International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

SARS-CoV-2 Mutations, Origin of Variants and Prediction of Mutations to Control Severe Diseases in Humans

Satyesh Chandra Roy

Former Professor and Head, Department of Botany and Co-ordinator, Centre of Advanced Study for Cell and Chromosome Research, UGC Emeritus Professor, University of Calcutta
Email: *scroyind[at]yahoo.com*

Abstract: ***With the increase of COVID-19 as pandemic throughout the world causing several deaths in several countries, researches are going on to find out the ways in the origin of mutations in SARS-CoV-2 genome. The process of mutation and origin of new variants have been discussed. Again the Codon Usage Pattern in the present virus has also been described. Lastly different modern methods for the prediction of mutations in the virus have been mentioned. New technologies are used to study genomic changes and changes in the sequence of amino acids of the viral protein are used to predict the origin of future mutations. The information for the prediction of mutations will definitely help to synthesize the specific drugs for the Coronavirus as well as in the manufacture of proper vaccine.***

Keywords: SARS-CoV-2 genome, Mutation, Codon Usage, Origin of Variants, Prediction of Mutations

1. Introduction

The widely spread pandemic of COVID-19 throughout the world is due to the Corona virus known as Sars-CoV-2 which was first detected in the Wuhan district of China at the end of 2019. With the increase of pandemic in 2020 all countries of the world are facing serious threats to public health and economic crisis. So far seven types of coronavirus have been detected such as SARS-CoV, MERS-CoV, HKU1, NL 63, OC 43, 229 E and recent one SARS-CoV-2. Of these coronaviruses, SARS-COV, MERS-CoV and SARS-CoV-2 cause severe diseases to humans and other viruses show mild symptoms.

Characteristic Features of SARS-CoV-2 GENOME

SARS-CoV-2 is under the family of coronaviruses having positive sense single stranded RNA virus with size of RNA (genomes) from 26 to 32 kb with 12 Open Reading Frames (ORFs) encoding 27 proteins. This single stranded RNA virus genome has 5` Cap structure and 3` poly A tail that acts as a mRNA for producing replicase polyproteins through process of translation. Most of the part of genome encodes two polyproteins of pp1a and pp1b. From these polyproteins 16 non-structural proteins are formed. There are four structural proteins such as Spike (S), Envelope (E), Membrane (M) and Nucleocapsid (N). Another important feature is the lack of proofreading or postreplicative repair mechanisms causing the accumulation of many deleterious mutations with increasing virulence and occurrence of many variants (Graepel et al 2017). About two thirds of the SARS-CoV-2 genome bears largest ORF (ORF 1a/b region) encoding 16 non-structural proteins (nsp1-16). The remaining parts of the genome comprises ORFs encoding structural and accessory proteins

(Rahimi et al 2021). The genome of SARS-CoV-2 encodes two open reading frames (ORFs) such as ORF1a and ORF 1ab. Both genes are encoding polyproteins. Of them ORF 1ab encodes replicase polyproteins and four structural proteins (Spike proteins, Envelope protein, Membrane protein (M) and Nucleocapsid protein (N).

From the comparative study of alpha- and beta coronaviruses, characteristic features of SARS-CoV-2 has been noted as i) SARS-Cov-2 is found to be optimized for binding to the human receptor ACE2 and ii) spike protein has a furin cleavage site at the S1 and S2 boundary leading to the presence of three O-linked glycans around the site. The presence of furin cleavage site and O-linked glycans at the S1 and S2 border has an important role in viral infectivity and capacity for host range (Andersen et al 2020)

Again the six amino acids of the Receptor binding domain (RBD) of spike protein of this virus are playing an important role in binding to ACE2 receptors with high affinity and in selecting the host range of the virus. This high affinity of binding of SARS-CoV-2 to human receptor ACE2 instead of animal may be due to natural selection of virus for their survival as high level of deforestation occurs everywhere. Generally the transmission of this type of virus, harbouring in Bats, occurs from animal to animal instead of animal to human. The occurrence of the type of change in transmission of virus from animal to human may be due to natural selection or due to manipulation done in the laboratory (Roy 2020 a, b).

With the occurrence of global pandemic of SARS-CoV-2, many researches are going on to understand the way of coronavirus mutation, evolution and emergence of this severe disease in human populations. Comparative studies of different related viruses like Murine hepatitis virus (MHV) and MERS with present SARS-CoV-2, it has been found that the high fidelity in infecting human cells is due to the occurrence of recombination in the spike protein sequence (Huang et al 2020; Gribble et al 2021) that happens during infections in humans. So the spike protein (S) is generally used as an antigen for the development of vaccine.

