

**Immunoinformatics and genomic characterization of SARS-CoV-2 helicase (nsp13)  
mutational profile; an attractive antiviral therapeutic target**

Niti Yashvardhini <sup>1\*</sup>, Deepak Kumar Jha<sup>2</sup>, Amit Kumar<sup>3</sup>, Kumar Sayrav<sup>4</sup>

<sup>1</sup> Department of Microbiology, Patna Women's College, Patna, 800 001, India

<sup>2</sup> Department of Zoology, S.M.P. Rajkiya Mahila Mahavidyalaya, Ballia, 277401, India

<sup>3</sup> Department of Botany, Patna University, Patna-800 005, India

<sup>4</sup> Department of Chemistry, V.K.S. University, Ara, 802301, India

\*<sup>1</sup> Corresponding author- [nitiyashvardhini@gmail.com](mailto:nitiyashvardhini@gmail.com)

## **ABSTRACT**

SARS-CoV-2 (Severe Acute Respiratory Syndrome coronavirus) poses an unprecedented public health threat to the mankind globally since December 2019. Considering the escalating number of positive cases, the World Health Organization announced public health crisis on 11<sup>th</sup> March 2020 worldwide. The pathogenesis of SARS-CoV-2 (causal agent of COVID-19) is not clearly understood till date and the role of antiviral therapeutics is yet to be established for this pandemic. The objective of our study was to compare the SARS-CoV-2 sequence of Wuhan virus with those of Indian SARS-CoV-2 isolates; only those mutations which occurred in the helicase (nsp13) region were addressed and used for subsequent study. For mutational characterization multiple sequence alignment was done followed by protein dynamics study using Chimera and Dynamut software. Altogether, 51 mutations were detected in the nsp13 of Indian SARS-CoV-2, out of them 7 mutants were used for subsequent study. Furthermore, prediction of secondary structure as well as protein dynamics study revealed that these mutations altered the structural stability and flexibility of helicase protein. In this *in silico* study, predictive tools of immunoinformatics were used for the prediction of B-cell, T-cell and MHC I epitopes. Two mutations have been detected in the predicted epitope region of SARS-CoV-2 helicase that might induce its conformational changes and considered as a major challenge in the development of vaccine. The present investigation was, therefore, undertaken to analyze one of the crucial drug target like helicase which is indispensable for the replication /transcription machinery of SARS-CoV-2.

**Keywords :** SARS-CoV-2 · Helicase · COVID-19 · Mutation · Genome variability. Vaccine

## **Introduction**

SARS-CoV-2, a novel betacoronavirus, belonging to the family Coronaviridae and order nidovirales is an enveloped, positive-sense, largest known RNA virus (genome size~30kb) and the causative agent of coronavirus disease 2019 (COVID-19) [1,2]. Coronavirus outbreak initiated from a small sea food market in Wuhan, China [3]. Although, several vaccines candidates are approved and are used on a mass scale, however the contagiousity of this virus are still a great threat to human beings whose ripple effect poses a huge human health crisis. As of 23<sup>rd</sup> December 2021, confirmed cases of COVID-19 have been reported 276,436,619, including 5,374,744 casualties, globally by WHO. Multiple mutations help in evolution of the virus as well as generates variability which enable viruses to evade host immunity [4]. RNA viruses, like SARS-CoV-2 most frequently showing high rate of mutation because of poor fidelity of its RNA polymerase, leading to antigenic variability [5,6] and these high rates are greatly linked to virulence and evolvability of this virus, traits prerequisite for viral adaptation [7].

SARS-CoV-2 genome comprises of 14 open reading frame (ORF) sequences which encode 29 proteins including 4 types of structural proteins; S (spike), E (envelope), M (membrane), and N (nucleocapsid) that are essential for the assembly of a complete virion particle. Additionally, Replication and transcription of the SARS-CoV-2 genome are accomplished by a complex RNA replication/transcription process, consisting of as many as 16 non structural proteins (nsps) including 9 accessory proteins [8,9]. Among the non structural proteins RNA dependent RNA polymerase (RdRp, nsp12), helicase (nsp13) as well as main proteases (mpro) or chymotrypsin like protease (3CL pro) together make up highly conserved region in SARS-CoV-2 and considered as a suitable drug target [10].

SARS-CoV-helicase (nsp13) plays important role in viral replication, catalyzing the unwinding of nucleic acids and also separates the double helical structure of complex RNA duplex structure into the single-stranded forms in ATP/NTP dependent manner. These vital characteristics of SARS-CoV-2 helicase protein describe its importance as one of the potential target for the development of anti human corona virus agents [11,12]. The SARS-CoV-2 helicase (nsp13) contain 600 amino acids which are located in the region of orflab polyprotein, ranging from 5325 to 5925 amino acids. The overall structure of SARS-CoV-2 nsp13 acquire a triangular pyramid like shape similar to that of SARS and MERS-nsp13 which is comprising of at least five domains. In the present study, we investigated the potential effects of mutations on the activity of helicase protein isolated from Indian SARS-CoV-2 and compare it with those of Wuhan isolates. Considerable alterations have been observed in the protein dynamics and predicted epitopes like B-cell, T-cell and MHC I of SARS-CoV-2 helicase. Since RNA helicase plays pivotal role in the life cycle of SARS-CoV-2 hence, it can be used as a major ant viral therapeutic target.

## **Methods**

### **Sequence Retrieval**

A total of approximately 500 SARS-CoV-2 protein sequences were downloaded from NCBI virus database, also the first reported virus sequence from Wuhan was also retrieved to be used as a reference (Accession number YP\_009724389) [9]. The samples of these SARS-CoV-2 viruses were submitted from India in the month of June 2020 to February 2021.

### **Sequence alignment and Phylogenetic analysis**

The sequences downloaded above were aligned using CLUSTAL Omega multiple sequence alignment programme with HMM profiling [13]. Jalview was used to visualize these aligned

files and the differences in the helicase / nsp13 region were recorded (600 amino acid length). Protein Variation Effect Analyzer known as PROVEAN v1.1.3 was used to detect the non-synonymous amino acid substitutions with cutoff predicted score of -2.50 [14]. To know the phylogeny, a neighbor joining phylogenetic tree was prepared using MEGAX (Molecular Evolutionary Genetics Analysis) with default parameters [15]. The sequence file was first aligned in MEGAX and then the tree was constructed using this alignment.

### **Secondary Structure prediction**

To know the formation or deletion of alpha helices, beta sheets and turns, the secondary structure prediction of the wild type and mutated helicase protein was done using an online server, CFSSP [16].

