Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure

Gyanendra Bahadur Chand 1, Atanu Banerjee 2, Gajendra Kumar Azad 1

1 Department of Zoology, Patna University, Patna, Bihar, India
2 Department of Zoology, Samastipur College, Samastipur, Bihar, India

Corresponding Author: Gajendra Kumar Azad
Email address: gkazadpatnauniversity@gmail.com

The rapid development of SARS-CoV-2 mediated COVID-19 pandemic has been the cause of significant health concern, highlighting the immediate need for the effective antivirals. SARS-CoV-2 is an RNA virus that has an inherent high mutation rate. These mutations drive viral evolution and genome variability, thereby, facilitating viruses to have rapid antigenic shifting to evade host immunity and to develop drug resistance. Viral RNA-dependent RNA polymerases (RdRp) perform viral genome duplication and RNA synthesis. Therefore, we compared the available RdRp sequences of SARS-CoV-2 from Indian isolates and ‘Wuhan wet sea food market virus’ sequence to identify, if any, variation between them. Our data revealed the occurrence of seven mutations in Indian isolates of SARS-CoV-2. The secondary structure prediction analysis of these seven mutations shows that three of them cause alteration in the structure of RdRp. Furthermore, we did protein modelling studies to show that these mutations can potentially alter the stability of RdRp protein. Therefore, we propose that RdRp mutations in Indian SARS-CoV-2 isolates might have functional consequences that can interfere with RdRp targeting pharmacological agents.
Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure

Gyanendra Bahadur Chand¹, Atanu Banerjee² and Gajendra Kumar Azad³*

¹Department of Zoology, Patna University, Patna-800005, Bihar, India
²Department of Zoology, Samastipur College, Samastipur-848134, Bihar, India
³Department of Zoology, Patna University, Patna-800005, Bihar, India

*Corresponding Author:
Gajendra Azad

Email address: gkazad@patnauniversity.ac.in

Keywords: COVID-19, SARS-CoV-2, RNA-dependent RNA polymerases (RdRp), Nsp12, Mutation, Indian geographical area
ABSTRACT

The rapid development of SARS-CoV-2 mediated COVID-19 pandemic has been the cause of significant health concern, highlighting the immediate need for the effective antivirals. SARS-CoV-2 is an RNA virus that has an inherent high mutation rate. These mutations drive viral evolution and genome variability, thereby, facilitating viruses to have rapid antigenic shifting to evade host immunity and to develop drug resistance. Viral RNA-dependent RNA polymerases (RdRp) perform viral genome duplication and RNA synthesis. Therefore, we compared the available RdRp sequences of SARS-CoV-2 from Indian isolates and 'Wuhan wet sea food market virus' sequence to identify, if any, variation between them. Our data revealed the occurrence of seven mutations in Indian isolates of SARS-CoV-2. The secondary structure prediction analysis of these seven mutations shows that three of them cause alteration in the structure of RdRp. Furthermore, we did protein modelling studies to show that these mutations can potentially alter the stability of RdRp protein. Therefore, we propose that RdRp mutations in Indian SARS-CoV-2 isolates might have functional consequences that can interfere with RdRp targeting pharmacological agents.

INTRODUCTION

SARS-CoV-2 (a member of Coronaviruses) outbreak occurred in Wuhan, China in December 2019 and it became a pandemic by spreading to almost all countries worldwide. The SARS-CoV-2 causes COVID-19 disease that has created a global public health problem. As of May 8, 2020, more than 3.9 million confirmed COVID-19 cases were reported worldwide with 0.27 million confirmed deaths. As the virus spread to new locations, they alter their protein sequence by the introduction of mutations in their genome that help them to survive better in the host. (Sackman et al., 2017). The β lymphocytes of host’s adaptive immune system eventually identify the specific epitopes of the pathogenic antigen and start producing protective antibodies, which in turn results in agglutination and clearance of the pathogen (Alcamí & Koszinowski, 2000;
Chaplin, 2010; Jensen & Thomsen, 2012). Being an efficient unique pathogen, a virus, often mutate its proteins in a manner that it can still infect the host cells, evading the host’s immune system. Even when fruitful strategies are discovered and engaged, the high rate of genetic change displayed by viruses frequently leads to drug resistance or vaccine escape (McKeegan, Borges-Walmsley & Walmsley, 2002).

