

Genome sequencing and genetic characterization of 17 SARS-CoV-2 viruses and spike protein analyses of isolates from Bangladeshi patients

Shahina Akter

Bangladesh Council of Scientific and Industrial Research <https://orcid.org/0000-0001-7776-6686>

Murshed Hasan Sarkar¹

Bangladesh Council of Scientific and Industrial Research

Tanjina Akhtar Banu

Bangladesh Council of Scientific and Industrial Research <https://orcid.org/0000-0003-3318-9287>

Barna Goswami

Bangladesh Council of Scientific and Industrial Research

Iffat Jahan

Bangladesh Council of Scientific and Industrial Research

Eshrar Osman

SciTech Consulting and Solution

Mohammad Samir Uzzaman

SciTech Consulting and Solution

Abu Sayeed Mohammad Mahmud

Bangladesh Council of Scientific and Industrial Research

Tasnim Nafisa

National Institute of Laboratory Medicine and Refferel Center

Maruf Ahmed Molla

National Institute of Laboratory Medicine and Refferel Center

Mahmuda Yeasmin

National Institute of Laboratory Medicine and Refferel Center

Ahashan Habib

Bangladesh Council of Scientific and Industrial Research

Salim Khan (✉ k2salim@yahoo.com)

Bangladesh Council of Scientific and Industrial Research <https://orcid.org/0000-0003-0603-6728>

Research Article

Keywords: Next-generation sequencing, BCSIR-NILMRC, Bangladesh, SARS-CoV-2, Spike, genomes, phylogeny, RT-PCR

Posted Date: October 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-968594/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

To better comprehend the genetic epidemiology of SARS-CoV-2, we utilized NextSeq 550 to conduct whole-genome sequencing of certain samples in Bangladesh. The samples were taken from the nasopharyngeal swabs of 17 persons who had tested positive for COVID-19 using RT-PCR. Following the sequencing, the data was uploaded to the GISAID database and evaluated by comparing their nucleotide alterations to the Wuhan SARS-CoV-2 reference genome. Based on phylogenetic and mutation studies, it was shown that 88% belong to the pangolin lineage B.1.1.25, whereas 11% corresponds to its parent lineage B.1.1. We also compared the spike protein sequences of two isolates (BCSIR-NILMRC-422 and BCSIR-NILMRC-424) to the reference genome and observed two mutations including, D614G and P681R, at positions 614 and 681, respectively. Although D614G being the most frequent, not present in every isolate. Using standard mode studies, we also looked at molecular dynamics and discovered that this SNP reduces the S1 domain's flexibility. This finding suggests that the P681R mutation is a distinguishing feature of this novel voice of concern's virological profile, which may be linked to viral pathogenicity in Bangladeshi patients.

Introduction

The present world has been facing a great problem due to Covid 19, which turned into an epidemic, where thousands of people are dying each day throughout the world. The number of new cases increased robustly, while the number of new deaths continued to increase, with over 20904838 confirmed cases, 4570946 confirmed deaths (on 7th September, 2021) reported globally (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>).

In Bangladesh, Covid 19 case was first reported in March 2020, and since then the country has experienced a steady rise in infections [1] resulting in 1517166 confirmed cases and 26628 deaths as of 7th September, 2021 (<https://corona.gov.bd/?gclid>). The samples of Bangladeshi patients detected as Covid 19 positives were collected from the National Institute of Laboratory Medicine and Referral Center (NILMRC) and transported to Bangladesh Council of Scientific and Industrial Research (BCSIR), for sequencing of SARS-CoV-2 isolates. After sequencing through the NGS method, the generated data were submitted and published in GISAID (Global initiative of sharing all influenza data) and NCBI. The isolates were characterized by complete genome sequencing and in silico analyses. These sequences were also compared with the available GenBank sequences and GISAID to monitor the mutations and understand their relation with SARS-CoV-2 of Wuhan, China as a reference available in the public database [2].

To get relief from this epidemic, vaccine or antiviral therapeutic is very important. For designing an antiviral therapeutics or vaccines, it is important to comprehend the genetic sequence, structure, and function of the viral active proteins especially the spike protein. When a virus attempts to be acclimatized in a new environment, in a new host, in a new geographical location, and a new population, it might make alterations to its genetic make-up which in turn carry slight modifications in viral proteins [2].

In SARS-CoV-2, the spike protein is a structural protein that forms a homotrimer on the surface of the viral lipid envelope [7]. In the spike protein of SARS-CoV-2, the receptor-binding domain (RBD) facilitates interaction with the host cell receptor [4]. This trimer is made up of monomers that contain S1 and S2 subunits. S1 subunit helps in attachment to the host cell receptor while S2 subunit helps infusion to host cell and entry. Consequently, the spike protein has been an area of interest for designing vaccines against Covid 19. It is imperative to acquire a comprehensive mutation profile of this protein from extensive genome sequencing and different geographical locations of the world, as this trimeric spike protein has a habit of getting mutated [7].

At this juncture, in the investigation, we report the molecular characterization of SARS-CoV-2 sequences from seventeen positive cases in Bangladesh. This study aims to reveal the most common mutations of SARS-CoV-2 viral isolates from Bangladesh in comparison to the reference sequence from Wuhan, China (NC_045512.1). The data in this study may contribute to understanding the molecular evolution of SARS-CoV-2 in Bangladesh by analyzing the mutation found in the isolates of SARS-CoV-2.

Material & Methods

Nasopharyngeal specimen of seventeen COVID-19 positive patients were collected from the National Institute of Laboratory Medicine and Referral Center (NILMRC), Bangladesh. RT-PCR assay for detection of SARS-CoV-2 virus was conducted by NILMRC. Preparation of library and next generation sequencing of 17 samples were conducted by Genomic Research Laboratory of BCSIR. The human research ethics committee of the National Institute of Laboratory Medicine and Referral Center (NILMRC) approved the whole-genome sequencing of SARS-CoV-2 in this study (project code- 224125200). PCR identification of SARS-CoV-2 was done from nasopharyngeal swab using Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic Kit by Sansure Biotech (Wuhan, China).

Viral RNA was extracted using PureLink™ Viral RNA/DNA Mini Kit (Cat.no: 12280050, Thermofisher Scientific, USA) [1]. The cDNA of all samples were used to prepare paired-end libraries with the Nextera™ DNA Flex Library Preparation kit according to the manufacturer's instructions (Illumina Inc., San Diego, CA). The Library pool of the samples was sequenced using the S4 flow cell of Illumina NextSeq 550 instrument in paired-end fashion (read length 151 bp; Illumina Inc.). FASTQ files were generated in Illumina BaseSpace Sequence Hub (<https://basespace.illumina.com>). The generated data after NGS sequencing for BCSIR-NILMRC-422 and BCSIR-NILMRC-424 were submitted in the NCBI as SRA submission and got the SRA number SRP336906 and SRP336782 (PRJNA762998 and PRJNA762670) respectively. Genomes of SARS-CoV-2 viruses were assembled using Basespace DRAGEN RNA Pathogen Detection V3.5.14 software with default parameters. From the fastq files, Consensus fasta files of SARS-CoV-2 genome were generated using Basespace Dragon RNA Pathogen Detection software version 3.5.1 (<https://basespace.illumina.com>) with default settings. SARS-CoV-2 isolate sequences from Bangladesh were compared to the reference SARS-CoV-2 sequence (NC_045512.1), through nucleotide substitutions. The positions of the nucleotides and amino acids were further confirmed from GeneBank reference sequences (NC_045512.1) [8].

