Covid-19 vaccine development through immunoinformatics guided multiple epitoping and demonstration of Drug Repurposing through molecular docking

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

Master of Science

In

Biotechnology

Submitted by:

Sandeep kumar

2K20/MSCBIO/25

Under the supervision of:

Prof. Yasha Hasija

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi - 11004

Abstract

Coronavirus-2 (SARS-CoV-2) is responsible for the symptoms of COVID-19. Almost every nation has been infected by SARS-CoV-2. The COVID-19 pandemic has prompted researchers to concentrate on the development of a vaccine and treatment techniques by becoming intimately familiar with the infection's biology. In order to minimize COVID-19 mortality and provide global immunity, a highly efficacious SARS-CoV-2 vaccine is imperative. The lengthy and costly process of developing vaccines could be sped up with immunoinformatics techniques. There have been advances in immunoinformatics tools used for reverse vaccinology to develop a SARS-CoV-2 vaccine, including Vaxijen, IEDB, NetCTL 1.2, PEP-FOLD, and studies of the development of MHC-I and II binding epitopes, among others. A drug repurposing strategy would reduce time and cost compared to drug discovery from scratch. It is an effective strategy for leveraging existing medications. Immunoinformatics may help identify T cell and B cell epitopes with more confidence, leading to fewer experiments and higher dependability for identifying vaccine candidates.

Contents

LIST OF FIGURES

List of Tables

Chapter 1

1.1 INTRODUCTION

A virus called SARS-COV-2 causes COVID-19, a viral illness that originated in China's Wuhan province and has since spread to 228 other nations, prompting a Pandemic [1][2]. Most commonly, the virus causes lungs to be damaged, resulting in flu-like symptoms, congestion, and, in more serious cases, pneumonia [3]. Deaths among older adults and people with chronic diseases are higher than normal according to statistical data. High-risk individuals may be struck by ARDS, respiratory failure, and inflammatory processes in addition to acute heart failure, heart damage, and multi-organ failure while infected with SARS-COV-2 [4]. Approximately 228 countries around the world have been affected by SARS-COV-2, which has been designated a pandemic. As of April 27, 2022, 510,887,255 Infected individuals . Approximately 6,241,104 have died. The number of people in serious need of breathing help has decreased from almost 464,142,621 to 42,203. The United States of America, India, Brazil, France and Germany are all top 5 impacted countries. It has been estimated that there have been 1,019,008 fatalities and nearly 82.789 million cases. The United States has the most cases (nearly 82.789 million) and the highest fatality rate. It has been estimated that there have been 1,019,008 fatalities and nearly 82.789 million cases. As the number of cases of COVID-19 continues to increase, the WHO has correctly designated it a worldwide medical crisis and pandemic [5]. SSRNA-encoded viruses, family of enclosed viruses with a deadly genome [6]. COVID-19, in contrast to SARS-CoV and MERS coronavirus , is triggered by SARS-CoV-2, a more virulent variant (MERS-CoV, 2013). To understand the pathophysiology of this virus and create successful therapies, comprehensive research is critical. [7].

There are 4 genera in which Coronaviruses are classified [8]: α Coronaviridae, β Coronaviridae, ϒ Coronaviridae, and δ Coronaviridae coronaviruses [9]. Approximately 2.6 is the reproduction number (R0) for each individual dissemination of SARS-CoV-2, indicating a rapid proliferation of infected cases [10].

The most abundant RNA viruses have genomes between 26 and 32 kb in size, called Coronavirus genomes [11]. Severe acute respiratory and MERS-CoV share approximately 82 percent sequence similarity, and key enzymes and matrix proteins share > 90 percent similarity. Detection and treatment of a common infection was made possible due to the greater extent of the gene. A SARS-CoV-2 structural protein has four structural proteins: spikes (S), envelopes (E), membranes (M), and nucleocapsids (N). There is a greater degree of sequence similarity between these proteins and the MERS-CoV and acute respiratory syndrome genes.

Coronaviruses attach their spike proteins to host cells' surface receptors upon entering the host cell. In order to interact with the ACE 2 receptor, the Spike protein subunits fuse to the cell surface. This is mediated by the S1 binding domain.

In order to develop SARS-CoV-2 successfully produced antiviral drugs, similar enzymes, including major $3CL^{pro}$, PL^{pro} , non-structural protein 12, & RdRP, might be addressed [14]. In order to better identify the genetic basis of disease, a comparing generation sequencing approach may be used to determine COVID-19-specific therapeutic approaches.

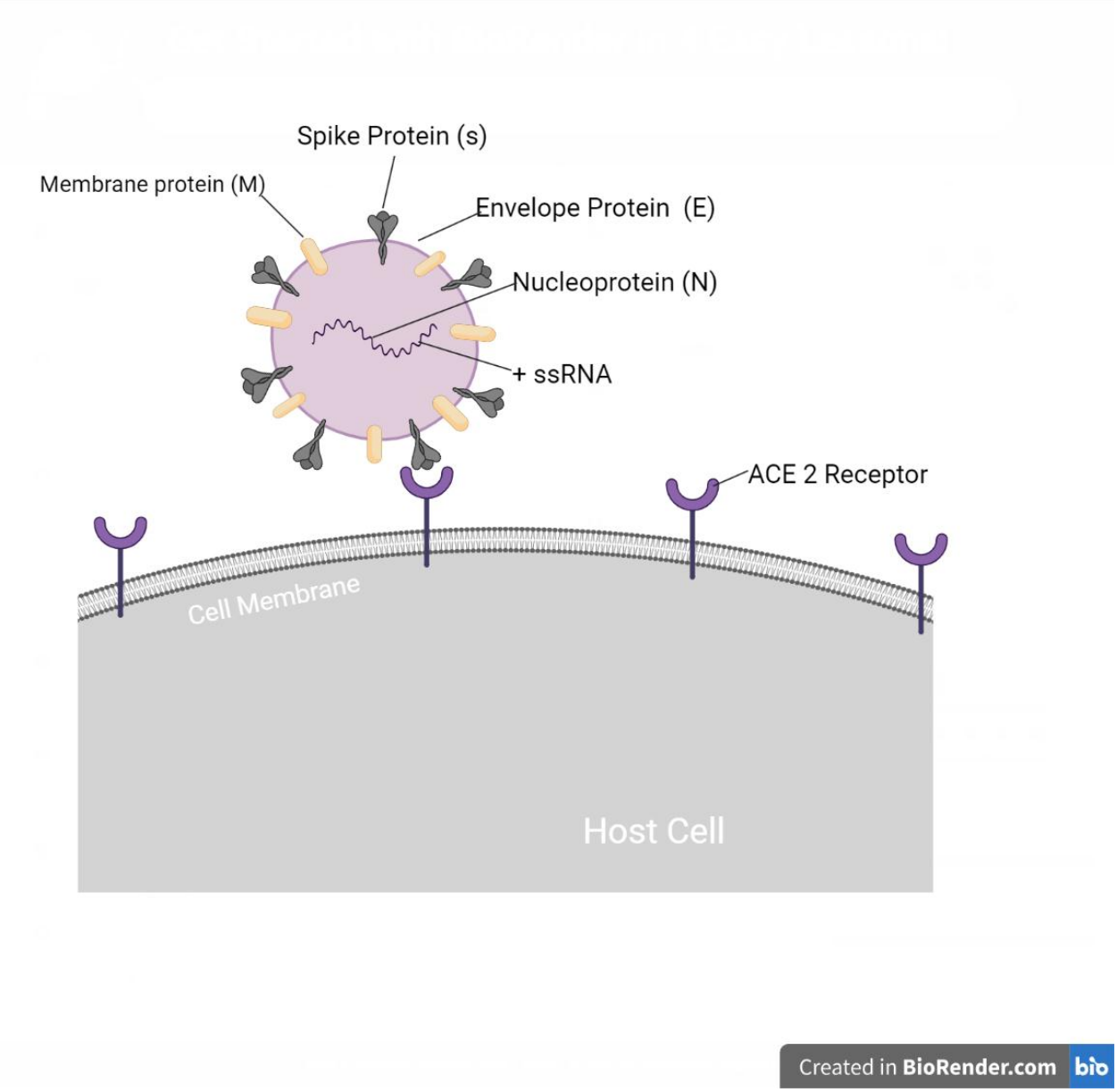

Fig. 1. Mechanism of attachment of spike protein with ACE-2 human receptor

The following 7 infectious CoVs that result in modest clinical symptoms include HCVNL63, HCV-229E, HCV-OC43, and HCV-HKU1. A person who is infected by SARS-CoV, MERS-CoV, or SARS-CoV2 can suffer from severe respiratory illness and even die. As a result of adaptive variations in the virus' genome, the covid-19 virus has become extremely dangerous and difficult to treat and vaccine [15].

1.2 Pathogenic mechanisms based on molecular mechanisms

Those who have contracted covid-19 have serious lung infections, like those with SARS & MERS. The symptoms induced by some cov phenotypes are very modest [16], such as HKU1, NL63, OC43, and 229E. ACE2 receptors attach to capsule particles and allow them to enter the human body.

This leads to the synthesis of massive polypeptides, which are then digested by proteases located in the rough endoplasmic reticulum of the host [13,17]. Protein cleavage, controlled by 3CLpro and PLpro [18,19], forms nascent virions by folding and assembling large polyprotein structures. A proteolytic enzyme found in the SARS-primary, vital role is played by Covid-19 in transcription and replication of the virus [20]. The replication of viral genomes requires another intriguing enzyme called RdRp [21].

Proteolytic proteins are important for the survival, reproduction, and propagation of viruses, which makes them a potential target in pharmacological drugs. It might be possible to develop more effective medicines if we can determine where sequence similarities and differences occur.

It is particularly important to develop and design small-molecule inhibitors for putative drug metabolizing enzymes. Molecular analyses revealed that these pocket regions of enzymes are evolutionarily conserved, and they align with their respective CoVs to a high degree. It is thought that the therapeutic compounds used to treat SARS-CoV and MERS-CoV could equally be used to treat COVID-19. As a result, there is a possibility that these pharmaceutical compounds could also be used to treat COVID-19. In contrast, SARS-CoV-2 exhibits significant structural differences from SARS-CoV spike RBD, specifically in two areas where it interacts with ACE2; as a result, previously licensed antibodies and therapy peptides targeting SARS-CoV did not

perform well for covid-19, requiring the growth of new structure upon drugs that address these structural changes. [24].

It has been suggested that CoVs have cytocidal effects and are also able to induce innate immunity in host cells [25]. The cytopathological effects of CoV infection such as cell lysis and apoptosis have been well documented. Synyctia are formed as a result of cellular fusion caused by the virus. The destruction of the Golgi complexes during virus assembly, the mobilisation of vesicles for replication, and the mobilization of vesicles for replication cause the above events in cells. It can cause cytotoxic effects in hepatocytes and syncytia development in lungs, in comparison with other coronavirus such as SARS-CoV & MERS-CoV.

The innate and adaptive immune systems play a role in the pathophysiology of SARS-CoV-2 [26]. Throughout the illness, T lymphocytes and signaling molecules such as cytokines exert an important influence on the course of the illness. Researchers have discovered that innate pattern recognition receptors (PRRs) substantially influence immune system reactions to covid-19 illness by identifying chemicals released by pathogens and secreted by injured cells

1.3 The genome of the SARS-COV-2 virus

ssRNA (+) is the only part of the covid-19 genome [27]. NCBI's database (NC 045512.2) contains the SARS-CoV-2 genome sequence, which has a size of 29.9 kilobytes [11]. The genetic composition of covid-19 consists of 13-15 ORFs, including 12 functional ORFs. There are Genes coding for 12 proteins and 11 expressed proteins in the genome, with 38% GC content. In terms of genomic arrangements, SARS & MERS share similarities 28 & 29. They are classified as replication and protease genes (1a-1b), and also include important proteins including S, E, M, and

N, in 5′3′ order of occurrence, which are regarded as vaccine targets. The genetic variations that underlie Fusion and entry of viruses, & persistence in the human are crucial [23].

Covid-19 is genetically identical to certain other Coronavirus to the tune of 89 percent. From GenBank, authors obtained the translational sequence for covid-19 polypeptide. A 7096-residue fusion protein accompanied by a number of structural and nonstructural proteins (NSPs) is found in the genomic structure of the entire SARS virus. Polyprotein pp1ab is translated by ORF 1-a via ribosomal frameshift under gene 1b, whereas pp1aa is encoded by ORF-1b via ribosomal frameshift underneath gene 1b. Virological genome proteinases break down these polyproteins even further, producing 16 proteins that are common to all the viruses in the same family.

Its genome contains 30119 nucleotides, which is much bigger than the genome of SARS-CoV-2. There are a few different kinds of genetic elements found in MERS, including 5′ components that support, poly(A) tails at the 3′ end, and 16 NSP genes from the 5′ end that are called nsp 1-nsp 16. An auxiliary component (ORF1, ORF2, ORF5, ORF8) is found at 3′ genome's end. Covid-19's genome consists of 4 structural genes (S, M, N, and E) & 5 AMG genes (ORF8, ORF5, ORF4b, ORF4a, ORF3). Covid-19 is significantly more contagious than SARSCoV and MERS-CoV because of its different epidemiological characteristics. It is likely that other vertebrates served as a "transition" or "exacerbating" host, with subsequent geographical isolation explaining the evolution of some or all of the changes essential for exposure to humans [30].

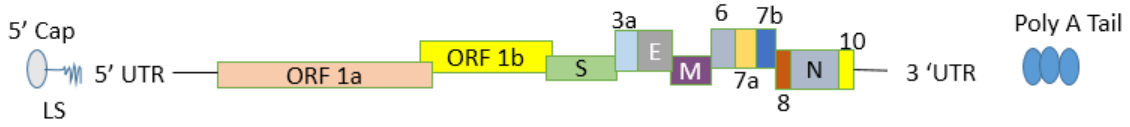

Fig 1.2 a poly-A tail at the 3′ end, and 5′ and 3′ UTRs make up the portrayal of 5′ covered mRNA.

There are strong similarities between the BAT-CoV genome and the Severe acute respiratory genome, suggesting a mammal source near Wuhan, China. It has been suggested, however, that bats are reservoirs naturally for Viruses, such as CoV-2 [32-34]. SARS-CoV2 may also have a Pangolin origin, establish on the strong identity of a specific gene [35]. The mechanism by which bat CoVs are transmitted to humans is still not known. Another study has found that dogs could be infected with SARS-CoV-2.

There are many sequence similarities between human ACE 2 and dog ACE 2 (13 of 18), which suggests that their interaction with SARS-CoV-2 spikes RBD is relatively similar, suggesting homologous transmission [36].

1.4 proteins for non structural function in sars-cov-2

In particular, a structural protein involved in capsid formation are NSPs present in the viral genome that play a variety of roles in viral assembly and dissemination [37]. Viral pathogenesis is controlled by these proteins by regulating early transcription, helicase activity, immune responses, gene expression, and antiviral defenses [38–40].

1.5 Spike Glycoprotein (S)

As a result, the spike glycoprotein plays a vital role in pathogenicity via its RBD binding to the human host [58]. The S protein attaches the virion to the host, starting the invasion process. Each subunit performs a specific function such as adhesion to the ACE 2 receptor. In effect, S1 helps viral particles adhere to cellular membranes by connecting with humans' ACE2 [17]. During the normal endocytosis process of the host, S protein undergoes structural changes [24]. A vaccine's development is critical to identifying structural changes within the target molecule that may alter immune function. Changing the structure of the protein, as well as its immunogenicity, might result from variations in the S protein. Despite the fact that the binding affinity area of SARS-S1 CoV-2 has changed multiple times, its ability to link with ACE2 remains unchanged in humans, pigs, civets, & bats, mouse ACE2 excepted [60–62].

Another components of the Spike protein, such as S2, function as heterodimers which help virus particles adhere to human cells. Three structural forms of the S2 protein can be seen during the fusion process: 1) the native state at pre-fusion, 2) transition state at prehairpin and 3) hairpin state at post-fusion. The topic is of great interest in determining how these structural variations affect viral entry into a cell's membrane, as they may lead to the design of more efficient therapies [59]. Fusion proteins are produced by residual cleaved S2' components of S proteins [63]. It was discovered that the stalk of SARS-spike CoV-2 is 99 percent identical to bat SARS-like and human Severe acute respiratory viruses, indicating that antiviral drugs targeting S2 regions of these viruses may be helpful in treating COVID-19SARS-COV-2. [15,64].

S2, other S proteins component, functions as a heterodimer that aids in virus particles fusing with the human cell surface. The S2 protein occurs in three primary structural forms throughout the fusion process: 1) pre-fusion native state, 2) prehairpin transitional assert, and 3) post-fusion hairpin condition. Understanding how well these variable structural states manage viral entrance

into the cellular membrane is intriguing because it could lead to the creation of efficient treatments [59]. The S protein's residual S2' cleaved component serves as a fusion protein [63]. As far as the stalk S2 nucleotide is concerned, SARS-spike CoV-2 is 99 percent identical to both bats and humans Severe acute respiratory, revealing a broad spectrum of antiviral drugs targeted against these viruses' S2 region, which may be effective in COVID-19 treatment [15,64].

Among the segments of SARS-CoV2 that can change the most is the RBD of spike protein. MARS-CoV has a distinct RBD composition compared to the originally recognized virus strains, based on an MSA analysis of the RBD of the S protein.

Virus attachment to human receptors occurs mainly through RBD's 90-amino-acid binding affinity domain, implying that many mechanisms are involved in disease. Disorder of covid-19 which has six residues that support adhesion to ACE 2, five of which diverge from Severe acute respiratory, a characteristic that should be taken into account in drug planning and administration. [16].

1.6 Envelope protein (E)

A subset of virus particles known as envelope membrane proteins aids in the synthesis and expulsion of the virus [65].

A possible therapeutic target for the SARS-CoV-2 protein is the E protein. Viral RNA assembly and development are controlled by the E protein [66]. Ionic conductance within cells is supported by nutrient pores generated by the E-protein which function as viroporins.

1.7 Membrane protein (M)

Protein complexes made up of M proteins has vital integral role in the packing of ssRNA, working in conjunction with N, S, and E proteins . These proteins share a common amino acid structure

due to their consistent lengths. A virus' characteristic form is determined by the proteins associated with membrane, which are the most prevalent proteins in Coronavirus. There are three envelope proteins in M proteins.

1.8 Nucleoprotein (N)

The nucleocapsid proteins (N) assist in the folding of viral RNA into ribonucleocapsids [68]. Coronaviruses with sequences 90% similar to SARS-CoV-2 share the peptide of SARS- CoV. Viral replication and transcription are aided by this protein by binding to viral genomes and M proteins. N proteins might therefore prove useful for therapy. Their RNA-binding sequence consists of 140 amino acids, found at their core, enables them to attach to viral RNA in a "bead on a string" fashion [65]. These three viruses exhibit substantial repetitions in their N protein MSA profiles. Because the Severe acute respiratory staphylococcus co-variant and SARS-CoV-2 N proteins have such a close sequence similarity, antibodies that detect SARS-CoV-2 N proteins are likely to detect SARS-CoV-2. There have been patches of minor sequence differences that indicate a separation in the evolution of the MERS-CoV strain.

1.9 Replicase polyprotein

It is also important to understand that replicase polyprotein has an impact on in improving degradation of host RNA &viral replication [21]. A majority of the viral genome's nucleotide content can be found in the non-structural ORFs 1a and 1b. Polypeptide repeat repeats, also known as replicase polyproteins, contribute to disease pathogenesis in a variety of ways [70]. Proteins like these contribute primarily to viral replication and transcription by assisting in the replication of viral RNA. There are a number of NSPs and proteases such as PLpro and 3CLpro, and each is

classified as either a NSP, a NSP2, a NSP3, or a NSP4. Virus RNA replication and transcription are fundamentally dependent on the RdRp domain in ORF1ab.

There are three domains that make up the replicase polyprotein: macro, papain-like, and major protease motifs. Furthermore, MSA of replicase polyprotein 1a reveals variable domains between all three domains on top of the minimal sequence comparison noted in the major protease. In all the major proteases of the SARS-CoV and SARSCoV-2 viruses, the sequence identity is higher (96%), suggesting they may have evolved together.. In addition to acting as an RNA-binding protein complex, it also plays a role in regulating gene expression, replicate proteolytic processing 1ab includes macro, papain-like, and major proteases. All four CoV strains are more conserved by MSA of replicase proteolytic processing 1ab.

CHAPTER 2

2.1 A description of the clinical features of COVID-19

Patients with covid-19 who are infected with COVID-19 are classified as mild, moderate, or severe. As disease severity increases, CD4+, CD8+, and B cells decrease in both relative and absolute numbers. Clinically, plasma proteins can contain activation of the proinflammatory cytokines while dendrite cell activation and B cell metabolism are suppressed. [71].

SARS CoV-2 disease often presents with lympopenia. A lymphocyte may migrate from circulation to lungs due to immunogenic activation, resulting in a lymphocyte shortage in peripheral blood. After a chronic & severe inflammatory response, the lymphocytes eventually apoptose and anergize. As a result, in different stages of disease, Lymphocyte Activity will vary dramatically [72].

A 20 percent percentage of COVID-19 patients present with severe problems resulting from an uncontrollable systemically hyperinflammatory immune reaction to severe acute respiratory infections, the so-called "cytokine storm." This illness is a result of a blood serum level of IL-6 that is too high. Anti-inflammatory and immuno-modulating treatments might be able to suppress this inflammation immune response in COVID-19.