The SARS-CoV-2 has a single stranded RNA genome of approximately 29.8Kb in length and accommodates 14 ORFs encoding 29 proteins that includes four structural proteins: Envelope (E), Membrane (M), Nucleocapsid (N) and Spike (S) proteins, 16 non-structural proteins (nsp) and 9 accessory proteins (Gordon et al., 2020a)(Wu et al., 2020) including the RNA dependent RNA polymerase (RdRp) (also named as nsp12). RdRp is comprised of multiple distinct domains that catalyze RNA-template dependent synthesis of phosphodiester bonds between ribonucleotides. The SARS-CoV-2 RdRp is the prime constituent of the replication/transcription machinery. The structure of the SARS-CoV-2 RdRp has recently been solved (Gao et al., 2020) and show three distinct domains.

For RNA viruses, the RNA-dependent RNA polymerase (RdRp) presents an ideal target because of its vital role in RNA synthesis and absence of host homolog. RdRp is therefore, a primary target for antiviral inhibitors such as Remdesivir (Gordon et al., 2020b) that is being considered as a potential drug for the treatment of COVID-19. Since, RNA viruses constantly evolve owing to the rapid rate of mutations in their genome, we decided to analyse RdRp protein sequence of SARS-CoV-2 from different geographical region to see if RdRp also mutate. Here, in the present study, we identified and characterised three mutations in RdRp protein isolated from India against that of the ‘Wuhan wet sea food market’ (Wu et al., 2020) SARS-CoV-2. Altogether, our data strongly suggests the prevalence of mutations in the genome of SARS-CoV-2 needs to be considered to develop new approaches for targeting this virus.

METHODS

Sequence retrieval:

We downloaded all SARS-CoV2 sequences from the NCBI virus database as shown in table 1. As of now, there are 28 SARS-CoV-2 sequences from India have been
deposited in this database, out of which, two sequences are not complete; therefore, we retrieved 26 datasets of SARS-CoV-2 (samples are from Indian). As a reference, we downloaded the sequence of SARS-CoV-2 that was first reported genome sequence deposited in the NCBI virus database from the ‘Wuhan wet sea food market area’ from the early days of COVID-19 pandemic (Wu et al., 2020). This virus was formerly called 'Wuhan seafood market pneumonia virus having the accession number YP_009724389.

Sequence alignments and structure:
All the RdRp protein sequences were aligned by multiple sequence alignment platform of CLUSTAL Omega (Madeira et al., 2019). Clustal Omega is a multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. The alignment file was carefully studied and differences in the amino acid changes were recorded.

Secondary structure predictions:
We used CFSSP (Ashok Kumar, 2013) (Chou and Fasman secondary structure prediction), an online server, to predict secondary structures of SARS-CoV-2 RdRp protein. This server predicts the possibility of secondary structure such as alpha helix, beta sheet, and turns from the amino acid sequence. CFSSP uses Chou-Fasman algorithm, which is based on analyses of the relative frequencies of each amino acid in secondary structures of proteins solved with X-ray crystallography.

RdRp dynamics study:
To investigate the effect of mutation on the RdRp protein structural conformation, its molecular stability and flexibility, we used DynaMut software (Rodrigues, Pires & Ascher, 2018) (University of Melbourne, Australia). To run this software we first downloaded the known protein structure of SARS-CoV-2 RdRp from RCSB (PDB ID: 6M71(Gao et al., 2020), 7BV1 and 7BV2(Yin et al., 2020) and used it for analysis. Next, the 6M71, 7BV1 and 7BV2 structure was uploaded on DynaMut software and effect of mutation in various protein structure stability parameters such as vibrational entropy; the atomic fluctuations and deformation energies were determined. DynaMut, a web server
implements well established normal mode approaches that is used to analyse and
visualise protein dynamics. This software samples conformations and measures the
impact of mutations on protein dynamics and stability resulting from vibrational entropy
changes and also predicts the impact of a mutation on protein stability.