### **Estimation of helicase protein physicochemical properties**

The physicochemical properties such as molecular weight, extinction coefficient, amino acid composition, instability index, estimated half life, aliphatic index and average of hydropathicity (GRAVY) was calculated using Protparam tool of Expasy online program. Also, the hydropathy plot of the helicase protein was prepared using ProtScale tool of expasy [17].

### **Helicase protein dynamics study**

Protein modeling of helicase protein of wild type and mutated was done using Chimera software [18]. Ramachandran plot was prepared using Swiss model homology modeling platform. The effect of mutation on the overall of dynamics of helicase protein was studied using Dynamut software [19]. Dynamut provides information on the stability, NMA analysis, flexibility, rigidness, conformation of mutated as well as wild type protein. Several parameters were

recorded like flexibility analysis, vibrational entropy, atomic and deformation energies using first 10 non-trivial modes of the structure. The change in intramolecular interactions was also recorded using Dynamut upon mutation.

### **B-cell Epitope prediction**

Helicase protein sequence was used to predict a probable epitope, if any in the sequence using Antibody Epitope Prediction software, Bepipred [20].

### **Prediction of SARS-CoV-2 T-cell epitopes and MHC allele identification**

IEDB [20] Tepitool server was used to predict the T-cell epitope binding alongwith the detection of MHC allele showing highest affinity for the T-cell epitope. This web server predicts epitopes restricted to a large number of MHC I and MHC II alleles.

### **MHC allele cluster analysis**

MHCcluster 2.0 online tool was used for the cluster analysis of MHC class I and MHC class II alleles which might interact with the epitopes leading to immune responses. This online server predicts epitopes and the allele binding in a phylogenetic way in form of clusters [21].

### **Antigenicity and allergenicity prediction**

The antigenicity of the helicase protein was evaluated using VaxiJen v2.0 server, which classifies antigen on the basis of auto cross-covariance (ACC) transformation of the protein sequences [22]. The allergenicity of the helicase protein was predicted using AllerTOP server, which evaluates protein allergenicity on auto cross variance ACC method [23].

## **Results**

## **Revelation of SARS-CoV-2 helicase mutation and phylogenetic analysis**

The study was initiated by acquiring sequences of SARS-CoV-2 helicase, 600 amino acid length. There were nearly 500 SARS-CoV-2 full length protein sequences submitted from India in the month of June 2020 to February 2021. The virus genome being first sequenced was used as a reference for analyzing mutants. These sequences were aligned using Clustal omega and visualized by Jalview to detect similarity and variation in the protein sequences of India to that of Wuhan. Those mutations which occurred in the helicase region (nsp13) of the ORF1ab polyprotein were identified and characterized further in this study. A total of 51 mutations were detected in the helicase of Indian SARS-CoV-2 among which seven were characterized further which are P53S, H164Y, Y205C, G206C, H245R, R442Q and F499L as shown in supplementary table 1. These seven SARS-CoV-2 helicase mutants were further characterized to study the impact of mutation on helicase structure and function.

### **Nonsynonymous mutation in SARS-CoV-2 helicase**

The seven nonsynonymous amino acid substitutions showed different impact on protein structure and function. Out of seven mutations Y205C, G206C and F499L were deleterious for the protein whereas others were neutral at -2.5 cutoff values of PROVEAN score (Table 1).

To detect the phylogenetic relationships of different SARS-CoV-2 isolates from India as with that of Wuhan, the helicase protein sequences were aligned using MEGAX and a neighbour joining phylogenetic tree was prepared. The analysis revealed the presence of Indian isolates on different cluster to that of Wuhan, showing the evolutionary variability in SARS-CoV-2 isolates (Fig. 1).

## **Determination of physicochemical parameters and hydropathy index of SARS-CoV-2 helicase protein**

The analysis using ProtParam (ExPasy) revealed that the helicase protein is 600 amino acids in length with a molecular weight of 66854.75 Da, GRAVY score of -0.096 and instability index 33.31 (Table 2). The hydropathy plot showed C-terminal amino acid to be more hydrophobic as compared to the N-terminal end (Fig. 2a).

### **Protein modeling and structure prediction**

The helicase protein sequence was modeled using Chimera software. The models of both wild type and mutated protein sequences were built as shown in Fig. 2b. Ramachandran plot analysis showed that most of the amino acid sequences were in the favored region of the plot (Fig. 2c) i.e., nearly 537 amino acids, making up to 89% of the total sequence present (Supplementary table 2).

### **Secondary structure alteration of helicase protein upon mutation**

Secondary structure analysis was done using CFSSP online program to detect the variation occurring in the alpha helix, beta sheet and turns upon mutation with respect to wild type helicase. The point mutations P53S, Y205C, H245R and F499L didn't show any impact on secondary structure however the other three H164Y, G206C and R442Q showed significant structural changes (Fig. 2d). The point mutation at position 164, where histidine is replaced by tyrosine in the helicase protein resulted in loss of beta sheet structure at positions 163, 164, 165 and 166.

Histidine being almost neutral amino acid favors the formation of beta sheet structure and its substitution resulted in loss of sheet secondary structure. Our analysis showed that point mutation at 206 position, where glycine is substituted by cysteine leading to loss of turn at position 207 as glycine, a non polar amino acid favors beta turn structure. This point mutation also resulted in formation of sheet structure at points 205, 206, 207 and 208 which may be due to cysteine moiety as its S group favors the incorporation of sheets. Further, the substitution of arginine by glutamine at position 442 incorporates sheet structure at 441 and 442 points. Glutamine is a polar uncharged amino acid, favors the formation of hydrogen bonds and hence beta sheets. Overall, the secondary structure analysis depicts significant changes in the formation and loss of helix, sheet and turn that can bring huge impact on helicase protein and hence leading to the SARS-CoV-2 multiplication.

### **Changes occurred in protein dynamics of helicase protein upon mutation**

Dynamut server was used for the study on protein dynamics of helicase protein. Dynamut software provides information about the flexibility or steadiness of a protein upon mutation as compared to the wild type as calculated by ENCoM, DUET, mCSM and others. The free energy difference,  $\Delta\Delta G$  between the wild and mutated protein sequences was calculated using Dynamut and the values showed a stabilizing mutation in all the mutated helicase proteins. The H164C mutant showed the highest  $\Delta\Delta G$  (1.945 kcal/mol), signifying the most stable mutant protein followed by F499L, Y205C, P53S and G206C, as shown in Table 3. H245R showed the least free energy difference amongst all mutants with 0.937 kcal/mol. Further, in this study, vibrational entropy energy ( $\Delta\Delta S_{Vib}$  ENCoM) was calculated, which provides information on the configurational entropy of the proteins with single minima of the energy landscape. The  $\Delta\Delta S_{Vib}$  ENCoM calculated for all the mutants revealed a negative value signifying the rigidification of

protein structure upon mutation. Most flexible protein amongst all was H245R with a  $\Delta\Delta S_{\text{vib}} \text{ENCoM}$  of  $-1.733 \text{ kcal.mol}^{-1} \text{ K}^{-1}$ , followed by G206C and R442Q. The most rigid structure upon point mutation was H164Y with  $\Delta\Delta S_{\text{vib}} \text{ENCoM}$  of  $-4.479 \text{ kcal.mol}^{-1} \text{ K}^{-1}$ . The visual representation of flexibility analysis depicted similar results, i.e., gain in rigidification upon mutation, shown by blue region in the Figure (Fig. 3a).