A computer simulation (in silico test) reveals that the degree of COVID-19 clinical manifestations, as well as the ability of the body to react to SARS-CoV-2 infection, may vary with genetic variances. Liver tissue infected with a coronavirus is attacked by antiviral signals. A connection is made between these signals and the activation of leukocytes, particularly cytotoxic T cells, to destroy the invading organism.

2.2 How the immune system responds to COVID-19

The HLA gene variants are known to influence a person's vulnerability to viral illness. It is possible that genetic diversity in the key HLA genes is responsible for COVID-19 prevalence. I believe that addressing this heterogeneity will help improve the chances of diagnosing those who are at high risk for developing the condition. SARS-CoV-2 screening and HLA typing can be combined in the normal community to increase risk assessment. A vaccine against SARS-CoV-2 should prioritize those with high-risk HLA alleles.

If HLA antigens bind to a virus or segment of a virus, they create a physical manifestation on the cell membrane, which indicates the cells are contaminated and encourages certain immune cells to remove them. HLA systems are able to detect a greater amount of virus peptides the stronger the inflammatory system is. Using computer modeling, it was found that certain HLA alleles bind to a large number of peptides found in SARS-CoV-2, while the others do not. Consequently, it was established that HLA variations should be considered important biologically, i.e., they reflect a degree of disease lethality disparities. Changes in HLA loci do not seem to be the only factor affecting COVID-19 prevalence. Several HLA haplotypes have been associated with varying vulnerability to diseases, because T cell receptors can recognize the antigen-binding domains on HLA molecules. Thus, the result of the immune response is to produce HLA proteins that attach better to SARV peptide on the membranes of APCs. [73].

Once the virus has invaded the target tissue (CTL), cytotoxic T cells recognize the antigens expressed by Hla alleles on viral proteins.

2.3 Vaccines to combat Covid-19 infection

They used reverse vaccination and immunoinformatics in their research to develop epitope-based subunit vaccines for covid-19. A bioinformatics approach is useful in reverse vaccine development for examining the pathogen's genome and proteome, as well as identifying and investigating its neoantigens. A vaccine can be created using this technique by targeting specific viral antigenic regions, in a process that is more effective, simple, efficient, and cost-effective than traditional vaccine development processes.

As a result of viral infection, virus particles degrade into tiny fragments inside the proteasomes of infected cells. According to previous information, viral peptides are delivered to T lymphocytes by HLA atoms and molecules of infected cells. A structural or nonstructural protein may contain epitopes that T lymphocytes recognize.

It is highly recommended to develop vaccines against extremely antigenic regions. SARS-CoV-2 variants may elicit Antigen-stimulated CD4+ T Helper cells that stimulate B cells to produce a large number of specific antibodies, but there is currently little evidence that has been shown.. As a result, vaccine creation is now using this method, as it saves both time and money over traditional "test and error" approaches like "wet" lab experiments (unlike dry labs, which involve in silico experiments using computers). Immune responses against Covid-19 can be initiated in the normal community by T cell epitopes that trigger a robust immune response. On the other hand, molecular biology laboratories must investigate and verify the ability of such epitopes to function as vaccine candidates. Therefore, the majority of in silico analyses used to select epitopes for traditional ("wet") lab testing only considered epitopes that were capable of influencing CD4+ and CD8+ T cell responses in longitudinal fashion [74].

2.4 Case studies for vaccine development

Epitopes on proteins such as S, E, N, M, and ORFs have been identified by immunoinformatics on the covid-19 reference sample by Kiyotani et al. [75]. Researchers identified alleles of HLA-A, HLA-B, and HLA-C that appeared in a Japanese population at a prevalence of over 5. Afterward, we sought peptide epitopes to be given by HLA I & II both proteins that would be highly specific for the SARS-CoV-2 proteins. ORF1ab2168-2176 and ORF1ab4089-4098, which were anticipated to be highly affine to HLA-A*24:02, HLA-A*02:01, and HLA-A*02:06 in T cells, with a size of 83.8 percent, have now been shown to be highly affine to the Japanese population. HLA compounds containing these peptides can be used to track CTLs effects in individual & people who are affected but not symptomatic. The above epitopes have not been altered during sequencing. They believe that these potential epitopes could aid in creating functionalized vaccines to protect against covid-19 [76].

Based on a detailed analysis using inside silico computer simulations of ligand binding between both the substances class I for the 145 HLA-A, -B, and -C genotypes and the whole SARS-CoV-2 structure, HLA-B*46:01 antigen actually works by binding a relatively small number of the expected SARS-CoV-2 proteins, according to the analysis of pass immunity from the four widely distributed coronaviruses. Based on this finding, those possessing this type of gene more exposed to SARS-COV-2 infection, as has been demonstrated repeatedly with SARS-CoV, and it is consistent with clinical evidence that this variant is linked to serious disease. SARS-CoV-2 peptide sequences that are evolutionarily conserved have been reported for the HLA-B*15:03 allele [77]. Different HLA genotypes may modulate T cell responses, altering disease outcome and transmission. These peptides are predicted to transit the proteasome system. It is not dependent on the prevalence of HLA alleles in the community that SARS-CoV-2 peptides provide antigenic information. The demographic evolution of SARS-CoV-2 might impact the viral epitope repertoire

displayed or affect HLA-independent epitopes. Several different viruses can be tested using this method [78].In order to construct an epitope-based vaccination for SARS-CoV-2 using immunoinformatic methods to analyze viral proteomes, Joshi et al. found in interactions with HLA-allegic groups, peptides have been identified that are antigenic, nontoxic, and nonallergenic. This epitope is a major contender as an anti-SARS-CoV-2 vaccine candidate due to its improved binding indices in HLA epitope compounds and comparatively good stability, lethality, and population coverage [79].Human cells have been used most frequently to determine which immunegenic epitopes are most suitable for ACE2 receptors by using spike proteins (S proteins). Consequently, it is thought that the S protein could be a candidate for vaccine. Bhattacharya et al. [80] have identified the antigenic epitopes were transformed into a unique vaccine ingredient by using linker peptide which enhanced the resiliency and durable construction of the modelled and verified vaccine component. Because TLR5 appears to dock with the vaccine component, the response element is more likely to trigger immune pathways that kill specific antigens by serendipitous means. In other words, the antigenic covid-19 epitopes selected are promising candidates for developing an immunomodulatory multi-epitope peptide vaccine. Researchers V. Baruah and S. Bose identified important epitopes in Covid-19's S protein using the same immunoinformatics technique. The anti-SARS-CoV-2 vaccine may contain some of these epitopes. An immunogen that produces long-lasting humoral immunity should also stimulate CTLs. These investigations have detected three of these epitopes specific to covid-19 [81].D. Santoni et al. [82] conducted analyses using immunoinformatic method that examines viral peptides evolved from humans that are much farther apart than three mutational steps, in the hypothesis if evolution increases the evolutionary distance from the patient, the likelihood of adhering to HLA antigens increases. In order to find viral peptides that are not found in humans

(nullomeres), the scientists looked for peptides derived from viruses. To reduce the danger of autoimmunity, it is essential to identify the most distant peptides from humans. All reported SARS-CoV-2 strains had 25 of the 27 nullomeres in common. Researchers found nine peptides as a result of experimental research. The immune system sites identified in silico by covid-19 peptides need to be tested clinically to determine whether they are immunogenic. It is possible to engage 11 and 10 distinct HLA molecules with the YVMHANYIF peptide, according to in silico analysis, and the YYHKNNKSW peptide may interact with 8 HLA alleles, according to in silico analysis.An epitope-based vaccine is distinguished from a homogenous vaccine by the combination of both - and T cell-specific epitopes to generate an immunogen that is capable of drawn a good immune reaction. For vaccine development, peptides and epitopes are advantageous due to their simplified production technologies, chemical inertness, and incapacity to transmit disease. Molecular dynamics analysis revealed that the NOM–TLR4 & NOM–HLA-A*11:01 binding models showed essentially the same behavior [83]. This is because the composite itself contains both - and T cell epitopes.A complete list of immunostimulating peptides from the SARS-CoV-2 virus has been compiled via a computer program. Sarkar et al. [84] generated projection and clustering algorithms to identify HLA alleles that might interact with the SARS-CoV-2 epitope. They determined how well these antigenic, non-allergenic, and harmless epitopes attached to HLA molecules by estimating their 3D structures. TSNFRVQPTESI peptide from spike protein was most effective in binding HLA class I epitopes, while GVLTESNKK peptide from S-protein was most effective in binding HLA class II epitopes.The epitopes CV-1, CV-2, and CV-3 used as templates for docking were evaluated as effective models for anti-SARS-CoV-2 vaccines. CV-1 was found to be the best docking model based on molecular docking parameters. CV-2 had the best binding efficiency when HLA-DRB3*02:02 and HLA-DRB1*03:01 were present. It is possible that SARS-CoV-2

immunization may be possible if these verification experiments show impressive outcomes. Computer simulations have confirmed that the vaccine is nonallergenic, stable, and capable of eliciting humoral and cell-mediated immunity. The sequence of the defensin (TLR3 agonist) serves as an adjuvant for binding, improving immunogenicity. The vaccine is still being manufactured and undergoing clinical assessment to determine how immunogenic it is. A vaccine of this type may one day be developed for use in population health [85].A RNA-based vaccine can be created by using the self-amplifying RNA in covid-19's cytoplasm. The RNA sequence of the recombinant target molecule is used instead of the sequence of the specific antibody. The immunostimulatory protein is transcribed from vaccine mRNA once it is transported through lipid nanoparticles to the cytosol. When an antigen-presenting cell releases the protein, APCs immediately gather it. mWith the ability to express chimeric viral genes in the cytosol, cytosolic expression of chimeric viral mRNA may improve the safety, efficacy, ease of production, and viability of protein-based vaccines, as well as prevent virus chromosomal integration. There is no better approach to vaccine development than RNA-based vaccines, since they are capable of producing a huge quantity and are more timely in terms of pandemic response. A vaccine RNA injection may result in immune cells digesting and converting the RNA into a specific protein, which then stimulates other immune cells and leads to antibody production [86].By using mRNA-1273 encapsulated in lipid nanoparticles (LNP) and RNA nanoparticles (LNPs), Moderna's Ltd (USA) developed a vaccine that encodes the SARS-CoV-2 S-protein. When sufficient antibodies titers form against S-protein, a two-fold therapeutic benefit can occur, i.e., the body's immune system could remove the antigen-antibody combination, boosting viral elimination and reducing contagion [87] [88].It is efficient to use prefusion-stabilized protein immunogens that contain suppressant epitopes to prevent enveloped viruses. The discovery of alterations caused by stable Betacoronaviruses has led to

research into their structure and how they boost glycoprotein production and sensitivity through prefusion state spikes. inPhase III research is now evaluating the effectiveness of mRNA-1273.Two proline substitutions (2Ps) in the S protein at locations 986 and 987 have been discovered and proved to stabilize the pre-fusion configuration of the S protein by K.S. Corbett et al. [89]. Researchers created in silico serological assays without doing any additional experimental evaluation [90]. S proteins and beta coronaviruses suffer similar effects from the 2P alteration, making it possible to develop general vaccines by using antigens for S proteins. For pandemic preparedness, generalization is essential [91].The SARs-CoV-2 (mRNA-1273) S-protein (2P) synthesis began simultaneously with the preclinical assessment. A pathogen-based vaccine development programme can be improved and sped up by applying emerging technologies, such as synthetic vaccines. The mechanism by which proteins are placed and function is called glycosylation. Posttranslational modifications involve glycosylation. A high glycosylation degree on Covid-19 has significant genetic implications and has hindered vaccine development. Glycosylation of structural proteins in viruses impacts viral proliferation and penetration of cells, allowing them to evade immune responses. The development of specific antibodies targeting S-protein, however, may be enabled by mRNA-based vaccination technologies irrespective of its glycosylation state. Because mRNA-based vaccines do not incorporate DNA, do not induce autoantibodies, are easy to manufacture in mass quantities, and exhibit high purity, they are a feasible option for combating SARS-COV-2 [91].

CHAPTER 3

3.1 The Imunoinformatics Toolbox

Human and other genomes sequenced model organisms has provided a vast number of data that has proven useful in immunology research. Currently, multiple types of scholarly journals and clinical records contain vast data amounts from clinical and epidemiological studies. Researchers seeking mechanisms of immunity and disease development have a gold mine of information at their fingertips. Immunoinformatics emerged from the need to manage an immunological resource that continues to expand.

The intersection of computational immunology and experimental immunology is indeed immunoinformatics, often referred to as immunoinformatics. Comprehend immunological data using computational methods and technologies. Data mining is not only useful for managing large amounts of information, but also for developing new theories about immune function.

A vertebrate has both an innate and an adaptive immune system, according to immunology's basic paradigm. Its evolutionarily more conserved conservation makes it faster, older, and more evolved than adaptive immunity. Adaptive immunity evolves from it. An immune system's innate components are less specific than its adaptive counterparts, and serve as a first line of defense [92]. As defined above, a protective barrier is made up of 4 kinds: applications (ehin five to six days after early exposure to a pathogen, vertebrates develop inflammatory responses. The process is orchestrated by specialised cells, using chemical interactions and intercellular communication chemicals like cytokines and chemokines on the surface of the cells. Because memory is retained, a response is higher and more precise with repeated exposure to the same pathogen [93]. Two essential elements of the adaptive immune system are cellular immunity and humoral immunity .

The matching receptor on B or T cells recognizes a small region of an antigen known as an epitope. B cell epitopes contain amino acids that are straight and interrupted. These are short, linear peptides that act as T cell epitopes. In general, T cells tend to belong to either of two clusters. There are two different glycoproteins on the surface of T cells, labelled as CD8 and Among the features of the immune system that contribute to its complexity are its hierarchical structure and the way it combines. In the process, huge amounts of data are being gathered on immune systems. It is imperative that immunological research address this complexity. Immunologists have long employed high-throughput experimental approaches, which have generated massive amounts of data detailing immunity, therapeutics, and epidemiology. In order to facilitate the storage and analysis of this data, computer-assisted methods must be used. This led to the creation of the field of immunoinformatics. [94].

. It may lead to the discovery of new vaccines by examining binding sites. There is a procedure called reverse vaccination [95]. Traditionally, pathogens must be cultivated and then antigenic proteins are extracted. A member of the immunoome is any gene or protein involved in immunological response, excluding any gene or protein expressed by cell types other than lymphocytes [96]. A study of immune responses is called immunomics [97]. Immunome responses arise as a result of interactions between allergens and hosts. A novel field in biology and genetics, immunomics studies immune system mechanisms by making use of high-throughput approaches.

Molecular manufacturing and gene cloning can be used to manufacture proteins, or to clone their genes into an expression system. It is cheaper and non-infectious to use engineered molecules compared to viruses or bacteria. Etiophores play an important role in understanding disease mechanisms, determining host-pathogen interactions, discovering antibiotic targets, and

developing new vaccines. MHC molecules and antigenic peptides can be used to predict epitopes via their binding affinity. Studies have shown that experimental procedures can be lengthy and complex.

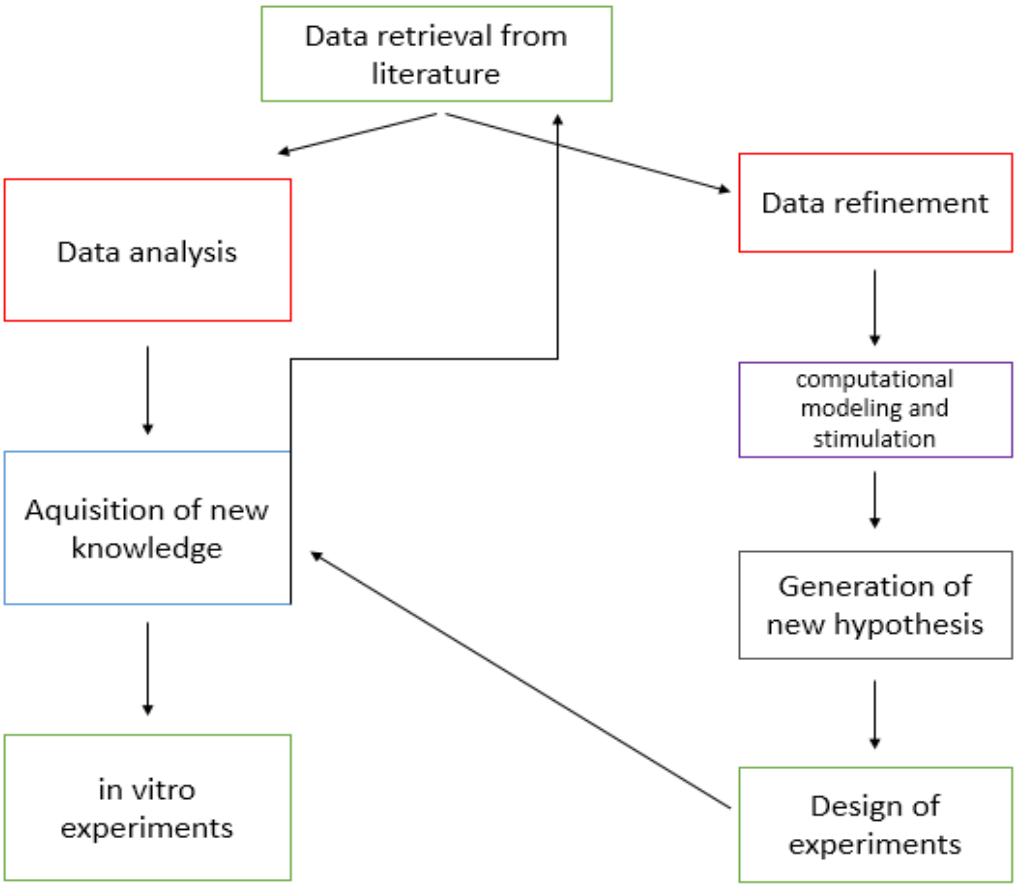

Fig 3.1 Work flow of immunoinformatics

3.2 Tool and algorithm for immunoinformatics

3.2.1Epitope prediction for B cells

The antigenic factors are referred to as B cell epitopes (BCRs) since they are found on the surfaces of diseases. 6 hypervariable loops varied lengths and aa content are found within the hydrophobic BCR binding site. Two types of epitopes are found on B cells: conformational epitopes and

continuous/linear/sequential epitopes. A typical epitope, on the other hand, is discontinuous, with residues at distant locations brought closer together by folding of the folded 3D protein structure. In most cases, linear epitopes are used. For discontinuous B cell epitopes, there is a prediction method available both for sequence-based and framework epitopes.

3.2.2 Epitopes as continuous B cells: Prediction of Methodology

Several approaches are used to predict continuous B cell epitopes, including sequence, amino acid affinity scaling , and machine-learning approaches.

3.2.3 Algorithms based on sequences

Sequence-based methods are often used to identify epitope region that reachable to antibodies. These methods can only predict continuous epitopes. Based on the identification of 2 protection epitopes in the HA1 [98], series approaches have been used to identify them. In rabbits, the 91–108 epitope (SKAFSNCYPYDVPDYASL) can evoke antibodies that improve flu-strain infection rates, which have been referred to as a defensive epitope [99]. A/Achi/2/68 (H3N2), protective against mice by the WTGVTQN epitope (127-133), is the second continuous antigen found in WTGVTQN.

3.2.4 Using a scale-based approach to analyze amino acids

In traditional approaches to detecting probable linear B cell epitopes from antigenic sequences, amino acid propensity scales are an important tool for detecting likely linear B cell epitopes. Protein sequences can be scored according to amino acid scales. Based on the i(n 1)/2 surrounding residues on either side of residue i, the score for protein I is computed. n proteins of the frame are scored according to the mean of scale values for residue i. The propensity scale approach of Pellequer [100] was tested on a database of 14 epitope-annotated peptides. Compared to other scales he evaluated [101], Parker et al. [102], Chou and Fasman [103], Levitt [104] and Emini [105] performed Improved. Two datasets were used to test the Na*ve Bayes classifier by El-Manzalawy et al. [106]: a probability dataset and a BCI-Pep dataset. Based on estimated protein turns, the Bepitope software tool [107] creates continuous epitopes. Using over 30

propensity scale variables, this version of PREDITOP [108] updates PREDITOP. The antibondy-antige interactions of proteins, which are predicted by both linear and conformational epitopes, are studied. To accommodate this, the AgAbDb [109] database that are created, Structures of antigen-antibody cocrystals are based on this mechanism.

3.2.5 Methodologies for machine learning

Machine-learning technologies and methods are used to recover epitope features. Following is a brief description of some of these methods. Using feed-forward and recurrent neural networks, Saha and Raghava [110] predicted Epitopes continuously expressed by B cells using ABCpred .Using COBEpro [111], biomarkers can be predicted for B cells using continuous B cells. During the first stage of COBEpro, fragment epitope propensity scores are assigned to protein sequence segments. It produces an epitopic propensity score in the second SVM scores of fragments are used to determine phase in the antigenic sequence[112][113]. Several propensity scale algorithms were examined on the Pellequer dataset [114], but in the end Levitt [115] came out as the best. They then evaluated these HMMs on the Pellequer database in order to identify the best parameters based on their predictions of linear B cell epitope positions. HMM was paired with Parker [116] and Levitt [117] as the two best approaches to propensity score estimation. Machine learning combined with amino acid scales can now predict continuous epitopes with an accuracy of 60–66%. It is possible to accurately track B cell epitopes by improving the quality of the current databases.