RESULTS:
Identification of mutations in RdRp protein present in Indian isolates
The SARS-COV-2 sequencing data was downloaded from NCBI (NCBI-Virus-SARS-
CoV-2 data hub). There are 26 SARS-CoV-2 sequences available from India. We
downloaded all Indian SARS-CoV-2 sequences and Wuhan SARS-CoV-2 sequence,
which was deposited for the first time after COVID-19 cases started to appear in Wuhan
province, China. We downloaded protein sequence of NSP12/ RdRp from the database.
All these RdRp sequences were first aligned in CLUSTAL Omega to check for
similarities or differences. The Clustal Omega algorithm produces a multiple sequence
alignment by producing pairwise alignments. Using this algorithm, we identified seven
mutations in isolates from Indian SARS-CoV-2 samples compared to Wuhan SARS-
CoV-2 RdRp sequence as shown in figure 1. We considered the original Wuhan
sequence as the wild type for this comparison. These mutations on RdRp are A97V,
A185V, I201L, P323L, L329I, A466V and V880I respectively. Out of these, one isolate
have triple mutation, two isolates have double mutation, and rest have single mutations
(figure 1).

P323L mutation causes the alteration in secondary structure of RdRp
Next, we studied the effect of these mutations on secondary structure of RdRp. Our
data revealed that mutation at four positions have no effect in secondary structure that
includes I201L, L329I, A466V and V880I (figure 2 C, E, F and G). However, mutation in
rest of the three sites causes change in secondary structure (A97V, A185V and P323L)
as shown in figure 2 (panel A, B and D). At two positions (97 and 185) the alanine
amino acid is substituted by valine. The valine side chain is larger than alanine, and
substitution of valine at position 97 and 185 impairs packing of the protein as revealed
by our secondary structure predications at these two sites (figure 2A and 2B, compare
panel i and ii). Our analysis showed that there is addition of two sheet at positions 97 and 98 due to mutation of A97V (figure 2A, compare panel i and ii). Similarly, in A185V mutant, there is a loss of turn at 184 and replacement of helix with sheet structure at 181, 182, 183 and 184 positions (figure 2B, compare panel i and ii). Further, our secondary structure prediction also showed changes in secondary structure when proline is substituted by leucine at 323 position (Figure 2D and H, compare panel i and ii). The detailed analysis revealed that the mutant RdRp (P323L) have attained considerable changes in secondary structure at the mutation site. There is a loss of turn structures from position 323 and 324 and addition of five sheets at positions 321, 322, 323, 324 and 327 (Figure 2D and H, highlighted in dashed box). Proline possesses a unique property, its side chain cyclizes back on to the backbone amide position, due to which it contributes to the secondary structure formation because of its bulky pyrrolidine ring that places steric constraints on the conformation of the preceding residue in the helix. The substitution of proline to leucine in the mutant RdRp might result in loss of the structural integrity provided by proline. Leucine is a hydrophobic amino acid and generally buried in the folded protein structure. Altogether, the substitution of alanine to valine at position 97, 185 and proline to leucine at position 323 in mutant RdRp is changing the secondary structure of the protein that might have functional consequences.

**P323L alters the stability dynamics of tertiary structure of RdRp.**

To understand the impact of mutations on tertiary structure of RdRp, we performed protein modelling using Dynamut software (Rodrigues, Pires & Ascher, 2018). The Dynamut software provides us the information about the alteration in protein stability and flexibility due to the mutations in the native protein structure. Our data revealed that there is change in vibrational entropy energy ($\Delta \Delta S_{\text{Vib\, ENCoM}}$) between the wild type (Wuhan isolate) and the mutant Indian isolate (Figure 3A). Vibrational entropy represents an average of the configurational entropies of the protein within single minima of the energy landscape (Goethe, Fita & Rubi, 2015). The negative $\Delta \Delta S_{\text{Vib\, ENCoM}}$ of mutant RdRp represents the rigidification of the protein structure and positive $\Delta \Delta S_{\text{Vib\, ENCoM}}$ represents gain in flexibility. Here, our data show that the
mutation at A185V and P323L lead to rigidification of mutant protein structure Figure 3B and D). However, the mutation at I201L leads to increase in flexibility (Figure 3C).