Genome assembly of the raw data was performed using the assembly toolkit in the EzCOVID19 cloud service, provided on the EzBioCloud website (<https://eztaxon-e.ezbiocloud.net/tools/sc2?id=5bbb155a-dd3e-4e9b-9b10-54b4e67926bd>). In every cases, the genome sequence consisted of almost 29,903 nucleotides and covered 99.84% % of the SARS-CoV-2 Wuhan reference genome (NC_045512.2), with an average GC content of 38%. Multiple sequence alignment platform of CLUSTAL Omega was used to align all the spike protein sequences. Before this, the sequence range of spike protein was selected and separate that portion from the whole genome sequences. The alignment FASTA file was viewed using MView and variations in the sequence or the amino acid changes were reported. To predict secondary structures of SARS-CoV2 spike protein CFSSP (Chou and Fasman secondary structure prediction) server was used. For template matching and the target sequence from one of the isolates carrying the specific mutation were done using Swiss-Model (<https://swissmodel.expasy.org/interactive/FGwJJj/templates>). The available SARS-CoV-2 spike pdb file was generated using this tool. The pdb files of two isolates BCSIR-NILMRC-422 and BCSIR-NILMRC-424 were uploaded on a web server designed by the University of Melbourne, Australia named DynaMut. The changes in vibrational entropy along with the atomic fluctuations and deformation energies caused by mutation were determined. Automatic calculations were accomplished by the software over the first ten non-trivial modes of the molecule for atomic fluctuation and deformation energy calculations [9].

Results And Discussion

More than 750 whole-genomes of SARS-CoV-2 strains collected from different divisions of Bangladesh have been sequenced using NGS (NextSeq 550) in BCSIR, Bangladesh and submitted to GISAID. From them seventeen cases hCoV-19/Bangladesh/BCSIR-NILMRC_392, hCoV-19/Bangladesh/BCSIR-NILMRC_398, hCoV-19/Bangladesh/BCSIR-NILMRC_376A, hCoV-19/Bangladesh/BCSIR-NILMRC_390, hCoV-19/Bangladesh/BCSIR-NILMRC_391, hCoV-19/Bangladesh/BCSIR-NILMRC_393, hCoV-19/Bangladesh/BCSIR-NILMRC_396, hCoV-19/Bangladesh/BCSIR-NILMRC_397, hCoV-19/Bangladesh/BCSIR-NILMRC_411, hCoV-19/Bangladesh/BCSIR-NILMRC_420, hCoV-19/Bangladesh/BCSIR-NILMRC_421, hCoV-19/Bangladesh/BCSIR-NILMRC_422, hCoV-19/Bangladesh/BCSIR-NILMRC_423, hCoV-19/Bangladesh/BCSIR-NILMRC_424, hCoV-19/Bangladesh/BCSIR-NILMRC_426, hCoV-19/Bangladesh/BCSIR-NILMRC_427 and hCoV-19/Bangladesh/BCSIR-NILMRC_428 were analyzed in this experiment. Almost complete (29,903 nucleotides) genomes of all cases were obtained. All these seventeen isolates were placed in GR clade according to GISAID analyses and 2B clade according to Nextclade. Except hCoV-19/Bangladesh/BCSIR-NILMRC_376A and hCoV-19/Bangladesh/BCSIR-NILMRC_390, all isolates are belongs to B.1.1.25 lineage (Table-1). This B.1.1.25 recognized as the Bangladeshi lineage (<https://cov-lineages.org/lineage.html?lineage=B.1.1.25>) and B.1.1 is its parent lineage.

The age interval of the patients from these Bangladeshi samples, from whom the viral isolates were taken, were between the age range of 22–74 years (Fig. 1a) and the sex distribution was 88.23% male and 11.76% female (Fig. 1b). But this range is very specific in these samples only, not the whole scenario of the country. All of the uploaded samples were came from Dhaka and Naraynganj of the country.

To identify clusters inside the diversity of Severe acute respiratory syndrome-related coronavirus SARS-CoV-2 (2019 outbreak), the coronavirus typing tool applies phylogenetic analysis. This typing tool has prudently selected reference sequences that denote the diversity of each well-defined phylogenetic cluster (detailed in nextstrain). Additionally, this tool executed extensive testing to be sure that their reference strains accurately classify other sequences [10, 11]. If the sequence is identified as part of the SARS-CoV-2 cluster, an additional phylogenetic analysis is performed to identify one of the following lineages: *B.1.351_501Y.V2_20H (Beta variant)*: A cluster of concerns first originated in South Africa (report-COVID-19 B.1.351 (501Y.V2) Variant of Concern – What We Know So Far; Public Health Ontario, 02.07.2021); *B.1.1.7_501.V1_201*: A cluster of concern first discovered in the UK by A. Rambaut *et al.* (report) *P.1 aka 501Y.V3 aka 20J*: A cluster of concerns first discovered in Brazil by N. Faria *et al.* (report); *Y453F.Cluster5_20B*: A cluster of concern first discovered in Denmark, believed to have spread from mink infections; *B.1.1.70_501Y_20B* [12].

Sequences, not part of any cluster of concern will be designated as *International AB Diversity* [10, 11]. It was only possible to make the original version of this tool weeks after the virus was first sequenced because the sequence of a novel coronavirus from Wuhan using data generated by the Shanghai Public Health Clinical Center & School of Public Health, the National Institute for Viral Disease Control and Prevention, the Institute of Pathogen Biology, and the Wuhan Institute of Virology shared via GISAID, was released before publication. We, from our deepest heart, acknowledge their contribution.

The B.1.351 (501Y.V2) a variant, which is out cluster of concern emerged in late 2020 in Eastern Cape Province, South Africa, and subsequently spread throughout South Africa and to over 40 countries worldwide, including Canada. One B.1.351 case has been detected as of February 4, 2021, in Ontario (report-COVID-19 B.1.351 (501Y.V2) Variant of Concern – What We Know So Far; Public Health Ontario, 02.07.2021). During the first wave of the pandemic, epidemiological and modeling studies show that, the B.1.351 variant is more transmissible compared to lineages circulating. At present, there is ambiguity concerning the ability of B.1.351 to impact COVID-19 severity. Phylogenetic analysis showed that all the cases of SARS-CoV-2 are belonging to GR Clade.

Assembly was performed by aligning reads to a pre-defined Wuhan reference genome (NC_045512.2). Genome assembly of the raw data was done by Assembly toolkit in the EzCOVID19 and EDGE COVID-19 cloud service provided on the EzBioCloud website and EDGE bioinformatics website respectively [13, 14]. To retain the exceptional variations of raw data using the EDGE COVID19 cloud service a consensus genome was generated [13]. The consensus genome was then compared against the same reference genome to calculate single nucleotide variations (SNV) with positions. The SNVs were compared against GISAID clade variation markers. The genome presented here belonged to B.1.1.25 lineage of GR clade. This result was again confirmed by analyzing the raw data using the default parameters of the coronavirus typing tool offered by Genome Detective [15].

Single-nucleotide variations between the input assembly and the SARS-CoV-2 reference NC_045512.2 (https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2) were shown in table-2. When the SARS-CoV-2

strains were isolated from Bangladeshi patients, compared to the reference viral sequence (NC 045512.1), Nonsynonymous and synonymous mutations were identified. These studies have been done using EZBioCloud (<https://www.ezbiocloud.net/tools/sc2/?id=4024a5b2-40bb-40b5-acd6-c4dd65b38d76>). The queried dataset covers 99.84% of the SARS-CoV-2 genomic sequence. The nucleotide sequences were indicated starting from the 5' UTR, while the corresponding amino acid changes were mentioned separately for each protein-coding region. In more than 30% of the viral isolates, nucleotide substitutions of C>T were observed, whereas, G>T, and G>A showed 20% and 15% respectively. Among these mostly seen substitutions, C>T, which are present in the ORF1ab, S gene, and ORF3a gene (Table-2). G>T is found in ORF1ab, ORF3a, and in the M gene whereas, G>A present in the S gene and maximum cases in the N gene.

