Vacunas

www.elsevier.es/vac

Original article

Analysis of SARS-CoV-2 mutations in the main viral protease (NSP5) and its implications on the vaccine designing strategies

Niti Yashvardhini[a], Amit Kumar[b], Deepak Kumar Jha[c,*,1]

[a] Department of Microbiology, Patna Women's College, Patna 800 001, India
[b] Department of Botany, Patna University, Patna 800 005, India
[c] Department of Zoology, P. C. Vigyan Mahavidyalaya, J. P. University, Chapra 841 301, India

ARTICLE INFO

ABSTRACT

SARS-CoV-2 (Severe Acute Respiratory Syndrome), an etiolating agent of novel COVID-19 (coronavirus 2019) pandemic, rapidly spread worldwide, creating an unprecedented public health crisis globally. NSP5, the main viral protease, is a highly conserved protein, encoded by the genome of SARS-CoV-2 and plays an important role in the viral replication cycle. In the present study, we detected a total of 33 mutations from 675 sequences submitted from India in the month of March 2020 to April 2021. Out of 33 mutations, we selected 8 frequent mutations (K236R, N142L, K90R, A7V, L75F, C22N, H246Y and I43V) for further analysis. Subsequently, protein models were constructed, revealing significant alterations in the 3-D structure of NSP5 protein when compared to the wild type protein sequence which also altered the secondary structure of NSP5 protein. Further, we identified 9 B-cell, 10 T-cell and 6 MHC-I promising epitopes using predictive tools of immunoinformatics, out of these epitopes some were non-allergenic as well as highly immunogenic. Results of our study, however, revealed that 10 B-cell epitopes reside in the mutated region of NSP5. Additionally, hydrophobicity, physiochemical properties, toxicity and stability of NSP5 protein were estimated to demonstrate the specificity of the multiepitope candidates. Taken together, variations arising as a consequence of multiple mutations may cause alterations in the structure and function of NSP5 which generate crucial insights to better understand structural aspects of SARS-CoV-2. Our study also revealed, NSP5, a main protease, can be a potentially good target for the design and development of vaccine candidate against SARS-CoV-2.

© 2021 Elsevier España, S.L.U. All rights reserved.

Análisis de las mutaciones del SARS-CoV-2 en la principal proteasa viral (NSP5) y sus implicaciones en las estrategias de diseño de vacunas

RESUMEN

Palabras clave:
SARS-CoV-2
Pandemia
Epítopos
NSP5
Mutación
Vacuna

El SARS-CoV-2 (Síndrome Respiratorio Agudo Severo), un agente etiológico de la nueva pandemia de COVID-19 (coronavirus 2019), se propagó rápidamente por todo el mundo y creó una crisis de salud pública sin precedentes a nivel mundial. El NSP5, la proteasa viral principal, es una proteína altamente conservada, codificada por el genoma del SARS-CoV-2 y juega un papel importante en el ciclo de replicación viral. En el presente estudio se detectaron un total de 33 mutaciones de 675 secuencias presentadas desde la India en el mes de marzo de 2020 a abril de 2021. De 33 mutaciones, se seleccionaron 8 mutaciones frecuentes (K236R, N142L, K90R, A7V, L75F, C22N, H246Y e I43V) para su posterior análisis. Posteriormente, se construyeron modelos proteicos que revelaron alteraciones significativas en la estructura 3D de las proteínas NSP5 en comparación con la secuencia de proteínas de tipo silvestre que también alteraron la estructura secundaria de la proteína NSP5. Además, se identificaron 9 epítopos prometedores de células B, 10 de células T y 6 de MHC-I, utilizando herramientas predictivas de inmunoinformática, algunos no alergénicos y altamente inmunogénicos. Los resultados de nuestro estudio, sin embargo, revelaron que 10 epítopos de células B residen en la región mutada de NSP5. Adicionalmente, se estimó la hidrofobicidad, propiedades fisicoquímicas, toxicidad y estabilidad de la proteína NSP5 para demostrar la especificidad de los candidatos multiepítopos. En conjunto, las variaciones que surgen como consecuencia de múltiples mutaciones pueden causar alteraciones en la estructura y función del NSP5 que generan conocimientos cruciales para entender mejor los aspectos estructurales del SARS-CoV-2. Nuestro estudio también reveló que el NSP5, una proteasa principal, puede ser un blanco potencialmente bueno para el diseño y desarrollo de la vacuna candidata contra el SARS-CoV-2.

Introduction

The rapid emergence of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), responsible for causing the ongoing pandemic of novel coronavirus disease 2019 (COVID-19), induces moderate to severe respiratory distress (such as cough, cold, dyspnea) in humans around the world.[1] The novel COVID-19 has been reported from the wildlife market in Wuhan city of Hubei province (China), in late December 2019.[2] SARS-CoV-2 has now affected 218 countries, posing devastating public health threat across the globe. As of May 21, 2021, almost after 18 months of the outbreak of this pandemic worldwide over 107,838,255 confirmed cases of COVID-19 have been reported to WHO including 2,373,398 casualties (WHO COVID-19 Dashboard).[3] As many as 60 different vaccines against coronavirus have reached various stages of clinical development and many of them have been approved for immunization purposes nowadays. Vaccinating vulnerable population to achieve herd immunity against SARS-CoV-2 infection is of great importance, however, due to the emergence of new variants of this virus it is very difficult to assess how long the available vaccine will remain durable and effective.[4,5]