Further, we also calculated the free energy differences, $\Delta \Delta G$, between wild-type and mutant. The free energy differences, $\Delta \Delta G$, caused by mutation have been correlated with the structural changes, such as changes in packing density, cavity volume and accessible surface area and therefore, it measures effect of mutation on protein stability (Eriksson et al., 1992). In general, a $\Delta \Delta G$ below zero means that the mutation causes destabilization and above zero represents protein stabilization. Here, our analysis showed positive $\Delta \Delta G$ for A185V and P323L mutations suggesting that P323L mutation is stabilising protein structure (Figure 3A); however, we observed negative $\Delta \Delta G$ for I201L mutation indicating its destabilising behaviour (Figure 3A). The $\Delta \Delta G$ for P323L mutant was 0.908 kcal/mol which is significantly higher than others. Therefore, we also did protein modelling of P323L mutant using two additional structures of RdRp (7BV1, 7BV2) (Yin et al., 2020) and analysed its impact on $\Delta \Delta G$. Our data show that in both the cases (7BV1 and 7BV2) the P323L mutation is leading to stabilisation of the protein structure with the values of $\Delta \Delta G$ as 0.530 kcal/mol and 0.460 kcal/mol respectively.

We further closely analysed the changes in the intramolecular interactions due to these three mutations in RdRp. Our data showed that it is affecting the interactions of the residues which are present in the close vicinity of alanine, isoleucine and proline. The substitution of wild type residues with mutant residues alters the side chain leading to change of intramolecular bonds in the pocket, where these amino acids resides as shown in figure 4A, B and C. Therefore, it can be conclusively stated that the three mutations namely, A97V, A185V and P323L, in large, are changing the stability and intramolecular interactions in the protein that might have functional consequences.

**DISCUSSION**

RNA viruses including coronaviruses exhibits high mutation rates that provide these viruses evolutionary advantage to better adapt and survivability (Domingo & Holland, 1997). Mutations derive the process of natural selection by selecting those viral strains that are more potent and fit (Duffy, 2018) that facilitates drug resistance and immune
The SARS-CoV-2 infected humans from Wuhan province in China and quickly spread to almost all countries worldwide. Since, this virus has already spread to different demographic areas having different climatic condition, temperature, humidity and seasonal variations; therefore, we can predict that this virus might be mutating to adapt to new environments. Towards this, we investigated the mutations in RdRp of SARS-CoV-2. We focused primarily on RdRp because it is an indispensable protein of this virus that helps in its replication and transcription. More importantly, there are many drugs which specifically target RdRp and are potent antivirals. Our study reports the RdRp mutations present in the Indian isolates of SARS-CoV-2 and, will help to understand the effect of variations on RdRp protein. Our data demonstrate that A97V, A185V and P323L mutations lead to significant changes in the protein secondary structure (Figure 2). P323L mutation lies in the interface domain (residues A250-R365) of the RdRp protein. This domain helps in the coordination of N and C terminal domains of RdRp; therefore, a mutation in the interface domain might have drastic impact on the function of RdRp. A recent virtual molecular docking studies, screened approximately 7500 drugs to identify SARS-CoV-2 RdRp inhibitor revealed several potential compounds (DOI: 10.20944/preprints202003.0024.v1) namely, Simeprevir, Filibuvir and Tegobuvir, etc. Same study also predicted that these drugs bind RdRp at a putative docking site (a hydrophobic cleft) that includes phenylalanine at 326th position. Interestingly, the mutation identified in our study is very close to the docking site (P323L and L329I). Therefore, it is reasonable that the substitution of these two amino acids at 323 and 329 might interfere with the interaction of these drugs with RdRp. Further, our data also revealed that P323L mutation is causing stabilisation of the protein structure. The mutations in RdRp have already been linked to drug resistance in different viruses. Such as a study on RdRp of influenza A virus demonstrate that a K229R mutation confers resistance to favipiravir (Goldhill et al., 2018). Similarly, mutation in hepatitis C virus RdRp at P495, P496 or V499 & T389 and have been linked to drug resistance (Delang et al., 2012). Altogether, it is conceivable that many RNA viruses acquire drug resistance through mutations in RdRp. However, the functional characterisation of RdRp mutations investigated in our study needs to be
carried out to understand the exact role of these mutations which will help scientific community to better therapeutic targeting of SARS-CoV-2.

