown in light green, depicting the interaction with surrounding residues. MTs seem to form more interactions than WT.

Figure 8. Mutation (N228K) in NSP5 (main protease) and its dynamic effect.
Table 4. Sociodemographic Information of SARS-CoV-2 Genomic Isolates

| virus name                         | accession id       | collection date | location | gender | age |
|-----------------------------------|--------------------|-----------------|----------|--------|-----|
| HCOV-19/Pakistan/NIH-HA5001/2020   | EPI_ISL_468163     | 02/06/2020      | islamabad | male   | 23  |
| HCOV-19/Pakistan/NIH-45579/2020    | EPI_ISL_468162     | 02/06/2020      | islamabad | female | 46  |
| HCOV-19/Pakistan/NIH-45090/2020    | EPI_ISL_468161     | 02/06/2020      | islamabad | female | 49  |
| HCOV-19/Pakistan/NIH-45143/2020    | EPI_ISL_468160     | 02/06/2020      | islamabad | female | 55  |
| HCOV-19/Pakistan/NIH-44905/2020    | EPI_ISL_468159     | 02/06/2020      | islamabad | male   | 87  |
| HCOV-19/Pakistan/KHI1/2020         | EPI_ISL_451958     | 16/03/2020      | karachi   | unknown| unknown |
| HCOV-19/Pakistan/GILGIT1/2020      | EPI_ISL_417444     | 04/03/2020      | gilgit    | female | 40  |
| HCOV-19/Pakistan/KPK.KUST-SJTU/2020 | EPI_ISL_513925   | 15/05/2020      | peshawar  | male   | 54  |

“Current genome (KPK).”

“Current genome (KPK).”

Muhammad Ishfaq — Centre forOMIC Sciences, Islamia College Peshawar, Peshawar 25120, Pakistan

Muhammad Irfan — Department of Oral Biology, College of Dentistry, University of Florida, Gainesville, Florida 32611, United States

Abdullah G. Al-Sehemi — Research Center for Advanced Materials Science (RCAMS), King Khalid University, Abha 61413, Saudi Arabia; Department of Chemistry, College of Science, King Khalid University, Abha 61413, Saudi Arabia

Shabir Muhammad — Research Center for Advanced Materials Science (RCAMS), King Khalid University, Abha 61413, Saudi Arabia; Department of Chemistry, College of Science, King Khalid University, Abha 61413, Saudi Arabia; State Key Laboratory of Microbial Metabolism, Shanghai—Island—Belgrade Joint Innovation Center on Antibacterial Resistances, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030, P. R. China; Peng Cheng Laboratory, Shenzhen, Guangdong 518055, P. R. China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c05163

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

D.Q.W. is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (grant no. 32070662, 61832019, and 32030063); the Science and Technology Commission of Shanghai Municipal (grant no.: 19430750600), the Natural Science Foundation of Henan Province (162300210060), and SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14). The computations were partially performed at the Pengcheng Lab and the Center for High-Performance Computing, Shanghai Jiao Tong University. The authors are also thankful to the Institute of Research and Consulting Studies at King Khalid University for funding this research through grant number 2-N-20/22 and the support of the Research Center for Advanced Materials Science is highly acknowledged.

REFERENCES

(1) Gorbatenya, A. E.; Baker, S. C.; Baric, R. S.; de Groot, R. J.; Drosten, C.; Gulyaeva, A. A.; Haagmans, B. L.; Lauber, C.; Leonovich, A. M.; et al. The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It SARS-CoV-2. Nat. Microbiol. 2020, 5, 536–544.

(2) Biswas, A.; Bhattacharjee, U.; Chakrabarti, A. K.; Tewari, D. N.; Banu, H.; Dutta, S. Emergence of Novel Coronavirus and COVID-19: Whether to Stay or Die Out? Crit. Rev. Microbiol. 2020, 46, 182–193.

(3) Fung, T. S.; Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. Annu. Rev. Microbiol. 2019, 73, 529–557.