Coronavirus (CoVs) is an enveloped, single-stranded, positive sense RNA virus of ~30 kb length .[6] The genome of SARS-CoV-2 encodes four types of structural (spike S, envelope E, membrane M and nucleocapsid N) and various conserved non-structural proteins ranging from NSP1to NSP16 including nine accessory proteins.[7,8] ORF1ab encodes non-structural proteins of SARS-CoV-2 which is crucial for the viral life cycle and pathogenesis. NSP5 (main viral protease, M^{pro}) has been found synonymous with 3C-like protease (3CL^{pro}), that mediates cleavage at 11 different sites of polyproteins to generate other non-structural proteins and also plays significant role in the viral replication cycle.[9,10] Due to its essential and conserved role in viral development, NSP5 is considered as promising antiviral therapeutic target against SARS-CoV-2 infections. NSP5 of coronavirus is a ~30 KDa protein, possessing structurally conserved three domain cysteine protease and acts as a main protease for proteolytic processing of viral replicase polyproteins such as pp1a and pp1b.[11–14] Interestingly, the yields of NSPs proteins gets affected by the inhibition of the NSP5-mediated cleavage and hence, the viral replication can also be prevented. Due to this reason, since the advent of this pandemic, several studies have been performed to identify various compounds, capable to antagonize the activity of NSP5 and also help in better understanding of the molecular mechanism behind the inhibition.[15,16]

The genome of SARS-CoV-2 is rapidly evolving by acquiring multiple mutations. As it is quite evident from numerous previous studies, NSP5 of coronavirus plays crucial role in the viral infection and pathogenesis. Present *in silico* study was, therefore, carried out to detect and characterize mutations of NSP5 of SARS-CoV-2. We identified a total of 33 mutations

from 675 sequences submitted from India in the month of March 2020 to April 2021 and compared with the first reported sequence from Wuhan, as a reference sequence. Subsequently, the impact of mutation on the secondary structure and protein dynamics was observed that help in designing therapeutics and/or vaccine to curb SARS-CoV-2 infections.

In addition to this, NSP5 of SARS-CoV-2 was explored to determine the potent antigenic epitopes of B-cell and T-cell with their MHC alleles to predict multiepitope vaccine (MEV) construct. Owing to their specificity, stability, less time-consuming and cost-effective properties as well as the ability to induce significant humoral and cellular immune responses, MEVs are found to be advantageous over single epitope or conventional vaccine development approach.[17] Further, several predictive tools of immunoinformatics were utilized to validate the non-allergenic, non-toxic, antigenicity, toxicity, structural stability/flexibility and physiochemical properties and hydrophobicity of the designed multi-epitopes vaccine candidate.

Materials and methods

Data mining

The full length protein sequence of ORF1ab polyprotein, 7096 amino acid long which encodes for non-structural proteins in SARS-CoV-2 were retrieved from NCBI virus database. NCBI virus database keeps a deposit of all SARS-CoV-2 sequences submitted from different parts of the world. As on April 29, 2021, 675 full length ORF1ab amino acid sequences were submitted from India which was used in this study. The first reported ORF1ab protein sequence with Accession number YP_009724389 was also downloaded to be used as a reference or wild type sequence in this study. From the full length ORF1ab polyprotein sequence, the sequence of NSP5 (SARS-CoV-2 protease) was procured being 306 amino acid long.

Identification of protease mutants from India

To detect the variations in the protease protein amino acid sequences, the NSP5 protein sequences from India were aligned with the first reported SARS-CoV-2 sequence from Wuhan. To align these polypeptides, Clustal Omega online platform[18] was used which creates 1000 of alignments based on HMM profile seeded guide trees. These alignments were viewed on Jalview to detect the variations occurring in the protease protein with reference to Wuhan type protease sequence. The non-synonymous amino acid variants were analyzed using Protein Variation Effect Analyzer known as PROVEAN v1.1.3 with cutoff predicted score of –2.50 to detect the effect of mutation on the NSP5 protein.[19] PROVEAN predicts the effect of amino acid substitution on the overall function of aa protein. A score namely delta alignment is calculated which are the PROVEAN scores of the substituted protein. The threshold limit for this score being –2.5 below or equal to which the mutation is deleterious and above this threshold limit the variation has neutral effect.

Calculation of physicochemical properties and hydropathy index of protease protein

Physicochemical properties of any protein includes its molecular weight, aliphatic index, composition of different amino acids including positively and negatively charged, atomic composition, estimated half life, instability index, hydrophobicity (GRAVY score) and other parameters. These parameters were calculated using Protparam tool of Expasy online platform. The hydropathy plot was prepared using Protscale tool, an expasy program.[20]

Secondary structure prediction

The secondary structure of the NSP5 protein was predicted using CFSSP (Chou and Fasman Secondary Structure Prediction) online software.[21] The analysis was done for both wild type and mutated protein sequences to study the alteration in the secondary structure of the protein such as changes in helix, turn and sheet formation due to mutation.

NSP5 protein dynamics study

Phyre2 online modeling platform was used to build the models of wild type and mutated NSP5 proteins.[22] Dynamut software was applied to detect the impact of mutation on the structure flexibility and dynamicity of NSP5 protein.[23] Dynamut computes information on the stability, NMA analysis, flexibility, rigidness, conformation of mutated as well as wild type protein. Several parameters were calculated like flexibility analysis, vibrational entropy, atomic and deformation energies using first 10 non-trivial modes of the structure. To check whether upon variation intramolecular interactions can change, Dynamut was used to predict the effect of mutation on intramolecular interactions.

Identification of lineal B-cell epitopes

IEDB was used to predict the lineal B-cell epitopes in the NSP5 protein of SARS-CoV-2.[24] IEDB webserver constructs epitopes based on estimation of parameters like flexibility, accessibility, hydrophilicity, turns, polarity and antigenic propensity of the protein using amino acid scales and HMMs.

MHC class I allele identification

The T-cell epitope binding alongwith the detection of MHC allele showing highest affinity for the T-cell epitope was predicted using IEDB Tepitool server.[24] This platform provides information on the binding of HLA allele with both type I and type II MHC molecules.

Antigenicity and allergenicity evaluation

To identify the antigenicity of the NSP5 protein, Vaxijen v2.0 server which predicts antigens according to the auto cross-covariance (ACC) transformation of the protein sequences was used.[25] The prediction of vaccine allergenicity was done using AllerTOP server, which evaluates protein allergenicity on auto cross variance (ACC method) that explains residues hydrophobicity, size, flexibility and other parameters.[26]

Results

Identification of mutation in protease of SARS-CoV-2 and detection of non synonymous mutants

Altogether 675 full length sequences of ORF1ab were submitted from India from March 2020 to April 2021. These 675 sequences were downloaded alongwith a reference sequence of Wuhan type virus from NCBI virus database (Supplementary table 1). The multiple sequence alignment was performed for all these ORF1ab sequences with reference to Wuhan type virus and the alignment file was viewed using Jalview. Those mutations which occurred in NSP5 were recorded and used for further analysis. A total of 33 point mutations were detected in this 306 amino acid long NSP5 protein of Indian isolates (Supplementary table 2). Amongst these point mutations, K236R, N142L, K90R, A7V, L75F, C22N, H246Y and I43V were the most frequently occurring mutations and hence used for further characterization in this study (Supplementary Fig. 1).