As large number of mutations occur in this virus during infections, so it is difficult to develop one-time vaccine, effective medicines and one time diagnostic tests. In addition to this high mutation rates of SARS-CoV-2, the virus gets the ability to adapt to changing environments

rapidly or in other words attain the strong power of adaptability. From the genetic point of view, it has been noted that the high adaptability of RNA virus to new hosts is polygenic in nature involving multiple mutations in a variety of genes leading to high fitness ability (Holmes 2007).

Mutations in SARS-CoV-2

One of the most important features of this virus is the high mutation rate which may be due to the lack of proofreading in having no Exonucleases. So the virus is constantly making new variants and has also increased the adaptive capacity. In getting mutations with high error rate, viruses may get an advantage to get large number of mutations in short time and they may select some beneficial mutations needed for adaptation. It has been noted that the genomic variability of SARS-CoV-2 is found in different geographical areas of the globe with difference in fatality rate varies in patients of different areas due to diverse demographic compositions.

The whole genomic data of SARS-CoV-2 is now available in public database of the Global Initiative on Sharing All Influenza Data (GISAID). The database of GISAID is made through compilation of 253 SARS-CoV-2 viral genomes (complete or partial) which has been contributed by clinicians and researchers from different parts of the world since December 2019 facilitating the scientists to understand the diversity and evolution of SARS-CoV-2 along with its phylogenetic relationships. This will help to trace the infective pathways and to take some preventive measures (Forster et al 2021). Detailed genomic studies of the virus have identified three major clades in SARS-CoV-2 genomes. These are Clade G (Variant of Spike protein S-D614G), Clade V (Variant of the ORF 3a coding protein NS3-G251) and Clade S (Variant ORF8-L84S) [Mercatelli and Giorgi 2020). From the database of GISAID up to September 2020, SARS-CoV-2 virus has a mutation rate of $8x10^{-4}$ nucleotides/genome per year (Rahimi et al 2020). In another study of genome sequence analysis of 220 from the GISAID database showed that the occurrence of mutations are higher in Europe and North America as compared to Asia indicating the relationship of mutation pattern with time and geographical areas in different parts of the world (Rahimi et al 2020). Several studies on genetic variations of SARS-CoV-2 showed the presence of different types of mutations like missense, synonymous, insertion, deletion and non-coding mutations (Rahimi et al 2020). The phylogenetic analysis of 749 genetic sequences of SARS-CoV-2, obtained from data of GISAID, showed that strains were distinguished by the missense mutation in the Spike protein (S) showing an amino acid change from Aspartate to Glycine residue at position 614 and this mutation of Spike protein called as D614G. This mutation was found more prevalent in Italian population when severe infection was found in Italy after China (Isabel et al 2020).

During the study of SNPs in SARS-CoV-2, another interesting finding has been noted such as Transitions (Purine to Purine and Ptrimidine to Pyrimidine) and Transversions (Purine to Pyrimidine and Pyrimidine to Purine) of the nucleotide in the codon. In other words it is C < T transition or A < G transition or G < C or A < T or U.

Another investigations after analyzing the nature of each mutation showed that the mutation is more frequent in Single Nucleotide polymorphisms (SNPs) than insertion or deletion types in different continents comprising of amino acid changing events (Mercatelli et al 2020). More research is needed to come to a conclusion to identify any specific mutation responsible for causing severity of the disease. However, there is a possibility that D614G mutation has a great impact on the infectivity of the disease as the Spike protein of the virus is the only protein that helps its entry to the human cell.

Another type of mutation called Frameshift mutation is also common in SARS-CoV-2 where deletion or insertion of nucleotides takes place in the coding regions (Missense codons) resulting in the alteration of sequence of amino acids to form altered proteins. Again the deletion or insertion of nucleotides may change the codons in such a way that do not form any amino acid at all (Non-sense codons) resulting in the formation of truncated proteins. These effects result in the synthesis of altered or no proteins and thus causing some phenotypic changes leading to the origin of many variants of SARS-CoV-2.It is known that virus particles penetrate the human cells through Spike protein to cause a disease called COVID 19. So the spike protein of the virus is very important as it mediates in preparing Vaccine and specific drugs. It has also been noted that the virus SARs-CoV-2 is changing during infections of the people due to the appearance of mutation again and again (Callaway 2020) in the host cells leading to origin of variants.