The emergence of mutations in SARS-CoV-2 evolution has been characterized by, in the context of 'variants of concern', that influence virus characteristics, including transmissibility and antigenicity. These mutations impact probably in response to the changing immune profile of the human population [16]. The amino acid change D614G in spike protein was noted to be increasing in frequency in April 2020 and to have emerged several times in the global SARS-CoV-2 population [16]. The coding sequence reveals a high dN/dS ratio, suggesting positive selection at the codon position 614 [17, 18]. Consequent studies indicated that, D614G confers a moderate advantage for infectivity and transmissibility. There is another mutations P681R in spike protein have now arisen and are discussed in this article. Similarly, Bioinformatics analyses reveal that the P681R mutation in the spike protein of certain isolates are highly conserved in this lineage. P681R mutations were highly conserved in the B.1.1.25 lineage, and remarkably, the P681R mutation was the most representative mutation in this lineage.

Each of the 17 viral isolates were also evaluated utilizing the mutations they harbor. Most of the viral isolates were found to have different amino acid substitution combinations (Table-3). However, the same amino acid substitutions combination Spike_D614G and Spike_P681R were observed for the samples BCSIR-NILMRC_393, BCSIR-NILMRC_396, BCSIR-NILMRC_421, BCSIR-NILMRC_422, and in BCSIR-NILMRC_424. In this study we have emphasized these two amino acid mutations and hence we will discuss elaborately with these two amino acid mutations in the spike protein later on. In every isolates Spike_D614G is the common mutation found in the spike protein (Table-3). The analyzed viral isolates were found to harbor 11 to 19 mutations per isolate compared to the reference sequence.

B.1.1.25 Lineage with S:P681R Reports are given in the figures number 3, 4 and 5. For this study we have used outbreak.info tool (https://outbreak.info/situation-reports?pango=B.1.1.25&mutS=S%3AP681R&selected=BGD&loc=USA&loc=USA_US-CA&loc=BGD). Concerns surrounding new strains of SARS-CoV-2 (hCoV-19), the virus behind the COVID-19 pandemic, have been developing. The prevalence of the B.1.1.25 Lineage with S:P681R in the world, how it is changing over time, and how its prevalence varies across different locations specially in Bangladesh, has been outlined in this report. All these reports are presented in the following figures (Fig. 3, supplementary figure 1 and 2). We depend on shared virus sequences from the GISAID to describe the prevalence of sets of mutations in Mutation Situation Reports. We also rely on the accuracy of the sequences and sample

metadata deposited in GISAID, while we apply filters to eliminate some low quality sequences and unreasonable metadata. These sequences are a sample of the total number of cases – often a biased sample and may not represent the true prevalence of the mutations in the population. In general, it is important to note that case numbers for any given lineage/mutation can be significantly affected by overall case numbers and rates of genomic sequencing at any given location (www.Outbreak.info) [19].

Lineage B.1.1.25 is a Bangladeshi lineage and B.1.1 is its parent lineage. Most common countries of this lineage are Bangladesh 50.0%, Canada 27.0%, United States of America 11.0%, United Kingdom 5.0%, and Australia 2.0% (<https://cov-lineages.org/lineage.html?lineage=B.1.1.25>). All the spike protein sequences along with reference sequence from Wuhan were aligned using the multiple sequence alignment platform of CLUSTAL Omega for Sequence alignments and structure. After finishing the alignment, the generated file was viewed using MView and variations in the sequence or the amino acid changes were noted. For the prediction of secondary structures of SARS-CoV2 spike protein, CFSSP (Chou and Fasman secondary structure prediction) server was used in this experiment. To check for the presence of similarities or differences, all these spike protein sequences were first aligned in CLUSTAL Omega. We found that at position 614, all the isolates from Bangladesh have a common mutation D614G whereas, in the reference genome (Wuhan) it is 'D' instead of 'G'. So, in this position for all the isolates amino acid D (aspartic acid) changed into G (Glycine). Also, at position 681, four of the isolates BCSIR-NILMRC_393, BCSIR-NILMRC_396, BCSIR-NILMRC_422, BCSIR-NILMRC_424 showing mutation. In this position within these four isolates amino acid P (Proline) changed into amino acid R (Arginine) (Fig. 4). For comparison, the original Wuhan sequence as the wild type were considered. We found these two significant differences at amino acid positions in spike protein that were mutated in these isolates overall (<https://www.britannica.com/science/amino-acid/Standard-amino-acids>).

In the S1 domain, one of these mutations was D614G. This mutation posited near the receptor-binding domain at a downstream position. Another mutation, P681R harboring further downstream in the S protein in the same domain. The D614G mutation which is a conserved mutation in every 17 isolates, in its spike (S) protein has emerged in the spring of 2020, a SARS CoV-2 derivative harboring and quickly become predominant [20]. The D614G-bearing variant has quickly swept out the original strain, as because the D614G mutation increases viral infectivity, fitness, and inter-individual transmissibility [21]. Interestingly, the P681R mutation in the spike protein is unique and recently recognized as the VOCs so far in Bangladesh and the P681R mutation is located in the proximity of the furin cleavage site of the SARS-CoV-2 S protein [22].

For secondary structure analyses of those mutant proteins, we have used CFSSP: Chou & Fasman Secondary Structure Prediction server (<http://www.biogem.org/tool/chou-fasman>). This is a very effective server which predicts secondary structure of protein from the amino acid sequence. In this server, Chou & Fasman algorithm has been implemented [23]. Secondary structure was analyzed for the mutation D614G of BCSIR-NILMRC-422 and mutation P681R of BCSIR-NILMRC-424. In the case of mutation D614G, Glycine (G) are non-polar with aliphatic R groups and Aspartate (D) are polar amino acids with a carboxylic acid on its side chain that gives it acidic (proton-donating) properties

(<https://www.britannica.com/science/amino-acid/Standard-amino-acids>) and have side chain while Glycine has no side chain. On the other hand, in mutation P681R, P (Proline) non-polar with aromatic R group changed into R (Arginine), which is a basic amino acid, occur with an overall charge of +1 at physiological pH. Another alteration from Proline to Arginine can thus potentially interrupt the local folding of that particular protein. In figure 5, the secondary structure prediction are showed changes in and around the site of mutations. In the spike region, where mutation occurred, there were loss of turn structures in positions 614 and 681. The alternation took place in secondary structure might cause a variation in the function of S1. In this region, S1 helps the fusion process of the spike protein and consequently mutation in S1 may have altered receptor spike interactions and hence infectivity [2].

To correlate, if changes in secondary structure are also reflected in the dynamics of the protein in its tertiary structure, we performed normal mode analyses and studied protein stability and flexibility. For this purpose we have used Dynamut (http://biosig.unimelb.edu.au/dynamut/results_prediction/16263266400). This tool provides users implementing two distinct, well established normal mode approach that can use to analyze and visualize protein dynamics. The impact of mutations on protein dynamics and stability resulting from vibrational entropy changes supposed to be done by sampling conformations and assessment.

Our mutations of interest D614G and P681R on spike proteins of SARS-CoV-2 isolates from Bangladeshi patients, of these, maximum are destabilizing ($\Delta\Delta G < 0.0$ kcal/mol) and few stabilizing ($\Delta\Delta G > 0.0$ kcal/mol). Change in vibrational entropy energy ($\Delta\Delta S_{Vib} ENCoM$) in case of mutation D614G between the wild type Wuhan isolate and the BCSIR-NILMRC- 422 isolate was 0.856 kcal.mol⁻¹.K⁻¹, while in case of mutation P681R it was 0.018 kcal.mol⁻¹.K⁻¹ (Fig. 6). In the case of mutation D614G the $\Delta\Delta G$ (Gibbs free energy) was -0.511 kcal/mol and the $\Delta\Delta G ENCoM$ was -0.684 kcal/mol whereas, in case of mutation P681R it was 0.224 kcal/mol (Stabilizing) and the $\Delta\Delta G ENCoM$ was -0.014 kcal/mol. All these findings recommended a stabilizing mutation in these types of the spike.