The three non-synonymous amino acid substitutions (N142, L75F and C22N) amongst the eight showed deleterious impact on the structure and function of NSP5 protein. All other five mutants showed neutral impact on the protein at −2.5 cutoff values of PROVEAN score (Supplementary table 3).

Estimation of physicochemical properties and hydropathy index of SARS-CoV-2 NSP5 protein

The physicochemical properties of SARS-CoV-2 protease protein were estimated using Protparam (ExPasy). The analysis revealed that the NSP5 protein is 306 amino acids in length with a molecular weight of 33,796.64 Da, instability index 27.65, aliphatic index 82.12 and GRAVY score of −0.019 (Table 1). The hydropathy plot showed C-terminal amino acid
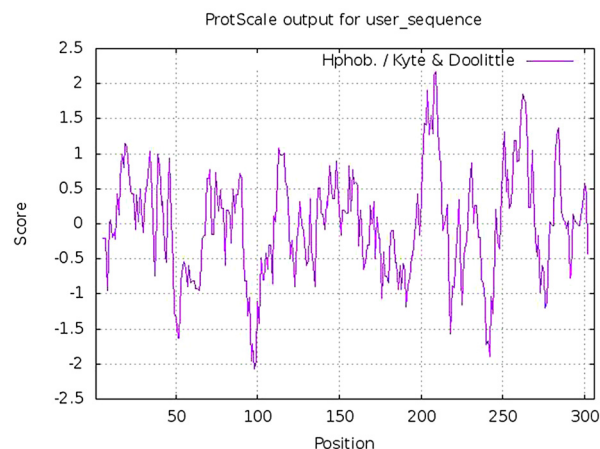

Fig. 1 – Hydropathy plot of wild type protease protein showing hydrophobic amino acid residues.

to be more hydrophobic as compared to the N-terminal end of NSP5 protein (Fig. 1).

Changes in secondary structure of NSP5 protein upon mutation

To detect the alteration in formation and loss of alpha helix, beta sheet and turns upon mutation in NSP5 protein secondary structure prediction was done using CFSSP online program with respect to wild type protein. The mutations K236R, N142L, K90R, A7V, C22N and H246Y showed significant secondary structural changes (Fig. 2a) and hence their effect was studied. The point mutation at position 236, where lysine is replaced by arginine in the NSP5 protein resulted in loss of helix structure at positions 235. Our analysis showed that the mutation at 142, where asparagine

Table 1 – Physicochemical properties of NSP5 protein (wild type).				
Physicochemical properties	Protease	Amino acid composition	No.	Percent composition (%)
Molecular weight	33,796.64	Ala (A)	17	5.6
No. of amino acids	306	Arg (R)	11	3.6
Theoretical pI	5.95	Asn (N)	21	6.9
Instability index	27.65	Asp (D)	17	65.
No. of negatively charged (Asp+ Glu)	26	Cys (C)	12	3.9
No. of positively charged (Arg + Lys)	22	Gln (Q)	14	4.6
aliphatic index	82.12	Glu (E)	9	2.9
Grand average of hydropathicity	−0.019	Gly (G)	26	8.5
Estimated half-life (mammalian reticulocytes, *in vitro*)	1.9 h	His (H)	7	2.3
Atomic composition		Ile (I)	11	3.6
C	1499	Leu (L)	29	9.5
H	2318	Lys (K)	11	3.6
N	402	Met (M)	10	3.3
O	445	Phe (F)	17	5.6
S	22	Pro (P)	13	4.2
Formula	$C_{1499}H_{2318}N_{402}O_{445}S_{22}$	Ser (S)	16	5.2
Total number of atoms	4686	Thr (T)	24	7.8
		Trp (W)	3	1.0
		Tyr (Y)	11	3.6
		Val (V)	27	8.8
		Phy (O)	0	0.0
		Sec (U)	0	0.0

Fig. 2 – (a) Secondary structure prediction of NSP5 protein. Effect of mutation at different sites on the secondary structure of protease protein (A–H). The first secondary structure in each (A–F) represents the Wuhan type sequence while the second represents the mutated one. The mutation location and respective secondary structures are marked with boxes. (b) Mutational effect on structural dynamics of protease protein. Blue represents rigidification, whereas red represents gain in flexibility upon mutation. (c) Effect of point mutation on interatomic interactions of NSP5 protein. Interatomic interactions were altered by mutations at different locations. Wild type amino acid residues are colored in light green and represented as stick with the surrounding residues where any interactions exist. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is replaced by leucine resulted in formation of helix and sheet at position 141 and loss of turn at 143. Asparagine being a polar uncharged amino acid favors formation of turn, whereas leucine being a non polar amino acid forms helix. Further, the substitution of lysine by arginine at position 90 resulted in loss of helix and sheet at points 91, 92 and 93. The A7V mutant resulted in formation of sheet at positions 3, 4, 5, 6 and 7 as valine has larger non-polar group compared to alanine and hence more tendency to form sheets. C22N mutant showed formation of turn at point 22, as asparagines favors turn formation. The substitution of histidine by tyrosine at 246 position resulted in loss of helix at 242 and 243 positions. Tyrosine being an aromatic amino acid has more tendencies to form sheets rather than helix. Overall, the secondary structure analysis depicts significant changes in the formation and loss of helix, sheet and turn that can bring huge

impact on NSP5 protein and hence leading to the SARS-CoV-2 multiplication and infection.