The high nucleotide identity of SARS-CoV-2 has been observed first in Bats and then to other animal hosts showing a high degree of adaptation to different hosts which ultimately leads to jump species boundaries and infect humans (Isabel et al 2020). The survey was made to search mutation of SARS-CoV-2 worldwide for finding out the reason for spreading the virus as pandemic giving stress on the mutation of Spike protein as it helps the virus to enter human cells. The most predominant mutation in the virus is the missense mutation of spike protein known as D614G. In the survey during December 2019 and March 2020 this D514G mutation is found in Europe in66% cases and 44% in other parts of the world (Isabel 2020). In the mutation D614G of Spike protein, the amino acid Aspartate (D used in Biochemistry) is replaced by Glycine (G) altering a single nucleotide in the RNA code of the virus. The rapid spread of D614G mutation was found in European countries may lead to the decision that the mutation D 614G in the gene of the Spike protein is the main factor for the wide transmission of the disease (Korber et al 2020). Thus the role of mutation and recombination may be the most important factor for the evolution of virus and origin of variants. However it is yet to be confirmed.

Codon Usage Pattern

The codon is a triplet of three nucleotides used for a specific amino acid residue in a polypeptide chain or as a stop codon for the termination of protein synthesis. Many amino acids are coded by more than one codon. It is also known that there are 61 codons for 20 amino acids. So the phenomenon of Codon Usage is useful to select particular codon from alternative codons available by any organism particularly

after point mutation. Different codons that are used for the same amino acid are also known as Synonymous Codons. Certain synonymous codons are preferred over other synonymous codons by certain organism and this phenomenon is called Codon usage bias.This is found in all organisms including viruses.

The codon usage refers to the frequency with which an organism uses the available codons in genes. There are preferences for the use of the alternative codons by different species of an organism. This is also called Codon Usage Bias. During molecular evolutionary studies it has been found that Codon bias is found in any organism after mutation to maintain balance between mutational and translational selection of such genes through the use of alternate codons (Codon Usage). This phenomenon is expected to find in an organism where the frequency of mutation is high such as in SARS-CoV-2 (Dilucca et al 2020). The random selection of synonymous codons varies from organism to organism and this method of selection (codon usage) by RNA virus may vary from host to host as the replication and translation of the virus take place in the host. It has been found that there is a specific pattern of codon usage by the virus which is different from the pattern used by the host. However, the virus is expressing their genes in the host using different codons. But it is not clear how the virus is synthesising their proteins in the host using different codon usage patterns. It is also known that there are several tRNAs that carry the same amino acid. These tRNAs are called isoacceptors (Jitobaom et al 2020). Several researches are going on to find out the reason of using different codons for the translation of viral proteins in the host. In this process, post-translational modifications of tRNA may also play an important role in translation, metabolism and sressresponse. It has also been noted that tRNA modifications are mostly located in the anticodon loop that is very crucial for mRNA decoding and fine tuning of the process (Rozov et al 2016). Generally the modifications in anticodon of tRNA are taking place in the third codon base of anticodon as this base has a certain amount of play or wooble. This third base position of anticodon as more than one position of pairing is possible as per Wooble Hypothesis of Francis and Crick (Rozov et al 2016).

Codon usage bias refers to the condition where specific codons are used in higher frequency than other synonymous codons during translation process. Codon usage bias is found in some species or organisms as a result of selection pressure from mutation. They use different synonymous codons in translation for the survival of mutants to maintain evolution of species. This phenomenon has been found more in RNA virus particularly in SARS-CoV-2 as the occurrence of mutation is very high in this virus. So the Codon usage bias may be the important driving force for the evolution of RNA viruses. Codon usage pattern has also been used by human genes in some important biological processes like Cell Cycle, the regulation of cell cycle process, cell division, microtubule cytoskeletal organization and RNA processing (Jitobaom et al 2020). Thus codon usage patterns of human genes have many similarities with RNA viruses. Although the replication of viral genome depends on the host cell translation processes but several viruses have their separate codon usage pattern.

Codon bias is measured in codon usage in a transcriptome wide manner using high throughput sequencing data (i.e, Ribo-seq) drom ribosomal activity. Codon usage bias is generally highly expressed here than other genes through RSCU (Relative Synonymous Codon Usage) calculation. This calculation was done from the protein coding sequences of the human genes and RNA viruses and then PCA (Principal Coefficient Analysis) was performed to assess the codon usage bias of a gene (Jitobaom et al 2020).

Comparative studies of the nucleotide compositions of different coronavirus genome it has been found that SARS-CoV-2 has a nucleotide composition similar to other coronaviruses but with a different trend U> A< G < C whereas in other viruses the trend is A> U > G> C. The GC content in SARS-CoV-2 is 0.37± 0.05. The most frequently used codons in SARS-CoV-2 are CGU (arginine, 2.34 times), GGU (glycine, 2.42) and least used are GGG (glycine), UGC (serine) [Dilucca et al 2020]. Thus SARs-CoV-2 has a high AU content and low GC content and mostly use codons ending with U. In SARS-CoV-2 genomic sequences C<U transitions are found in high frequency. There are mutational biases towards U rather than A in genomes. From the analysis of ENC (Effective number of Codons Analysis) and Forsdyke plots it has been noted that three genes N (nucleocapsid), RdRp (RNA dependent RNA polymerase) and S (spike protein) mutate and evolve faster than other two genes of SARS-CoV-2 (Dilucca et al 2020).