The protein dynamics investigation extended further, atomic fluctuations and deformation energies were calculated. Atomic fluctuation calculates the amplitude of the absolute atomic motion performed over first 10 non-trivial modes of the molecule; whereas deformation energy measures the amount of local flexibility present in the protein calculated over first 10 non-trivial modes of the molecule. The magnitude of the fluctuations in the protein structure calculated is shown by thin to thick tube colors in which blue shows low, white shows moderate and red shows high fluctuations. The visual changes occurring in the atomic fluctuation and deformation energies are shown in Fig. 3b and 3c.

This study also included the impact of mutation on the differences occurring in the intramolecular interactions caused by mutation in helicase protein. This estimation is based on all covalent and non-covalent interactions predicting intramolecular interactions. All the mutations in the helicase protein presented here, causes significant alterations in hydrophobic interactions, hydrogen bonds, ionic interactions, and other metal complex interactions. The substitution mutation in the wild type protein changes the side chain which ultimately results in altered intramolecular bonds (Fig. 3d). This study suggests that the mutation in proline, histidine, tyrosine, glycine, arginine and phenylalanine residues causes significant alterations in the intramolecular interactions with the neighboring molecules. From these results it can be concluded that the helicase protein mutation can not only change the flexibility or overall

dynamics of the protein but can also interrupt with its intramolecular interaction with the neighboring molecules.

### **B-cell epitope prediction**

The epitope for B-cell were predicted using helicase protein sequence and those with a threshold above 0.5 were selected. There were several epitopes predicted for this protein but one epitope falls in the region of mutated protein i.e., from 200-206 (Fig. 4a). In our analysis two mutations were found in this epitopic region, one at 205 and the other at 206. These mutations can change the antigenic property of the virus and therefore can help in host evasion, being not recognized by the host's antibody.

### **T-cell epitope prediction and identification of MHC binding allele**

We identified 5 T-cell binding epitopes showing different allele binding affinity as shown in Table 4. Moreover, MHC allele binding affinity was also predicted for the T-cell epitopes as shown in Table 5. The MHC class I immunogenicity of the helicase molecules (Table 6). The predicted epitopes can show immunogenicity and therefore produce cytokines to reduce the infection.

### **MHC cluster analysis**

The MHC class I cluster analysis of helicase is shown in Fig. 4b while that of class II allele is shown in Fig. 4c, in which the red zones indicate strong interaction of the HLA allele with the epitopes of helicase protein whereas yellow depicts weak interaction. We analysed the binding ability of all the HLA alleles with the helicase protein epitopes.

### **Assessment of antigenicity and allergenicity**

VaxiJen v2.0 server was used to predict the antigenicity of the helicase protein. This analysis indicates that the helicase protein sequence is antigenic with potent antigenicity of 0.5 with a threshold of 0.4%. A protein sequence to be used as a vaccine should be non- allergen as it should not show IgE mediated antibody response in the host system. The allergenicity of the helicase protein was predicted using Allertop tool, according to which helicase protein is non- allergen and hence can be a candidate protein for vaccine development.

## **Discussion**

Due to its ability to show high rates of point mutations, leading to enormous genetic variations, genome of coronavirus acquire considerable potential to evolve rapidly in response to changing host environments. As a consequence of high rates of mutation, SARS-CoV-2 easily evades the immunity of host and developed drug resistance properties. Helicase of SARS-CoV-2, plays an essential role in the viral RNA replication as it catalyzes the unwinding of duplex RNA structure into single stranded form so that the RNA dependent RNA polymerase (RdRp) can easily performed novel RNA synthesis. Thus, SARS-CoV-2 helicase accompanies the function of RdRp protein.

They are categorized into 6 superfamilies ranging from SF1 to SF6 and are supposed to be critically involved in the various process associated with the metabolism of nucleic acid [24]. Duplex unwinding of SARS-CoV-2 helicase appears to be an energy dependent process which utilizes the energy derived from the hydrolysis of ribonucleoside triphosphate [25,26]. Yu et al., 2012 [27] reported that SARS-CoV helicase protein showed inhibitory effects in vitro, against two potent antiviral agents such as scutellerein and myricetin by disrupting ATPase activity. Due to the presence of conserved sequences among all strains of coronaviruses, SARS-CoV-2

helicase might be used as a crucial antiviral target for therapeutic intervention [28]. Mutagenesis in the SARS-CoV-2 helicase and its structural characterization might be a potential concern in deep understanding of designing antiviral therapeutics.

The findings of the present study, emphasizes mainly on the occurrence of recurrent mutations in the helicase protein of Indian SARS-CoV-2 isolates; by comparing it with that of Wuhan SARS-CoV-2 isolate. Helicase is indispensable for viral replication along with RdRp and also involved in the process of multiplication and pathogenicity. The results of multiple sequence alignment revealed the presence of seven mutations in the SARS-CoV-2 helicase protein. Our data clearly demonstrate that H164Y, G206C and R442Q mutations significantly alters the protein secondary structure. The modeling and Ramachandran plot analysis showed the presence of major amino acids in the favored region of the quadrangle. The result of the phylogenetic tree exhibits the presence of Wuhan isolate and Indian isolates on different clusters of the tree, being evolutionarily different. Similarly, the nonsynonymous mutational study predicted Y205C, G206C and F499L mutations to be deleterious while others were found neutral. Furthermore, the protein dynamics study revealed the stabilizing mutation which alters the flexibility of the protein. These point mutations not only modify the dynamics of the protein but also interfere with its intramolecular interactions. Two mutations have been detected in the predicted epitope region of SARS-CoV-2 helicase that might induce its conformational changes and considered as major challenges in the designing of vaccine to curb SARS-CoV-2 infections.