Protein modeling and structure prediction

The 3D model of NSP5 protein was built using Phyre2 online modeling software, which performs modeling on the basis of template search. The template being used for protease protein was d2duca1 with 100% similarity coverage. The models of both wild type and mutated NSP5 protein sequences are shown in Supplementary Fig. 2.

NSP5 protein flexibility and stability change upon mutation

The impact of mutation on the dynamics of protease protein was estimated using Dynamut software.[23] Dynamut software

K236R

N142L

K90R

A7V

L75F

C22N

H246Y

I43V
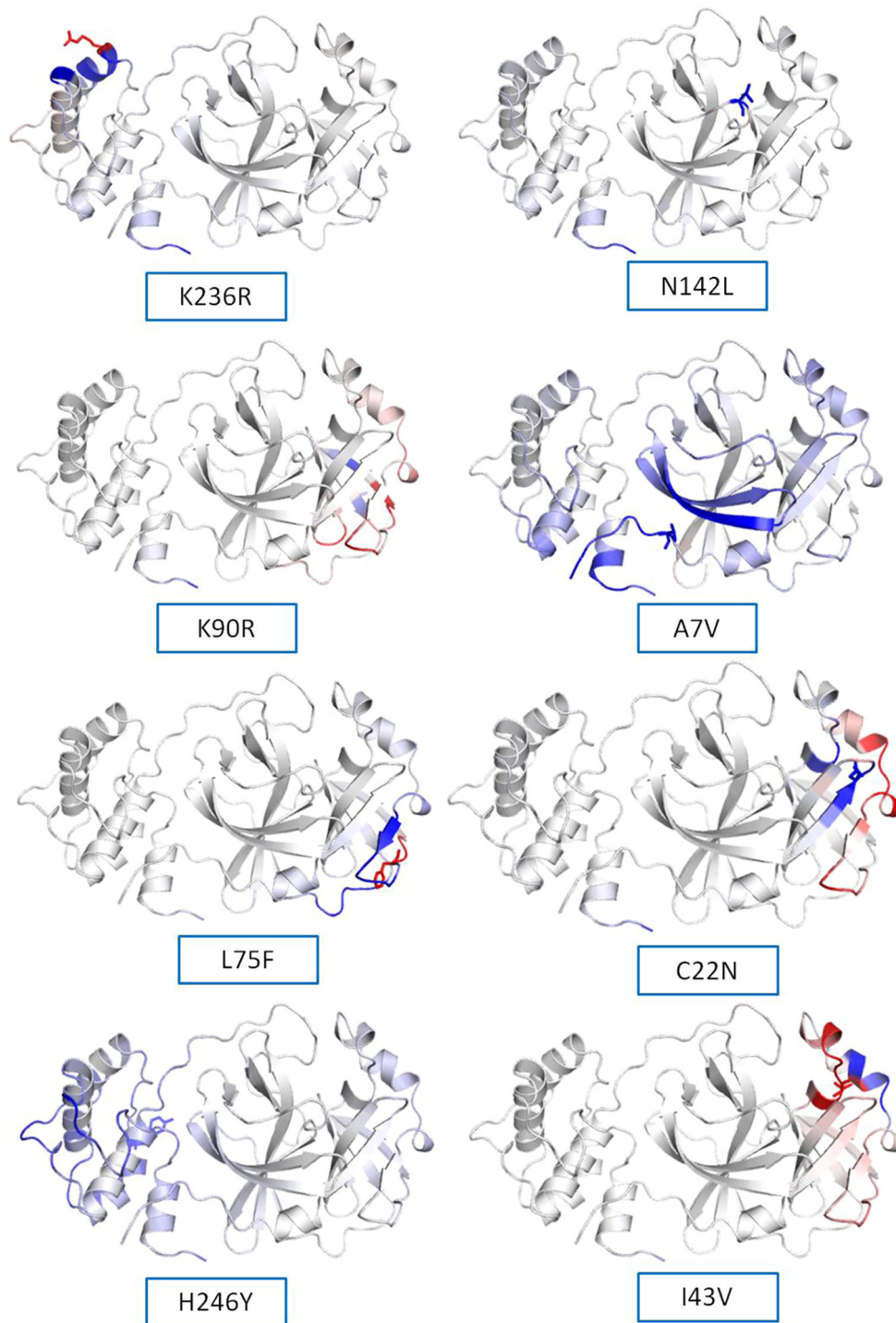
estimates the flexibility or steadiness of a protein upon mutation as compared to the wild type as calculated by ENCoM, DUET, mCSM and others. Negative value of $\Delta\Delta G$ denotes destabilization of protein upon mutation, whereas a positive value signifies stabilization. The free energy difference, $\Delta\Delta G$ between the wild and mutated protein sequences was calculated using Dynamut and the values showed a stabilizing mutation in five mutants of NSP5 protein as indicated by their