(4) Cascella, M.; Rajnik, M.; Cuomo, A.; Dulebohn, S. C.; Di Napoli, R. Features, Evaluation and Treatment Coronavirus (COVID-19). StatPearls; StatPearls Publishing: Treasure Island (FL), 2020.

(5) Contini, C.; Di Nuzzo, M.; Barp, N.; Bonazza, A.; De Giorgio, R.; Tognon, M.; Rubino, S. The Novel Zoonotic COVID-19 Pandemic: An Expected Global Health Concern. J. Infect. Dev. Countries 2020, 14, 254–264.

(6) Yang, T.; Liu, Y.; Deng, W.; Zhao, W.; Deng, J. SARS-CoV-2 Trajectory Predictions and Scenario Simulations from a Global Perspective: A Modelling Study. Sci. Rep. 2020, 10, 18319.

(7) Cui, J.; Li, F.; Shi, Z.-L. Origin and Evolution of Pathogenic Coronaviruses. Nat. Rev. Microbiol. 2019, 17, 181–192.

(8) Báez-Santos, Y. M.; St. John, S. E.; Meseçar, A. D. The SARS-CoV Papiain-like Protease: Structure, Function and Inhibition by Designed Antiviral Compounds. Antiviral Res. 2015, 115, 21–38.

(9) Hilgenfeld, R. From SARS to MERS: Crystallographic Studies on Coronavirus Proteases Enable Antiviral Drug Design. FEBS J. 2014, 281, 4085–4096.

(10) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlom, M.; Hilgenfeld, R.; Rao, Z. The Crystal Structures of Severe Acute Respiratory Syndrome Virus Main Protease and Its Complex with an Inhibitor. Proc. Natl. Acad. Sci. U.S.A. 2003, 100, 13190–13195.

(11) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus Main Protease (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. Science 2003, 300, 1763–1767.

(12) Ratia, K.; Pegan, S.; Takayama, J.; Sleeman, K.; Coughlin, M.; Bajli, S.; Chaudhuri, R.; Fu, W.; Prabhakar, B. S.; Johnson, M. E.; et al. A Noncovalent Class of Papain-like Protease/Deubiquitinase Inhibitors Blocks SARS Virus Replication. Proc. Natl. Acad. Sci. U.S.A. 2008, 105, 16119–16124.

(13) Gorbalenya, A. E.; Enjuanes, L.; Ziebuhr, J.; Snijder, E. J. Nidovirales: Evolving the Largest RNA Virus Genome. Virus Res. 2006, 117, 17–37.

(14) Snijder, E. J.; Breidenbeek, P. J.; Dobbe, J. C.; Thiel, V.; Ziebuhr, J.; Poon, L. L. M.; Guan, Y.; Rozanov, M.; Spaan, W. J. M.; Gorbalenya, A. E. Unique and Conserved Features of Genome and Proteome of SARS-CoV-1, an Early Split-off from the Coronavirus Group 2 Lineage. J. Mol. Biol. 2003, 331, 991–1004.

(15) Thiel, V.; Ivanov, K. A.; Putics, A.; Hertzig, T.; Schelle, B.; Bayer, S.; Weißbrich, B.; Snijder, E. J.; Rabenau, H.; Doerr, H. W.; Gorbalenya, A. E.; Ziebuhr, J. Mechanisms and Enzymes Involved in SARS Coronavirus Genome Expression. J. Gen. Virol. 2003, 84, 2305–2315.
Acute Respiratory Syndrome Coronaviruses Disrupts the Functional Expression of Open Reading Frame 8. J. Virol. 2007, 81, 13876–13888.

(36) Han, M. V.; Zmasek, C. M. PhylXML: XML for Evolutionary Biology and Comparative Genomics. BMC Bioinf. 2009, 10, 356.

(37) Robinson, O.; Dylus, D.; Dessimoz, C. Phylrio: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. Mol. Biol. Evol. 2016, 33, 2163–2166.

(38) Korber, B.; Fischer, W. M.; Gnannakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. Cell 2020, 182, 812–827.