In this study, various predictive tools of computational biology have been used for the prediction of B-cell, T-cell and MHC I epitopes. More importantly, peptide vaccines development involves the identification of immunodominant B and T-cell epitopes that are capable to induce specific immune responses [29]. Furthermore, for the epitope based peptide vaccine to be

immunogenic, its B-cell epitope of a target molecule must interact with a T-cell epitope. The T-cell epitopes often made up of 8-20 amino acids (a short stretch of peptide) and seems to be more propitious and hence generate prolong immune response mediated by CD8<sup>+</sup> T-cells [30] whereas the B-cell epitopes are consist of linear or discontinuous chain of amino acid which can be a protein [31,32,33]. Similarly, MHC class-I epitopes, exhibits interaction with the several HLA alleles and found to be antigenic. Two mutations have been detected in the predicted epitope region of SARS-CoV-2 helicase that might induce its conformational changes and considered as a major challenge in the development of vaccine

Taken together, the findings of our study suggest that the SARS-CoV-2 helicase is a prominent enzyme or necessary components for coronavirus replication and can be used as potential antiviral solution. Earlier studies have shown that the viral helicases appears as a challenging druggable targets [34,35], however, targeting different ATPase has been found to be successful recently and hence specific and potent inhibitors of helicase (nsp13) could be developed [36,37]. Interestingly, helicase protein is found to be non-allergen as well as antigenic, therefore, SARS-CoV helicase might be used as a suitable target for anti-viral drug discovery, as there are paucity of available information regarding designing of SARS-CoV-2 helicase (nsp13) inhibitors.

## **Conclusion**

The outbreak of SARS-CoV-2 pandemic poses an unprecedented threat to mankind globally. Our study reveals the emergence of seven human SARS CoV-2 helicase mutations in Indian isolates. The ongoing evolution and rapid multiplications help SARS-CoV-2 to better adapt in the host and survive in different geographical locations that warrant urgent investigation. Present work will provide new insights for developing antiviral drugs to curb this pandemic because these

mutations might have functional consequences which need to be incorporated in future research work.

## References

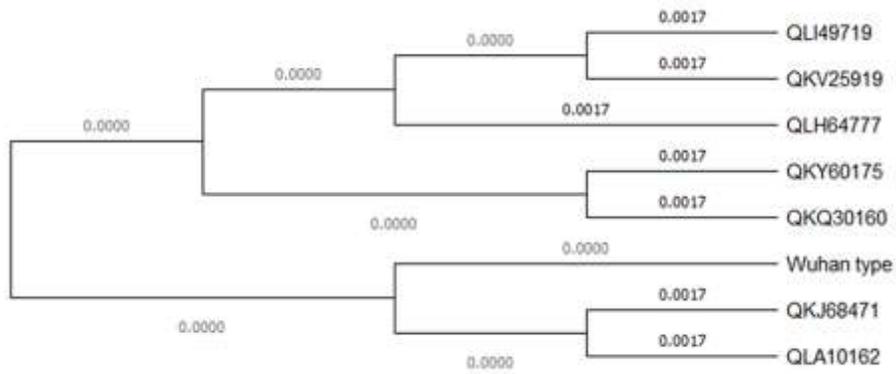
1. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395(10224):565-574. doi:10.1016/S0140-6736(20)30251-8
2. Yashvardhini N, Jha DK, Bhattacharya S. Identification and characterization of mutations in the SARS-CoV-2 RNA-dependent RNA polymerase as a promising antiviral therapeutic target. *Arch Microbiol*. 2021;203(9):5463-5473. doi:10.1007/s00203-021-02527-9
3. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-733. doi:10.1056/NEJMoa2001017
4. Sackman AM, McGee LW, Morrison AJ, et al. Mutation-Driven Parallel Evolution during Viral Adaptation. *Mol Biol Evol*. 2017;34(12):3243-3253. doi:10.1093/molbev/msx257
5. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol*. 1997;51:151-178. doi:10.1146/annurev.micro.51.1.151
6. Jha DK, Yashvardhini N, Kumar A. Identification of recurrent mutations in exonuclease (nsp14); a potential drug target in SARS-CoV-2. *Indian J Pathol Microbiol*. 2021;64(4):771-775. doi:10.4103/0377-4929.328516
7. Duffy S. Why are RNA virus mutation rates so damn high?. *PLoS Biol*. 2018;16(8):e3000003. Published 2018 Aug 13. doi:10.1371/journal.pbio.3000003

8. Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020;583(7816):459-468. doi:10.1038/s41586-020-2286-9
9. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China [published correction appears in *Nature*. 2020 Apr;580(7803):E7]. *Nature*. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3
10. Bacha U, Barrila J, Velazquez-Campoy A, Leavitt SA, Freire E. Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry*. 2004;43(17):4906-4912. doi:10.1021/bi0361766
11. Gao Y, Yan L, Huang Y, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. 2020;368(6492):779-782. doi:10.1126/science.abb7498
12. Furuta Y, Gowen BB, Takahashi K, Shiraki K, Smee DF, Barnard DL. Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Res*. 2013;100(2):446-454. doi:10.1016/j.antiviral.2013.09.015
13. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641. doi:10.1093/nar/gkz268
14. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745-2747. doi:10.1093/bioinformatics/btv195
15. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35(6):1547-1549. doi:10.1093/molbev/msy096