The interatomic interactions of mutation D614G in isolate BCSIR-NILMRC-422 have been shown in Figure 7. Studies of atomic fluctuations and deformation energies for D614G in case of isolate BCSIR-NILMRC-422 showed noticeable changes (Fig. 9). Atomic fluctuations compute the measure of absolute atomic motion whereas deformation energies detect the measure of flexibility of a protein. Figure 9 shows the photographic representations of the atomic fluctuation and deformation energies where positions that could be visibly detected to be altered have been marked for D614G in isolate BCSIR-NILMRC-422.

The interatomic interactions in case of mutation P681R in isolate BCSIR-NILMRC-424, have been shown in Figure 8. Analyses of atomic fluctuations and deformation energies for P681R in isolate BCSIR-NILMRC-424 showed detectable changes (Fig. 10). Deformation energies identify the amount of flexibility of a protein and atomic fluctuations calculate the measure of absolute atomic motion. The visual representations of the atomic fluctuation and deformation energies were shown in Figure 10, where positions that could be visibly identified to be different have been marked in mutation P681R in isolate BCSIR-NILMRC-424.

The mutations of P681R is the first report for Bangladesh in the isolates of the BCSIR-NILMRC-393, BCSIR-NILMRC-393, BCSIR-NILMRC-422 and BCSIR-NILMRC-424 and additional sequencing followed by adequate annotation would help to expand the acquaintance about mutation of spike protein in human SARS-CoV-2. These variations might lead to the diversification of particular virus and the subsequent emergence of variants, strains, serotypes and antibody escape mutants. Similarly, mutations might help the virus to adjust to the host environment better and expand its tissue tropism.

Conclusion

We demonstrated that, the B.1.1.25 lineage in some isolates from Bangladesh has acquired a distinct approach to expedite infection and evade antiviral immunity. The highly conserved mutation P681R that is present in four isolates among 17 Bangladeshi isolates of SARS-CoV-2, the efficacy to viral fusion and further accelerates its speed of action. In this study for prediction, we have used a web server DynaMut, which is an integrated computational method. This consensus approach allows for more accurate and reliable prediction of both stabilizing and destabilizing mutations in positions 614 and 681. Even though the impact of P681R mutation and the association between the COVID-19 severity and unusual symptoms caused by SARS-CoV-2 infection remains unclear, hence, elaborate studies on mutation supposed to be done. That would in turn help in better therapeutic targeting and vaccine preparation. The mutation P681R in spike protein substituting viral infection mode, may relate to the severity or unusual outcome of viral infection, therefore, the epidemic of the SARS-CoV-2 variants harboring the P681R mutation should be analyzed intensely.

Declarations

Data availability

The sequences of SARS-CoV-2 genome from Bangladesh were submitted in the GISAID database accession no. EPI_ISL_603221, EPI_ISL_603222, EPI_ISL_603223, EPI_ISL_603224, EPI_ISL_603225, EPI_ISL_603238, EPI_ISL_603239, EPI_ISL_603240, EPI_ISL_603241, EPI_ISL_603242, EPI_ISL_603243, EPI_ISL_603244, EPI_ISL_603245, EPI_ISL_603246, EPI_ISL_603247, EPI_ISL_603249, EPI_ISL_603250.

SRA number:

BCSIR-NILMRC_422 SRA- SRP336906 (PRJNA762998)

BCSIR-NILMRC_424 SRA- SRP336782 (PRJNA762670)

Code availability

The software used are:

<https://coronavirus.jhu.edu/map.html>

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>

<https://corona.gov.bd/?gclid>

<https://basespace.illumina.com>

<https://eztaxon-e.ezbiocloud.net/tools/sc2?id=5bbb155a-dd3e-4e9b-9b10-54b4e67926bd>

<https://swissmodel.expasy.org/interactive/FGwJJj/templates>

<https://cov-lineages.org/lineage.html?lineage=B.1.1.25>

https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2

<https://www.ezbiocloud.net/tools/sc2/?id=4024a5b2-40bb-40b5-acd6-c4dd65b38d76>

<https://outbreak.info/situation-reports?>

[pango=B.1.1.25&muts=S%3AP681R&selected=BGD&loc=USA&loc=USA_US-CA&loc=BGD](https://outbreak.info/situation-reports?pango=B.1.1.25&muts=S%3AP681R&selected=BGD&loc=USA&loc=USA_US-CA&loc=BGD)

www.Outbraek.info

<https://www.britannica.com/science/amino-acid/Standard-amino-acids>

<http://www.biogem.org/tool/chou-fasman>

<https://www.britannica.com/science/amino-acid/Standard-amino-acids>

http://biosig.unimelb.edu.au/dynamut/results_prediction/16263266400

Author contributions

AS did the analyses and wrote the manuscript. AS, SHM, BAT, GB, JI and KS performed the library preparation and sequencing. OE, USM, MMSA, HA, KS, administered the research work. NT, MAM and YM conceived and implemented the RT-PCR study. SA and KS edited and approved the manuscript. All authors are agreed to submit the manuscript to this journal.

Ethics approval

This study include experiments with human and the human research ethics committee of the National Institute of Laboratory Medicine and Referral Center (NILMRC) approved the whole-genome sequencing of SARS-CoV-2 in this study (project code- 224125200).

Declarations

The submitting research article "***Genome sequencing and genetic characterization of 17 SARS-CoV-2 viruses and spike protein analyses of isolates from Bangladeshi patients***" for publication in your journal

of repute, is a unique article and nobody do it earlier.

Funding

Government of People's Republic of Bangladesh

Acknowledgements

Authors acknowledge the encouragement and support extended by the project which is financed by Government of People's Republic of Bangladesh. We would like to express special thanks of gratitude to Architect Yeafesh Osman, honorable minister Science and Technology. We thank Dr. Abul Khair Mohammad Shamsuzzaman (late), Director, National Institute of Laboratory Medicine and Referral Centre (NILMRC) for providing us samples and many supports. We are very much thanked full to the secretary, ministry of science and technology. We are very much grateful to the Chairman of BCSIR Professor Dr. Md. Aftab Ali Shaikh for his continuous support to the team.

Conflict of interest

The authors declare that they have no conflicts of interest.

Consent to participate

All the authors are agreed to participate with this article

Consent for publication

Consent were taken from all the participates for this publication.

References

1. Akter S, Banu T A, G Barna, Osman E, Uzzaman S M, Habib M A, Jahan I, Mahmud A S M, Sarker M H, Hossain S, Shamsuzzaman A K M, Nafisa T, Molla M A, Yeasmin M, Ghosh A K, Din S S, Ray U C, Sajib S A, Hossain M, Khan S: Coding-Complete Genome Sequences of Three SARS-CoV-2 Strains from Bangladesh. *Micro. Res. Anno.*, **DOI:** 10.1128/MRA.00764-20.
2. Begum F, Mukherjee D, Thagriki D *et al.*: Analyses of spike protein from first deposited sequences of SARS-CoV2 from West Bengal, India [version 1; peer review: awaiting peer review] *F1000Research* 2020, 9:371 <https://doi.org/10.12688/f1000research.23805.1>
3. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol.* 2020;92:401–2. [PMCID: PMC7166628] [PubMed: 31950516]
4. Tai W, He L, Zhang X, *et al.*: Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol.* 2020.