(39) Islam, O. K.; Al-Emran, H. M.; Hasan, M. S.; Anwar, A.; Jahid, M. I. K.; Hossain, M. A. Emergence of European and North American Mutant Variants of SARS-CoV-2 in South-East Asia. Transboundary Emerging Dis. 2020, 1.

(40) Rodrigues, H. C.; Pires, D. E.; Archer, D. B. DynaMut: Predicting the Impact of Mutations on Protein Conformation, Flexibility and Stability. Nucl. Acids Res. 2018, 46, W350–W355.

(41) Cárdenas-Conejo, Y.; Lilian-Rico, A.; García-Rodríguez, D. A.; Centeno-Leija, S.; Serrano-Posada, H. An Exclusive 42 Amino Acid Signature in Pp1ab Protein Provides Insights into the Evolutive History of the 2019 Novel Human-Pathogenic Coronavirus (SARS-CoV-2). J. Med. Virol. 2020, 92, 688–692.

(42) Wang, R.; Chen, J.; Hozumi, Y.; Yin, C.; Wei, G.-W. Decoding Asymptomatic COVID-19 Infection and Transmission. J. Phys. Chem. Lett. 2020, 11, 10007–10015.

(43) Chand, G. B.; Banerjee, A.; Azad, G. K. Identification of Novel Mutations in RNA-Dependent RNA Polymerases of SARS-CoV-2 and Their Implications on Its Protein Structure. PeerJ 2020, 8, e9492.

(44) Hassan, S. S.; Ghosh, S.; Attirish, D.; Choudhury, P. P.; Seyer, M.; Pizzol, D.; Adadi, P.; El-Aziz, T. M. A.; Soares, A.; Kandimalla, R.; Lundstrom, K.; Tumbwala, M.; Aljabali, A. A. A.; Lal, A.; Azad, G. K.; Uversky, V. N.; Scherhan, S. P.; Baetas-da-Cruz, W.; Uhal, B. D.; Rezaei, N.; Brufsky, A. M. A. Unique View of SARS-CoV-2 through the Lens of ORF8 Protein. 2020, bioRxiv:2020.08.25.267328.

(45) Koyama, T.; Platt, D.; Parida, L. Variant Analysis of SARS-CoV-2 Genomes. Bull. W. H. O. 2020, 98, 495–504.

(46) Tan, Y.; Schneider, T.; Leong, M.; Aravid, L.; Zhang, D. Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis Mechanisms of SARS-Related Coronavirus. 2020, bioRxiv:2020.04.09.077736.

(47) Gordon, D. E.; Jing, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O’Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; et al. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. Nature 2020, 583, 459–468.

(48) Hachum, A.; Kaviani, N.; Cohen, A. C.; Chiu, A. W. H.; Chu, D. K. W.; Mok, C. K. P.; Tsang, O. T. Y.; Yeung, Y. C.; Perera, R. A. P. M.; Poon, L. L. M.; et al. ORF8 and ORF3b Antibodies Are Accurate Serum Markers of Early and Late SARS-CoV-2 Infection. Nat. Immunol. 2020, 21, 1293–1301.

(49) Zhang, Y.; Zhang, J.; Chen, Y.; Luo, B.; Yuan, H.; Huang, F.; Yang, T.; Yu, F.; Liu, J.; Liu, B.; et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I. 2020, bioRxiv:2020.05.24.111823.

(50) Hosseini Rad Sm, A.; McLellan, A. D. Implications of SARS-CoV-2 Mutations for Genomic RNA Structure and Host MicroRNA Targeting. Int. J. Mol. Sci. 2020, 21, 4807.
(53) Hurst, K. R.; Koetzner, C. A.; Masters, P. S. Identification of In Vivo-Interacting Domains of the Murine Coronavirus Nucleocapsid Protein. *J. Virol.* 2009, 83, 7221–7234.

(54) Liang, Y.; Wang, M.-L.; Chien, C.-S.; Yarmishyn, A. A.; Yang, Y.; Lai, W.; Luo, Y.-H.; Lin, Y.-T.; Chen, Y.-J.; Chang, P.-C.; Chiou, S.-H. Highlight of Immune Pathogenic Response and Hematopathologic Effect in SARS-CoV, MERS-CoV, and SARS-CoV-2 Infection. *Front. Immunol.* 2020, 11, 10221.