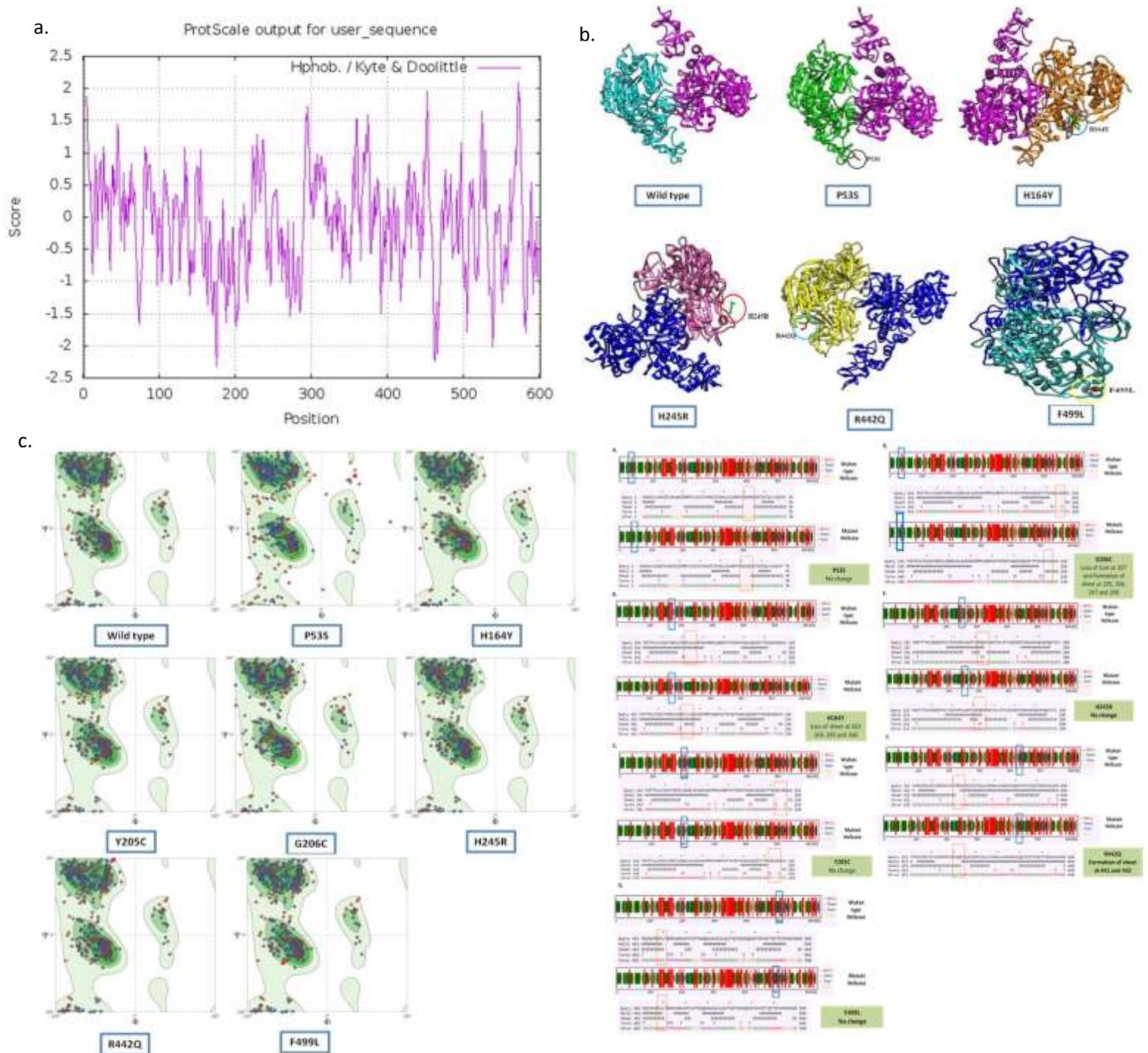
16. Ashok, KT. CFSSP: Chou and Fasman secondary structure prediction server, Wide. Spec. 2013; 1:15–19 DOI 10.5281/zenodo.50733
17. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M R, Appel RD, Balroch A. Protein Identification and Analysis Tools on the ExPASy Server. Prot. Proto. Hand. 2005 571-607
18. Goddard TD, Huang CC, Ferrin TE. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. Structure. 2005;13(3):473-482. doi:10.1016/j.str.2005.01.006
19. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res. 2018;46(W1):W350-W355. doi:10.1093/nar/gky300
20. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2:2. Published 2006 Apr 24. doi:10.1186/1745-7580-2-2
21. Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M. MHCcluster, a method for functional clustering of MHC molecules. Immunogenetics. 2013;65(9):655-665. doi:10.1007/s00251-013-0714-9
22. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8:4. Published 2007 Jan 5. doi:10.1186/1471-2105-8-4
23. Dimitrov I, Flower DR, Doytchinova I. AllerTOP--a server for in silico prediction of allergens. BMC Bioinformatics. 2013;14 Suppl 6(Suppl 6):S4. doi:10.1186/1471-2105-14-S6-S4

24. Jankowsky E. RNA helicases at work: binding and rearranging. *Trends Biochem Sci.* 2011;36(1):19-29. doi:10.1016/j.tibs.2010.07.008
25. Seybert A, Hegyi A, Siddell SG, Ziebuhr J. The human coronavirus 229E superfamily 1 helicase has RNA and DNA duplex-unwinding activities with 5'-to-3' polarity. *RNA.* 2000;6(7):1056-1068. doi:10.1017/s1355838200000728
26. Seybert A, Ziebuhr J. Guanosine triphosphatase activity of the human coronavirus helicase. *Adv Exp Med Biol.* 2001;494:255-260. doi:10.1007/978-1-4615-1325-4\_40
27. Yu MS, Lee J, Lee JM, et al. Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsp13. *Bioorg Med Chem Lett.* 2012;22(12):4049-4054. doi:10.1016/j.bmcl.2012.04.081
28. Jia Z, Yan L, Ren Z, et al. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* 2019;47(12):6538-6550. doi:10.1093/nar/gkz409
29. Yashvardhini N, Kumar A, Jha DK. Immunoinformatics Identification of B- and T-Cell Epitopes in the RNA-Dependent RNA Polymerase of SARS-CoV-2. *Can J Infect Dis Med Microbiol.* 2021;2021:6627141. Published 2021 Apr 20. doi:10.1155/2021/6627141
30. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet.* 2020;395(10223):514-523. doi:10.1016/S0140-6736(20)30154-9
31. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.* 2020;395(10223):507-513. doi:10.1016/S0140-6736(20)30211-7