3.2.6 Developing a predictive method for discontinuous B cell epitopes

To portray and anticipate B cell epitopes accurately, The design of the protein in 3D should be known; accordingly, the forecast task is more difficult than for T cell epitopes. Changes in epitope number are related with changes in protein collapsing [118]. The most reliable technique for

deciding the epitope of B cells is X-beam crystallography. Andersen et al. introduced DiscoTope, which consolidates amino corrosive information, geological data, and surface openness data [119]. Roughly 15.5 percent of deposits found in intermittent epitopes are perceived, with 95% explicitness. This is the primary technique that is professed to perform better compared to strategies dependent just upon grouping information for anticipating irregular B cell epitopes. By joining the qualities got from amino corrosive and half-circle openness numbers [120], PEPITO predicts conformational epitopes in light of side chain directions and dissolvable availability of amino corrosive buildups. Also, the authors assert that it is more accurate than DiscoTope.For mapping structural B cell epitopes, Bublil and colleagues developed Mapitope [121]. Initially, Mapitope proposed that even the simplest of epitope segments was an AAP formed by folding the residues inside an epitope. A collection of fondness peptides was obtained after screening phage polypeptide libraries for antibodies. Following this collection, the algorithm returned one to three epitope options on the surfaces of the crystalline nuclei of the antigens.Using a computer technique, Solner et al. [122] told the best way to consequently choose and rank peptides that initiate in any case practically modified antibodies. Combining B cell epitope predictions with antigen variability and pattern conservancy was done for predicting posttranslational modifications (PTMs). Based on their findings, they saw that strong antigenicity, low variation, and a low probability of PTM were important characteristics for identifying biorelevant locations. This web interface predicts conformational and sequential epitopes as well as antigenic determinants [123]). The PDB was examined to acquire non-excess datasets of numerous three-layered designs of antigens and antigen-immune response buildings, recently gathered by Ponomarenko [124].Because protein antigens are not 3D primarily demonstrated, the server's worth is restricted. As of late, the arbitrary woodland calculation (RF) was applied to anticipate conformational B cell

epitopes, underlining because of inner buildups, various commitments from adjoining deposits, and an imbalanced informational collection, remembering an increment for nonepitope buildups contrasted and epitope buildups.

3.2.7 Methodology based on mimotopes

For protein–protein interaction research (e.g., antigen–antibody interactions), for identifying protein functions, and for therapeutic and vaccine research, phage display libraries are typically used [74]. Using a phage display library of randomized peptides, Pizzi et al. [125] show how to map B cell epitopes by searching for mimotopes that adhere to a target antibody strongly.This panel of mimotopes is thought to mimic the physicochemical characteristics and spatial structure of real epitopes [126, 127, and 128]. Both mimotopes and antigens are perceived by a similar counter acting agent paratope. The mimicked component of the epitope is the mimotope. There is a chance mimotope has any knowledge about epitope. The mimotope may, however, bear no similarity to the natural epitope of the antigen. Physiochemically and spatially, these organisms mimic one another [129]. The aforementioned MIMOP tool [130] was used to anticipate an epitope using mimotope pools. A degenerated alignment analysis is used for MimAlign, while a consensus identifier is used for MimCons. It maps a solitary mimotope or a gathering of mimotopes to an antigen design similarly as MIMOX [131] . It then looks for bunches of buildups that could frame the local epitope. Mapitope [132] (a high level server for mimotope-based epitope forecast techniques) and Pepitope [133] (a high level assistance for mimotope-based epitope expectation) contain Pepitope [133]'s high level help on mimotope-based epitope forecast strategies. These attributes are mapped onto each mimotope's surface. Pepitope's process is different from MIMOX's as a mimotope must be aligned first. Mimotopes rather than 3D structural information is identified at the alignment stage as opposed to the 3D structure being examined right away. The

mimotope's replicated surface area can be identified using it.These mimotopes were selected from the same phage display peptide libraries that had been screened with appropriate antibodies. They were sometimes mistaken for conformational epitopes. A software program developed by Scriber and colleagues [134] allows peptide sequences to be identified inside 3D protein complexes utilizing 3DEX. The method takes into account both the surface exposure of amino acids and the physiochemical proximity of C- or C- atoms. According to the authors, the authors found mimotopes within the 3D structure of Gp120 in HIV-positive plasma.

3.2.8 Ensemble prediction using Hybrid (Ensemble)

Gathering approaches frequently outflank and frequently surpass single indicators in biomolecular sequencing and underlying order examinations [135]. Presently, there are three different ways of joining various indicators, S, into a solitary agreement or meta-indicator: (1) larger part casting a ballot, (2) weighted straight, and (3) meta-learning. A few closest neighbor and choice tree-based classifiers are prepared by using various arrangements of preparing information attributes for developing a group of straight B cell epitope classifiers.

3.2.9 Epitopes that are specific to T cells

The current challenge is to accurately predict interactions between molecules in immunological forecasting software. Various MHC molecules have been used to estimate binding affinity for the most widely used approaches. To be perceived by cytotoxic T cells, antigenic peptides should be bound to MHC. To precisely foresee T cell epitopes, recognizing MHC-restricting peptides is of fundamental significance. Different strategies are used to find MHC-restricting peptides, including lattices, stowed away Markov models, fake brain organizations, support vector machines, and peptide structure.In request to precisely anticipate T cell epitopes, recognizing MHC-restricting

peptides is of vital significance. Different techniques are used to find MHC-restricting peptides, including grids, stowed away Markov models, fake brain organizations, support vector machines, and peptide structure.

3.2.10 Methods based on matrices

Involving BLOSUM and amino corrosive based vectors for direct T cell epitope forecast, Huang and Dai analyzed a clever peptide coding system for epitope expectation.Specifically, it substituted the diagonal columns of the BLOSUM matrix with the values in each nonzero element in the amino acid indication vector. MMBPred [136] predicts MHC binding peptides with very high affinities. Matrix data is used to predict the conformation of polypeptides, rather than polypeptide conformation. MHC alleles' matrix data is used to predict polypeptide conformation.It is crucial for preparing and presenting MHC I antigens that use transfer-associated protein (TAP). TAP transporters are capable of transporting amino acid peptides between 8 and 40 amino acids into the endoplasmic reticulum (ER). With PREDTAP [137], it is possible to predict the binding of peptides to hTAP. An activation function based on sigmoid function and a three-layer back propagation network was used. As inputs, nonamer peptides were represented by binary strings. Second-order HMMs were also employed. Hidden Markov Model-Based Method produces sensitive as well as specific results.

3.2.11 Automated Neural Networks Based Method

With ANNs, the relationship between neighbouring amino acid residues can be detected in a putative epitope. Based on associated input sequences and outputs, such as the MHC molecule's ability to bind that specific sequence, an ANN is trained for that particular molecule [138]. In theory, a trained artificial neural network (ANN) can predict binding affinity for new epitopes. By

combining neural networks and prior knowledge, Neilson et al. [139] have improved the predictive ability of T cell class I epitopes.

An HMM-based input was used along with scant encoding and BLOSUM coding. Buus et al. [140] determined the binding affinity of 528 amino acid peptides to HLA I molecule A*0204 by using the Buus et al. method. NetCTL [141] (http://www.cbs.dtu.dk/services/NetCTL/) can predict TAP transport efficiency and peptide MHC class I binding. NetMHC 3.0 [142] uses ANNs and weight matrices (http://www.cbs.dtu.dk/ services/NetMHC/). 67 HLA allele position-specific score matrices (PSSMs) and 55 MHC proteins (43 human and 12 nonhuman) were used to train the model.

Many elements add to the trouble of foreseeing MHC class II restricting peptides, including the length of announced peptides, the unclear center region for every peptide, and the quantity of amino acids utilized as anchors. Brusic and associates created PERUN [143], a crossover strategy to anticipate MHC class II restricting peptides. This approach uses both existing trial information and skill in restricting themes, developmental strategies, and counterfeit brain organizations (ANNs). PlaNet adaptation 5.6 [144] was utilized to plan and prepare a three-layered full-connected feed-forward ANN. A coordinated model for displaying the comprehensive corruption cycle of MHC class I ligands and their show, EpiJen [145] (http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm), has been created.

3.2.12 Methods other than machine learning

A subterranean insect province search framework, or ACS, has been demonstrated to be valuable in the arrangement of intricate streamlining issues, as well as in the distinguishing proof of different alignments.The point of an ACS [146] is to observe the best arrangement involving a

quest strategy for some random arrangement of peptides. The TAPPRED study [147] (http://www.imtech.res.in/raghava/tappred/) concentrated on nine amino corrosive qualities to distinguish a connection between amino corrosive restricting liking and physiochemical attributes. Cascade SVM has been shown to be more reliable than SVM for predicting peptide TAP binding affinity. Compared to other existing methods, cascade SVM is more effective. Based on data described using SVM and support vectors (SV), Nanni [148] demonstrated how T cell epitope can be predicted. In experiments, it has been shown that immunoproteasomes are involved in the production of MHC class I ligands. A tool called Pcleavage [149] developed for anticipate exactly how protein antigens will cleave to meet this purpose. Weka (Waikato Environment for Knowledge Analysis) (101), PEBLS (Parallel Exemplar Based Learning) [150], and SVMs [151] are employed in this system.

3.2.13 Predicting structure using structure-based methods

For foreseeing T cell epitope, latest methodologies depend on the peptide's partiality to tie to MHC. To concentrate on the hidden course of resistant acknowledgment, to foster fruitful peptide-based antibodies, and to plan immunotherapies, exactly recognizing peptides that tight spot to MHC atoms is basic [152]. The problem of peptide-MHC binding was solved with CoMSIA, an advanced 3D QSAR method. A set of aligned three-dimensional peptide structures is utilized to characterize binding. A T cell epitope [153] is predicted in silico by TEPITOPE [154]. TEPITOPE compares pocket profiles with HLA II for a particular peptide, finding similar HLA II with different binding capacities for that peptide.The only HLA-DR molecules of approximately 700 that are known can be used with it. The TEPITOPE skillet program, a web instrument to quantify peptide buildup similarity with the limiting pockets of HLA-DR particles, extrapolates from the known restricting locales of HLA-DR atoms. FudT Epitope Designer [155], a web application for

catching explicit peptide buildup secures, utilized virtual restricting pockets to work out peptide buildup similarity with restricting pockets. Zhao et al. [156] used 29 MHCp crystal structures to develop a model based. It offers a cumulative evaluation of how each peptide residue interacts with the virtual pocket-forming residue in the MHC molecule. The ElliPro [157] tool is an online tool capable of predicting antibody epitopes and plotting protein sequences and structures. Three approaches are used here: the ellipsoid-like shape is approximated, the residue protrusion index (PI) is calculated, and the residues that are nearby are clustered based on their PI values. It utilizes the 2D QSAR strategy for expectation of peptide-MHC allele particularity and works with both class I and class II MHC alleles. It is accepted that peptides that tight spot to the MHC on cancer cells can incite a safe reaction against the growth. By consolidating PeSSI (peptide-MHC restricting primary expectation through solvated interfaces), Schiewe et al. [158] fostered a mechanized technique for anticipating the adaptation of peptides to MHC atoms. Antigens from disease testicles called KU-CT-1 were used, which bound to HLA-A2.Using a known three-layered development of few MHC-peptide edifices, MHC class I groupings, known restricting energies for MHC-peptide buildings, as well as a bigger paired dataset with data about solid folios and non-fasteners, Jojic and teammates [159] fostered a viable design put together model based with respect to the information from past work. A twofold stringing strategy was utilized, in which the stringing model's boundaries could be learned and MHC and peptide groupings can be strung onto the construction of different alleles. Using a technique developed by Furman et al. [160], they analyzed a large number of MHC class I alleles. Candidate peptides are threaded in this procedure, and their binding compatibility is assessed using pairwise statistics. To analyze pairwise potential, Miyazawa and Jernigan used Miyazawa and Jernigan's table.Peptide vaccines targeting T cells are being developed using immunodominant peptides. [161] Altuvia et al. zeroed in on the antigenic

peptides recognized by cytotoxic T lymphocytes. To observe the peptide arrangements that fit into the MHC groove best, the it was utilized to string technique Pred observes MHC class II restricting destinations in antigenic protein successions utilizing a graphical internet based application. For 51 HLA-DR alleles, they used Sturniolo et al.'s pocket profile database . A wide variety of MHC class I and II ligand prediction algorithms as well as methods for predicting minor histocompatibility antigens can be found on the EpiToolKit. In addition, mutations can affect epitopes on T cells.

S.no	Author	Conformational & linear B cell epitopes	MHC I & II binding prediction	Toxicity	Population coverage analysis	Antigenicity	Allergencity
1.	Rakib et al.[121]	Linear – BepiPred & IEDB	NetCTL 1.2	ToxinPred	IEDB	VaxiJen,IEDB	AllerTop
2.	Tohidinia et al.[122]	Ellipro for both MHC I and II	ProPred I & II	ToxinPred		Vaxijen (>0.4 threshold) & Bcepred	AllergenFP
3.	Yazdani et al.[123]	BepiPred & IEDB, Discontinous-ElliPro	CTL Pred,RANK PEP			VaxiJen v2.0(Threshold at 0.5)	AllergenFP v.1.0
4.	Lin at al.[124]	IEDB and BcePred & Discontinous - DiscoTope2.0	IEDB	ToxinPred		VaxiJen v2.0 (Threshold >1)	Allergen FP 1.0
5.	Akhand at al.[125]	Linear-Bepipred	TMHMM & IEDB	ToxinPred	IEDB	VaxiJen	AllergenFP , AllerTOP , Allermatch

TABLE 1. Analyzes of immunoinformatics tools for SARS-CoV-2 vaccine development.

CHAPTER 4

4.1 Molecular Docking

Molecules dock in three dimensions when they are positioned together. In molecular docking, the relationship between molecules is explored by exploring the way they interact. Molecular docking is taken to refer to the interactions between proteins. Molecular docking has several forms when it comes to protein interactions.

When proteins interact with one another, they are called protein-protein interactions. Similarly, proteins bind to DNA in a protein-DNA interaction. Interactions between proteins and their ligands are called protein-ligand interactions.

The interaction of two proteins with similar sines is called Protein-Protein Docking. A thorough understanding of the structure of protein-protein complexes in three dimensions (3D) is required to understand how molecular systems work. A great many proteins are oriented towards their interactions with one or more partners. Protein-protein complexes participate in a variety of cellular processes, and the chemistry and biology of their constituents pose significant scientific challenges. The docking of protein-protein complexes formed by two or more proteins without experimental measurements of their molecular structure. Protein-protein interactions are normally more rigid, although the interfaces can be altered to improve binding and ease movement. As a result, the interface between the two molecules tends to be flatter and smoother than those involved in protein-ligand interactions. There are steric constraints that limit the ability of the body to modify and these are said to be rigid.

Known as protein-protein docking, protein-protein interaction happens when two proteins have comparable sines. Knowing the three-dimensional (3D) structure of protein-protein complexes is crucial to understanding molecular systems. It is common for microbes to be grouped based on how they interact with one or more partners. In addition to their involvement in cellular processes, protein-protein complexes play significant roles in regulating various biochemical and biological processes. While there does not yet exist a general recognition code, the DNA's global conformation can determine how the eventual interaction surface will be modulated. DNA-Protein complexes can be predicted using a silico approach through protein-ligand docking, a process that checks the structure, position, and orientation of proteins.

A protein motif binds tightly to its ligand, and is referred to as a lock and key mechanism when it interacts with small molecules like ligands. Specifically, this work deals with computer-aided drug design problems, which arise from issues associated with the design of bioactive compounds. This module further divides Protein-Ligand Docking into five segments for convenience to the user, in order to predict and rank the structures that will occur as a result of the association of a given ligand with a given target protein.

- Ligand Preparation
- Preparation of receptors
- An analysis of the binding site
- Analysis of docks

AUTODOCK is used for docking. Spike protein is the receptor molecule in all variants and PUBCHEM molecules. We report the molecules with a higher binding energy than the natural duck molecule for further investigation.

It is costly and time-consuming to develop new therapeutics. Most of the time, new therapeutics do not become available for several years after they are developed. It is possible to find new uses for pre-existing medicines that have been tested in a variety of therapeutic contexts or with a variety of defined molecules, even if they were unsuccessful treatments at the time [161]. It is essential that existing medications be used for COVID-19 treatment since it has spread rapidly around the globe.

The average time to develop a new drug is 10 to 15 years. The majority of compounds examined did not advance to advanced stages. Among 22 phase 1 trials since 1995, 10% of compounds made it to the clinical stage. In order to detect and reject targets at any stage of the drug development process, computational tools like molten docking are now being used. This lengthy process can be greatly simplified using computers.

Accelerating the process can be achieved by repositioning medications. Repositioning of medication or repositioning of drugs (Rudrapal, Khairnar et al. 2020) is also known as medication repositioning. It is intended to uncover new applications for existing treatments that have not been documented in the medical record through repurposing pharmaceuticals. This leads to shorter time to market, a more established distribution network, a cheaper price, and significant improvement in clinical data due to the reduction in pipeline. In preclinical and clinical testing, scientific data are very important since they enable a faster process. As 45 percent of drug leads fail due to toxicity and safety, this technique would also increase chances of success.

There are, however, challenges associated with relocating medications. Many medications are only licensed to be used for certain illnesses. An expanded clinical study of a repurposed pharmaceutical

is necessary, as well as determining the appropriate dosage. Delivery systems are also of concern. The targeting technique of a particular chemical may need to be altered in order to reach a certain tissue, which would require several pipeline stages. Drug repurposing requires looking at how a drug interacts with other drugs used with the new drug as well. Although repurposing poses difficulties, it can expedite the development of a drug and save lives.

Computer technologies are the most efficient way to increase the possibility of repurposing drugs. This can be accomplished through many different computational methods. Medicines with known interactions with specific disease proteins can be found using virtual screening. Medicine molecules may be compared to all crystal structures in the database to see if they react with any crystal structure. By searching molecular libraries, one can identify likely, less effective, or free analogues to treatment targets. Utilising networks allows us to link illnesses with biological processes, and we may understand the efficacy of drugs in one context by using them in another. A molecular docking technique is often employed in such circumstances.

4.2 Methods and Procedure

4.2.1 Docking protocol

Step 1. **Preparing target protein**

Input proteins must contain polar hydrogens and have all water molecules removed before they can be docked . We will get errors if this is not the case.

- ref.pdb file is opened in ADT

  File > Read Molecule

- The polar hydrogens are added to the objective structureFile>Save>Write PDB >ref1.pdbStep 1.* Merge the nonpolar hydrogens

- Edit>Hydrogens>merge
- Nonpolar* Kollman charges are added to the peptideEdge>Charges>Add Kollman Charge* The construction is savedFile>Save>Write PDB >ref1.pdb
- *Introducing the docking tool.

  ligand>Input->;Open...

  change the file type to '.pdb', pick Cefuroxime saved in '.pdb' format> Open

  ADT establishes the docking record and adds gasteigercharges to the ligand structure. In the case of peptide ligands, Kollman charges would have been added.

- Predict the root of the ligand from its Torsion Tree.
- Ligand>Torsion Tree>Detect Root...

  View rotatable bonds; Select number of twists

  The Ligand is saved in PDBQT record design as 'Ligand.pdbqt'.

  Ligand>Output>Save as PDBQT

Step 2. Grid map preparation

Grid is placed on the objective dynamic site and should contain everything a ligand is likely to interact with. ADT opens the document dialogue to save the initialized target protein in '.pdbqt' format as soon as the record opens. * A target protein is selected where the ligand will dock.

Grid>Macromolecule>Choose..(ref1)

Select the protein target prepared in sync 1. Immediately after the document opens ADT opens the document save dialogue to save the initialized target protein in

'.pdbqt' format. This construction is saved in thedirectory in which we will run AutoDock.

- Setting the Maptypes that will be utilized for the grid.

  Grid>Set Map Types>Choose Ligand.(Ligand)·

- Selecting thebinding site residues.

  Select>Select From String Enter the accompanying subtleties in the 'Select FromString'

  discoursebox:Molecule>ref1Chain>AResidue>ASN331,ILE332,THR333,ASN334,LEU335,ALA520,PRO521,ALA522THR523,VAL524

  Atoms>;*>Add·

  Setting the GridBox position and size.

  Grid Box...Setting every one of the elements of the lattice box to 60points.

  Using the (181.568 186.477 192.798) x, y, z coordinatesfor the focal point of the crate and dispersing 0.402.·

- Saving the currentgrid positioning

  File>Close Saving Current· Saving Grid fileGrid>Output>Save grid.gpf

- In the document save exchange expressly type "grid.gpf" tosave the framework boundary file.·

  Running AutoGridRun>Run AutoGrid..

Step 3. Arrangement mooring boundary file· Setting theprotein focus to be docked

Docking>Macromolecule>Set filename>ref1.pdbqt·

- Ligand is selectedwhich is to be docked

Docking>Ligand>Choose > Cefuroxime.pdbqtA

ligand boundaries window opens and all boundaries at the default values are selected.· Setting search parameters

Docking>Search Parameters>Genetic Algorithm...