CONCLUSIONS

Altogether, our data strongly suggests that SARS-CoV-2 is acquiring mutations as it is spreading to new locations. Most likely, these mutations are helping SARS-CoV-2 to adapt better inside host and new geographical locations. One of the mutations identified in our study (P323L) might have functional consequences that need to be addressed in future studies.

ACKNOWLEDGEMENTS

We would like to thank Patna University, Patna for providing necessary infrastructural support.

FIGURE AND TABLE LEGENDS

Figure 1: Multiple sequence alignment of Wuhan SARS-CoV-2 RdRp protein with sequences obtained from India. The mutations are highlighted in red font. Only those sequences are shown that have variations, rest of the sequences are identical among all samples.

Figure 2: Effect of mutations on secondary structure of RdRp. Figure A, B, C, D, E, F, G and H demonstrate seven mutations observed in Indian isolates. Panel (i) represents sequence of Wuhan isolate and panel (ii) represents sequence of Indian isolates. The small rectangular box shows the mutant residue. The difference of secondary structure between Wuhan and Indian isolates are highlighted with dashed box in respective panels.

Figure 3: Analysis of RdRp dynamicity and flexibility. A) The table shows the values of change in ΔΔS ENCoM and ΔΔG due to the mutation. B) Δ Vibrational Entropy Energy between Wild-Type and Mutant RdRp, amino acids are colored according to the
vibrational entropy change as a consequence of mutation of RdRp protein. **BLUE** represents a rigidification of the structure and **RED** a gain in flexibility.

Figure 4: Interatomic interactions mediated by A185V, I201L and P323L of RdRp. (A-B) represents alanine to valine substitution at 185<sup>th</sup> position, (C-D) represents isoleucine to leucine substitution at 201<sup>st</sup> position, (E-F) represents proline to leucine substitutions at 323<sup>rd</sup> position. Wild-type and mutant residues are colored in light-green and are also represented as sticks alongside with the surrounding residues which are involved on any type of interactions.

**Table 1:** Details of SARS-CoV-2 sequences used in this study

**REFERENCES**

Alcami A, Koszinowski UH. 2000. Viral mechanisms of immune evasion. *Trends in Microbiology*. DOI: 10.1016/S0966-842X(00)01830-8.

Ashok Kumar T. 2013. CFSSP: Chou and Fasman Secondary Structure Prediction server. *Wide Spectrum*. DOI: 10.5281/zenodo.50733.

Chaplin DD. 2010. Overview of the immune response. *Journal of Allergy and Clinical Immunology*. DOI: 10.1016/j.jaci.2009.12.980.

Delang L, Froeyen M, Herdewijn P, Neyts J. 2012. Identification of a novel resistance mutation for benzimidazole inhibitors of the HCV RNA-dependent RNA polymerase. *Antiviral Research*. DOI: 10.1016/j.antiviral.2011.10.012.

Domingo E, Holland JJ. 1997. RNA VIRUS MUTATIONS AND FITNESS FOR SURVIVAL. *Annual Review of Microbiology*. DOI: 10.1146/annurev.micro.51.1.151.

Duffy S. 2018. Why are RNA virus mutation rates so damn high? *PLoS Biology*. DOI: 10.1371/journal.pbio.3000003.

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*. DOI: 10.1126/science.1553543.

Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J,
Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. DOI: 10.1126/science.abb7498.

Goethe M, Fita I, Rubi JM. 2015. Vibrational entropy of a protein: Large differences between distinct conformations. *Journal of Chemical Theory and Computation*. DOI: 10.1021/ct500696p.

Goldhill DH, Te Velthuis AJW, Fletcher RA, Langat P, Zambon M, Lackenby A, Barclay WS. 2018. The mechanism of resistance to favipiravir in influenza. *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1811345115.

Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, O&M#039;Meara MJ, Guo JZ, Swaney DL, Tummino TA, Huttenhain R, Kaake RM, Richards AL, Tutuncuoglu B, Foussard H, Batra J, Haas K, Modak M, Kim M, Haas P, Polacco BJ, Braberg H, Fabius JM, Eckhardt M, Soucheray M, Bennett MJ, Cakir M, McGregor MJ, Li Q, Naing ZZC, Zhou Y, Peng S, Kirby IT, Melnyk JE, Chorba JS, Lou K, Dai SA, Shen W, Shi Y, Zhang Z, Barrio-Hernandez I, Memon D, Hernandez-Armenta C, Mathy CJP, Perica T, Pilla KB, Ganesan SJ, Saltzberg DJ, Ramachandran R, Liu X, Rosenthal SB, Calviello L, Venkataramanan S, Liboy-Lugo J, Lin Y, Wankowicz SA, Bohn M, Sharp PP, Trenker R, Young JM, Cavero DA, Hiatt J, Roth T, Rathore U, Subramanian A, Noack J, Hubert M, Roesch F, Vallet T, Meyer B, White KM, Miorin L, Rosenberg OS, Verba KA, Agard D, Ott M, Emerman M, Ruggero D, Garc&amp;amp;amp;iacute-Sastre A, Jura N, von Zastrow M, Taunton J, Ashworth A, Schwartz O, Vignuzzi M, d&amp;#039;Enfert C, Mukherjee S, Jacobson M, Malik HS, Fujimori DG, Ideker T, Craik CS, Floor S, Fraser JS, Gross J, Sali A, Kortemme T, Beltrao P, Shokat K, Shoichet BK, Krogan NJ. 2020a. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv*. DOI: 10.1101/2020.03.22.002386.

Gordon CJ, Tchesnokov EP, Woolner E, Perry JK, Feng JY, Porter DP, Gotte M. 2020b. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *Journal of Biological Chemistry*. DOI: 10.1074/jbc.ra120.013679.
Irwin KK, Renzette N, Kowalik TF, Jensen JD. 2016. Antiviral drug resistance as an adaptive process. *Virus Evolution*. DOI: 10.1093/ve/vew014.

Jensen S, Thomsen AR. 2012. Sensing of RNA Viruses: a Review of Innate Immune Receptors Involved in Recognizing RNA Virus Invasion. *Journal of Virology*. DOI: 10.1128/jvi.05738-11.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*. DOI: 10.1093/nar/gkz268.

McKeegan KS, Borges-Walmsley MI, Walmsley AR. 2002. Microbial and viral drug resistance mechanisms. *Trends in Microbiology*. DOI: 10.1016/S0966-842X(02)02429-0.

Rodrigues CHM, Pires DEV, Ascher DB. 2018. DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*. DOI: 10.1093/nar/gky300.

Sackman AM, McGee LW, Morrison AJ, Pierce J, Anisman J, Hamilton H, Sanderbeck S, Newman C, Rokyta DR. 2017. Mutation-driven parallel evolution during viral adaptation. *Molecular Biology and Evolution*. DOI: 10.1093/molbev/msx257.

Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. *Nature*. DOI: 10.1038/s41586-020-2008-3.
Figure 1

Multiple sequence alignment of SARS-CoV-2 RdRp protein

Multiple sequence alignment of Wuhan SARS-CoV-2 RdRp protein with sequences obtained from India. The mutations are highlighted in red font. Only those sequences are shown that have variations, rest of the sequences are identical among all samples.
Figure 2

Prediction of secondary structure of RdRp protein

Effect of mutations on secondary structure of RdRp. Figure A, B, C, D, E, F, G and H demonstrate seven mutations observed in Indian isolates. Panel (i) represents sequence of Wuhan isolate and panel (ii) represents sequence of Indian isolates. The small rectangular box shows the mutant residue. The difference of secondary structure between Wuhan and Indian isolates are highlighted with dashed box in respective panels.
Figure 3

Effect of mutations on structural dynamics of RdRp protein

Analysis of RdRp dynamicity and flexibility. A) The table shows the values of change in $\Delta \Delta S$ ENCoM and $\Delta \Delta G$ due to the mutation. B) $\Delta$ Vibrational Entropy Energy between Wild-Type and Mutant RdRp, amino acids are colored according to the vibrational entropy change as a consequence of mutation of RdRp protein. **BLUE** represents a rigidification of the structure and **RED** a gain in flexibility.