$\Delta\Delta G$ values (Table 2). The mutants N142L, H246Y and I43V were destabilizing for NSP5 protein with -$\Delta\Delta G$ values. The most stable mutant amongst all was L75F showing highest positive value of $\Delta\Delta G$ (1.200 kcal/mol), followed by C22N (0.884 kcal/mol) and A7V (0.653 kcal/mol) as shown in Table 4. The highest negative value of $\Delta\Delta G$ was shown by I43V (−1.002 kcal/mol) followed by N142L (−0.029 kcal/mol) and H246Y (−0.028 kcal/mol). The vibrational entropy change ($\Delta\Delta S_{Vib}$ ENCoM) provides
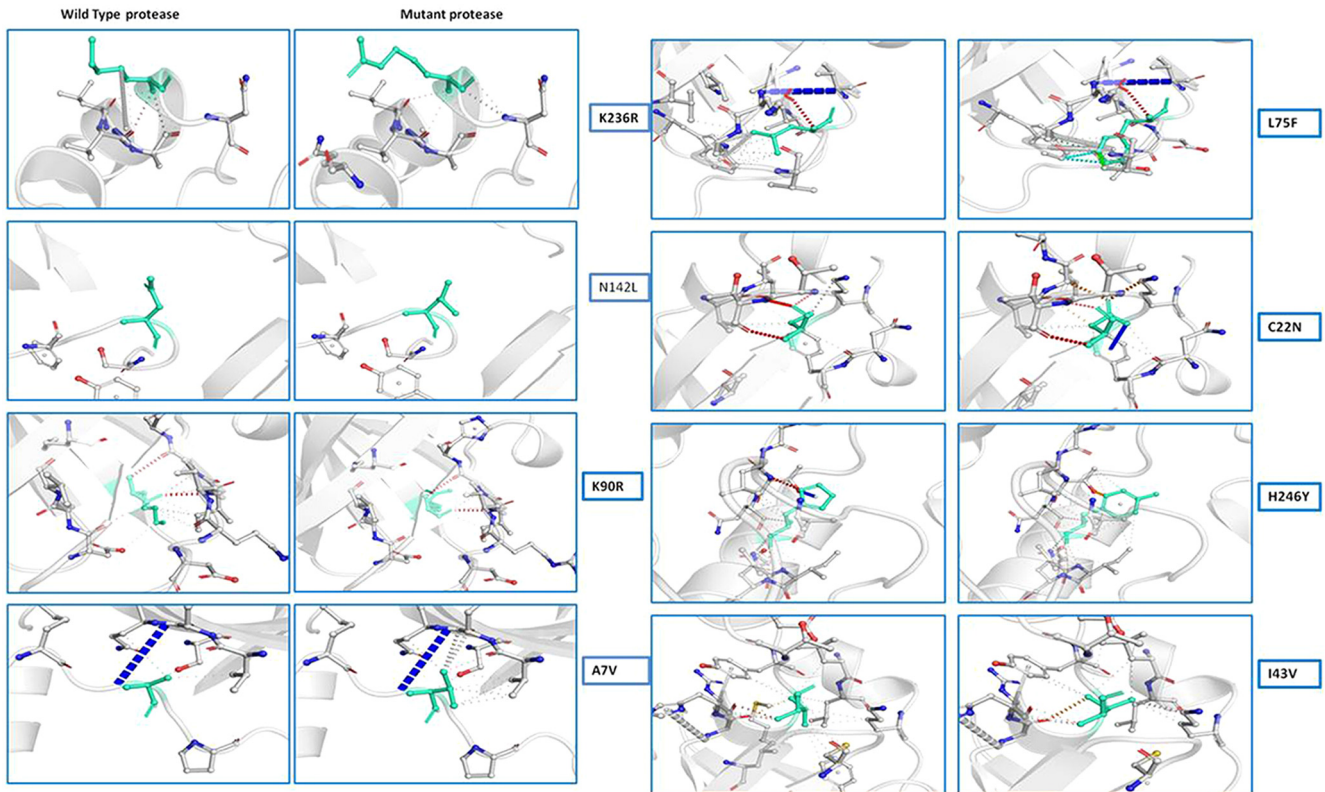
information on the configurational entropy of the proteins with single minima of the energy landscape. The $\Delta\Delta S_{Vib}$ ENCoM was calculated for the mutant and wild type protease protein to calculate the vibrational entropy energy change between wild type and mutant. The $\Delta\Delta S_{Vib}$ ENCoM calculated for all the protease mutants revealed a negative value signifying the rigidification of protein structure upon mutation except for I43V and K90R mutant which have positive values of $\Delta\Delta S_{Vib}$ ENCoM, signifying gain of flexibility upon mutation in NSP5 protein. The visual representation of flexibility analysis depicted similar results, of gain in rigidification upon mutation shown by blue region in all the NSP5 mutants except for I43V and K90R mutants, shown by red color region in Fig. 2c.

Further, the findings of our study dealt with the detection of variation in intramolecular interactions of NSP5 protein with its neighboring molecules upon mutation. All the NSP5 mutants studied here showed significant changes in intramolecular interactions that occurred in NSP5 proteins upon mutation (Fig. 2d). The mutation caused significant alterations in the interactions like hydrogen bonds, ionic interactions, hydrophobic interactions and other metal complex interactions. The substitution in side chain of the amino acids changes due to mutation hence disrupting neighboring interactions. This study predicts that the mutation in leucine, asparagines, lysine, cysteine, alanine residues causes significant alterations in the intramolecular interactions with the neighboring molecules (Fig. 2d). From these results, it can be concluded that the NSP5 protein mutation not only changes the overall dynamics of the protein but can also interrupts its intramolecular interaction.

B-cell epitope prediction with its antigenicity and allergenicity

Lineal B-cell epitopes were predicted for NSP5 protein using NSP5 protein sequence as query and threshold value of 0.5 was selected. A total of eight B-cell epitopes predicted for this protein above the threshold value which are shown in Table 3 (Fig. 3). Out of these nine epitopes, the epitopes KMAFPSGKV, EDMLNPNYEDL, QNGMNG and EFTPFDVVR were highly antigenic as well as non-allergic, whereas some epitopes were immunogenic but allergic. These five predicted epitopes can be a good candidate in vaccine production against SARS-CoV-2. In our analysis, 9 mutations out of 33 were found in the epitopic region of protease protein. These mutations not only change its epitopic region rather changes its overall antigenicity and therefore can help in host evasion.

Prediciton of T-cell epitope and MHC class I immunogenicity

Altogether 9 T-cell binding epitopes were predicted for NSP5 protein showing different allele binding affinity. The sequence of these epitopes along with its position is shown in Table 4. Out of these nine T-cell epitopes only two were allergenic and others were immunogenic as well as non-allergic. The MHC class I immunogenicity of the NSP5 molecules is shown in Table 5. A total of six peptides were predicted with a potential of MHC class I immunogens. These epitopes can induce immunogenicity and hence increase cytokine production in cells to combat the infection.

Table 2 – Effect of mutation on the structural dynamics of protease protein as shown by ΔΔS ENCoM and ΔΔG values.