- Following changes are made to the parameters:
- Number of runs - >; 20Population size - > 150

  Rest of the boundaries left at the default values· Docking parametersare set to default· Setting theDocking yield boundary file

  Docking>Output>Lamarckian GA.
- Save the record as "Dock.dpf".·
- Running AutoDockRun>Run AutoDock.Step 4. View Docked compliance energies· Open the dock.dlgfile·
- Check theCLUSTERING HISTOGRAM and RMSD TABLE·
- Choose the 1 stconformation - note down the limiting energy· After the AutoDockcompleted its run, the outcome is seen in ADT.
- Analyze >Dockings->;Open>Dock.dlg· Macromolecule isloaded into viewAnalyze>Macromolecule>Choose..(ref)·
- Selecting the macromolecule

  Select>Direct Select>Molecule List·
- Viewing the conformations
- Analyze>Conformations>Play, positioned by energy…

  Select the & symbol

  Select > build H-bonds

Select > Show info

Select >Build Current

- Goto edit > Delete Molecule >select all >except the conformation lastly saved >delete
- File > Save > Write PDB > ligand_conf3.pdb

4.3 Result and Discussion

Several of the drugs evaluated have the potential to bind variant spike protein at the RBD, according to an in silico binding study. In humans, through the ACE-2 receptor, spike protein interacts with the receptor. In addition to ASN331, ILE332, THR333, ASN334, LEU335, ALA520, PRO521, ALA522, THR523, VAL524 are sites that are examined [162]. Docking research indicates that ampicillin and cefuroxime bind strongly to RBDs. In Table 1, we list the RBD residues involved in the hydrogen bonding, Hydrophobic interaction, binding affinity and salt bridge.

Infections are spreading so rapidly, drug repurposing is critical to developing new treatments as soon as possible. To determine if any of these medications might be utilized to treat variant, this study analyzed medications for potential use. We have demonstrated that drug repurposing is an effective technique for identifying lead compounds for future exploration of the Spike glycoprotein of variant using this overview of how to repurpose drugs.

Compounds repurposed in this study have been identified as potential targets for Variant's S protein. Computational docking studies have found Umifenovir and Ampicillin to be highly affinity for the variant spike protein RBD. Medications such as these may be able to prevent spike protein from attaching to the human receptor (ACE-2) , which could prevent infection. The result

may be that these drugs prevent viral replication by inhibiting viral enzymes' proteases. A further in vitro validation of the medicines' binding potential to the target proteins is required since the study relies on in vitro methodologies.

Table 2. Experimental Molecule Structures and Interactions of Wuhan_HU

Ligand	Binding energy	Hydrogen bonding	Hydrophobic Interaction	Salt bridge
CEFUROXIME	-5.57	ASN331 ASN331 ALA522 ALA522 ASN544 GLN564	ASN331	-
UMIFENOVIR	-4.66	ASN331 ASN331 ASN331	PRO330 ASN331 ILE332 PRO521 THR53 GLN54 LEU582	-
CEFOTAXIME	-5.78	ASN331 ASN360 ALA522 ALA522 ASN544 GLN564 PRO579	PRO521 GLN54	
CEFTRIAXONE	-6.45	PRO330 ASP364 SER366 ASN388 PRO527 LYS529 SER530		ASP364 ASP364
AMPICILLIN	-6.60	LYS537 CYS538 PRO589 HIS625	LYS537 VAL622 ALA623	-

TABLE 3. Experimental Molecule Structures and Interactions of Alpha variant

Ligand	Binding energy	Hydrogen bonding	Hydrophobic Interaction	Salt bridge

CEFUROXIME	**-6.63**	**ASP361 SER527**	**PHE326 PHE326 VAL359 LYS529**	**ASP361**
UMIFENOVIR	**-6.78**	**LEU332 PH335 GLY336 GLU337**	**LEU332 LEU332 PHE335**	**-**
CEFOTAXIME	**-4.24**	**THR330 CYS522 CYS522 SER527 SER527**	**PRO327 ILE29**	**-**
CEFTRIAXONE	**-5.56**	**PRO327 ILE329 CYS358 LYS525 LYS525 SER527**	**PRO524**	**-**
AMPICILLIN	**-7.74**	**ILE329 CY522 LYS525 SER527**	**PRO327 PR327 ASN328 ILE329 ILE329 VAL359**	**-**

TABLE 4. Experimental Molecule Structures and Interactions of Delta variant

Ligand	**Binding energy**	**Hydrogen bonding**	**Hydrophobic Interaction**	**Salt bridge**
CEFUROXIME	**-7.89**	**ALA520 LYS526 \LYS526**	**ILE330 ILE330 LYS526**	**-**
UMIFENOVIR	**-6.36**	**ASN358 THR521**	**ILE330 VA360 VAL360**	**-**
CEFOTAXIME	**-7.38**	**PRO328 ALA520 CYS523 LYS526 LYS526**	**VAL360 THR521**	**-**
CEFTRIAXONE	**-7.24**	**PRO328 THR331 ASN358 THR521 CYS523 GLY524 LYS526**	**-**	**-**

AMPICILLIN	**-6.12**	**ALA520 CYS523 GLY524 LYS526**	**PRO328 ASN332 LYS526**	**-**

TABLE 5. Experimental Molecule Structures and Interactions of omicron variant

Ligand	Binding energy	Hydrogen bonding	Hydrophobic Interaction	Salt bridge
CEFUROXIME	**-6.23**	**LYS525 SE527**	**PHE326 VAL359**	**ASP361**
UMIFENOVIR	**-7.48**	**LEU332 PH335 ASP36 GL337**	**LEU332 PH335 PHE335 ASP336 LEU365**	**ASP336 ASP336**
CEFOTAXIME	**-6.68**	**ILE329 CY358 CYS358 CYS522 GLY523 LYS525 SER527**	**ASN328 LYS526**	**-**
CEFTRIAXONE	**-6.78**	**LEU332 ASP336 GLU337 LYS553 SER556 SER556**	**-**	**GLU337**
AMPICILLIN	**-7.90**	**PRO327 GLY523 LYS525 SER527**	**PHE326 PRO327 ASN328 ILE329 VAL359 LYS526**	**-**