| S. No | Wuhan Isolate | Indian Isolate | Amino-acid Position | $\Delta \Delta S$ ENCoM | $\Delta \Delta G$ DynaMut |
|-------|---------------|----------------|----------------------|-------------------------|-------------------------|
| 1     | A             | V              | 185                  | -0.42 kcal.mol$^{-1}$.K$^{-1}$ | 0.134 kcal/mol           |
| 2     | I             | L              | 201                  | 0.066 kcal.mol$^{-1}$.K$^{-1}$ | -0.047 kcal/mol          |
| 3     | P             | L              | 323                  | -0.277 kcal.mol$^{-1}$.K$^{-1}$ | 0.908 kcal/mol           |

**A185V**

**I201L**

**P323L**
Figure 4

Effect of amino acid substitution on interatomic interactions

Interatomic interactions mediated by A185V, I201L and P323L of RdRp- (A-B) represents alanine to valine substitution at 185th position, (C-D) represents isoleucine to leucine substitution at 201st position, (E-F) represents proline to leucine substitutions at 323rd position. Wild-type and mutant residues are colored in light-green and are also represented as sticks alongside with the surrounding residues which are involved on any type of interactions.
Table 1 (on next page)

Details of SARS-CoV-2 sequences used in the analysis

The accession number of SARS-CoV-2 sequences used in this study are mentioned in the table.
Table 1: Details of SARS-CoV-2 sequences used in the analysis

| S.No | Accession Number | Authors | Institute/ University |
|------|------------------|---------|-----------------------|
| 1    | YP_009724389     | Wu,F., Zhao,S., et al., 2020 | Fudan University, Shanghai, China |
| 2    | QHS34545         | Yadav,P.D., et al., 2020 | NIV, Pashan, Pune, Maharashtra 411021, India |
| 3    | QIA98582         | Potdar,V., et al., 2020 | NIV, Pashan, Pune, Maharashtra 411021, India |
| 4    | QJC19489         | Pandit,R.,Shah,T., et al., 2020 | GBRC, Gandhinagar, Gujarat 382010, India |
| 5    | QJF11810         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 6    | QJF11822         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 7    | QJF11834         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 8    | QJF11846         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 9    | QJF11858         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 10   | QJF11870         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 11   | QJF11882         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 12   | QJF77844         | Muttineni,R., et al., 2020 | VRL, Dept. of Zoology, Osmania University, Hyderabad-500007,India |
| 13   | QJF77856         | Muttineni,R., et al., 2020 | VRL, Dept. of Zoology, Osmania University, Hyderabad-500007,India |
| 14   | QJF77868         | Muttineni,R., et al., 2020 | VRL, Dept. of Zoology, Osmania University, Hyderabad-500007,India |
| 15   | QJF77880         | Muttineni,R., et al., 2020 | VRL, Dept. of Zoology, Osmania University, Hyderabad-500007,India |
| 16   | QJQ27840         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 17   | QJQ27852         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 18   | QJQ27864         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 19   | QJQ27876         | Pattabiraman,C., et al., 2020 | NIMHANS, Bangalore, Karnataka 560029,India |
| 20   | QJQ28343         | Trivedi, P., et al., 2020 | GBRC, Gandhinagar, Gujarat 382011, India |
| 21   | QJQ28355         | Hinsu, A., et al., 2020 | GBRC, Gandhinagar, |
|   | Manuscript ID | Author(s)          | Location                                      |
|---|---------------|--------------------|-----------------------------------------------|
| 22| QJQ28367      | Sabara, P., et al., 2020 | GBRC, Gandhinagar, Gujarat 382011, India      |
| 23| QJQ28379      | Puvar, A., et al., 2020 | GBRC, Gandhinagar, Gujarat 382011, India      |
| 24| QJQ28391      | Raval, J., et al., 2020 | GBRC, Gandhinagar, Gujarat 382011, India      |
| 25| QJQ28403      | Gandhi, M., et al., 2020 | GBRC, Gandhinagar, Gujarat 382011, India      |
| 26| QJQ28415      | Shah, T., et al., 2020  | GBRC, Gandhinagar, Gujarat 382011, India      |
| 27| QJQ28427      | Pandya, M., et al., 2020 | GBRC, Gandhinagar, Gujarat 382011, India      |