S. no.	Wuhan isolate	Indian isolates	Amino acid position	ΔΔG Dynamut	ΔΔS ENCoM	ΔΔG ENCoM	Mutation type
1.	K	R	236	0.441 kcal/mol	-0.138 kcal.mol^{-1} K^{-1}	0.110 kcal/mol	Stabilizing
2.	N	L	142	-0.029 kcal/mol	-0.052 kcal.mol^{-1} K^{-1}	0.041 kcal/mol	Destabilizing
3.	K	R	90	0.456 kcal/mol	0.125 kcal.mol^{-1} K^{-1}	-0.100 kcal/mol	Stabilizing
4.	A	V	7	0.653 kcal/mol	-0.472 kcal.mol^{-1} K^{-1}	0.377 kcal/mol	Stabilizing
5.	L	F	75	1.200 kcal/mol	-0.322 kcal.mol^{-1} K^{-1}	0.258 kcal/mol	Stabilizing
6.	C	N	22	0.884 kcal/mol	-0.030 kcal.mol^{-1} K^{-1}	0.024 kcal/mol	Stabilizing
7.	H	Y	246	-0.028 kcal/mol	-0.108 kcal.mol^{-1} K^{-1}	0.087 kcal/mol	Destabilizing
8.	I	V	43	-1.002 kcal/mol	0.253 kcal.mol^{-1} K^{-1}	-0.202 kcal/mol	Destabilizing

Table 3 – List of lineal B-cell epitopes for NSP5 protein with their sequence, length, site, antigenicity and probable allergenicity.

No.	Start	End	Peptide	Length	Antigenicity	Allergenicity
1	5	13	KMAFPSGKV	9	0.6043 (Probable antigen)	Non-allergen
2	47	57	EDMLNPNYEDL	11	1.091(Probable antigen)	Non-allergen
3	93	109	TANPKTPKYKFVRIQPG	17	0.145(Probable non-antigen)	Non-allergen
4	170	196	GVHAGTDLEGNFYGPFVDRQTAQAAGT	27	0.2846(Probable non-antigen)	Allergen
5	225	228	TTLN	4	0(Probable non-antigen)	Non-allergen
6	236	247	KYNYEPLTQDHV	12	0.9135(Probable antigen)	Allergen
7	273	278	QNGMNG	6	1.1867(Probable antigen)	Non-allergen
8	290	298	EFTPFDVVR	9	1.6049(Probable antigen)	Non-allergen

Cluster analysis of MHC alleles

The cluster analysis of MHC class I allele is shown in Fig. 3c while that of class II allele is shown in Fig. 3d, where the red zone denotes strong interaction of the HLA allele with the epitopes of NSP5 protein, whereas yellow depicts weak interaction. We analyzed the binding ability of all the alleles with the protease epitopes.
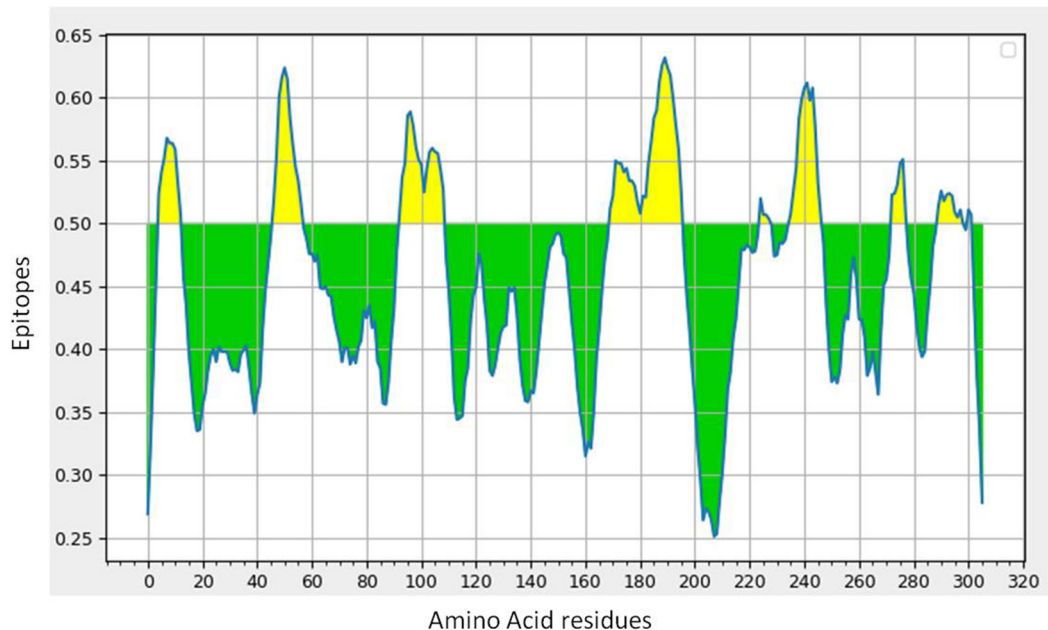

Fig. 3 – (a) B-cell epitope prediction of NSP5 protein. The threshold cutoff is 0.5 above which the residues are epitopes. (b) The results of MHC cluster analysis. (A) Heat map of MHC class I cluster, (B) tree map of MHC class I cluster. (c) The results of MHC cluster analysis. (A) Heat map of MHC class II cluster, (B) tree map of MHC class II cluster.

A.

B.

Assessment of antigenicity and allergenicity

VaxiJen v2.0 server was used to predict the antigenicity of the protease protein. The property of antigenicity depends on the ability of the vaccine to bind to the B-cell and T-cell receptors and increase the immune response in the cell. This analysis indicates that the NSP5 protein sequence is antigenic with potent antigenicity at a threshold of 0.4%. A good immunogen should not show allergic response in the host cell. The allergenicity of B-cell epitopes of the NSP5 protein was predicted using Allertop tool as many B-cell and T-cell epitopes were non-allergenic and hence can be a candidate protein for vaccine development.

Discussion

Coronavirus poses an unprecedented threat for human health globally. Considering its contagiousity, World Health

A.

B.