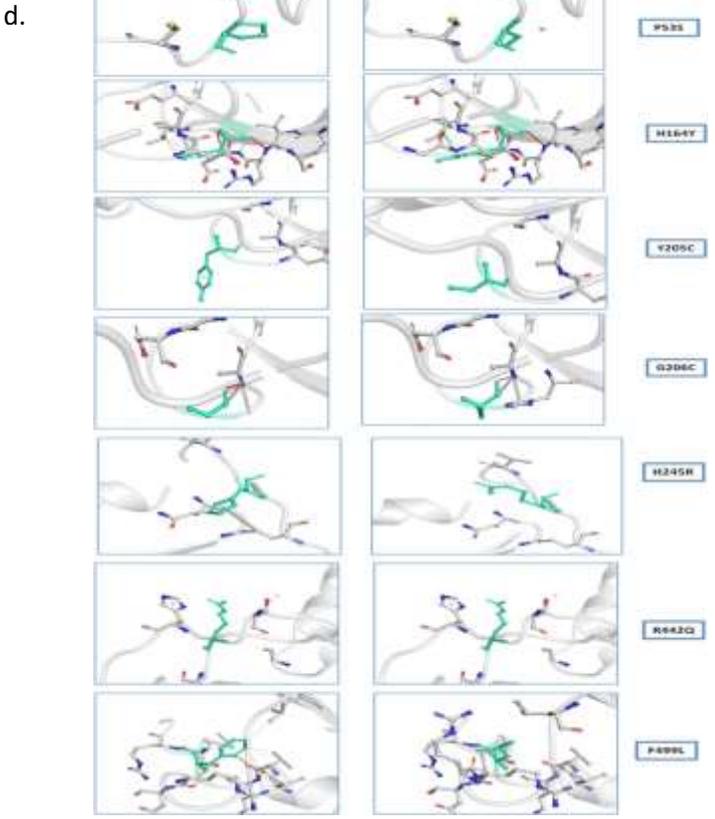
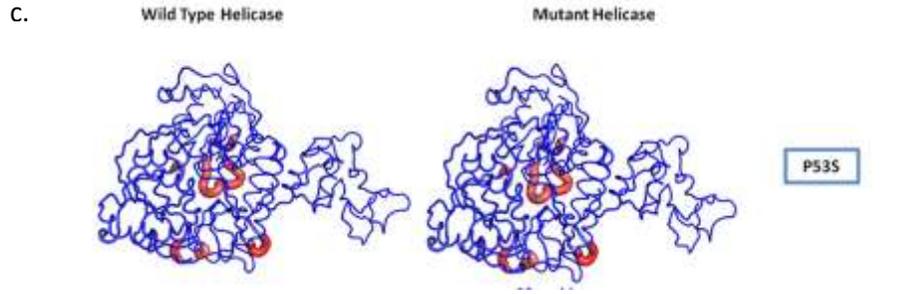
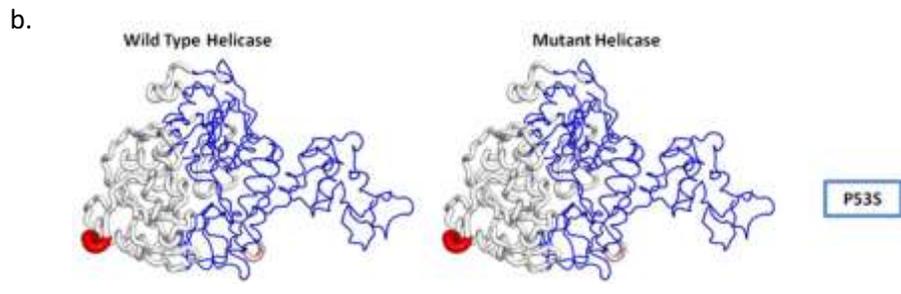
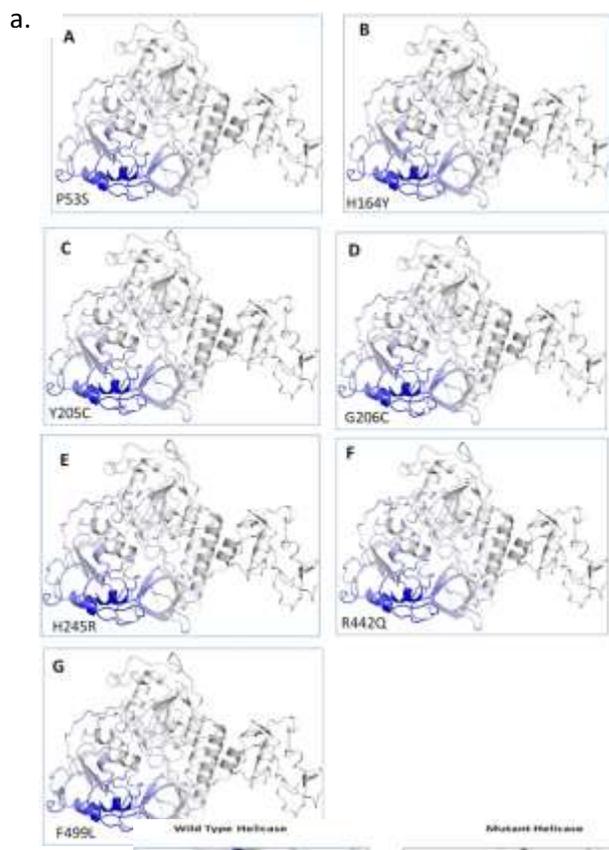
32. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med.* 2020;382(13):1199-1207. doi:10.1056/NEJMoa2001316
33. Yashvardhini N, Kumar A, Jha DK. Analysis of SARS-CoV-2 mutations in the main viral protease (NSP5) and its implications on the vaccine designing strategies. *Vacunas.* 2021c. DOI: [10.1016/j.vacun.2021.10.002](https://doi.org/10.1016/j.vacun.2021.10.002)
34. Shadrack WR, Ndjomou J, Kolli R, Mukherjee S, Hanson AM, Frick DN. Discovering new medicines targeting helicases: challenges and recent progress. *J Biomol Screen.* 2013;18(7):761-781. doi:10.1177/1087057113482586
35. Datta A, Brosh RM Jr. New Insights Into DNA Helicases as Druggable Targets for Cancer Therapy. *Front Mol Biosci.* 2018;5:59. Published 2018 Jun 26. doi:10.3389/fmolb.2018.00059
36. Kleymann G, Fischer R, Betz UA, et al. New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. *Nat Med.* 2002;8(4):392-398. doi:10.1038/nm0402-392
37. Crute JJ, Grygon CA, Hargrave KD, et al. Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease. *Nat Med.* 2002;8(4):386-391. doi:10.1038/nm0402-386



**Fig. 1** Phylogenetic tree of Indian SARS-CoV-2 and Wuhan SARS-CoV-2 isolates with reference to helicase protein.



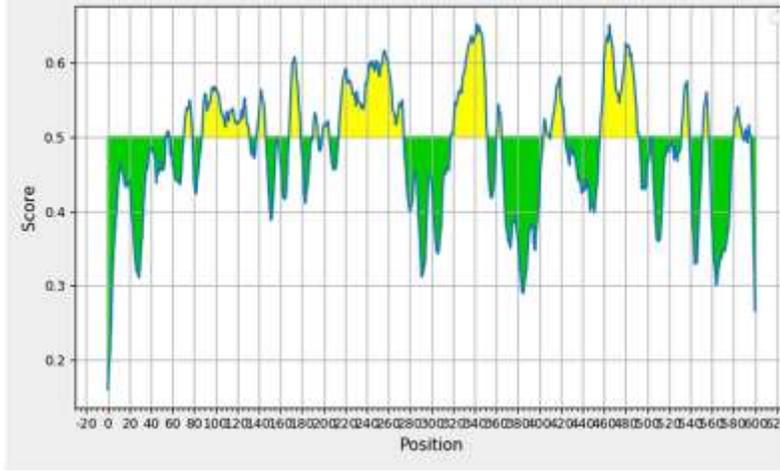
**Fig. 2a** Hydropathy plot of wild type helicase protein showing hydrophobic amino acid residues. **Fig. 2b** Protein modeling of wild type and mutant helicase protein. Models were prepared by Chimera software. **Fig. 2c** Ramachandran plot analysis of template and mutated protein, showing 89% of amino acid residues in favored region, rest in outlier region. **Fig. 2d** Secondary structure prediction of helicase protein. Effect of mutation at different sites on the secondary structure of helicase protein, (A-G) represent seven mutations observed in Indian isolates. The first secondary structure in each (A-G) represents the Wuhan type sequence while the second represents the mutated one. The mutation location and respective secondary structures are marked with boxes.