Most of the mutations in SARS-CoV-2 take place in the host cell during replication using the host cellular enzymes. These enzymes are Adenosine deaminases acting on RNA (ADAR) which promotes A>G mutations and another enzyme of the family Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide like (APOBEC). The latter enzyme promotes C<U mutations. These are the viable replicating enzymes of the host in the virus (Simmonds and Ansari 2021). Viruses also show number of changes in viral proteins and genomic sequences with excess C<U changes may be the cause for adapting to the new hosts through mutations. This mutational bias towards U in SARS-CoV-2 may play an important role in spreading global pandemic of Most of the mutations in SARS-CoV-2 take place in the host cell during replication using the host cellular enzymes. These enzymes are Adenosine deaminases acting on RNA (ADAR) which promotes A>G mutations and another enzyme of the family Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide like (APOBEC). The latter enzyme promotes C<U mutations. These are the viable replicating enzymes of the host in the virus (Simmonds and Ansari 2021). Viruses also show number of changes in viral proteins and genomic sequences with excess C<U changes may be the cause for adapting to the new hosts through mutations. This mutational bias towards U in SARS-CoV-2 may play an important role in spreading global pandemic of the virus. It has also been observed that the ability of emergence of SARS-CoV-2 in human population may also appear through recombination within the spike protein sequence (Gribble et al 2021). Thus the recombination in SARS-CoV-2 may be associated with increased spread, severity of the disease and many recombinant populations or variants with higher frequencies of defective viral genomes. Again the high frequency of mutation in Spike protein may

be the cause of reducing the durability of vaccine of SARS-CoV-2.

Origin of Variants of SARs-CoV-2

There are many variants of SARS-CoV-2 such as Alpha, Beta, Gamma, Delta and Omicron. The Alpha lineage is B.1.1.7, Beta lineage is B.1.351, Gamma lineage refers to P.1, Delta lineage is B.1.617.2 and Omicron`s lineage is B.1.1.529. The lineage refers to the closely related with that strain of virus that has been determined by sequence analysis.

The accumulation of mutations in the genomes of SARS-CoVp2 virus is one of the most important factors in the origin of many variants depending on the mutation rates and impacts of mutation within and between hosts leading to the emergence and spread of variants within human populations. But SARS-CoV-2 has a proof-reading domain (Exo-N) so some mutations become non-functional and so mutation rates are reduced compared to other viruses like Influenza, HIV and Hepatitis C viruses (Otto et al 2021). It has been noted that SARS-CoV-2 genome shows 1.87 x 10^{-6} nucleotide substitutions per site per day which is 5 times lower than influenza A/H3N2 of 10.9x 10^{-6} nucleotide substitutions per site per day indicating the occurrence of about 20 genetic changes per year within a lineage (Otto et al 2021). Of all mutations, Synonymous mutations do not change anything but the non-synonymous mutations can change the amino acid followed by protein and can persist with probability. It has been noted that selection has little time to eliminate deleterious mutations accumulated over the pandemic`s short time frame. (Otto et al 2021 ; Wang et al 2020). So this type of deleterious mutations may persist in SARS-CoV-2 for some time and can produce new variants during this short period. The variant is defined to a virus with different sequence from other viruses while Lineage refers to viruses that are closely related in their sequence analysis. The Strain is called when the viruses show difference in their properties as in case of SARS-CoV-1 and SARS-CoV-2.

The World Health Organisation (WHO) has differentiated variants of SARS-CoV-2 into two types as Variant of Interest (VOI) and Variant of Concern (VOC). The former one (VOI) contains mutations that may alter its phenotypic characters with community transmission in impacting human health. The latter (VOC) has mutations causing high transmission rate with severe impact on human health and also can reduce the efficacy of vaccines and drugs (Otto et al 2021). The patients suffering for longer time with viral infections gave an opportunity to virus for many replications, increased mutation rate resulting in the weakening of the immune system of the patient. The sequence analysis of this virus showed the occurrence of many alterations in their genomes that may lead to the origin of new variants. So the variants of SARS-CoV-2 are now classified as Variety of Interest (VOI) and Variety of Concern (VOC).

The former class (VOI) has some special characters attained through mutations that are associated with some changes that can increase the receptor binding capacity of the virus leading to high transmissibility of infections without much severity of the disease. For example, these are B.1.526, B.1.525 and P.2 and are probably originated from D614G mutants.