Table 4 – T-cell epitope prediction of SARS- CoV-2 protease and its allergenicity.			
Peptide	Start position	Score	Allergenicity
MLNPNYEDL	49	1.197	Non-allergen
IRKSNHNFL	59	1.128	Non-allergen
VLAWLYAAV	209	1.122	Non-allergen
AMRPNFTIK	129	1.117	Allergen
TPFDVVRQC	292	1.048	Allergen
GSPSGVYQC	120	1.025	Non-allergen
TLNDFNLVA	226	0.948	Non-allergen
FLNRFTTTL	219	0.889	Non-allergen
ITVNVLAWL	200	0.855	Non-allergen
TVNVLAWLY	201	0.780	Non-allergen

Organization on March 11, 2020 has declared public health emergency internationally (WHO 2020). SARS-CoV-2 is a member of RNA viruses and has remarkable capacity to mutate their genome in a very short period of time.[27] Notably, majority of viral mutation shows harmful effects. Moreover, a mutation is essential for viral evolution and adaptability, these traits are considered as the key determinants for viruses to survive in the dynamic environment of host and also enabling them to evade the pre-existing immunity of host and most often acquire drug resistance. SARS-CoV-2 infections emerged from Wuhan, China, soon began to spread globally. Rapid transmission of coronavirus infection depends on various factors such as polymerase fidelity, different

Table 5 – Showing class I immunogenicity of NSP5 protein of SARS-CoV-2.

Peptide	Length	Score
FYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE	60	1.51334
SGVTFQ	6	0.16646
PLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQC	60	0.1167
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIR	60	−0.01126
SPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGN	60	−0.11804
KSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNG	60	−0.9389

geographical areas and population density, as well as poor health care system, climatic and environmental variations.[28] Mutational analysis of SARS-CoV-2 provides better understanding of its epidemiology, pathogenesis and to devise antiviral therapeutic strategies against COVID-19.

The results of our study revealed, a total of 33 mutations identified from 675 sequences of NSP5 (main viral protease) from India. Amongst these mutations, three were non-synonymous amino acid substitutions (N142, L75F and C22N), whereas others showed deleterious impact on the structure and function of NSP5 protein. The mutations K236R, N142L, K90R, A7V, C22N and H246Y showed significant alterations in the secondary structure of NSP5 protein. The mutations N142L, H246Y and I43V were destabilizing and possess -ΔΔG values. All NSP5 mutants except for I43V and K90R mutant (positive values) showed negative values of $\Delta\Delta S_{Vib}$ ENCoM and hence resulted in rigidification of protein structure. Due to these mutations, considerable alterations were observed at several positions that also affect its stability and dynamicity which in turn altered the function of NSP5. Roe et al[29] have reported that NSP5 are capable to make associatation with several other components of replication complex. Earlier studies have also revealed that important intra- and intermolecular interaction exist between the main viral protease NSP5 and other replicase gene, with mutation in the NSP5 domain as well as in the NSP3 and NSP10 which negatively affecting the activity of NSP5.[29–31] The design and development of vaccine gained much attention nowadays including the multiepitope, DNA as well as RNA-based vaccines for various infectious diseases (such as influenza virus, Ervebo virus), using predictive tools of immunoinformatics have become the major research priority. The conventional methods of vaccine designing strategies include experimental identification, establishing immunological correlation with the coronavirus to develop potential vaccine construct. For the structural activities of SARS-CoV-2, proteins are supposed to be important constituent involved in the viral infection, entry and replication. The findings of earlier studies suggested that protein could be a very good target for developing vaccine against SARS-CoV-2.[32–34] Additionally, for a peptide vaccine to be highly immunogenic B-cell epitope of its target molecule must interact with a T-cell immune epitope. The T-cell epitopes is made up of short fragments of peptide and hence appeared as more propitious, which generate long-term immune response mediated by CD8+ T-cells.[6] In contrast, the B-cell epitopes consists of lineal chain of amino acid.[35,36]

The epitope selection based on immunogenic features like antigenicity, allergenicity and toxicity. Similarly, the predicted antigenic determinants (epitopes) of MHC class-I showed interaction with the several HLA alleles and,

therefore, found to be antigenic. The hydropathy index and physiochemical properties of SARS-CoV-2 NSP5 protein were also estimated which revealed that protein is stable and can form non-covalent bonds (such as hydrogen bonds) with other protein molecules. The present *in silico* study was found consistent with the previous studies based on immunoinformatics approach for the design and developments of novel therapeutic intervention and/or vaccine against COVID-19.[37–40]

In this study, we investigated the NSP5, as a potent immunogenic epitopes which elevates prolonged humoral (B-cell) as well as cell-mediated (T-cell) immune response to counteract viral particles, and hence serves as a potential candidate vaccine. A total of eight B-cell and T-cell epitopes were predicted for NSP5 proteins, amongst which the epitopes KMAFPSGKV, EDMLNPNYEDL, QNGMNG and EFTPFDVVR were highly immunogenic as well as non-allergenic. Primarily, the efficacy of vaccine candidates relies on the selection of its antigen molecules.[41] The data obtained from our study also corroborates the previous findings. Earlier studies on SARS-CoV and MERS-CoV have shown that the S glycoprotein can induce antibodies to neutralize virus infection by blocking virus binding as well as its fusion to the host cell.[41,42] Yashvardhini et al[17] have also reported that multiepitopes-based peptide vaccines are safe and specific that need adjuvants to show high levels of immunogenicity.

In the present study, occurrence of recurrent mutations in the main viral protease (NSP5) of coronavirus elucidates structural alteration that might affect its functions. Using predictive tools of computational biology, we also predicted promising epitope based vaccine candidates that are capable to stimulate both humoral (B-cell) as well as cellular (T-cell) immune responses. However, our *in silico* designed vaccine construct showed high efficacy and, therefore, suggested as good candidate against SARS-CoV-2 infections. Moreover, further *in vivo* and *in vitro* studies are mandatory to validate the durability and efficacy of designed vaccine candidate.

Conclusion

Occurrence of recurrent mutations in the NSP5 of SARS-CoV-2 provides a deep insight in the identification and magnitude of virulence properties. The study also suggests continues molecular surveillances of novel coronavirus that might be useful in the development of ongoing biomedical intervention to curb this contagious disease. For the design and development of candidate vaccine, NSP5 of coronavirus has been chosen as potentially ideal target molecule because NSP5 is

the main viral protease of SARS-CoV-2 and plays an important role in the viral replication cycle. Moreover, our study sheds light on, high efficacy and durability of designed epitopes vaccine candidate applying various predictive tools of immunoinformatics; further, *in vivo* and *in vitro* studies are mandatory to validate designed candidate vaccine.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vacun.2021.10.002.