4.3.2 Ligand-Protein Visualization through Pymol :

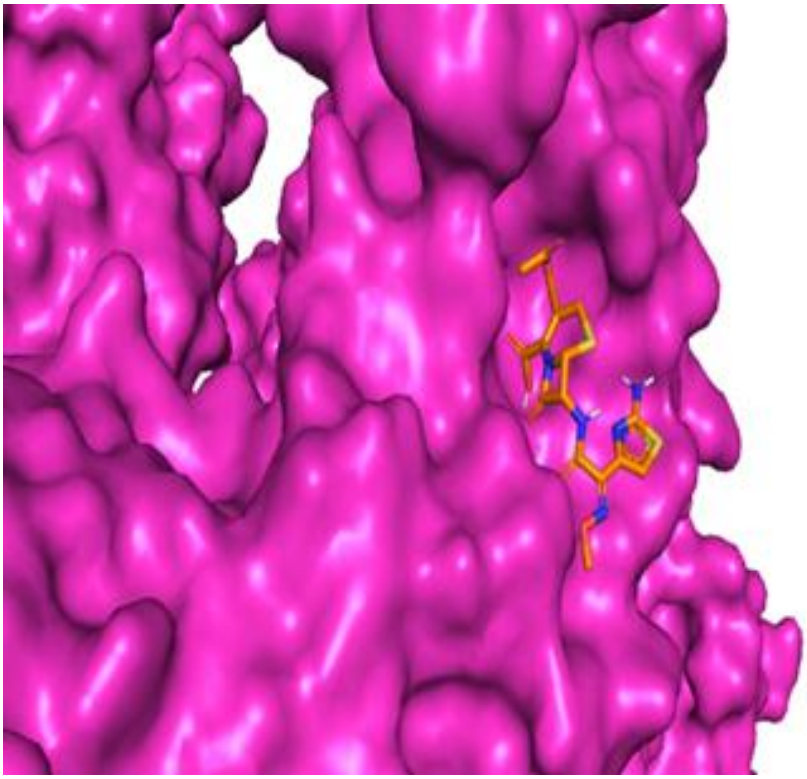

Fig 4.1 Molecular Interactions between Cefotaxime and Reference

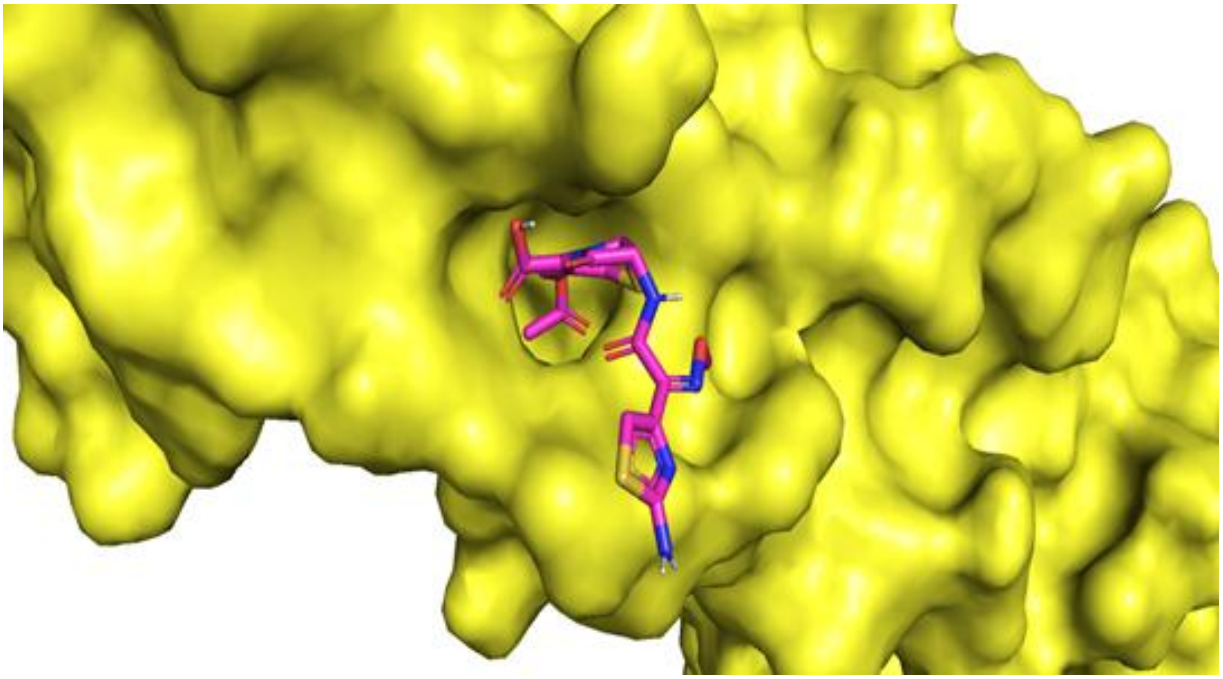

Fig 4.2 Molecular Interactions between Cefotaxime and Alpha

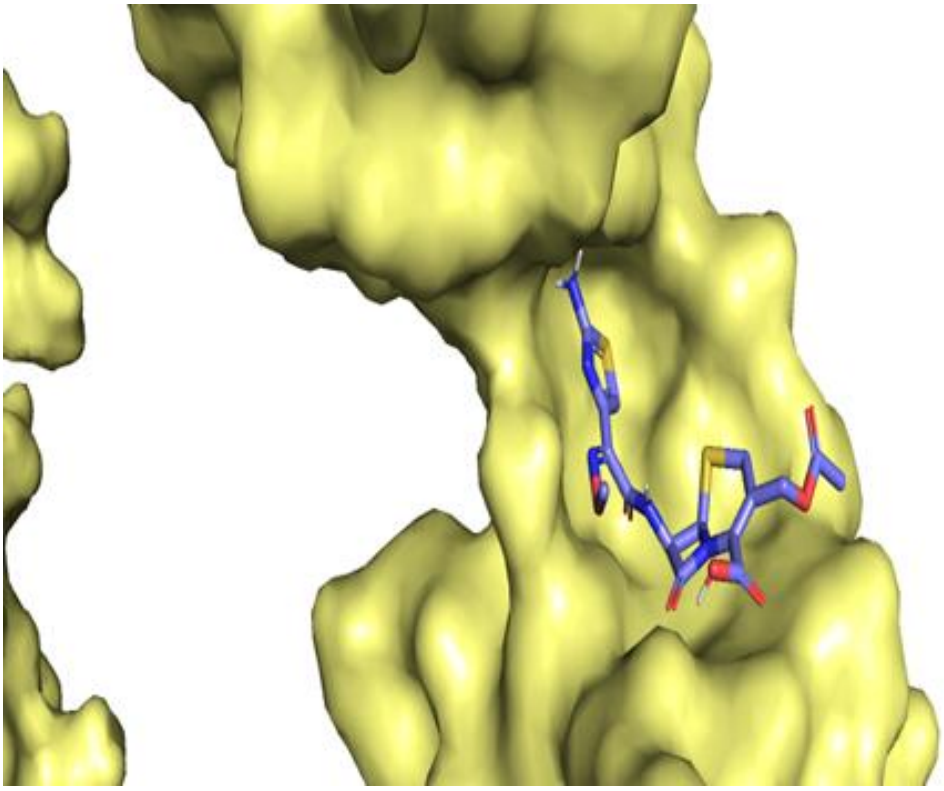

Fig 4.3 Molecular Interactions between Cefotaxime-Delta

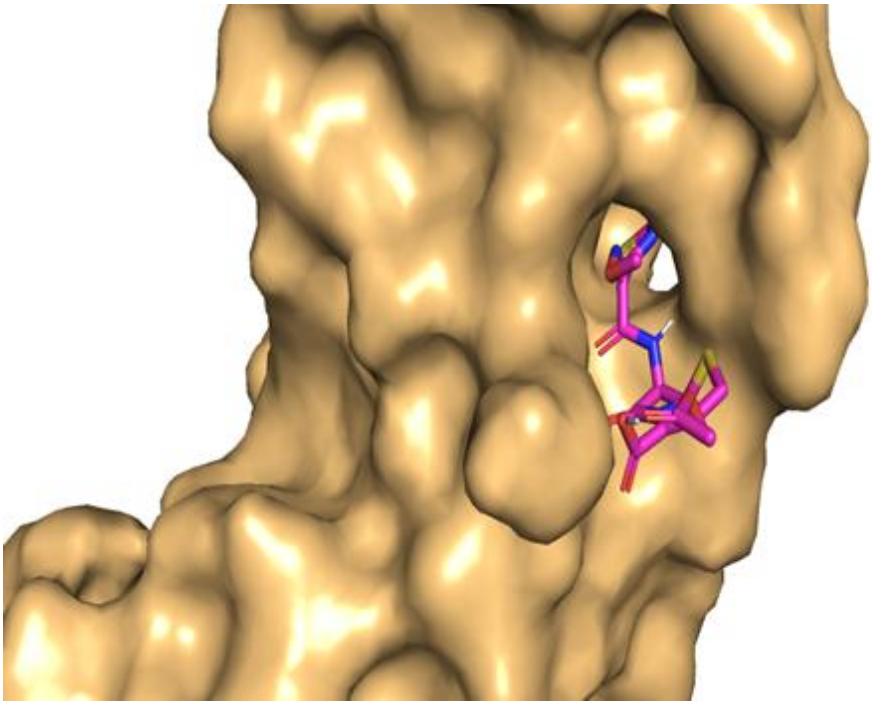

Fig 4.4 Molecular Interactions between Cefotaxime-Omicron

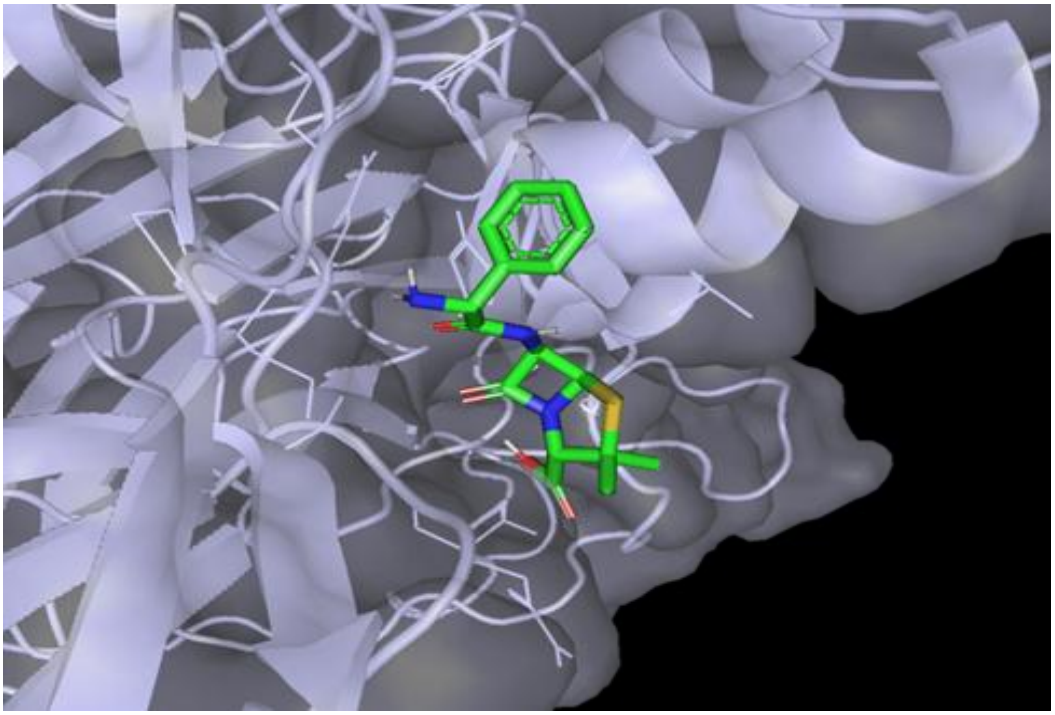

Fig 4.5 Molecular Interactions between Ampicilin-Reference

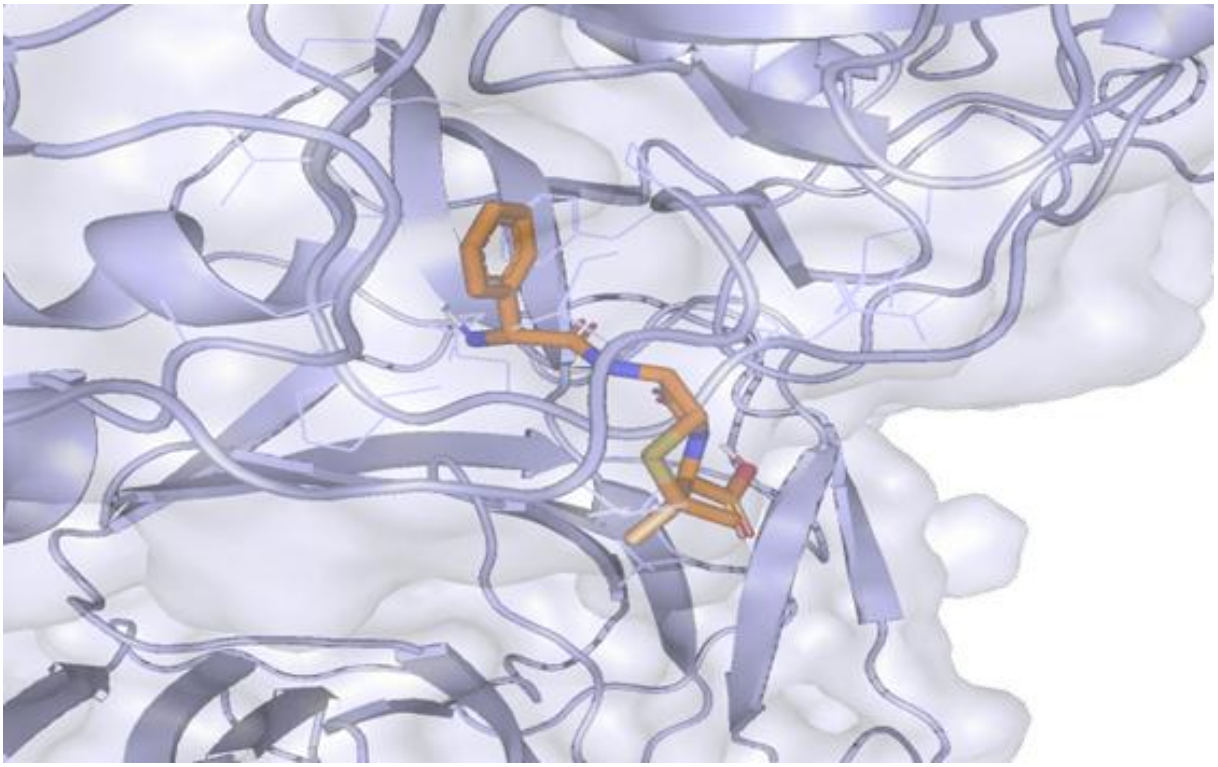

Fig 4.6 Molecular Interactions between Ampicilin-Alpha

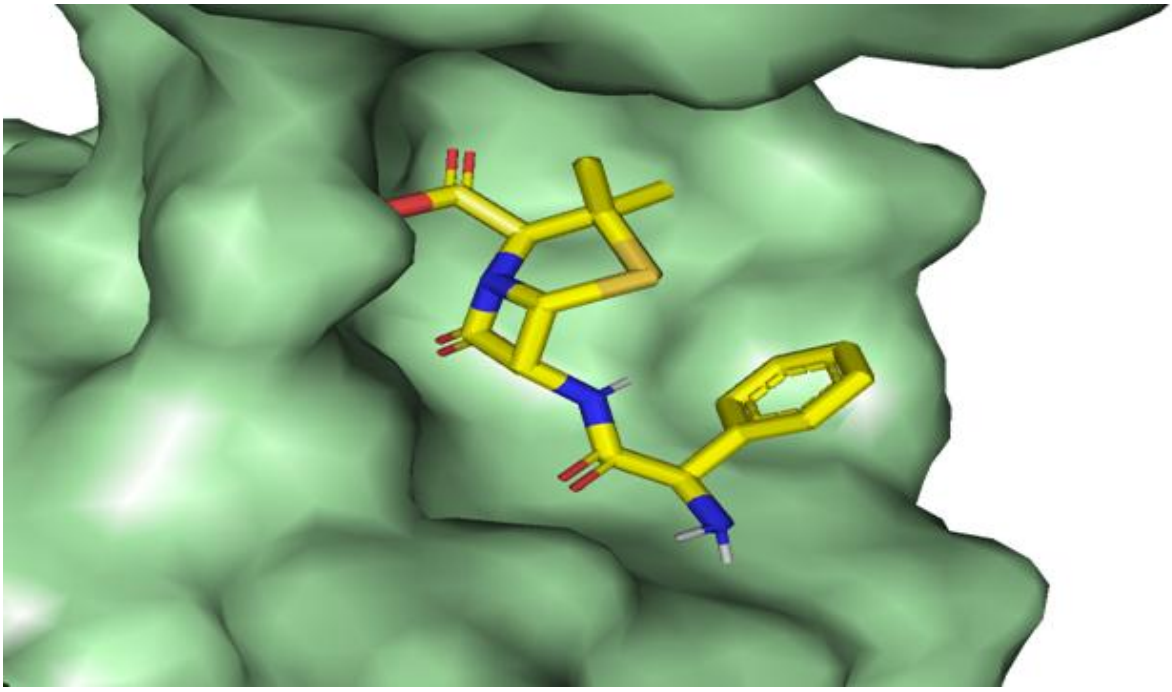

Fig 4.7 Molecular Interactions between Ampicilin-Delta

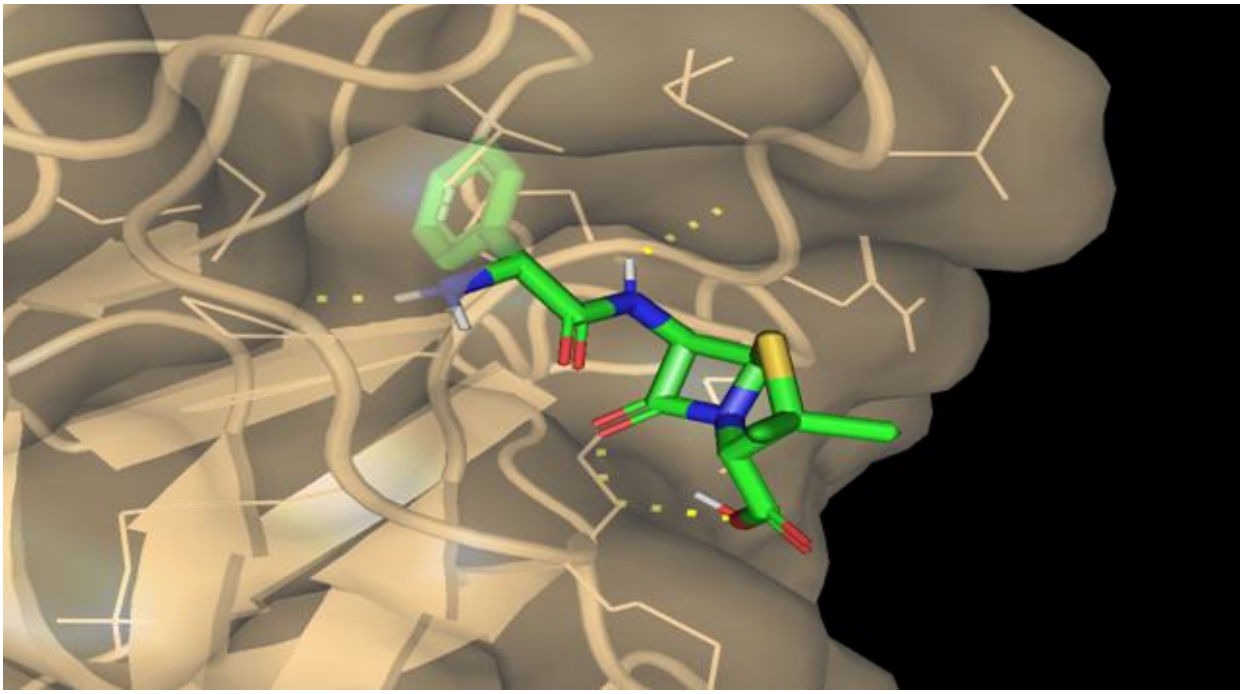

Fig 4.8 Molecular Interactions between Ampicilin-Omicron

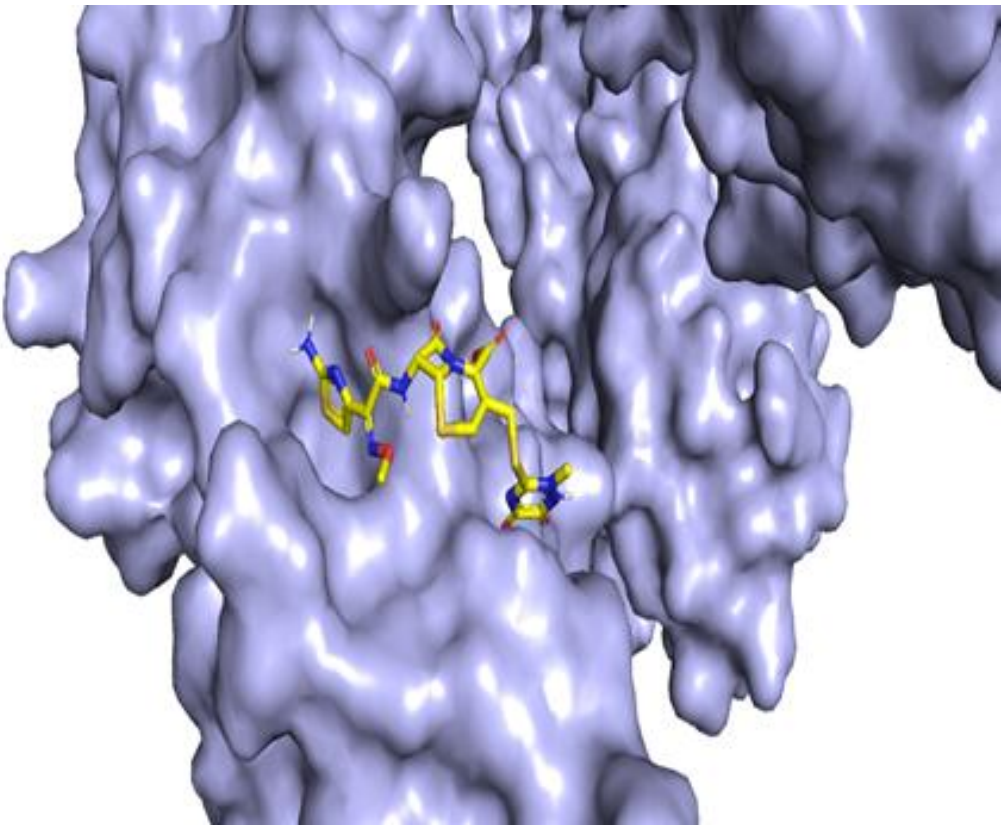

Fig 4.9 Molecular Interactions between Ceftriaxone- Reference

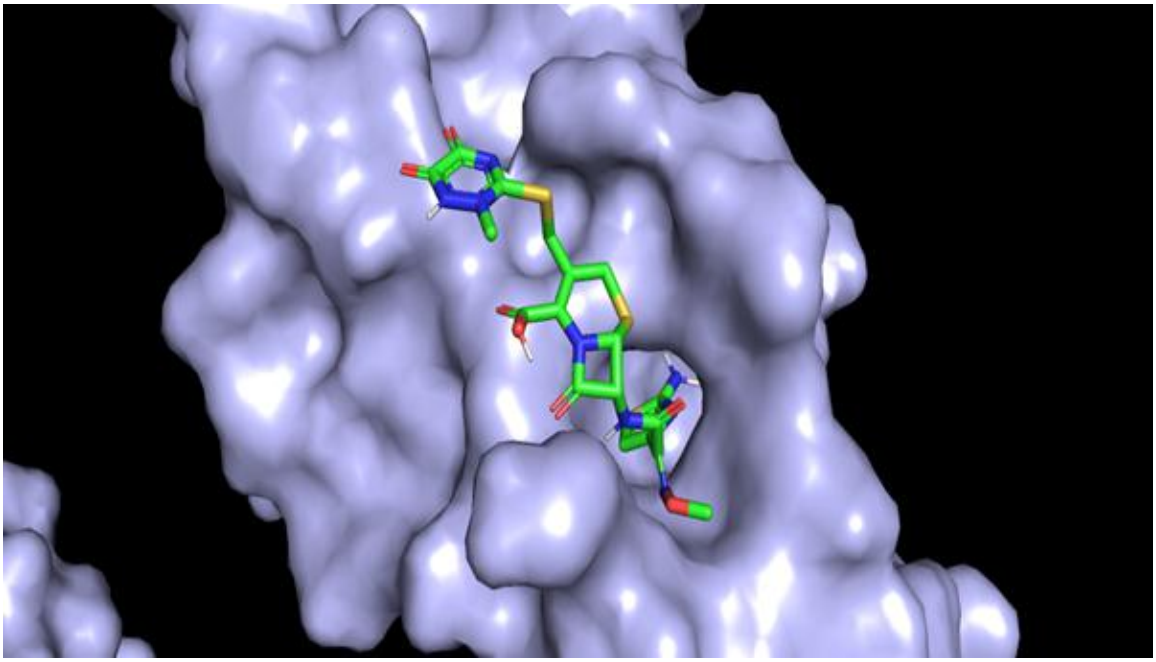

Fig 4.10 Molecular Interactions between Ceftriaxone- Alpha

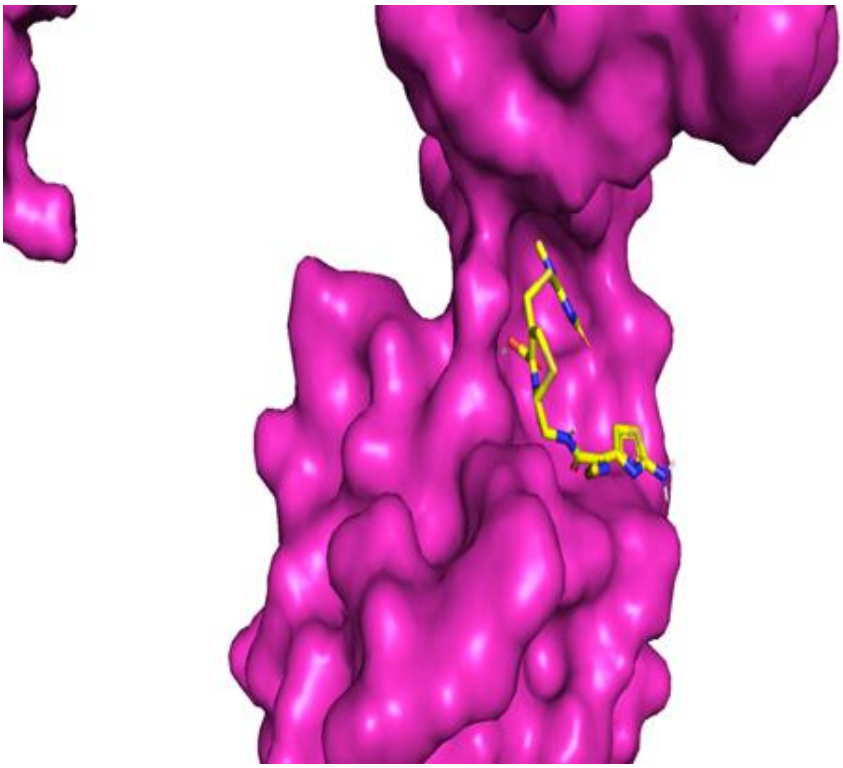

Fig 4.11 Molecular Interactions between ceftriaxone - Delta

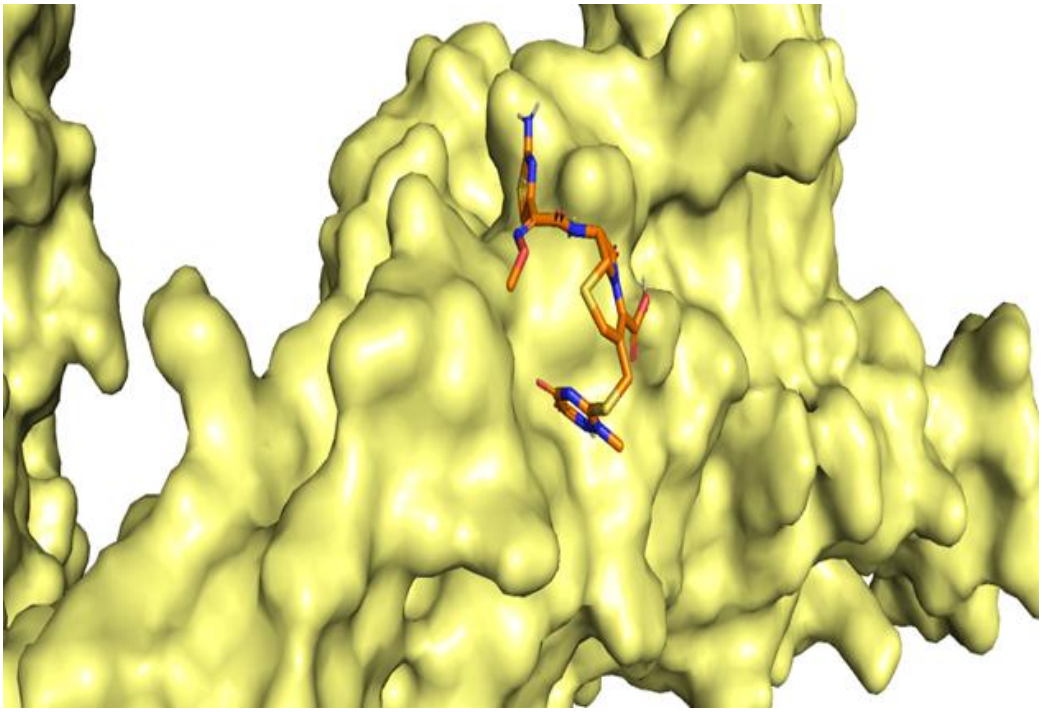

Fig 4.12 Molecular Interactions between ceftriaxone - Omicron

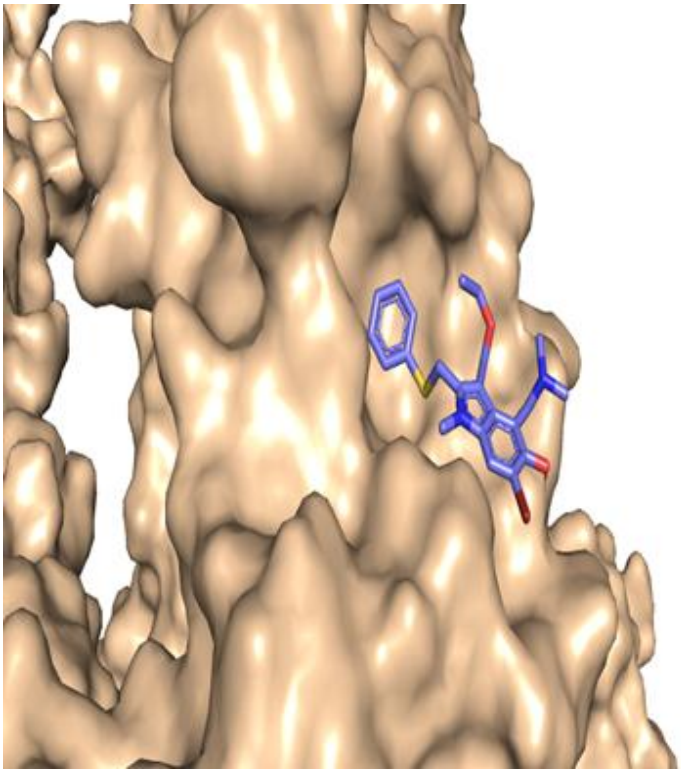

Fig 4.13 Molecular Interactions between umifenovir - Reference

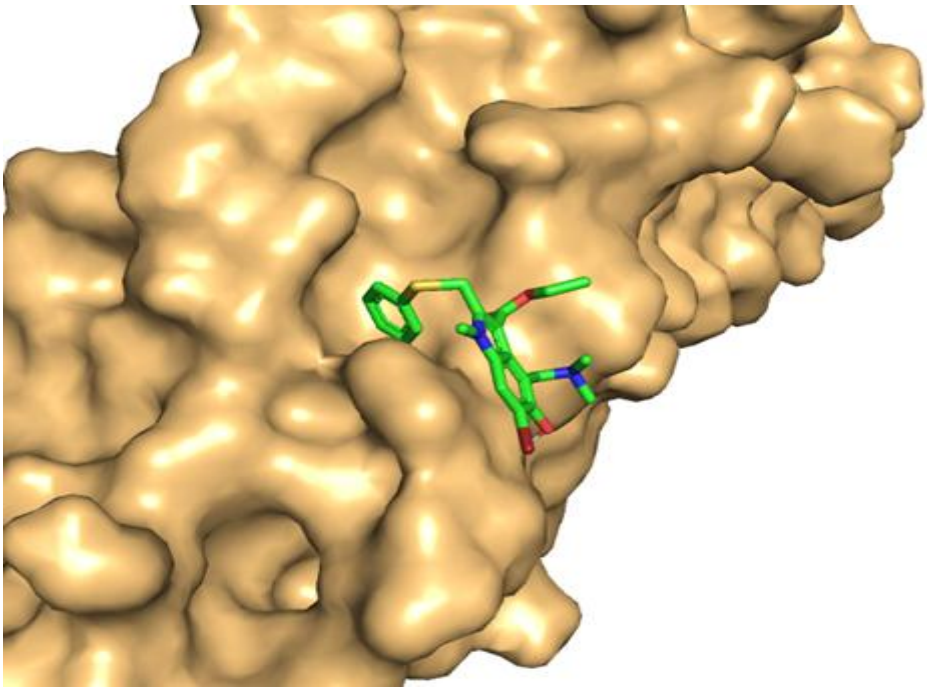

Fig 4.14 Molecular Interactions between umifenovir - Alpha

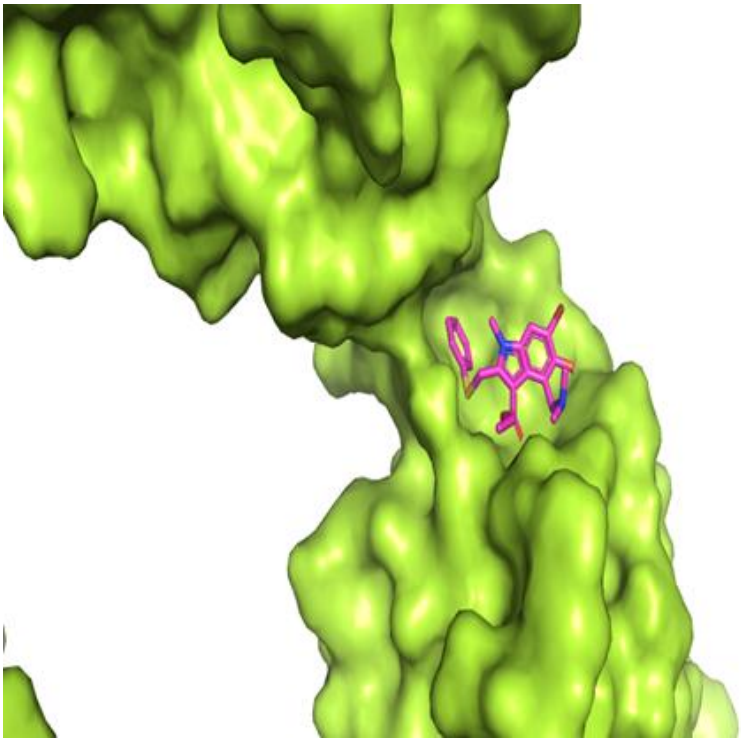

Fig 4.15 Molecular Interactions between umifenovir - Delta

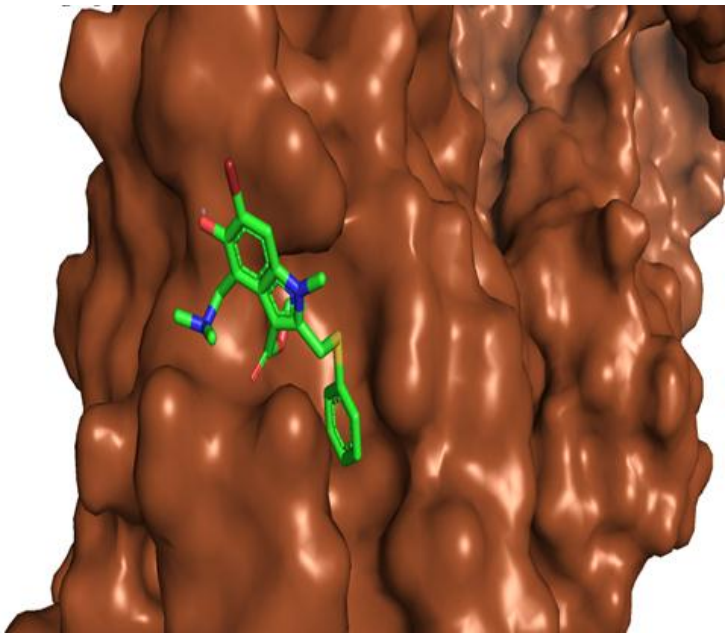

Fig 4.16 Molecular Interactions between umifenovir - Omicron

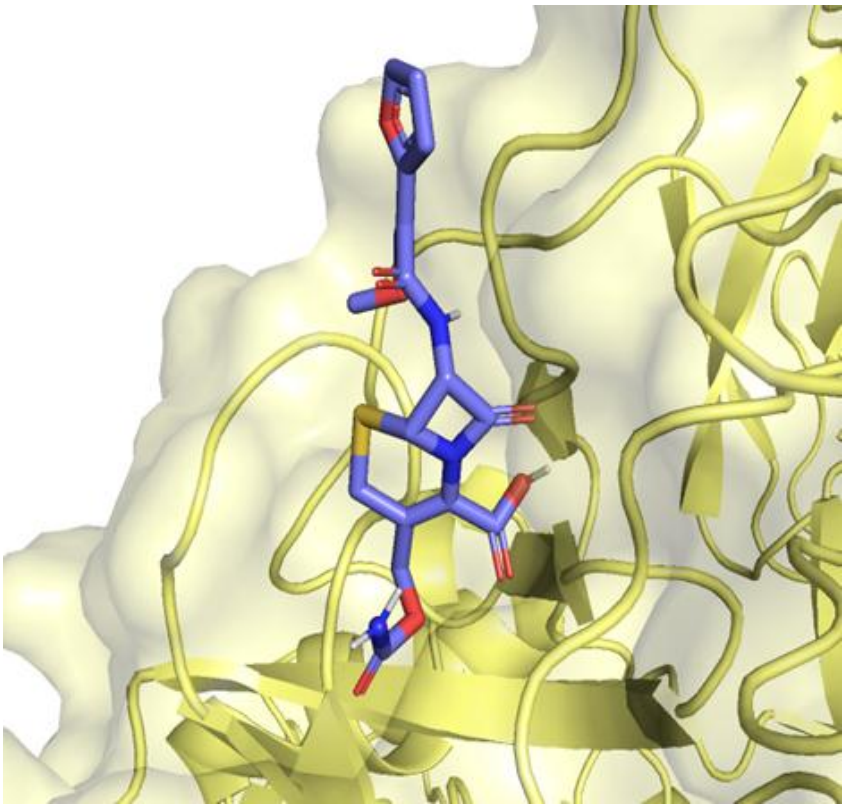

Fig 4.17 Molecular Interactions between cefuroxime -Reference

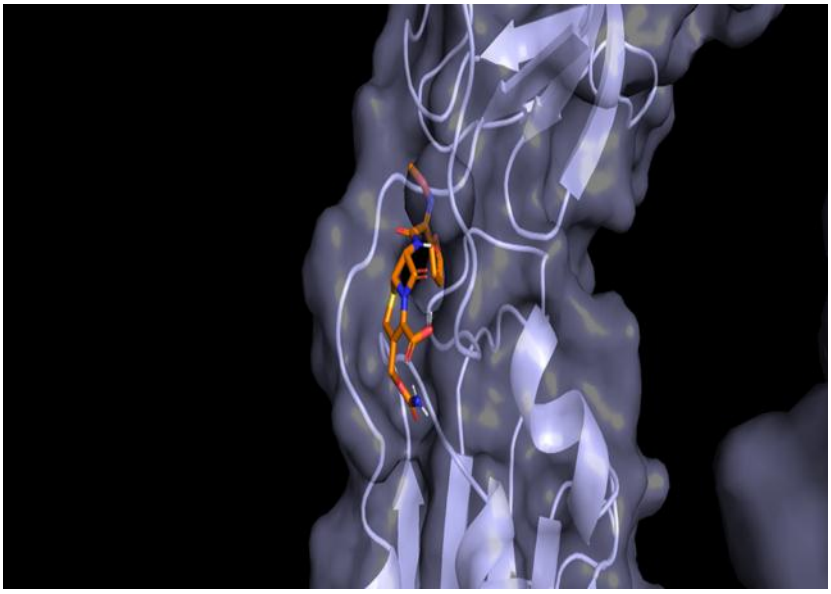

Fig 4.18 Molecular Interactions between cefuroxime - Alpha

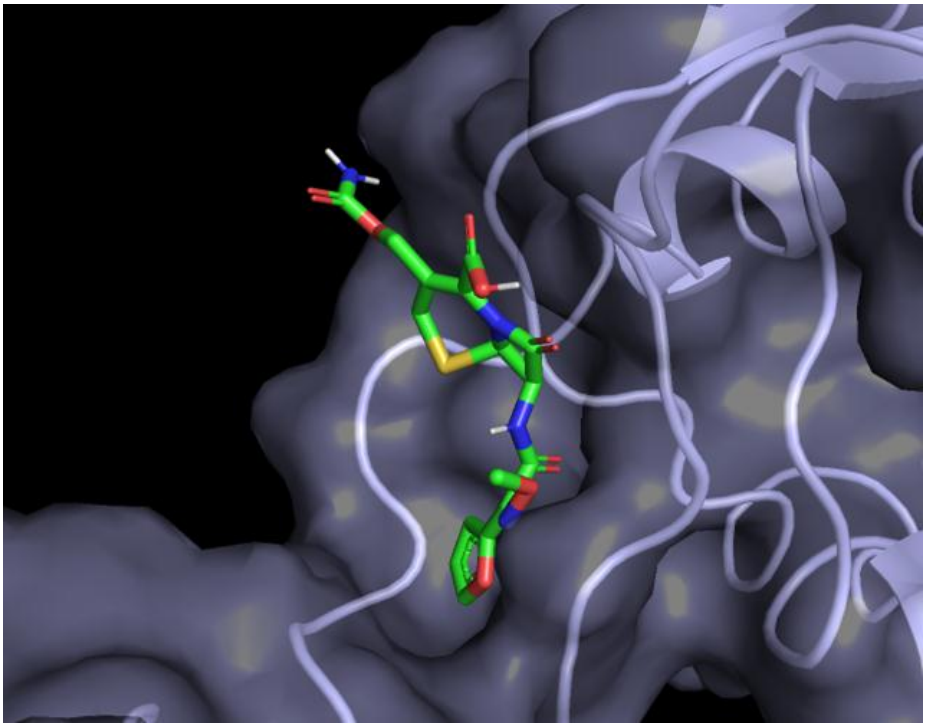

Fig 4.19 Molecular Interactions between cefuroxime - Delta

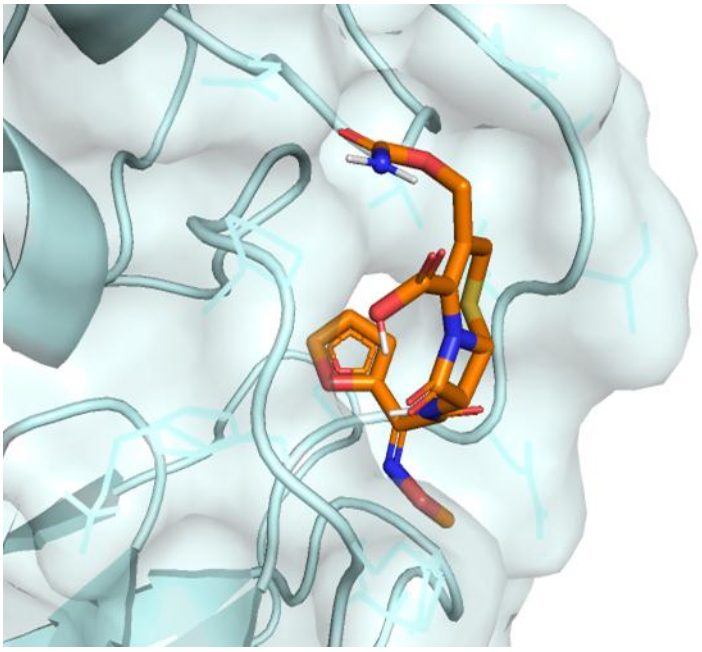

Fig 4.20 Molecular Interactions between cefuroxime – Omicron

Conclusion

From the above, we concluded the ease of binding (highest to lowest) Spike Protein of variants with drugs (Cefuroxime,Umifenovir, Cefotaxime, Ceftriaxone and Ampicillin)

1. reference - Ampicillin>Ceftriaxone>Cefotaxime>Cefuroxime>Umifenovir

2. Alpha - Ampicillin>Umifenovir>cefuroxime>ceftriaxone>cefotaxime

3. Delta - cefuroxime>cefotaxime>ceftriaxone>umifenovir>Ampicillin

4. Omicron - Ampicillin > umifenovir>ceftriaxone>cefotaxime>cefuroxime

Thus, we say that Ampicillin may be a good candidate as a drug for acting on spike protein for 3 out of 4 variants except for Delta one.

CONCLUSION

SARS-COV-2 is responsible for COVID-19, a viral illness that has spread to 228 countries across the globe since it began in China's Wuhan region. There is a growing number of cases of COVID-19, which has been categorized by the WHO as a global clinical emergency. Rapid development of a convincing and safe antibody is essential for reducing global mortality. Currently, there are no effective or authorised treatments, and routine vaccines can take years to develop. A more precise vaccination regimen can be created in silico through the use of bioinformatics, vaccinogenomics, immunoinformatics, atomic simulations, and underpinning science. Immunoinformatics techniques were employed to predict antigenic epitopes against SARS-CoV-2 in order to develop the Covid vaccine.

REFERENCES

[1] Thomas K, Goldsby J, Osborne RA, Barbara A, Kuby J (2006) Kuby immunology, 6th edn. Freeman and Co., WH

[2] Kimbrell DA, Beutler B (2001) The evolution and genetics of innate immunity. Nat Rev Genet 2:256–267

[3] Korber B, LaBute M, Yusim K (2006) Immunoinformatics: comes of Age. PLoS Comput Biol 2:0484–0492

[4] Gardy JL, Lynn DJ, Brinkman FSL, Rew H (2009) Enabling a systems biology approach to immunology: focus on innate immunity. Trends Immunol 30:249–262

[5] Davies MN, Flower DR (2007) Harnessing bioinformatics to discover new vaccine. Drug Discov Today 12:389–395

[6] Ortutay C, Vihinen M (2009) Immunome Knowledge base (IKB): An integrated service for immunome research. BMC Immunol 10

[7] Sette A, Fleri W, Peters B, Sathiamurthy M, Bui HH (2005) A roadmap for the immunomics of category A-C pathogens. Immunity 22: 155–161

[8] Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A (2007) Ab and T cell epitopes of infl uenza A virus, knowledge and opportunities. Proc Natl Acad Sci U S A 104:246–251

[9] Muller GM, Shapira M, Arnon R (1982) Antiinfl uenza response achieved by immunization with a synthetic conjugate. Proc Natl Acad Sci U S A 79:569–573

[10] Naruse H, Ogasawara K, Kaneda R, Hatakeyama S, Itoh T, Kida H, Miyazaki T, Good RA, Onoe K (1994) A potential peptide vaccine against two different strains of infl uenza virus isolated at intervals of about 10 years. Proc Natl Acad Sci U S A 91:9588–9592

[11] Pellequer J, Westhof E, Regenmortel MV (1991) Predicting the location of structure of continuous epitopes in proteins from their primary structure. Methods Enzymol 203: 176–201

[12] Parker J, Guo D, Hodges R (1986) New hydrophilicity scale derived from highperformance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray- derived accessible sites. Biochemistry 25:5425–5432

[13] Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol Relat Areas Mol Biol 47:45–148

[14] Levitt M (1978) Conformational preferences of amino acids in globular proteins. Biochemistry 17:4277–4285

[15] Emini E, Hughes J, Perlow D, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus specifi c synthetic peptide. J Virol 55:836–839

[16] EL-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting protective linear b-cell epitopes using evolutionary information. In: IEEE International Conference on Bioinformatics and Biomedicine 289–292 59. Odorico M, Pellequer JL (2003) BEPITOPE: predicting the location of continuous epitopes and patterns in protein. J Mol Recognit 16: 20–22

[17] Pellequer JL, Westhof E (1993) PREDITOP: a program for antigenicity predictions. J Mol Graph 11:204–210

[18] Saha S, Raghava GPS (2004) BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties. In: Nicosia G, Cutello V, Bentley PJ, Timis J (eds.) ICARIS Springer, LNCS 3239:197–204

[19] Ghate AD, Bhagwat BU, Bhosle SG, Gadepalli SM, Kulkarni-Kale UD (2007) Characterization of antibody-binding sites on proteins: development of a knowledgebase and its applications in improving epitope prediction. Protein Pept Lett 14:531–535

[20] Saha S, Raghava GPS (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65:40–48 64. Sweredoski MJ, Baldi P (2009) COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng Des Sel 22:113–120

[21] Larsen JEP, Lund O, Nielsen M (2006) Improved method for predicting linear B cell epitopes. Immunome Res 2

[22] Toseland CP, Clayton DJ, McSparron H et al (2005) AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. Immunome Res 1

[23] Anderson P, Nielsen M, Lund O (2006) Prediction of residues in discontinuous B cell epitopes using protein 3D structures. Protein Sci 15:2558–2567

[24] Sweredoski M, Baldi P (2008) PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. Bioinformatics 24: 1459–1460

[25] Hamelryck T (2005) An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. Proteins 59: 38–48

[26] Bublil EM, Mayrose NTFI, Penn O, Berman AR (2007) Stepwise prediction of conformational discontinuous B-cell epitopes using the mapitope algorithm. Proteins 68:294–304

[27] Sollner J, Grohmann R, Rapberger R, Perco P, Lukas A, Mayer B (2008) Analysis and prediction of protective continuous B cell epitopes on pathogen proteins. Immunome Res 4

[28] Kale KU, Bhosle S, Kolaskar AS (2005) CEP: a conformational epitope prediction server. Nucleic Acids Res 33:W168–W171

[29] Zhang W, Xiong Y, Zhao M, Zou H, Ye X, Liu J (2011) Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. BMC Bioinformatics 12:341