5. Eden JS, Rockett R, Carter I, Rahman H, De Ligt J *et al.* (2020). An emergent clade of SARS-CoV-2 linked to returned travelers from Iran. *Virus Evolution* 6 (1): 1–4. doi: 10.1093/ve/veaa027.
6. CH Rodrigues, DE Pires, DB Ascher: DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 2018; 46(W1): W350–W355.
7. S Cleemput, W Dumon, V Fonseca, W A Karim, M Giovanetti, Luiz Carlos Alcantara, Koen Deforche, Tulio de Oliveira; Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics*, btaa145, DOI: 10.1093/bioinformatics/btaa145.
8. M Vilsker, Y Moosa, S Nooij, *et al.*: An Automated System for Virus Identification from High-throughput sequencing data. *Bioinformatics*, bty695, <https://doi.org/10.1093/bioinformatics/bty695>.
9. CDC (2021). "SARS-CoV-2 variant classifications and definitions (updated June 10, 463 2021)". <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>.
10. EzCOVID19. Available at: <https://eztaxon-e.ezbiocloud.net/tools/sc2?id=5bbb155a-dd3e-4e9b-9b10-54b4e67926bd>. (Accessed: 31st May 2021)
11. Yoon S, H. *et al.*: Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617, 2017.
12. Cleemput S, *et al.* Genome detective coronavirus typing tool for rapid identification and characterization of novel coronavirus genomes. *Bioinformatics* 36, 3552–3555 (2020).
13. Frost, D W, Magalis B R, & Kosakovsky P, S L. Neutral theory and rapidly evolving viral pathogens. *Mol. Biol. Evol.* 35, 1348–1354, 2018.
14. Zhan X-Y, *et al.* Molecular evolution of SARS-CoV-2 structural genes: evidence of positive selection in spike glycoprotein. Preprint at *bioRxiv*, 2020. <https://doi.org/10.1101/2020.06.25.170688>.
15. Korber B, *et al.* Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827 e819 (2020).
16. Alaa A, Latif J, Mullen L, Manar A, *et al.*: Bangladesh Mutation Report and the Center for Viral Systems Biology. outbreak.info, (available at <https://outbreak.info/location-reports?loc=BGD&dark=false>).
17. Korber B, Fischer WM, *et al.* (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827.
18. Zhou B, Thi NT, T Hoffmann, D Taddeo, *et al.* (2021). SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 592, 122–127.
19. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A and Li F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci USA* 117, 11727–11734.
20. Ashok K, T. (2013). CFSSP: Chou and Fasman Secondary Structure Prediction server. *WIDE SPECTRUM: Research Journal.* 1(9):15–19.

Tables 1-3

Tables 1-3 are not available with this version.

Figures

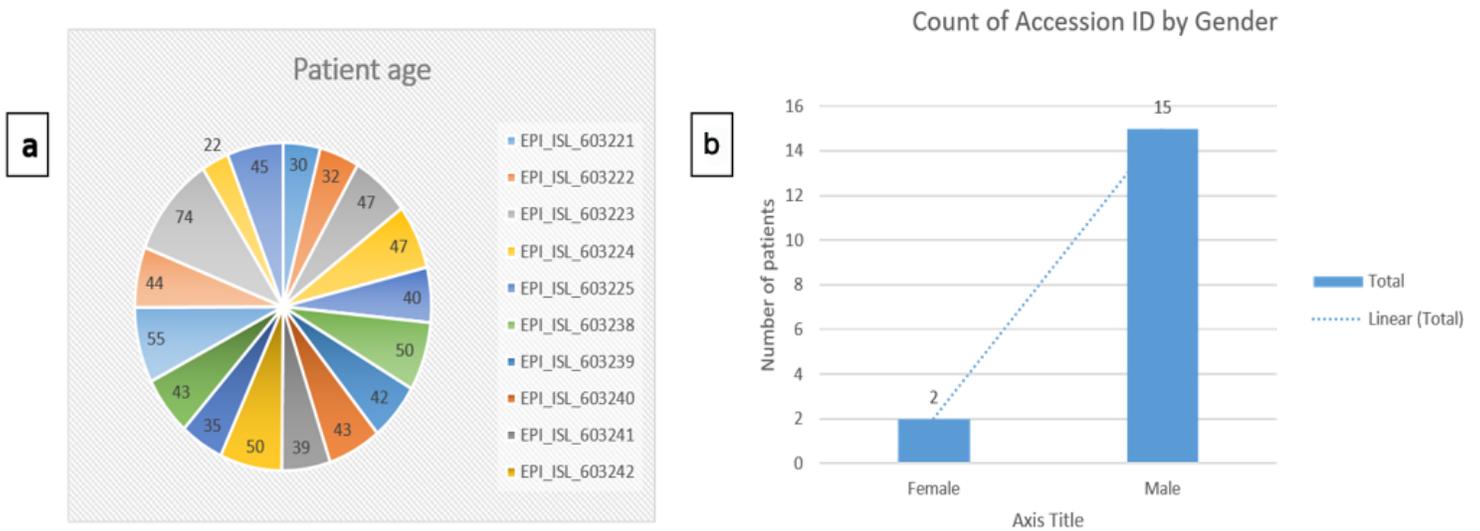


Figure 1

a) The age, and b) Sex of the 17 Bangladeshi patients from whom the SARS-CoV2 virus isolates were taken.

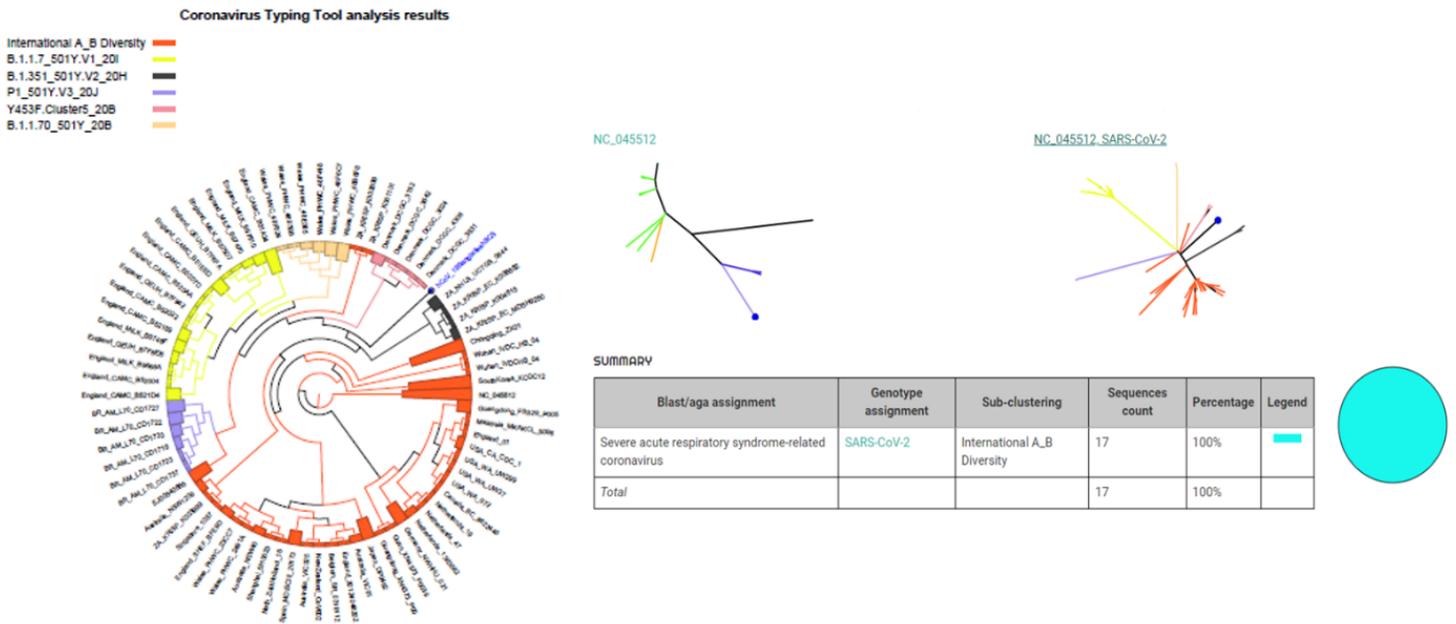


Figure 2

Schematic polar diagram displaying the phylogenetic tree of the 17 isolates of SARS-CoV2 sequenced in BCSIR, Bangladesh and comparing with the reference sequence NC_045512. The FASTA file was

uploaded to the Genome Detective Coronavirus Typing Tool version 1.17. The toolkit uses BLAST and phylogenetic methods to assign a lineage to the virus sequence according to a method described by Cleemput et al. (7). The genome of interests are of the B.1.351_501Y.V2_20H lineage according to Coronavirus Typing Tool (<https://www.genomedetective.com/app/typingtool>)