(55) Kang, S.; Yang, M.; Hong, Z.; Zhang, L.; Huang, Z.; Chen, X.; He, S.; Zhou, Z.; Zhou, Z.; et al. Crystal Structure of SARS-CoV-2 Nucleocapsid Protein RNA Binding Domain Reveals Potential Unique Drug Targeting Sites. 2020, bioRxiv:2020.03.06.977876.

(56) Zeng, W.; Liu, G.; Ma, H.; Zhao, D.; Yang, Y.; Liu, M.; Mohammed, A.; Zhao, C.; Yang, Y.; et al. Biochemical Characterization of SARS-CoV-2 Nucleocapsid Protein. *Biochem. Biophys. Res. Commun.* 2020, 527, 618–623.

(57) Hassan, S. S.; Attrish, D.; Ghosh, S.; Choudhury, P. P.; Uversky, V. N.; Uhal, B. D.; Lundstrom, K.; Rezaei, N.; Aljabali, A. A.; et al. Notable Sequence Homology of the ORF10 Protein Introspects the Architecture of SARS-COV-2. 2020, bioRxiv:2020.09.06.284976.

(58) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauererling, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* 2020, 368, 409–412.

(59) Lim, L.; Shi, J.; Mu, Y.; Song, J. Dynamically-Driven Enhancement of the Catalytic Machinery of the SARS 3C-like Protease by the S284-T285-I286/A Mutations on the Extra Domain. *PLoS One* 2014, 9, No. e101941.

(60) Shi, J.; Song, J. The Catalysis of the SARS 3C-like Protease Is under Extensive Regulation by Its Extra Domain. *FEBS J.* 2006, 273, 1035–1045.

(61) Luminex Corporation. Luminex Provides Update on SARS-CoV-2 Validation Testing Efforts. https://www.prnewswire.com/news-releases/luminex-provides-update-on-sars-cov-2-validation-testing-efforts-301022782.html (accessed Dec 4, 2020).

(62) Sieker, J. T.; Horowitz, C.; Hu, C.-T. K.; Lacombe-Daphnis, M.; Chirokas, B.; Pina, C.; Heger, N. E.; Rabson, A. R.; Zhou, M.; Bogen, S. A.; Horowitz, G. L. Analytic Sensitivity of 3 Nucleic Acid Detection Assays in Diagnosis of SARS-CoV-2 Infection. *J. Appl. Lab. Med.* 2020, jaal187.

(63) Moran, A.; Beavis, K. G.; Matushek, S. M.; Ciaglia, C.; Francois, N.; Tesci, V.; Love, N. Detection of SARS-CoV-2 by Use of the Cepheid Xpert Xpress SARS-CoV-2 and Roche Cobas SARS-CoV-2 Assays. *J. Clin. Microbiol.* 2020, 58, No. e00772-20.

(64) Federman, D. G.; Gupta, S.; Stack, G.; Campbell, S. M.; Peaper, D. R.; Dembry, L. M.; Fisher, A.; Tarabar, A. F.; Kozal, M.; Ruser, C. B. SARS-CoV-2 Detection in Setting of Viral Swab Scarcity: Are MRSA Swabs and Viral Swabs Equivalent? *PLoS One* 2020, 15, No. e0237127.

(65) Chong, Y. M.; Sam, I.-C.; Ponnampalavanar, S.; Syed Omar, S. F.; Kamalazaman, A.; Munusamy, V.; Wong, C. K.; Jamaluddin, F. H.; Gan, H. M.; Chong, J.; Teh, C. S. J.; Chan, Y. F. Complete Genome Sequences of SARS-CoV-2 Strains Detected in Malaysia. *Microbiol. Resour. Announce.* 2020, 9, No. e00383-20.

(66) Tippmann, H.-F. Analysis for Free: Comparing Programs for Sequence Analysis. *Briefings Bioinf.* 2004, 5, 82–87.