[30] Mayrose I, Penn O, Erez E et al (2007) Pepitope: Epitope mapping from affi nity-selected peptides. Bioinformatics 23:3244–3246

[31] Pizzi E, Cortese R, Tramontano A (1995) Mapping epitopes on protein surfaces. Biopolymers 36:675–680

[32] Moreau V, Granier C, Villard S, Laune D, Molina F (2006) Discontinuous epitope prediction based on mimotope analysis. Bioinformatics 22:1088–1095

[33] Huang J, Gutteridge A, Honda W, Kanehisa M (2006) MIMOX: a web tool for phage display based epitope mapping. BMC Bioinformatics 7

[34] Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, Sharan R, Pupko T (2007) Epitope mapping using combinatorial phagedisplay libraries: a graph-based algorithm. Nucleic Acids Res 35:69–78

[35] Huang YX, Bao YL, Guo SY, Wang Y, Zhou CG, Li YX (2008) Pep-3D-Search: a method for B-cell epitope prediction based on mimotope analysis. BMC Bioinformatics 9:538

[36] Schreiber A, Humbert M, Benz A, Dietrich U (2005) 3D-Epitope-Explorer (3DEX): Localization of conformational epitopes within three-dimensional structures of proteins. J Comput Chem 26:879–887

[37] Caragea C, Sinapov J, Silvescu A, Dobbs D, Honavar V (2007) Glycosylation site prediction using ensembles of support vector machine classifi ers. BMC Bioinformatics 8:438

[38] EL-Manzalawy Y, Honavar V (2010) Recent advances in B-cell epitope prediction methods. Immunome Res 6(Suppl 2):2

[39] Sollner J (2006) Selection and combination of machine learning classifi ers for prediction of linear B-cell epitopes on proteins. J Mol Recognit 19:209–214

[40] Huang L, Dai Y (2006) Direct prediction of T-cell epitopes using support vector machines with novel sequence encoding schemes. J Bioinform Comput Biol 4:93–107

[41] Bhasin M, Raghava GPS (2003) Prediction of promiscuous and high-affi nity mutated MHC binders. Hybrid Hybridomics 22:229–234

[42] Zhang GL, Petrovsky N, Kwoh CK, August JT, Brusic V (2006) Pred TAP: a system for prediction of peptide binding to the human transporter associated with antigen processing. Immunome Res 2

[43] Buus S, Lauemoller SL, Worning P, Kesmir C, Frimurer T, Corbet S, Fomsgaard A, Hilden J, Holm A, Brunak S (2003) Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artifi cial neural network approach. Tissue Antigens 62: 378–384

[44] Neilsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O (2003) Reliable prediction of T-cell epitopes using networks with novel

sequence representations. Protein Sci 12: 1007–1017 89. Buus S, Stryhn A, Winther K, Kirkby N, Pedersen LO (1995) Receptor–ligand interactions measured by an improved spun column chromatography technique. A high effi ciency and high throughput size separation method. Biochim Biophys Acta 1243:453–460

[45] Larsen MV, Lundegaard C, Lamberth K, Buss S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8

[46] Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M (2008) NetMHC-3.0: accurate web accessible predictions of human mouse and monkey MHC class I affi nities for peptides of length 8-11. Nucleic Acids Res 36:W509–W512

[47] Brusic V, Rudy G, Honeyman M, Hammer J, Harrison L (1998) Prediction of MHC class II-binding peptides using an evolutionary and artifi cial neural network. Bioinformatics 14: 121–130 93. Miyata J (1991) A User's Guide to PlaNet Version 5.6.

[48] Doytchinova IA, Guan P, Flower DR (2006) EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics 7:131

[49] Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern B 26:29–41

[50] Bhasin M, Raghava GPS (2004) Analysis and prediction of affi nity of TAP binding peptides using cascade SVM. Protein Sci 13: 596–607

[51] Nanni L (2006) Machine learning algorithms for T-cell epitopes prediction. Neurocomputing 69:866–868

[52] Bhasin M, Raghava GPS (2005) Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. Nucleic Acids Res 33:W202–W207

[53] Joachims T (1999) Marking large-scale support vector machine learning practical. In: Advances in Kernel methods: support vector learning. MIT Press, Cambridge, MA, pp 169–184 100. Cost S, Salzberg S (1993) A weighted nearest neighbor algorithm for learning with symbolic features. Mach Learn 10:57–78

[54] Witten IH, Frank E (1999) Data mining: practical machine learning tools and techniques with java implementations, 2nd edn. Morgan Kaufman, San Francisco

[55] Flower DR (2003) Towards in silico prediction of immunogenic epitopes. Trends Immunol 24:667–674

[56] Bian H, Hammer H (2004) Discovery of promiscuous HLA restricted T cell epitope with TEPITOPE. Methods 34:468–475

[57] Zhang L, Chen Y, Wong HS, Zhou S, Mamitsuka H, Zhu S (2012) TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. PLoS One 7:e30483

[58] Kangueane P, Sakharkar MK (2005) T epitope designer: HLA peptide binding prediction server. Bioinformation 1:21–24

[59] Zhao B, Mathura VS, Ganapathy R, Moochhala S, Sakharkar MK, Kangneane P (2003) A novel MHCp binding prediction model. Hum Immunol 64:1123–1143

[60] Ponomarenko JV, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B (2008) ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9

[61] Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) MHCPred: a server for quantitative prediction of peptide-MHC binding. Nucleic Acids Res 31:3621–3624

[62] Schiewe AJ, Haworth IS (2007) Structure based prediction of MHC-peptide association: algorithm comparison and approach to cancer vaccine design. J Mol Graph Model 26:667–675

[63] Jojic N, Gomez MR, Heckerman D, Kadle C, Furman OS (2006) Learning MHC-I peptide binding. Bioinformatics 22:e227–e235

[64] Furman OS, Altuvia Y, Sette A, Margalit H (2000) Structure-based prediction of binding peptides to MHC class I molecules: Application to a broad range of MHC alleles. Protein Sci 9:1838–1846

[65] Miyazawa S, Jernigan RL (1996) Residueresidue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 256:623–644

[66] Altuvia Y, Margalit H (2004) A structurebased approach for prediction of MHCbinding peptides. Methods 34:454–459

[67] Singh H, Raghava GPS (2001) Propred: prediction of HLA-DR binding sites. Trends Immunol 17:1236–1237

[68] Sturniolo T, Bono E, Ding J et al (1999) Generation of tissue-specifi c and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 17:555–561

[69] Feldhahn M, Thiel P, Schuler MM, Hillen N, Stevanovic S, Rammensee HG, Ohlbacher O (2008) EpiToolKit-a web server for computational immunomics. Nucleic Acids Res 1:W519–W522

[70] Flower DR, Phadwal K, Macdonald IK, Coveney PV, Davies MN, Wan S (2010) T-cell epitope prediction and immune complex simulation using molecular dynamics: state of the art and persisting challenges. Immunome Res 6(Suppl 2):S4

[71] Zhang C, Anderson A (1998) DeLisi C: structural principles that govern the peptidebinding motifs of class I MHC molecules. J Mol Biol 281:929–947

[72] Wan S, Coveney PV, Flower DR (2005) Molecular basis of peptide recognition by the TCR: affi nity differences calculated using large scale computing. J Immunol 175: 1715–1723 120. Stadler MB, Stadler BM (2003) Allergenicity prediction by protein sequence. FASEB J 17: 1141–1143

[73] Kong W, Tan TS, Tham L, Choo KW (2006) Improved prediction of allergenicity by combination of multiple sequence motifs. In Silico Biol 7:77–86 122. Bjorklund AK, Atmadja SD, Zorzet A, Hammerling U, Gustafsson MG (2005) Supervised identifi cation of allergenrepresentative peptides for in silico detection of potentially allergenic proteins. Bioinformatics 21:39–50

[74] Zorzet A, Gustafsson M, Hammerling U (2002) Prediction of food protein allergenicity: a bioinformatic learning systems approach. In Silico Biol 2:525–534

[75] Saha S, Raghava GPS (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34: W202–W209

[76] Fiers MWEJ, Kleter GA, Nijland H, Peijnenburg AACM, Peter NJ, Ham RCHJV (2004) Allermatch TM, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics 5

[77] FAO/WHO: Allergenicity of Genetically Modifi ed Foods. http://www.who.int/ foodsafety/publications/biotech/en/ec_jan 2001.pdf, 2001.