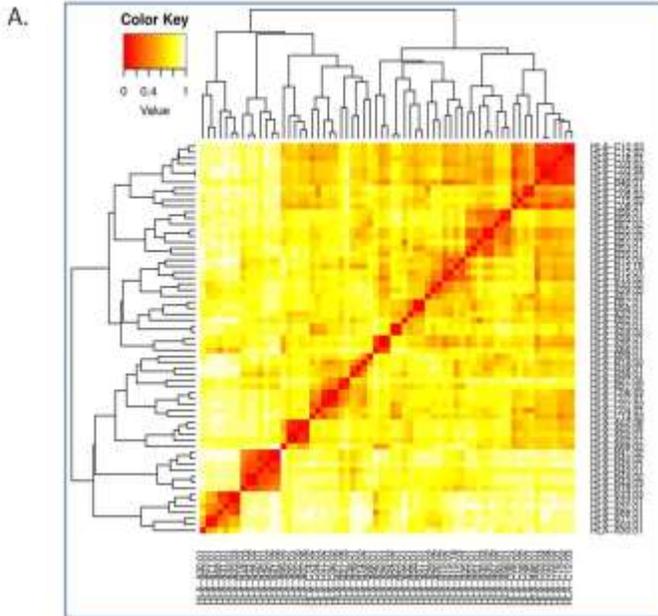
**Fig. 3a** Mutational effect on structural dynamics of helicase protein. Vibrational entropy energy between wild type and mutant helicase in which amino acids are colored according to vibrational entropy change in mutant with reference to wild type. Blue represents rigidification whereas red represents gain in flexibility upon mutation. **Fig. 3b** Visual representation of Atomic Fluctuations and **Fig. 3c**. Deformation Energies. The atomic fluctuation of both wild type and mutant helicase proteins is shown. The magnitude of fluctuation is represented by thin to thick tube colored blue (low), white (moderate) and red (high). The changes in fluctuation of mutant with reference to wild type are marked by red, green and yellow arrows. **Fig. 3d** Effect of point mutation on interatomic interactions of helicase protein. Interatomic interactions were altered by mutations at locus P53S, H164Y, Y205C, G206C, H245R and F499L as shown in figure. Wild type amino acid residues are colored in light green and represented as stick with the surrounding residues where any interactions exist.

and F499L as shown in figure. Wild type amino acid residues are colored in light green and represented as stick with the surrounding residues where any interactions exist.

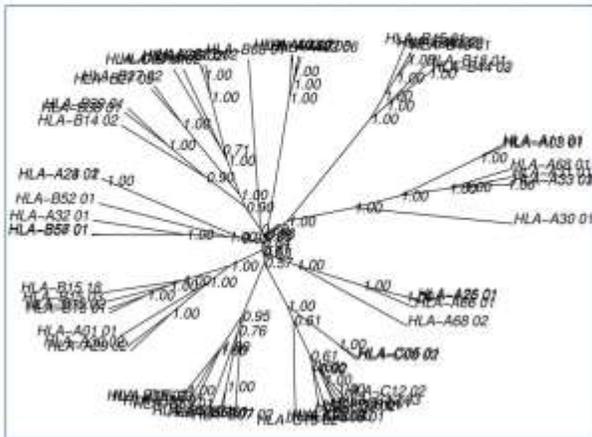
a.



b.

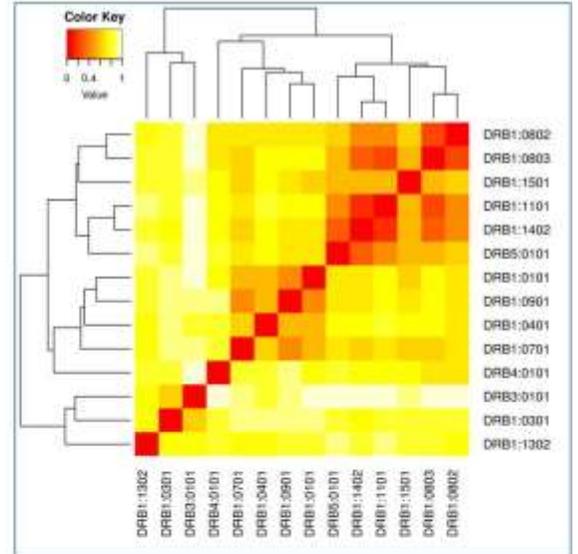


B.

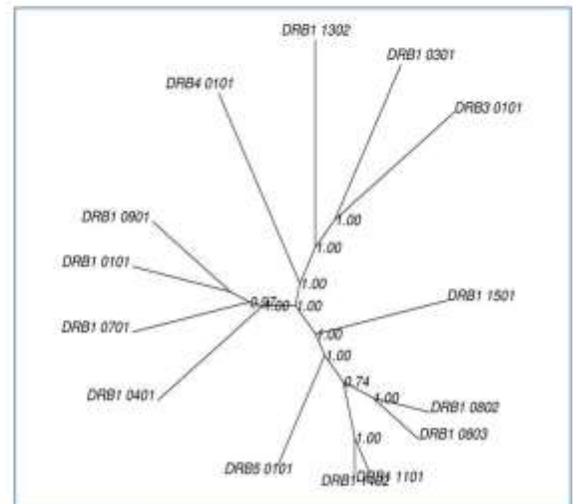


c.

A.



B.



**Fig. 4a** B-cell epitope prediction of helix protein. The threshold cutoff is 0.5 above which the residues are epitopes. Out of seven mutations detected in this study, two are predicted to be epitopes. **Fig. 4b** The results of MHC cluster analysis. **A.** tree map of MHC class I cluster, **B.** heat map of MHC class I cluster, **Fig. 4c** The results of MHC cluster analysis. **A.** tree map of MHC class II cluster **B.** heat map of MHC class II cluster.