The latter one i.e., VOC shows also high transmission rate of the disease and severity of infections including deaths. It also reduces immunities of patients, effectiveness of drugs and also may reduce efficacy of vaccines. The examples of these variants are B.1.1.7, P.1, B.1.351, B.1.427 and B.1.429. These are first detected in USA. It is interesting to note that all variants have a common mutation of D 614G (Vasireddy et al 2021).

In most of the countries, the most dominant variants found are B.1.1.7, B.1.351 and P.1. In UK the most common variant found is B.1.1.7 that falls under VOC. About 23 mutations in the Spike of this variant (B.1.1.7) have been detected in UK such as N501Y and P681H causing severe infections during September 2020 to February 2021. A new dataset analysis was done linking with 2, 245, 236 positive SARS-CoV-2 community tests and 17, 452 deaths in COVID 19 in England from November 2020 to February 2021. The presence or absence of the variant B.1.1.7 can be identified because mutation caused in this variant prevent PCR amplification of the Spike gene (S) target (Davies et al 2021). This is known as S gene Target Failure (SGTF). It has also been noted that the hazard of death associated with SGTF is 55% higher than in cases without SGTF (Davies et al 2021).

Another new variant detected is that of Omicron`s B.2 variants. The World Health organization (WHO) has recently announced a new subvariants of Omicron as BA.2 that causes COVID worldwide in March 22, 2022 (Schmidt 2022). It is not known till now that whether this variant will be more infectious than previous variants. Jeffrey Shaman and other scientists of School of Public Health, Columbia University stated that the genome of SARS-CoV-2 has still more fragility to infect human cells and to ignore the immune system of human body (Schimdt 2022). Bette Korber, Computational Biologist, working with viral diseases at the Los Alamos National Laboratory, New Mexico gives a new information that this new variant may spread in rural areas for producing more mutations which other varieties did not get any opportunity to spread in these interconnected communities. Then it will spread to other countries like Europe, Africa, USA and Asia (Schimdt 2022) with a chance of producing new variants. But further research is needed to confirm this view. As mutations are going on constantly in different strains of SARS-CoV-2, the new strain Omicron (BA.2) has a higher transmission potential and the ability to evade the immune response of the human body amd so this called Stealth Omicron (Mishra 2022). WHO is called this variant as Variant of Concern. This BA.2 is also called subvariant of Omicron strain known as BA.1. Now there is a report of another new strain called XE which has been probably originated from the hybrid of two strains BA.1 and BA.2

Prediction of Mutations

It has been noted that the unique property of producing high frequency of mutations in SARS-CoV-2 due to low fidelity

of its RNA polymerase, is the important factor for the origin of new variants with new characters in enhanced transmission as well as causing severity of disease with increased fatal rate. Several researches are going on to find out the method of predicting mutations in the virus, so that forecasting may help in designing of vaccine and drugs as well as the surveillance of virus and awareness to public health in advance. It is known that evolutionary path in any organism including viruses is generally determined by natural selection. Natural selection generally removes the lethal and deleterious mutations from the population through negative selection. The positive selection of beneficial mutations increases the fitness of population. Understanding the selective forces in selecting mutations in viruses is important to predict mutations (Dolan et al 2018).In addition to high rates of mutation in RNA viruses (SARS-CoV-2), they undergo frequent recombination and re-assortment creating novel genotypes leading to new variants (Dolan et al 2018). Of all mutations arising in the virus, mutations in the viral spike protein are very important as it is the target of antibody-mediated immunity. This spike protein is also important in producing vaccine as scientists are mostly used spike protein as antigen for designing vaccine (Maher et al 2022). Mutations expressed generally in three ways like;

i) Silent (no significant effect);

ii) Loss of Function or virulence and iii) Gain of Function or virulence in case of virus.

New technologies are used to study genomic changes or changes in the sequence of amino acids in the protein of virus to predict future mutations.