[78] FAO/WHO: Codex Principles and Guidelines on Foods Derived from Biotechnology ftp:// ftp.fao.org/codex/standard/en/Codex TextsBiotechFoods.pdf , 2003.

[79] Cui J, Han LY, Li H, Ung CY, Tang ZQ, Zheng CJ, Cao ZW, Chen YZ (2007) Computer prediction of allergen proteins from sequence-derived protein structural and physicochemical properties. Mol Immunol 44:514–520

[80] Barrio AM, Atmadja DS, Nistr A, Gustafsson MG, Hammerling U, Rudloff EB (2007) EVALLER: a web server for in silico assessment of potential protein allergenicity. Nucleic Acids Res 35:694–700

[81] Soeria-Atmadja D, Lundell T, Gustafsson MG, Hammerling U (2006) Computational detection of allergenic proteins attains a new level of accuracy with in silico variable-length peptide extraction and machine learning. Nucleic Acids Res 34:3779–3793

[82] Forouzesh, M., Rahimi, A., Valizadeh, R., Dadashzadeh, N., & Mirzazadeh, A. (2020). Clinical display, diagnostics and genetic implication of novel Coronavirus (COVID-19) epidemic. Eur Rev Med Pharmacol Sci, 24(8), 4607-4615

[83] Wang, F., Kream, R. M., & Stefano, G. B. (2020). An evidence based perspective on mRNA-SARS-CoV-2 vaccine development. Medical science monitor: international medical journal of experimental and clinical research, 26, e924700-1.

[84] Joshi, A., Joshi, B. C., Mannan, M. A. U., & Kaushik, V. (2020). Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. Informatics in Medicine Unlocked, 19, 100338.

[85] Wang, F., Hou, H., Luo, Y., Tang, G., Wu, S., Huang, M., ... & Sun, Z. (2020). The laboratory tests and host immunity of COVID-19 patients with different severity of illness. JCI insight, 5(10).

[86] Alijotas-Reig, J., Esteve-Valverde, E., Belizna, C., Selva-O'Callaghan, A., Pardos-Gea, J., Quintana, A., ... & Miró-Mur, F. (2020). Immunomodulatory therapy for the management of severe

COVID-19. Beyond the anti-viral therapy: A comprehensive review. Autoimmunity reviews, Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., ... & Melino, G. (2020). COVID-19 infection: the perspectives on immune responses. Cell Death & Differentiation, 27(5), 1451-1454.19(7), 102569.

[87] Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., ... & Melino, G. (2020). COVID-19 infection: the perspectives on immune responses. Cell Death & Differentiation, 27(5), 1451-1454.

[88] Li, X., Geng, M., Peng, Y., Meng, L., & Lu, S. (2020). Molecular immune pathogenesis and diagnosis of COVID-19. Journal of pharmaceutical analysis, 10(2), 102-108.

[89] Kiyotani, K., Toyoshima, Y., Nemoto, K., & Nakamura, Y. (2020). Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. Journal of human genetics, 65(7), 569-575.

[90] Zhang, S., Gan, J., Chen, B. G., Zheng, D., Zhang, J. G., Lin, R. H., ... & Yan, W. H. (2020). Dynamics of peripheral immune cells and their HLA-G and receptor expressions in a patient suffering from critical COVID-19 pneumonia to convalescence. Clinical & Translational Immunology, 9(5), e1128.

[91] Wang, W., Zhang, W., Zhang, J., He, J., & Zhu, F. (2020). Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). Hla, 96(2), 194-196

[92] Thevarajan, I., Nguyen, T. H., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C. E., ... & Kedzierska, K. (2020). Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. Nature medicine, 26(4), 453-455

[93] Jamilloux, Y., Henry, T., Belot, A., Viel, S., Fauter, M., El Jammal, T., ... & Sève, P. (2020). Should we stimulate or suppress immune responses in COVID-19? Cytokine and anti-cytokine interventions. Autoimmunity reviews, 19(7), 102567.

[94] Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., ... & Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. Journal of medical virology, 92(6), 618-631.

[95] Baruah, V., & Bose, S. (2020). Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. Journal of medical virology, 92(5), 495-500.

[96] Santoni, D., & Vergni, D. (2020). In the search of potential epitopes for Wuhan seafood market pneumonia virus using high order nullomers. Journal of immunological methods, 481, 112787

[97] Enayatkhani, M., Hasaniazad, M., Faezi, S., Gouklani, H., Davoodian, P., Ahmadi, N., ... & Ahmadi, K. (2021). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. Journal of Biomolecular Structure and Dynamics, 39(8), 2857-2872

[98] Lee, C. H., Pinho, M. P., Buckley, P. R., Woodhouse, I. B., Ogg, G., Simmons, A., ... & Koohy, H. (2020). Potential CD8+ T cell cross-reactivity against SARS-CoV-2 conferred by other coronavirus strains. Frontiers in immunology, 2878.

[99] Sarkar, B., Ullah, M. A., Johora, F. T., Taniya, M. A., & Araf, Y. (2020). Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology, 225(3), 151955

[100] Kalita, P., Padhi, A. K., Zhang, K. Y., & Tripathi, T. (2020). Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. Microbial Pathogenesis, 145, 104236

[101] Anderson, E. J., Rouphael, N. G., Widge, A. T., Jackson, L. A., Roberts, P. C., Makhene, M., ... & Beigel, J. H. (2020). Safety and immunogenicity of SARS-CoV-2 mRNA-1273 vaccine in older adults. New England Journal of Medicine, 383(25), 2427-2438.

[102] Widge, A. T., Rouphael, N. G., Jackson, L. A., Anderson, E. J., Roberts, P. C., Makhene, M., ... & Beigel, J. H. (2021). Durability of responses after SARS-CoV-2 mRNA-1273 vaccination. New England Journal of Medicine, 384(1), 80-82

[103] Tikhonova, A. N., Dolgalev, I., Hu, H., Sivaraj, K. K., Hoxha, E., Cuesta-Domínguez, Á., ... & Aifantis, I. (2019). The bone marrow microenvironment at single-cell resolution. Nature, 569(7755), 222-228.

[104] Widge, A. T., Rouphael, N. G., Jackson, L. A., Anderson, E. J., Roberts, P. C., Makhene, M., ... & Beigel, J. H. (2021). Durability of responses after SARS-CoV-2 mRNA-1273 vaccination. New England Journal of Medicine, 384(1), 80-82.

[105] Corbett, K. S., Edwards, D. K., Leist, S. R., Abiona, O. M., Boyoglu-Barnum, S., Gillespie, R. A., ... & Graham, B. S. (2020). SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. Nature, 586(7830), 567-571.

[106] Graham, B. S., & Corbett, K. S. (2020). Prototype pathogen approach for pandemic preparedness: world on fire. The Journal of Clinical Investigation, 130(7), 3348-3349

[107] Forouzesh, M., Rahimi, A., Valizadeh, R., Dadashzadeh, N., & Mirzazadeh, A. (2020). Clinical display, diagnostics and genetic implication of novel Coronavirus (COVID-19) epidemic. Eur Rev Med Pharmacol Sci, 24(8), 4607-4615.

[108] Wang, F., Kream, R. M., & Stefano, G. B. (2020). An evidence based perspective on mRNA-SARS-CoV-2 vaccine development. Medical science monitor: international medical journal of experimental and clinical research, 26, e924700-1.

[109] Joshi, A., Joshi, B. C., Mannan, M. A. U., & Kaushik, V. (2020). Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. Informatics in Medicine Unlocked, 19, 100338.

[110] Wang, F., Hou, H., Luo, Y., Tang, G., Wu, S., Huang, M., ... & Sun, Z. (2020). The laboratory tests and host immunity of COVID-19 patients with different severity of illness. JCI insight, 5(10).

[111] Alijotas-Reig, J., Esteve-Valverde, E., Belizna, C., Selva-O'Callaghan, A., Pardos-Gea, J., Quintana, A., ... & Miró-Mur, F. (2020). Immunomodulatory therapy for the management of severe

COVID-19. Beyond the anti-viral therapy: A comprehensive review. Autoimmunity reviews, 19(7), 102569.

[112] Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., ... & Melino, G. (2020). COVID-19 infection: the perspectives on immune responses. Cell Death & Differentiation, 27(5), 1451-1454.

[113] Li, X., Geng, M., Peng, Y., Meng, L., & Lu, S. (2020). Molecular immune pathogenesis and diagnosis of COVID-19. Journal of pharmaceutical analysis, 10(2), 102-108.

[114] Kiyotani, K., Toyoshima, Y., Nemoto, K., & Nakamura, Y. (2020). Bioinformatic prediction of potential T cell epitopes for SARS-Cov-2. Journal of human genetics, 65(7), 569-575.

[115] Zhang, S., Gan, J., Chen, B. G., Zheng, D., Zhang, J. G., Lin, R. H., ... & Yan, W. H. (2020). Dynamics of peripheral immune cells and their HLA-G and receptor expressions in a patient suffering from critical COVID-19 pneumonia to convalescence. Clinical & Translational Immunology, 9(5), e1128.

[116] Wang, W., Zhang, W., Zhang, J., He, J., & Zhu, F. (2020). Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). Hla, 96(2), 194-196.

[117]

[118] Thevarajan, I., Nguyen, T. H., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C. E., ... & Kedzierska, K. (2020). Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. Nature medicine, 26(4), 453-455.

[119] Jamilloux, Y., Henry, T., Belot, A., Viel, S., Fauter, M., El Jammal, T., ... & Sève, P. (2020). Should we stimulate or suppress immune responses in COVID-19? Cytokine and anti-cytokine interventions. Autoimmunity reviews, 19(7), 102567.

[120] Nguyen, A., David, J. K., Maden, S. K., Wood, M. A., Weeder, B. R., Nellore, A., & Thompson, R. F. (2020). Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2. Journal of virology, 94(13), e00510-20

[121] Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., ... & Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. Journal of medical virology, 92(6), 618-631.

[122] . Baruah, V., & Bose, S. (2020). Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. Journal of medical virology, 92(5), 495-500.

[123] Santoni, D., & Vergni, D. (2020). In the search of potential epitopes for Wuhan seafood market pneumonia virus using high order nullomers. Journal of immunological methods, 481, 112787.

[124] Enayatkhani, M., Hasaniazad, M., Faezi, S., Gouklani, H., Davoodian, P., Ahmadi, N., ... & Ahmadi, K. (2021). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. Journal of Biomolecular Structure and Dynamics, 39(8), 2857-2872.

[125] Lee, C. H., Pinho, M. P., Buckley, P. R., Woodhouse, I. B., Ogg, G., Simmons, A., ... & Koohy, H. (2020). Potential CD8+ T cell cross-reactivity against SARS-CoV-2 conferred by other coronavirus strains. Frontiers in immunology, 2878.

[126] Sarkar, B., Ullah, M. A., Johora, F. T., Taniya, M. A., & Araf, Y. (2020). Immunoinformatics-guided designing of epitope-based subunit vaccines against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology, 225(3), 151955.

[127] Kalita, P., Padhi, A. K., Zhang, K. Y., & Tripathi, T. (2020). Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2. Microbial Pathogenesis, 145, 104236.

[128] Anderson, E. J., Rouphael, N. G., Widge, A. T., Jackson, L. A., & Roberts, P. C. (2020). Makhene et al. Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. N Engl J Med, 383(25), 2427-2438.

[129] . Widge, A. T., Rouphael, N. G., Jackson, L. A., Anderson, E. J., Roberts, P. C., Makhene, M., ... & Beigel, J. H. (2021). Durability of responses after SARS-CoV-2 mRNA-1273 vaccination. New England Journal of Medicine, 384(1), 80-82.

[130] Corbett, K. S., Edwards, D. K., Leist, S. R., Abiona, O. M., Boyoglu-Barnum, S., Gillespie, R. A., ... & Graham, B. S. (2020). SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. Nature, 586(7830), 567-571.

[131] Graham, B. S., & Corbett, K. S. (2020). Prototype pathogen approach for pandemic preparedness: world on fire. The Journal of Clinical Investigation, 130(7), 3348-3349

[132] Thomas K, Goldsby J, Osborne RA, Barbara A, Kuby J (2006) Kuby immunology, 6th edn. Freeman and Co., WH

[133] Kimbrell, D. A., & Beutler, B. (2001). The evolution and genetics of innate immunity. Nature Reviews Genetics, 2(4), 256-267.

[134] Korber, B., LaBute, M., & Yusim, K. (2006). Immunoinformatics comes of age. PLoS Computational Biology, 2(6), e71.

[135] Gardy, J. L., Lynn, D. J., Brinkman, F. S., & Hancock, R. E. (2009). Enabling a systems biology approach to immunology: focus on innate immunity. Trends in immunology, 30(6), 249-262.

[136] Davies, M. N., & Flower, D. R. (2007). Harnessing bioinformatics to discover new vaccines. Drug discovery today, 12(9-10), 389-395.

[137] Ortutay, C., & Vihinen, M. (2009). Immunome knowledge base (IKB): an integrated service for immunome research. BMC immunology, 10(1), 1-5.

[138] Sette, A., Fleri, W., Peters, B., Sathiamurthy, M., Bui, H. H., & Wilson, S. (2005). A roadmap for the immunomics of category A–C pathogens. Immunity, 22(2), 155-161.

[139] Bui, H. H., Peters, B., Assarsson, E., Mbawuike, I., & Sette, A. (2007). Ab and T cell epitopes of influenza A virus, knowledge and opportunities. Proceedings of the National Academy of Sciences, 104(1), 246-251.

[140] Müller, G. M., Shapira, M., & Arnon, R. (1982). Anti-influenza response achieved by immunization with a synthetic conjugate. Proceedings of the National Academy of Sciences, 79(2), 569-573.

[141] Wang, Y., Di, Y., Ye, J., & Wei, W. (2021). Study on the public psychological states and its related factors during the outbreak of coronavirus disease 2019 (COVID-19) in some regions of China. Psychology, health & medicine, 26(1), 13-22.

[142] Pellequer, J. L., Westhof, E., & Van Regenmortel, M. H. V. (1991). [8] Predicting location of continuous epitopes in proteins from their primary structures. Methods in enzymology, 203, 176-201.

[143] Parker, J. M. R., Guo, D., & Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry, 25(19), 5425-5432.

[144] Garnier, J. (1978). Prediction of the secondary structure of proteins from their amino-acid sequence. J. Mol. Biol., 120, 97-120.

[145] Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. Biochemistry, 17(20), 4277-4285.

[146] Emini, E. A., Hughes, J. V., Perlow, D., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology, 55(3), 836-839.

[147] Yasser, E. M., Dobbs, D., & Honavar, V. (2008, November). Predicting protective linear B-cell epitopes using evolutionary information. In 2008 IEEE International Conference on Bioinformatics and Biomedicine (pp. 289-292). IEEE.

[148] Pellequer, J. L., & Westhof, E. (1993). PREDITOP: a program for antigenicity prediction. Journal of molecular graphics, 11(3), 204-210.

[149] Saha, S., & Raghava, G. P. S. (2004, September). BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In International conference on artificial immune systems (pp. 197-204). Springer, Berlin, Heidelberg.

[150] Ghate, A. D., Bhagwat, B. U., Bhosle, S. G., Gadepalli, S. M., & Kulkarni-Kale, U. D. (2007). Characterization of antibody-binding sites on proteins: development of a knowledgebase and its applications in improving epitope prediction. Protein and Peptide Letters, 14(6), 531-535.

[151] Saha, S., & Raghava, G. P. S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Structure, Function, and Bioinformatics, 65(1), 40-48.

[152] Larsen, J. E. P., Lund, O., & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. Immunome research, 2(1), 1-7.

[153] Toseland, C. P., Clayton, D. J., McSparron, H., Hemsley, S. L., Blythe, M. J., Paine, K., ... & Flower, D. R. (2005). AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. Immunome research, 1(1), 1-12.

[154] Haste Andersen, P., Nielsen, M., & Lund, O. L. E. (2006). Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Science, 15(11), 2558-2567.

[155] Sweredoski, M. J., & Baldi, P. (2008). PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. Bioinformatics, 24(12), 1459-1460.

[156] Hamelryck, T. (2005). An amino acid has two sides: a new 2D measure provides a different view of solvent exposure. Proteins: Structure, Function, and Bioinformatics, 59(1), 38-48.

[157] Bublil, E. M., Freund, N. T., Mayrose, I., Penn, O., Roitburd-Berman, A., Rubinstein, N. D., ... & Gershoni, J. M. (2007). Stepwise prediction of conformational discontinuous B-cell epitopes using the Mapitope algorithm. Proteins: Structure, Function, and Bioinformatics, 68(1), 294-304.

[158] Sollner, J., Grohmann, R., Rapberger, R., Perco, P., Lukas, A., & Mayer, B. (2008). Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins. Immunome Research, 4(1), 1-17.

[159] Kulkarni-Kale, U., Bhosle, S., & Kolaskar, A. S. (2005). CEP: a conformational epitope prediction server. Nucleic acids research, 33(suppl_2), W168-W171.

[160] Zhang W, Xiong Y, Zhao M, Zou H, Ye X, Liu J (2011) Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. BMC Bioinformatics 12:341

[161] Mayrose, I., Penn, O., Erez, E., Rubinstein, N. D., Shlomi, T., Freund, N. T., ... & Pupko, T. (2007). Pepitope: epitope mapping from affinity-selected peptides. Bioinformatics, 23(23), 3244-3246.

[162] Pizzi, E., Cortese, R., & Tramontane, A. (1995). Mapping epitopes on protein surfaces. Biopolymers: Original Research on Biomolecules, 36(5), 675-680.

Immunoinformatics Tools: A boon in vaccine Development Against Covid-19

Sandeep kumar

Department of Biotechnology Engineering, Delhi Technological University, Delhi 110042, India
Sandeepkumar_2k20mscbio25@dtu.ac.in

Yasha Hasija *

Department of Biotechnology Engineering, Delhi Technological University, Delhi 110042, India
yashahasija06@gmail.com

* Corresponding Author

***Abstract* - The first evidences of SARS Covid 19 virus were reported from Labs in Wuhan, China's Hubei Province, at the end of 2019. It spread very quickly throughout China, leading in an epidemic and a global pandemic. A large population was affected and died due to the pandemic in 2019. It shares genetic similarities with SARS-CoV-2 and MERS-COV. The development of an effective SARS-CoV-2 vaccine is important for reducing COVID-19 deaths and giving immunological protection to the worldwide community. The lengthy and expensive process of vaccine production can be shortened by using immunoinformatics approaches. immunoinformatics tools such as Vaxijen, IEDB,NetCTL 1.2, PEP-FOLD etc have previously been used in reverse vaccinology for SARS-CoV-2 vaccine development in areas such as antigen selection, toxicity, predicting vaccine targets ,allergenicity prediction and selection of MHC-I and II binding epitopes etc. In this review, we summarize some of the most useful immunoinformatics tools like vexijen, Bepipred 2.0,SVMTrip, FNepitope etc and their role in the development of covid 19 vaccines. The characteristics of such tools have been thoroughly reviewed, and which may provide experimental biologists with prediction insights that may enhance active research attempts to identify therapies for the infectious COVID-19 illness**.