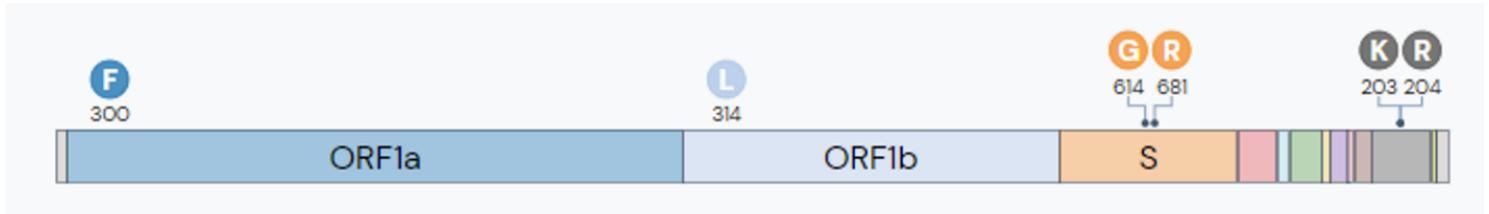


Figure 3

Characteristic mutations in variant. S protein from SARS-CoV-2 variant B.1.1.25 with mutations D614G and P681R. Mutations in at least 75% sequences.



Figure 4

Mutation analysis of 17 isolates from Bangladeshi patients. Multiple sequence alignment of Spike protein sequence of Bangladeshi isolates with sequences obtained from Wuhan. Sites of mutation are showed in Red.

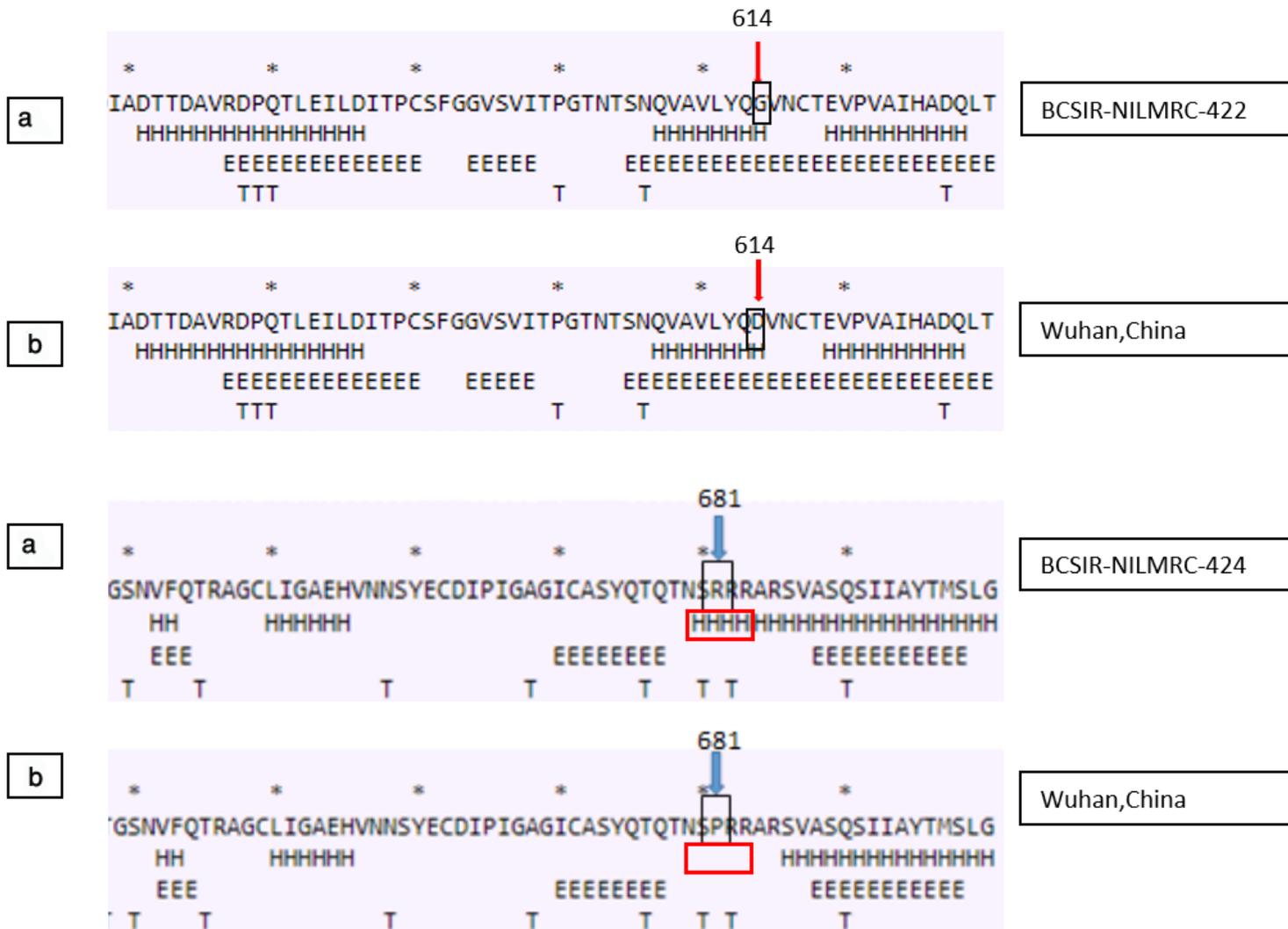


Figure 5

Effect of mutation at position 614 and 681 on the secondary structure of Spike protein in an isolate from Bangladeshi patient. (a) Secondary structure of Spike protein of Bangladeshi isolate showing the effect of mutation on the secondary structure; (b) Secondary structure of Spike protein of Wuhan isolate (the area around the residue 614 and 681 has been shown).