i) Method of Machine Learning

One type of Machine Learning method is called Deep learning which is based on artificial neural networks (ANN) with multiple layers between the input and output layers or nodes just like the human brain containing neurons. This method is called deeper based on number of layers. It is known that single neuron of brain receives thousands of signals which are not found in other neurons. So this method is also called Deep Neural (DeepNEU). Machine learning methods for prediction of mutations were used for many years in viral genetics. As SARS-CoV-2 is spreading very fast throughout the world with the increase of fatality rate in many countries several researches are going on to predict the mutation in this virus using Machine Learning methods to identify therapeutic targets and awareness for further spread of infection and future outbreaks (Esmail and Danter 2020). With the help of this method (DeepNEU v.5.0) the prediction on the impact of Gain of Functions (GOF) and Loss of Functions (LOF) in SARS-CoV-2 genome has been done as well as the virulence potential of SARS-CoV-2 mutations well ahead of the occurrence in nature (Esmail and Danter 2020). They first created computer simulations of human induced pluripotent stem cells (aiPSC) and lung cells (ai LUNG). After standardisation iLUNG simulations were exposed to simulated SARS-CoV-2 viremia (presence of viruses in the bloodstream) by turning on the Spike-RBP (Receptor binding Domain) in the presence of Transmembrane Serine Protease 2 (TMPRSS2). Then the ai Lung and LUNG COVID 19 simulations were done to evaluate set of factors from the known SARS-CoV-2 genome (Esmail and Danter 2020).

The identification of amino acid mutations in the virus may help to predict new mutations of the virus for the origin of variants. This can be done with the help of Alanine scanning method to reveal the hot spots on the virus to identify future mutations for selecting proper drugs and vaccine. The major hot spot amino acids are Glu 35, Tyr 83, Asp38, Lys 31, Glu 37, His 34, residues of Ace2 receptor and Gln 493, Gln 498, Asn 487, Tyr 505 and Lys 417, Ala 684 residues in Spike protein RBD (Veeramachaneni 2021).All residues are located at the interaction interfaces of the Spike protein of virus and ACE 2 of human cell.

ii) Alanine Scanning Method

Alanine scanning method is used to find out simultaneously the functional contributions of the side chains of amino acid at the interface. Alanine is used as it is chemically inert with methyl functional group that mimics the secondary structures of many other amino acids (Taken from Wikipedia). Alanine scanning is Site directed mutagenesis technique used to find of the role of specific residue in the analysis of mutation. In this method all residues are mutated in such a way that Alanine is present in all mutants.

It is known that there is a binding affinity between viral S-RBD (S protein – Receptor binding domain) of SARS-CoV-2 and the host cell receptor ACE2 (Angiotensin converting enzyme 2) of the human cell during the viral attachment to the cell for the progress of infection.

With the study of this protein interaction (Spike protein and ACE2) through computer modelling, it has been noted that some hot spot residues are potentially involved in the interaction and there must be some energetic importance of the contact or binding (Laurini et al 2020). The in silico (experiments conducted through computer modelling or computer simulation) studies through alanine scanning interaction entropy method showed that there are some peptide inhibitors whose sequences are extracted from ACE 2 domain that help in the binding of Spike protein with the receptor of the host cell. These peptide inhibitors may be used as blockers of Spike protein and receptor protein interaction leading to the control of SARS-CoV-2 infection (Han and Kral 2020). These studies may help in the development of vaccines and use of protein inhibitors to control the present and future pandemic.

It is known that protein-protein interface between ACE 2 and the Receptor Binding domain (RBD) of the Spike protein of SARS-CoV-2 binding gas plated an important role in making COVID 19. Through Computer simulation and Alanine

Scanning Mutagenesis method it has been revealed that there is a variation in the free energy during binding of two proteins. Again some hot spot residues like D38, K31, E37, K353 and Y41 on ACE2 and Q498, T500 and R403 on the RBD of Spike protein of SARS-CoV-2. These residues are very important in shaping and in determining stability of the

protein-protein binding or the binding strengths (Laurini et al 2020). In other words the binding strengths of the two proteins as well as the calculation of change in free energy during binding may be useful for assessing the intensity of the spread of infection to the people. Thorough and accurate free energy calculations may be helpful to predict new mutations and the transmission ability of the virus (Zou et al 2020). In addition to the calculation of change of free energy during binding, the exploration of a variety of molecular properties like van der Waals energy, internal energy of bond, dihedral angle between two bonds, electrostatic energy, solvent accessible surface area, polar solva-tion free energy and non-polar solva-tion free energy are important to predict accurately thefuture mutations in SARS-CoV-2 (Mendis et al 2021). See predictions try Figure of Alanine Scanning Wikipedia

iii) Neural Network System

Neural network has been developed from the study of the function of Neurons in the human brain taking the way that neurons send signal to one another. Again the neurons of brain is different from other neurons present in the body as single neuron of the brain can receive many signals and then send it to different organs for proper function. That means neuron of the brain has many input layers and one or more output layers. Taking this principle brain, Computer scientists have developed Neural Networks or Artificial Neural Networks (ANN) using Machine learning or Deep learning Algorithms. Each artificial neuron has multiple layers like in brain and a specific threshold value. When that value is reached it becomes activated and sends data to the next layer of network like that of neurons of human brain. This process of passing data from one layer to another or from one neuron to another is sometimes called Feed forward Network when they flow in one direction only that is from input to output. The oldest neural network of single neuron was developed by Frank Rosenblatt in 1958. Generally there are two types of Artificial network like i) Convolutional Neural Networks (CNN) and ii) Recurrent Neural Networks (RNN). The former is utilized for the recognition of image and patterns within the image. The latter (Recurrent Neural Network) is used to make predictions and future outcomes that have already been used in predictions of stock market or forecasting sales. Again Artificial Intelligence has been developed with this principle with many applications.