Trial registration number (if clinical trial)

None.

Funding

Nil.

Declaration of competing interest

No conflict of interest.

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. The Lancet. 2020;395(10224):565–74.
2. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. New Engl J Med. 2020;382:727–33.
3. WHO. Coronavirus disease (Covid-19) pandemic. 12.02.2021]; Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019; 2021.
4. Andreano E, et al. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma Proc Natl Acad Sci U S A. 2021;118(36). https://doi.org/10.1073/pnas.2103154118 e2103154118, 2020.12.28.424451.
5. Greaney AJ, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe. 2021;29(3):463–76.
6. Chan JF, Kok KH, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020;9:221–36. https://doi.org/10.1080/22221751.2020.1719902.
7. Gordon D, Jang G, Bouhaddou M, et al. A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. bioRxiv. 2020. https://doi.org/10.1101/2020.03.22.002386.
8. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579 (7798):265–9
9. Muramatsu T, Takemoto C, Kim Y, et al. SARS–CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. Proc Natl Acad Sci U S A. 2016;113:12997–3002. https://doi.org/10.1073/pnas.1601327113.

10. Yoshimoto FK. The proteins of severe acute respiratory syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. Protein J. 2020;39:198–216.
11. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol. 2015;1282:1–23.
12. Ziebuhr J, Gorbalenya AE, Snijder EJ. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol. 2000;81:853–79.
13. Stobart CC, Sexton NR, Munjal H, Lu X, Molland KL, et al. Chimeric exchange of coronavirus NSP5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. J Virol. 2013;87:12611–8.
14. Lu X, Lu Y, Denison MR. Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. Virology. 1996;222:375–82
15. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. Science. 2020;368:409–12. https://doi.org/10.1126/science.abb3405.
16. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. Nature. 2020;582:289–93. https://doi.org/10.1038/s41586-020-2223-y.
17. 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. https://doi.org/10.1155/2021/6627141 2021 Article ID 6627141, 8 pages.
18. Madeira F, Park Y, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucl Acids Res. 2019;47 (W1):W636–41.
19. Choi Y, et al. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinfo. 2015;31(16):2745–7
20. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. Prot Proto Hand. 2005:571–607.
21. Ashok Kumar T. CFSSP: Chou and Fasman secondary structure prediction server. Wide Spec. 2013;1:15–9. https://doi.org/10.5281/zenodo.50733.
22. Kelley L, Mezulis S, Yates C, Wass M, et al. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Proto. 2015;10(6):845–58.
23. Rodrigues CHM, Pires DEV, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucl Acid Res. 2018;46(W1):W350–5.
24. Kim Y, Ponomarenko J, Zhu Z, et al. Immune epitope database analysis resource. Nucl acid Res. 2012;40(W1):W525–30.
25. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinfo. 2017;8(1):4. https://doi.org/10.1186/1471-2105-8-4.
26. Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. BMC Bioinfo. 2013;14(Suppl 6):S4. https://doi.org/10.1186/1471-2105-14-S6-S4.
27. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. J Med Virol. 2020;92(4):455–9.
28. Wang MA, et al. Temperature significant change COVID-19 transmission in 429 cities. medRxiv. 2020.
29. Roe MK, Junod NA, Young AR, Beachboard DC, Stobart CC. Targeting novel structural and functional features of coronavirus protease nsp5 (3CLpro, Mpro) in the age of COVID-19. J Gen Virol. 2021;102(3). https://doi.org/10.1099/jgv.0.001558 Epub 2021 Jan 28. PMID: 33507143.
30. Donaldson EF, Graham RL, Sims AC, Denison MR, Baric RS. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J Virol. 2007;81:7086–98.

31. Stobart CC, Lee AS, Lu X, Denison MR. Temperature-sensitive mutants and revertants in the coronavirus nonstructural protein 5 protease (3CLpro) defne residues involved in long-distance communication and regulation of protease activity. J Virol. 2012;86:4801–10.

32. Stokes HL, Baliji S, Hui CG, Sawicki SG, Baker SC, et al. A new cistron in the murine hepatitis virus replicase gene. J Virol. 2010;84:10148–58.

33. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579 (7798):265–9

34. Tai W, He L, Zhang X, Pu J, 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. 2020:1–8.

35. Chen N, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507–13.

36. Li Q, et al. Early transmission dynamics in Wuhan, China, of novel coronavirusinfected pneumonia. New Engl J Med.. 2020;382(13):1199–207. https://doi.org/10.1056/NEJMoa2001316.

37. Dangi M, Kumari R, Singh R, et al. Advanced in silico tools for designing of antigenic epitope as potential vaccine candidates against coronavirus. Bioinfor Seq Struct Phylogeny. 2018:329–57.

38. Jr Slingluff CL, Lee S, Zhao F, et al. A randomized phase II trial of multiepitope vaccination with melanoma peptides for cytotoxic T cells and helper T cells for patients with metastatic melanoma (E1602). Clin Cancer Res. 2013;19 (15):4228–38.

39. Toledo H, Baly A, Castro O, et al. A phase I clinical trial of a multi-epitope polypeptide TAB9 combined with Montanide ISA 720 adjuvant in non-HIV-1 infected human volunteers. Vaccine. 2001;19(30):4328–36.

40. Khan A, Khan DM, Saleem S, et al. Phylogenetic analysis and structural perspectives of RNA-dependent RNA-polymerase inhibition from SARs-CoV-2 with natural products. Interdis Sci Comput Life Sci. 2020;12(2):335–48.

41. Chiou SS, Fan YC, Crill WD, et al. Mutation analysis of the cross-reactive epitopes of Japanese encephalitis virus enve-lope glycoprotein. J Gen Virol. 2012;93:1185–92.

42. Du L, He Y, Zhou Y, et al. The spike protein of SARSCoV — a target for vaccine and therapeutic development. Nat Rev Microbiol. 2009;7:226–36.