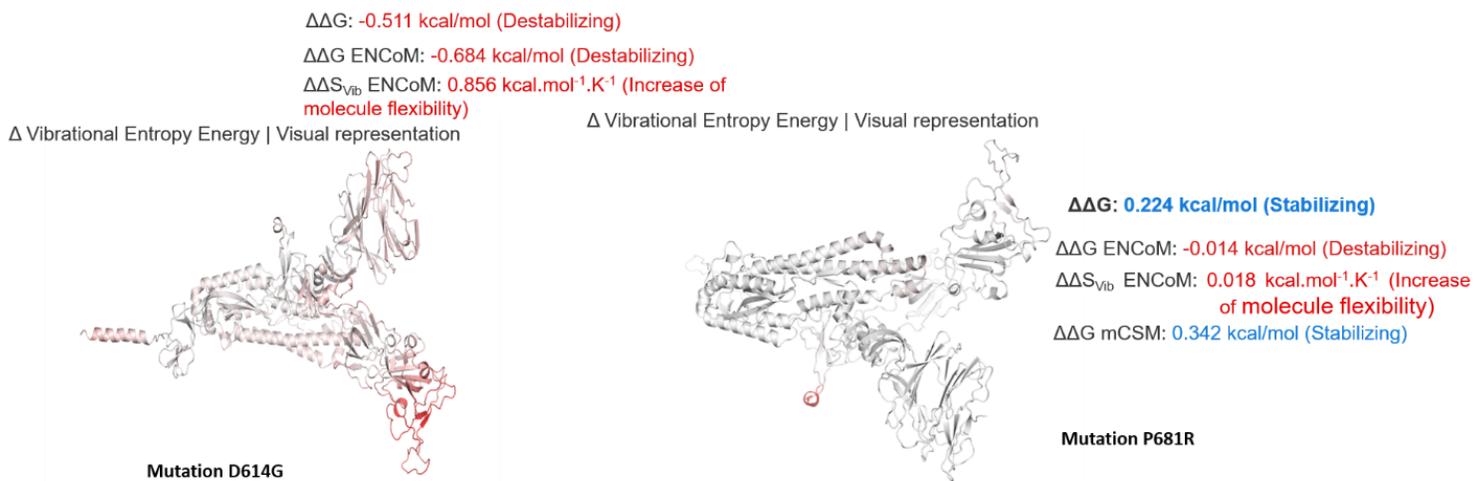


Figure 6

Vibrational Entropy Energy between Wild-Type and Mutant for mutation D614G and P681R in BCSIR-NILMRC-422. Amino acids colored according to the vibrational entropy change upon mutation. BLUE represents a rigidification of the structure and RED a gain in flexibility.

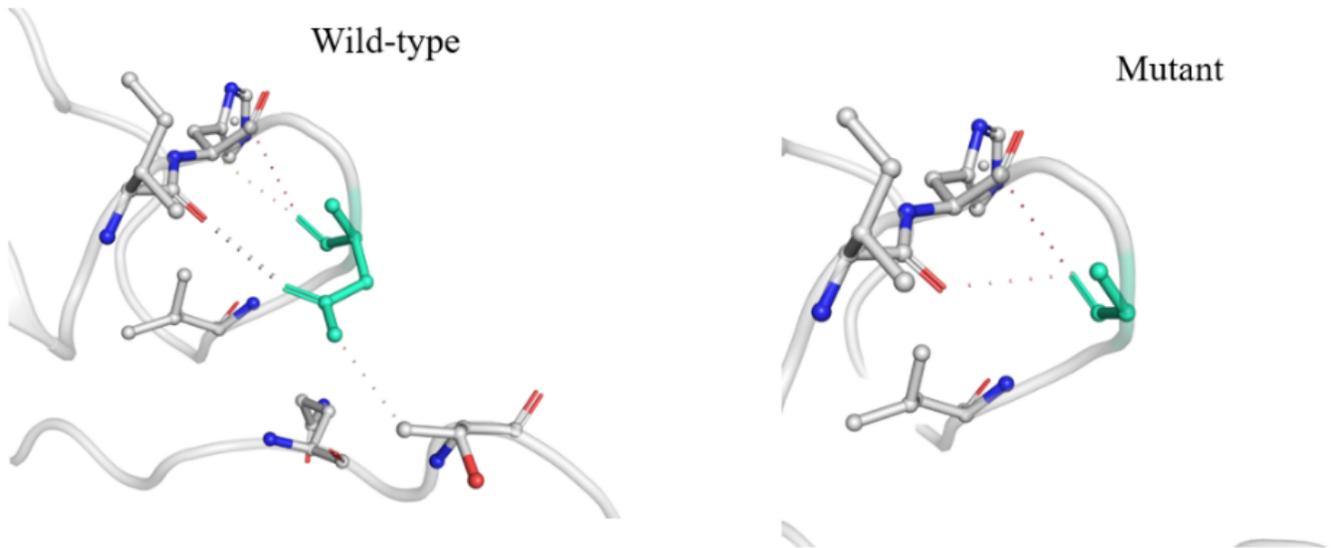


Figure 7

Interatomic interaction of D614G in isolate BCSIR-NILMRC-422. Wild-type and mutant residues are coloured in light-green and are also represented as sticks alongside with the surrounding residues which are involved on any type of interactions.

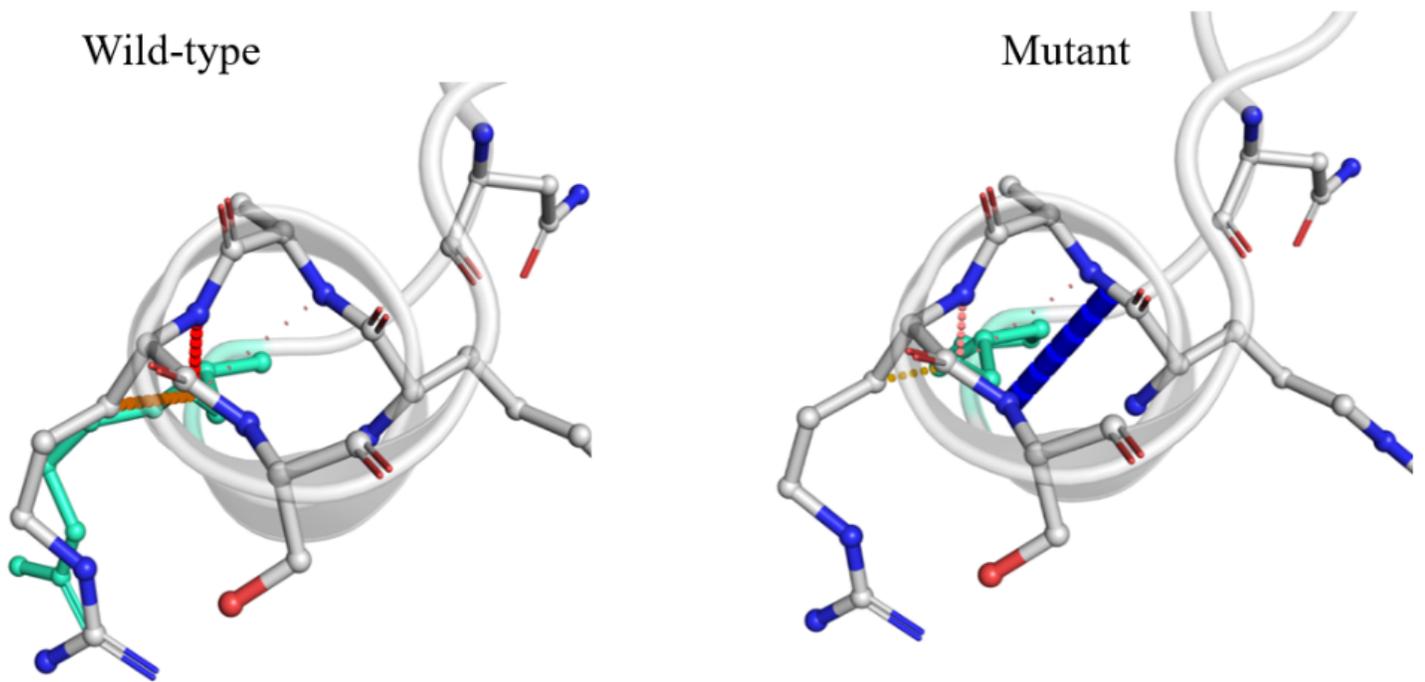


Figure 8

Interatomic interaction of mutation P681R in isolates BCSIR-NILMRC-424. Wild-type and mutant residues are coloured in light-green and are also represented as sticks alongside with the surrounding residues which are involved on any type of interactions.