With the prediction of Age Related Macular Degeneration successfully with accuracy of 0.82 in human by using Deep Learning Recurrent Neural Network, researches are going on for the prediction of mutations in SARS-C0V-2 with the help of Recurrent Neural Network of Long Short Term Memory (LSTM). This technique is known as COVID Deep Predictor technique (Saha et al 2021). The method of Data preparation was done by giving the i) sequences of the virus as Input followed by ii) creating descriptors of sequences using *k*-mer technique (Alignment free technique) and iii) by creating Bag of Unique-Descriptors (BoUDs) as vocabulary using Bag of Descriptors (BoDs) and embedded representation for each virus sequence was prepared. Then a model of COVID-Deep Predictor was made with the use of LSTM. The model validation was done using K-fold cross validation as well as other test sets of SARS-CoV-2 sequences (Saha et al 2021). The results were compared with other viruses like SARS-CoV-1, MERS-CoV, Ebola, Dengue and Influenza. However, more researches are needed to confirm the process. With the use of Recurrent Neural Network LSTM model predictions were attempted by analysing mutation rate of both nucleotide mutation and codon mutation separately (Pathan et al 2020). Neural network with Artificial Intelligence (AI) is also under investigation. Thus several researches are going on in this line in many laboratories of the world which will ultimately open a new horizon in the prediction of mutations in SARS-CoV-2 leading to manufacture of specific Vaccine and Drug for the recent pandemic.

References

[1] Alanine Scanning. Wikipedia. en.m.wikipedia.org.
[2] Andersen, KG et al. 2020, The Proximal Origin of SARS-CoV-2. Nature Medicine. 26: 450.455.
[3] Callaway, Ewen. 2020. Making Sense of Coronavirus Mutations. Nature 585: 174-17.
[4] Davies, NG. and Jarvis, CL., CMMID COVID 19 Working Group, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. 2021. Increased mortality in Community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature 593 (7858): 270-274.
[5] Dilucca, M., Sergio Forcelloni, Alexondros G. Georgakilas, Andrea Giansanti land Athanasia Pavlopoulou. 2020. Codon Usage and Phenotypic Divergences of SARS-CoV-2 Genes. Viruses 12: 498 - 519.
[6] Dolan, P.T., Zachary J. Whitefield and Rahul Andino. 2018. Mapping the Evolutionary Potential of RNA Viruses. Cell Host & Microbe 23 (April11) : 435 – 446.
[7] Esmail, Sally and Danter, Wayne R. 2020. DeepNEU: a Machine learning Stem cell simulation platform for evaluating the impact of Loss of Function and Gain of Function mutations in the SARS-CoV-2 genome. Research Square 123genetix https://orcid.org/0001-9595-4779.
[8] Forster, P., Lucy Forster, Colin Renfrew and M. Forster. 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. PNAS 117 (18) : 9241-9243.
[9] Graepel, KW., Xiaotao, Lu, James Brett Case, N.R. Sexton et al 2017. Proofreading Coronaviruses Adapt for Increased Fitness over Long Term Passage without Reversion of Exoribonuclease Inactivating Mutations. Amer. Soc. Of Microbiology. 8 (6):e01503-17.https://doi.org'10.1128/mBio 01503-17.
[10] Gribble, J., Laura J. Stevens., Maria L. Agostini, J. Anderson-Daniels, et al. 2021. The Coronavirus proofreading exoribnuclease mediates extensive viral recombination. PLOS Pathogens! https:// doi.org./10.1371/journal ppat.1009226.
[11] Han, Y and Kral, P. 2020. Computational Design of ACE 2 Based Peptide inhibitors of SARS-CoV-2. ACS Nano 14: 5143- 5147.
[12] Holmes, Edward C. 2006. The Evolution of Viral Emergence. PNAS. 103 (13): 4803-4804.
[13] Huang, J.M., Jan, S.S., Wei, X., Wan, Y. And Ouyang, S. 2020. Evidence of the Recombinant origin and