Table 1 List of 7 nonsynonymous amino acid substitutions in helicase protein

<b>Amino acid substitution in the helicase region</b>	<b>PROVEAN score</b>	<b>Variation effect on protein</b>
P53S	-1.633	Neutral
H164Y	1.312	Neutral
Y205C	-3.256	Deleterious
G206C	-3.754	Deleterious
H245R	1.779	Neutral
R442Q	0.467	Neutral
F499L	-5.883	Deleterious

Table 2 Physicochemical properties of helicase protein (wild type)

<b>Physicochemical properties</b>	<b>Helicase</b>	<b>Amino acid composition</b>	<b>No.</b>	<b>Percent composition (%)</b>
Molecular weight	66854.75	Ala (A)	46	7.7
No. of amino acids	600	Arg (R)	30	5.0
Theoretical pI	8.66	Asn (N)	30	5
Instability index	33.31	Asp (D)	29	4.8
No. of negatively charged (Asp+ Glu)	52	Cys (C)	26	4.3
No. of positively charged (Arg+ Lys)	64	Gln (Q)	18	3.0
aliphatic index	84.49	Glu (E)	23	3.8
Grand average of hydropathicity	-0.096	Gly (G)	32	5.3
Estimated half-life (mammalian reticulocytes, in vitro)	4.4 hours	His (H)	12	2.0
Atomic composition		Ile (I)	29	4.8
C	2981	Leu (L)	50	8.3
H	4670	Lys (K)	34	5.7
N	800	Met (M)	8	1.3
O	878	Phe (F)	24	4.0
S	34	Pro (P)	30	5.0
Formula	C <sub>2981</sub> H <sub>4670</sub> N <sub>800</sub> O <sub>878</sub> S <sub>34</sub>	Ser (S)	40	6.7
Total number of atoms	9363	Thr (T)	50	8.3
		Trp (Y)	34	5.7
		Val (V)	53	8.8
		Phy (O)	0	0.0
		Sec (U)	0	0.0

Table 3 Effect of mutation variability on the structural dynamics of helicase protein as shown by  $\Delta\Delta S$  ENCoM and  $\Delta\Delta G$  values.

S. No.	Wuhan Isolate	Indian Isolates	Amino acid position	$\Delta\Delta G$ Dynamut	$\Delta\Delta S$ ENCoM	$\Delta\Delta G$ ENCoM
1.	P	SC	53	1.485 kcal/mol	-2.089 kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.671kcal/mol
2.	H	Y	164	1.945 kcal/mol	-4.479 kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.877kcal/mol
3.	Y	C	205	1.497 kcal/mol	-2.160kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.728kcal/mol
4.	G	C	206	1.192 kcal/mol	-1.927kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.541kcal/mol
5.	H	R	245	0.937 kcal/mol	-1.733kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.387kcal/mol
6.	R	Q	442	1.314 kcal/mol	-2.058kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.646kcal/mol
7.	F	L	499	1.864 kcal/mol	-2.176kcal.mol <sup>-1</sup> K <sup>-1</sup>	1.741kcal/mol

Table 4 T-cell epitope prediction of SARS- CoV-2 helicase protein sequence from India and its allergenicity

<b>peptide</b>	<b>Start position</b>	<b>score</b>	<b>Allergenicity</b>
STSHKLVLS	36	1.00	Non-allergen
TTYKLNVDG	215	1.00	Non-allergen
VYTACSHAA	305	1.00	Allergen
VYDNKLKAH	456	1.00	Allergen
<b>RKAVFISPY</b>	507	1.00	Non-allergen

Table 5 T-cell epitope prediction of SARS- CoV-2 helicase protein sequence from India and its MHC restriction

MHC Restriction of CTL Epitope	
STSHKLVLS	HLA-A*0206
STSHKLVLS	HLA-Cw*0401
STSHKLVLS	H2-Db
STSHKLVLS	H2-Dd
STSHKLVLS	H2-Kb
STSHKLVLS	H2-Kd
STSHKLVLS	H2-Ld
STSHKLVLS	HLA-G
STSHKLVLS	H-2Qa
STSHKLVLS	Mamu-A*01
TTYKLVGD	HLA-Cw*0401
TTYKLVGD	H2-Db
TTYKLVGD	H2-Dd
TTYKLVGD	H2-Kb
TTYKLVGD	H2-Kd
TTYKLVGD	H2-Ld
TTYKLVGD	HLA-G
TTYKLVGD	H-2Qa
TTYKLVGD	Mamu-A*01
VYTACSHAA	HLA-A24
VYTACSHAA	HLA-A*2402
VYTACSHAA	HLA-Cw*0401
VYTACSHAA	H2-Db
VYTACSHAA	H2-Dd
VYTACSHAA	H2-Kb
VYTACSHAA	H2-Kd
VYTACSHAA	H2-Ld
VYTACSHAA	HLA-G
VYTACSHAA	H-2Qa
VYDNKLKAH	HLA-A24
VYDNKLKAH	HLA-A*2402
VYDNKLKAH	HLA-Cw*0401
VYDNKLKAH	H2-Db
VYDNKLKAH	H2-Dd
VYDNKLKAH	H2-Kb
VYDNKLKAH	H2-Kd
VYDNKLKAH	H2-Ld
VYDNKLKAH	HLA-G
VYDNKLKAH	H-2Qa
RKAVFISPY	HLA-A3
RKAVFISPY	HLA-Cw*0401
RKAVFISPY	H2-Db
RKAVFISPY	H2-Dd
RKAVFISPY	H2-Kb
RKAVFISPY	H2-Kd
RKAVFISPY	H2-Ld
RKAVFISPY	HLA-G
RKAVFISPY	H-2Qa
RKAVFISPY	Mamu-A*01
RKAVFISPY	HLA-B*2902
RKAVFISPY	HLA-A*3301

---

RKAVFISPY

HLA-A\*6801

---

Table 6 Showing Class I immunogenicity of helicase protein of SARS-CoV-2

peptide	length	score
PPISFPLCANGQVFGLYKNTCVGSDNVTDNFNAIATCDWTNAGDYILANTCTERLKLFAAE	60	0.75021
EYTFEKGDYGDVAVYRGTTTYKLNVDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTL	60	0.45794
LCEKALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTVNALPETTADIVVFDEI	60	0.42682
NVNRFNVAITRAKVGILCIMS DRDLYDKLQFTSLEIPRRNVATLQ	45	0.38009
TLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFTGYRVTKNSKVQIG	60	0.28073
SMATNYDLSVFNARLRAKHVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDM	60	0.02156
REFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFTQTTETAHSC	60	-0.11189
AVGACVLCNSQTSLRC	16	-0.40248
FLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVITHDVSSAINRPQIGVV	60	-0.77844
NISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYPSARIVYTACSHAAVDA	60	-0.87495
GACIRRPFLCCKCCYDHVISTSHKLVLVSNPYVCNAPGCDVTDVTQLYLGMSYYCKSHK	60	-1.0761