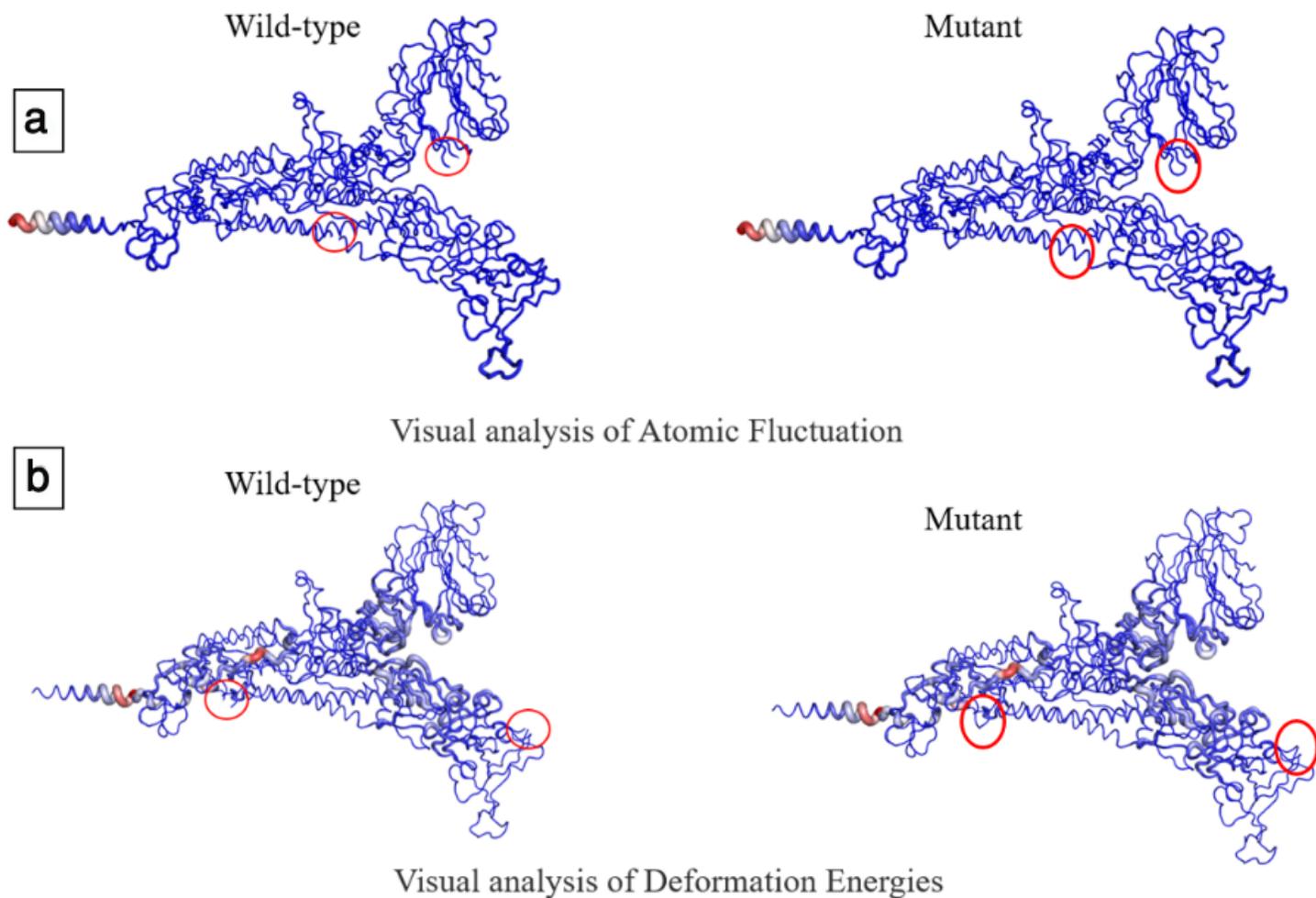


Figure 9

Visual analysis of fluctuations and deformation energies of mutation D614G in BCSIR-NILMRC-422. Magnitude of (a) atomic fluctuation and (b) deformation has been shown using thin to thick lines coloured blue (low), white (moderate) and red (high).

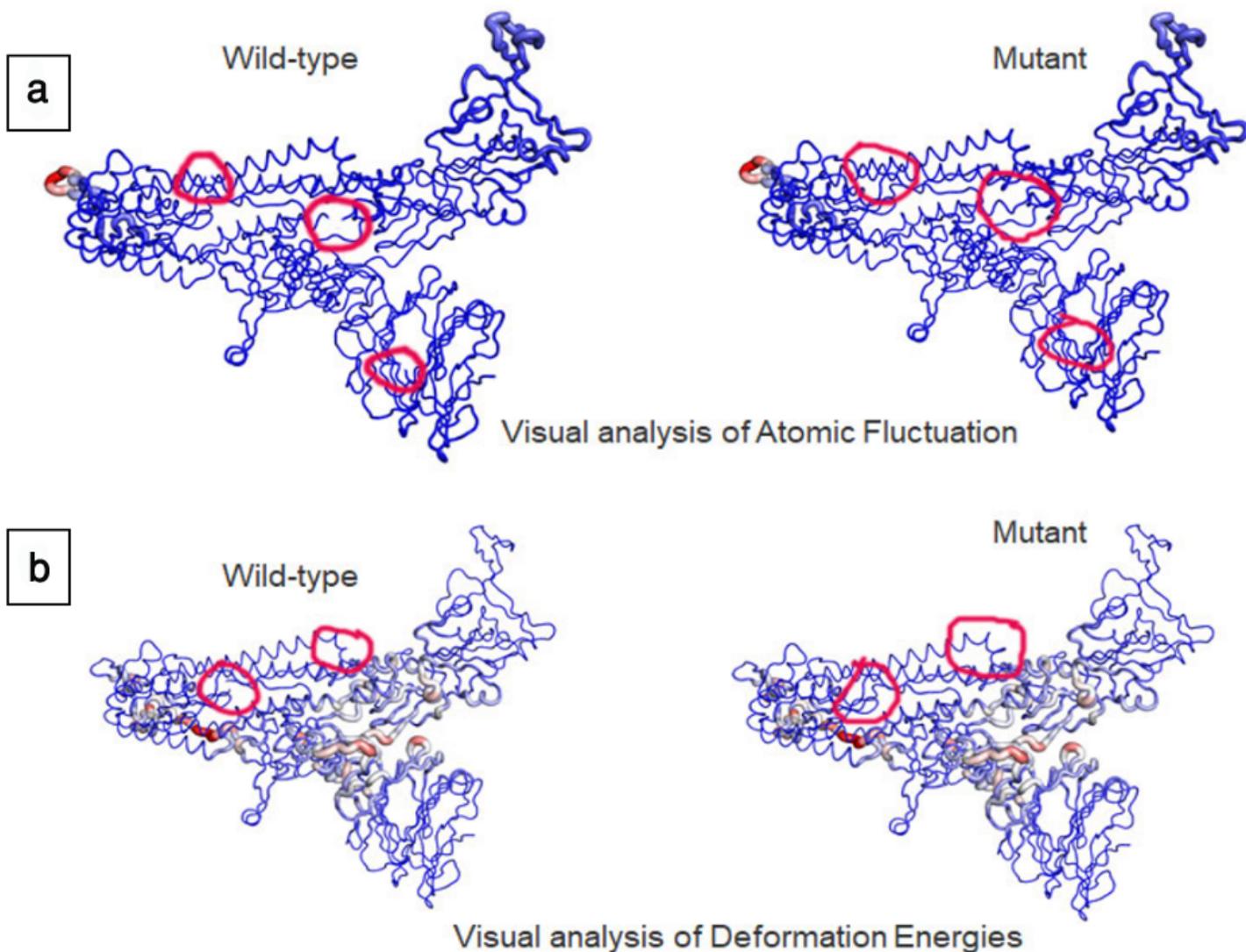


Figure 10

Visual analysis of fluctuations and deformation energies of mutation P681R in BCSIR-NILMRC-424. The magnitude of (a) atomic fluctuation and (b) deformation has been shown using thin to thick lines colored blue (low), white (moderate), and red (high).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfig1.tif](#)
- [Supplementaryfig2.tif](#)