Ongoing Mutations in Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). bioRxiv.2020 Mar 19 :2020. 03.16.993816.
[14] Isabel, Sandra, Lucia Grana-Miraglia, Jahir M. Gutierrez et al. 2020. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. Scientific Reports Nature Research. 10: 14031. https:// doi.org./10. 1038/ s41698-020-70827.z
[15] Jitobaom, K., Supinya Phakaratsakul, Thanyaporn Sirihongthong, et al. 2020. Codon usage similarity between viral and some host genes suggests a codon-specific translational regulation. Heliyon 6 : 1-13. e03915
[16] Korber et al 2020. Making Sense of Coronavirus Mutations. Nature 585: 174-177.
[17] Laurini, Erik, D. Marson, Suzana Aulic, Maurizio Fermeglia and Sabrina Pricl. 2020. Domputational Alanine Scanning and Structural Analysis of the SARS-CoV-2 Spike protein /Angiotensin – Converting Enzyme 2 complex. ACS Nano 1: 11821 -11830.
[18] Maher Cyrus, M., Istvan Bartha, Steven Weaver, Julia di lullo, Elena Ferri et al.2022. Predicting the mutational drivers of future SARS-CoV-2 variants of concern. Sci. Translational Medicine 14 (23 February). eabk3445
[19] Mendis, Jenny, Ekrem Kaya and Tugba G. Kucullal. 2021. Identification of Hotspot Residues in binding of SARS-CoV-2 Spike and Human ACE2 proteins. J. of Computational Biophys. And Chemistry 20 (7): 729-739.
[20] Mercatelli, Danielle and Federico M. Giorgi. 2029. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Frontiers in Microbiology, 22 July, 11: Article 1800.
[21] Mishra, Sanjay. 2022. A Stealth Omicron subvariant is now spreading, worrying experts. Nat, Geog.. Fenruary 3.
[22] Otto, Sarah P, Troy Day, Julien Arino, Caroline Colizn, Jonathan Dushoff, Michael Li, S. Mechai et al. 2021. The Origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. Current Biology 31: R 918- R 929, July 26.
[23] Pathan, Refat Khan, Munmun Biswas and Mayeen Uddin Khandaker. 2020, Chaos, Solitons and Fractals. 138.
[24] Rahimi, Azadeh, Azin Mirzadeh and Sohil Tavakolpour. 2021. Genetics and Genomics of SARS-CoV-2: A Review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. Genomics 113 : 1221-1232.
[25] Roy, SC. 2020 a. Corona Virus – Origin, Replication and Remedy for Future Threat. Science and Culture 85 (5-6) : 137- 143.
[26] Roy, SC. 2020 b. Mystery of Corona Virus – A Review. Int. J.of Engineering Applied Sciences and Technology 5 (2) : 413-418.
[27] Rozov. A, Natalia Dameshkina, Iskainder Khusainov et al. 2016. Novel base-pairing interactions at the tRNA wooble position crucial for accurate reading of the genetic code. Nature Communications. 7: 10457. Doi : 10.1038. ncomms 10457.
[28] Saha, I, N. Ghosh, Debasree Maity. A. Seal and Dariusz Plewczynski. 2021. COVID Deep Predictor: Recurrent Neural Network to Predict SARS-CoV-2 and Other Pathogenic Viruses. Frontiers in Genetics 12 (Article 569120). doi: 10.3389/ f.gene2021.569120.
[29] Schmidt, Charles. 2022. What we know about Omicron`s BA.2 Variant so far? Sci. Amer. April 2022.
[30] Simmonds, P and M. Azim Ansari. 2021. Extensive C< U tramsition biases in the genomes of a wide range of mammalian RNA Viruses ; potential associations with transcriptional mutations, damage- or host – mediated editing of viral RNA. PLOS Pathogens 17 (8): e1009596. https :// doi. Org./10. 1371/journal.ppat.1009596.
[31] Vasireddy, Deepa, Rachana, Vanaparthy, Gisha Mohan, Srikrishna Varun Malayala and Paavani Atluri. 2021. Review of COVID-19 Variants and COVID Vaccine Efficacy : What the Clinician should know? Clin. Med. Res.3 (6): 317-325.
[32] Veeramachaneni, G.K.. 2021. Structural and simulation analysis of hotspot residues interactions of SARS-CoV-2 with human ACE 2 receptor. J. of Biomolecular structure and Dynamics 39911) : 4015-4025.
[33] Wang, H, Pipes L and Nielsen, R. 2020. Synonymous mutations and the molecular evolution of SARS-CoV-2 origins. Virus Evol. 7. Veaa 098.
[34] Zou, Junjie, Jian Yin, Lei Fang, Mingjun Yang, T. Wang, et al. 2020. Computational Prediction of Mutational Effects on SARS-CoV-2 Binding by Free Energy Calculations. J. Chem. Information and Modeling 60: 5794- 5802.