Keywords:- Immunoinformatics, SARS-CoV-2, Vaccine, IEDB,VaxiJen,IFNepitope,NetCTLpan1.1,C-Immsim

I. INTRODUCTION

After multiple deaths in Wuhan, China, an unexpected pneumonia illness epidemic was reported in late December 2019 [1]. On March 11, 2020, the World Health Organization proclaimed the COVID-19 disease to be a pandemic. The epidemic spread quickly through Wuhan to numerous nations, including India, with thousands of people affected and most of them dying within months of its first dissemination[1][2] . There are about 263,842,124 cases and over 5,244,379 confirmed deaths as of December 02, 2021, impacting 224 nations (Data is obtained from https://www.worldometers.info/coronavirus/). The coronavirus(COVID-19) is a positive-sense single stranded RNA virus that belongs phylogenetically to the Coronaviridae family , order Nidovirales, Subgenus Sarbecovirus and the genus Betacoronavirus[3]. Severe acute respiratory syndrome coronavirus(SARS-CoV) [4][5][6] and Middle East respiratory disorder coronavirus (MERS-COV) [7][8] both share highly genetic similarity and designated as β-coronavirus. The COVID-19 disease virus has a unique protein structure on cleavage site that separates it from other corona viruses such as SARS or MERS. SARS-CoV-2 genomic RNA consist of the gene of RdRp (RNA dependent RNA polymerase), structural protein like nucleocapsid (N)(N), membrane (M), envelope(E), spike (S) as well as non suctrutural replicase polyproteins (nsp1-nsp16). S glycoprotein of virion can get attached to ACE 2 (angiotensin-converting enzyme 2) receptor [9]which is mostly expressed on the epithelia of the lung and small intestine though which SARS-CoV-2 enter into the host cell[10]. There are also other organs that show a high level of ACE 2 protein expression like the Brain, testis, gastrointestinal system, Heart, and adipose tissue [11]. In SARS-CoV-2, S-protein is made up of two functional components i.e S1 and S2. Receptor binding domain is found in S1, and it interacts directly to the cell receptor ACE-2 while conserved non-RBD regions in S2 is responsible for stable binding of S1 protein to the host's cell membrane. After the attachment process of S1 and S2 protein, Furin protease cleaves the multibasic site of S1/S2 which is essential for viral entry [12]. In the host cells infected by SARS-CoV-2,ACE-2 expression is significantly suppressed. Other than facilitating entry of

2022 IEEE Delhi Section Conference (DELCON) | 978-1-6654-5883-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/DELCON54057.2022.9753152

SARS-CoV-2 into cell,Ace 2 receptor has role in lowering blood pressure and metabolism of angiotensin 2 in heart.

As the situation worsened, the necessity for an appropriate peptide vaccine component against SARS-CoV-2 became more important. To successfully respond to novel SARS-CoV-2, technology that can swiftly generate efficient, cheaper, safer vaccines must be refined and established. Reverse vaccinology is a revolutionized methodology of vaccine development which is based upon genomic information of bacteria, virus, parasites, cancer cells, or allergens derived from in silico analysis (genomics, proteomics, bioinformatics)[13] .

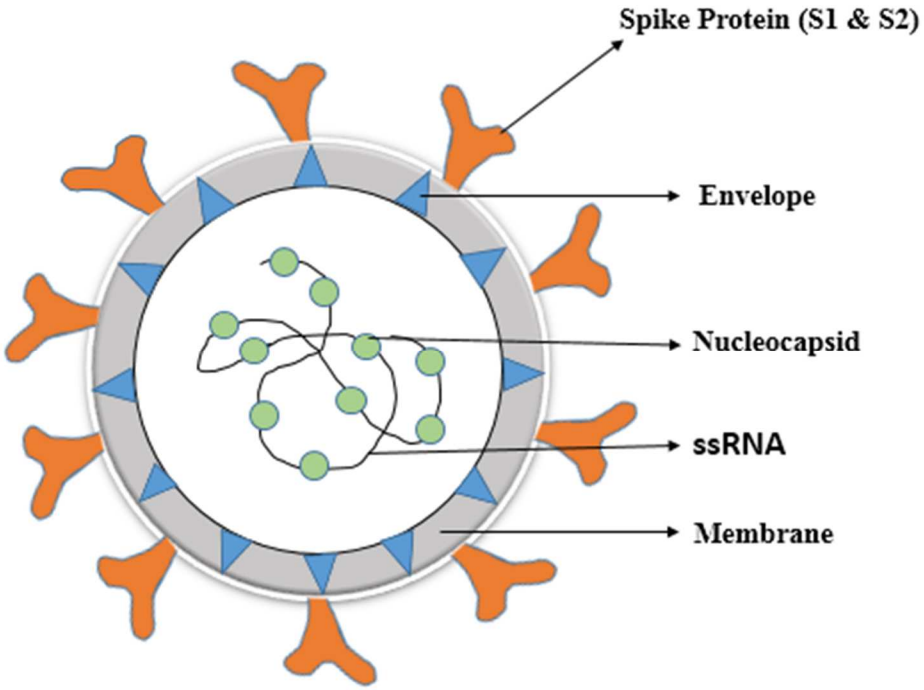

Fig. 1. Structural biology of SARS-CoV-2

Immunoinformatics play a vital role in vaccine development by identifying a multiepitope, targeting virion particles, appropiate linker etc [14]. However, despite the fact that immunology research is both expensive and time-consuming, massive volumes of data are frequently created. Immunoinformatics tools are the only way to examine such data with great speed and accuracy. In comparison to traditional vaccine development, Genome sequencing and in vitro B-cell validation is completed in a relatively short time rather than years as with traditional vaccine development [15]. With the help of Immunoinformatics tool, examining the peptide sequence of viral protein vaccine target epitopes can be discovered.

II. STEPS INVOLVED IN VACCINE DEVELOPMENT THROUGH IMMUNOINFORMATICS

Chukwudozie et al. have been developed several platforms to design peptide vaccines as well as numerous techniques' for verifying them using in silico approaches in simplified and detailed manner [16]. **Figure 2** shows a flow chart that summarises the protocols for Vaccine development using immunoinformatics tool.

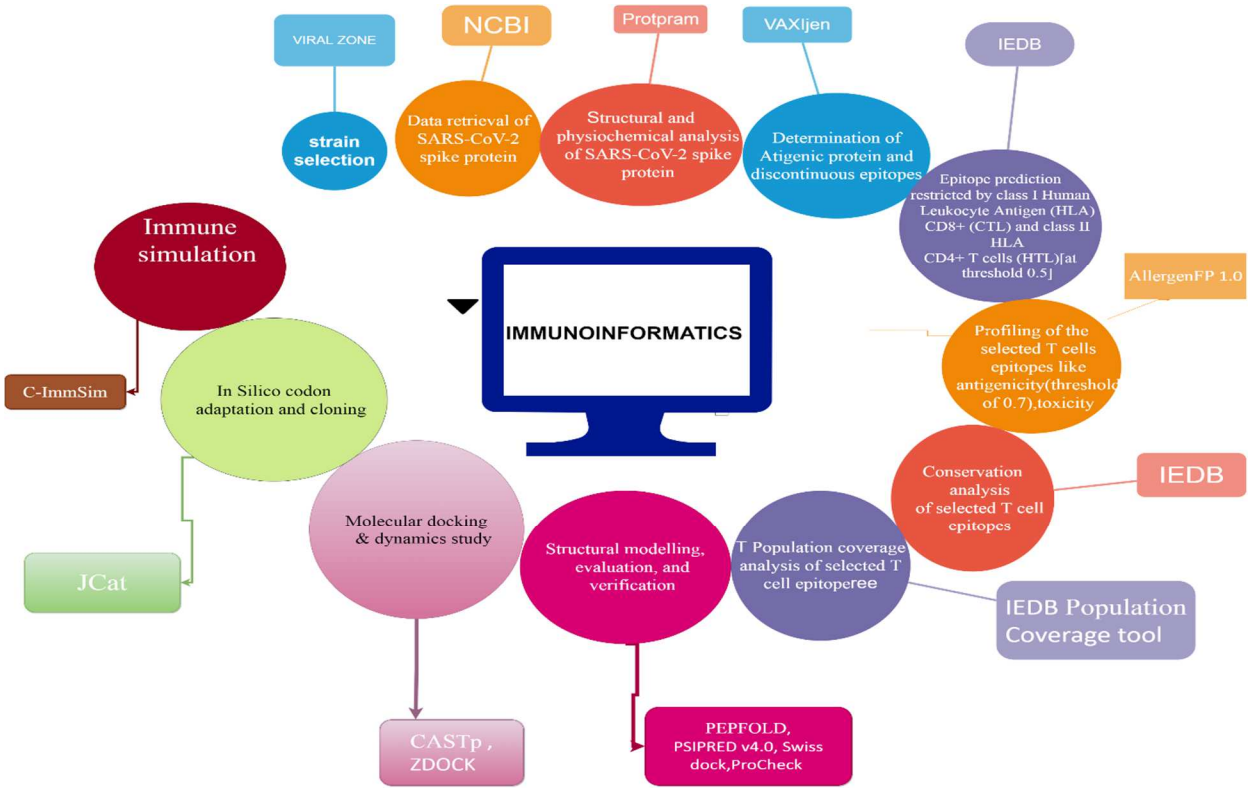

Fig. 2. Flow chart summarizing Protocol for vaccine development (epitope based) using Immunoinformatics approaches (Starting from Blue to brown circle).

III. IMMUNOINFORMATICS TOOLS USED IN COVID-19 VACCINE DEVELOPMENT

The immune system is classified into cellular or humoral and either cellular or humoral immune response can be produced based on disease. For example- For knowledge of cellular response or reaction triggered by a vaccine, researcher can use software for T Cell identification like Tepredict, Rankpep,Ctlpred,Propred,IEDB and if study of humoral response is required,software for identification of B cell antigen like Bepipred , Igpred, Bcipep, Mimox, Pepsurf, ellipro, BEST,EPCES etc can be used.

1) IEDB (https://www.iedb.org/): - IEDB (Immune Epitope Database) Consists set of Tools which allows us to retrieve different information about epitopes, MHC I binding predictions, funded by NIAID (National Institute of Allergy and Infectious Diseases). Over 22,536 references have been collected as of December 05, 2021, with over 1,184,242 epitopes (both Peptidic and non peptidic) , 1,165,695 B cell, 424,795 T cell, 3,318,008 MHC binding, and MHC ligand elution experiments in the database .

A. Case Studies using IEDB tool

Bhattacharya et al. designed COVID-19 Vaccine by targeting spike protein of HCOV and B-Cell epitopes of COVID-19 which was done through IEDB server [17]. He used BepiPred 2.0 prediction module from IEDB server for identifying the linear B-cell epitopes of SARS-CoV-2. Their results revealed that within spike protein of COVID-19, there are almost 34 successive linear B-Cell epitopes of varied length which include 13 HLA 1 epitopes as well as 3 HLA II epitopes that might be employed as vaccine candidates. As part of the study, these authors included immunoinformatics modelling utilizing PROCHECK for prediction of vaccine component structure. Despite this, researchers also used the short peptide linker adapter (EAAAK) to combine the highly antigenic epitopes into a single epitope vaccine component.As a result, they identified the T cell epitope, which was determined to be 64.29 % conserved and had identical conservation with other 5 potential T cell epitopes . The KSSTGFVYF epitope had the most interactions with HLA alleles, although possessing the same conservancy as the other 5 epitopes.

IEDB methods have been utlilized in several studies to identify B and T cell epitopes for SARS-CoV-2 vaccine development (Dong et al.[18]; Bhatt et al [19]; jhakr et al.[20]; Shehata et al.[21])

2) Vaxijen (https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) : - Vaxijen is a first alignment free approach to predict antigenicity of Bacterial, viral and tumour proteins based upon Auto Cross Covariance (ACC) transformation method. ACC is a method to measure the relationship between two features of a protein and transforms different length of protein sequences into uniform length vector. Vaxigen server consists of of protein characteristics simulations that have been developed from pre pre-processing ACC.

A. Case Studies using vaxijen tool

Anand et al. utilized Vaxijen 2.0 to predict antigenic epitope of T cell SARS-CoV-2. As a result, authors selected 38 epitopes, by maintaining critical threshold of antigenicity, they picked up only that showing prediction accuracy >0.4 were thought to be potent immunogens . These authors found vaxijen score or antigenicity score for selected 3 epitope are – for ORF3a "HVTFFIYNK" score is 0.9862, for M protein "LTWICLLQF" score is 1.1393, for nucleocapsid protein "KTFPPTEPK"score is 0.7571 [22]. Highest antigenicity score was observed in case of Orf 6 protein is 0.6131,whereas surface glycoprotein had the lowest score 0.4646 .

Chukwudozie et al. also utilized Vaxijen 2.0 tool to design multimeric epitope conjugate based SARS-CoV-2 vaccine by targeting spike protein of virus. For antigenicity test the selected T cell epitopes were tested at a suggested threshold of 0.7. The conjugated vaccine was significantly antigenic just before administration of the Outer membrane protein A as a adjuvant, which showed the score of 0.85 on the Vaxijen website, indicating that the vaccine is capable of generating both cellular as well as humoral immune responses despite the use of an adjuvant. As a result,on the basis of Maximum no. of epitopes binding to the HLA-I molecule, antigenicity and allergenicity test ,16 epitopes were found. All 3 cytokines, namely IFN- , IL-4, and IL-10, could only be stimulated by the peptide GYFKIYSKHTPINLV which was taken further into consideration for development of vaccine [16].

In design of multiepitope based 10 peptide vaccine from E protein of HCOV Abdelmageed et al. [23] who have used the vaxijen 2.0 server to predict the antigenicity of mutated protein (N, S, E, M) at threshold of 0.4. At the end, he concluded that out of 4 proteins, the best immunogenic target was reported to be the envelope (E) protein with antigenicity index of 0.6025. From the obtained results,researchers found out the antigenicity score of the proteins to be as follow- for Protein E is 0.6025, Protein M (0.5102), Protein S(0.4646),Protein N (0.5059).

ul Qamar et al. [24] design Multi-epitope vaccines (MEV), against SARS-CoV-2 utilising Vaxijen v2.0, have a great predicting potential of antigenicity, tumour antigens and subunit vaccines . He was used vaxijen 2.0 tool server to test the antigenicity of B cell & T cell epitopes with threshold of 0.5 . Only highly antigenic epitopes that had been screened were chosen for further research. ORF10 was determined to be the most antigenic protein, followed by E (0.6502) , M (0.6441), ORF6 (0.6131), ORF7a (0.6025) , ORF8 (0.5102), and N (0.5059) proteins with their antigenicity score . Because the antigenic values of ORF1, S, and ORF3 proteins were less than 0.5, they were removed from any further investigation.

3) NetCTLpan 1.1 (https://services.healthtech.dtu.dk/service.php?NetCTLpan-1.1) :- NetCTLpan 1.1 is a platform that allows users to anticipate CTL epitopes based on any MHC molecule with a predefined sequence of a protein (only for 8-11mer peptides) and also performs the predcition of proteasomal cleavage , Ligand combined score & Transport efficiency of TAP (transporter associated with antigen processing). Since the beginning of the year 2020, numerous laboratories have been developing multi-epitope based vaccines against spike protein of SARS-CoV-2. NetCTLpan 1.1 is believed to beat all existing epitope predicting tools for Cytotoxic T lymphocytes (CTLs).

A. Case studies using NetCTLpan 1.1

Ayyagari et al.[25] utilized NetCTLpan 1.1 platform for predicting end to-end cytotoxic T cell (CTL) epitope. For MHC class-I epitope detection threshold value is 1.0 set by researchers . Peptides with a percentage Rank of < 1.0 were deemed acceptable for further investigation. NetCTLpan detected total 37 peptides with high binding affinity for selected HLA-I by method of precentage score analysis. The ultimate vaccine consists of 223 amino acids with six HLA-1 molecules.

For designing multiepitope SARS-CoV-2 vaccine some authors like safavi et al.[26] utlizing NetCTLpan 1.1 tool for determining the most immunodominant epitope with the threshold of greater than 0.75. At the result, they found that the immunogenicity of Spike protein portion that is interacting with ACE2 receptor is good for multiepitope vaccine designing. these regions contained several number of epitopes with a greater potential for binding to numerous HLA-I and HLA-II alleles. Multiple CTL and CD4 + epitopes were also found in these locations . Both CTL and helper T cells can be activated by the overlapping epitopes. In the end researchers conclude,S protein's immunogenic region was chosen for potential vaccine development as several CTL and CD4+ epitopes were found in this region and activation of CTL and helper T cells can be achieved by the overlap epitopes.

NetCTLpan1.1 was shown to be more accurate than NetMHCpan 4.1, DeepLigand, PickPocket1.1, and MHCflurry2.0.and is seen applied in mishra et al. [27], Quiros-Fernandez [28] and Bukhari et al. [29].

4) IFNepitope (https://webs.iiitd.edu.in/raghava/ifnepitope/developer.php): - IFNepitope is a web-based prediction tool that allow users to predict, identify and design the Peptides that induce IFN-ϒ (gamma) cytokine. By stimulating macrophages and natural killer cells, IFN-ϒ cytokine stimulates both native and targeted type of immune responses. It consists of three main modules:-

- **Predict** – The IFN-ϒ triggering peptide from the collection of peptides (submitted by users) can be predicted using this module .
- **Design** – This server helps users to find out minimum mutated epitope that has a greater potency from an enlisted epitope.
- **Scan** - It allows users to find out which parts of protein are capable of inducing IFN-ϒ.

A. Case study using IFNepitope tool

Dong et al. [18], in their study, attempted to design a multiepitope subunit vaccine against SARS-CoV-2 using IFNepitope server for the prediction of IFN-gamma epitopes. The IFN-epitope server was used to enter HTL epitopes with poor scores). positive IFN epitope can be predicted by using SVM (support vector machine) technique. IFN-induction and MHC Class II binding, both of which promote T-helper cell activation, were used to select the final HTL epitopes. At the result based on their binding potential, IC50 < 500 nm and IFN- ϒ induction, researchers selected 14 HTL epitope for further analysis.

Shehata et al. [21]Used IFNepitope for epitope-based candidate vaccine production as it predicted IFN-gamma inducing epitopes.On the basis of Results obtained by the authors,negative scores showing negative predictions were eliminated .Six peptides were selected as possible linear neutralising epitopes for vaccine design as they scored significantly higher than the average score of >0.2 that indicates their potential to induce an IFN- response.These peptide epitopes are expected to induce IFN and cross react with B cells to mediate a humoral immune response.

5) C-ImmSim (https://kraken.iac.rm.cnr.it/C-IMMSIM/): - One of the most essential tools for immunoinformatics is cimmsim. C-ImmSim is a set of model relying on the Celada-Seiden model made by P.E Seiden and F.Celada to facilitate in silico prediction of immune stimulation (cell-mediated, humoral, or both) of peptides. C-ImmSim enables users to model deterministic system(DIfferential equation,integration etc) while it eliminates the mathematical assumptions and accounts for spatial patterns impact.

A. Case study using C-ImmSim Tool

C-Immsim tool was applied by Dong et al [18] for prediction of immunological stimulation of the final developed vaccine to assess the immune response profile for developing the multi-epitope vaccine. According to the data,3 doses were given at one-month periods and the vaccinations were given four weeks apart. The simulation volume was kept to 1,000, simulation steps 1000 were used, the random seed was adjusted at 12,345, and lipopolysaccharide was eliminated from vaccine injection. IgG1 + IgG2 and IgM, as well as the lowering antigen content IgG + IgM, were used for identification of the development of the secondary and tertiary immune responses. The immune simulation result confirmed the activation of immune response immunization. After vaccination, the B cell population was greatly increased. Additionally, there was an increase in number of CD8+ T cells and CD4 + T cells which indicated the emergence of secondary and tertiary immune responses

TABLE I : Overview of reviewed immunoinformatics Tools used in SARS-CoV-2 vaccine development studies.

S.no	Author	Conformational & linear B cell epitopes	MHC I & II binding prediction	Toxicity	Population coverage analysis	Antigenicity[a]	Allergencity
1.	Joshi et al.[30]	-	NETMHC 4.0 & NETMHC II PRED 3.2	Toxin Pred	IEDB	VaxiJen (threshold[b] of > 1.0)	AllergenFP 1.0
2.	Yashvardhini et al. [31]	Linear - Bepipred 2.0	Tepitool	-	IEDB	Vaxijen v2.0	AllerTOP
3.	Dar et al .[32]	Discontinous - Ellipro	NetCTL 1.2 & NetMHCPan	-		ANTIGEN Pro	AllergenFP 1.0
4.	Khairkhah et al. [33]	Linear - BepiPred-2.0, Discontinous - ElliPro	NetMHCpan4.1& NetMHCIIpan 4.0	-	IEDB	-	PA^3P
5.	Mitra et al.[34]	Linear- ABCPred	MHCPred), SYFPEITHI ,NetMHCIIpan 3.2, NetMHC 4.0	Toxin Pred	-	VaxiJen v2.0	AllerTop

[a.] Antigenicity Score (*In the process of vaccine development, antigenicity score is a score for identification of the most probable antigenic protein which indicates its capacity to induce an immune response.*)

[b.] Threshold parameter (*Those parameters that indicate a higher likelihood of eliciting an immune response are designated as threshold parameters for the experiment. The threshold parameter varies from experiment to experiment.*)

IV. IMMUNOINFORMATICS LIMITATION IN VACCINE DEVELOPMENT

Though immunoinformatic tools are excellent sources for vaccine development , there are some limitations associated with them in vaccine development and design are as follows :-

1. To generate raw data for the vaccine development, immunoinformatics relies on wet lab research work.
2. The predictions obtained by Immunoinformatics approach do not proof the concepts formally and thus cannot replace the traditional experimental research methods which involve the actual test of hypotheses.
3. In vaccine designing, the accuracy of output data varies from the complexity of the immunoinformatics tool utilized. subsequent findings of the analysis will be incorrect as well [35].
4. Proteins, linear and discontinuous epitopes, but not other biomolecules such as polysaccharide, can be employed for vaccine production in an immunoinformatic approach for vaccine development.

5. Antigen residues have been shown to form epitopes under certain situations, affecting prediction algorithms [36]. Improved prediction methods should be require in immunoinformatic tool.

V. CONCLUSION

A prophylactic vaccine is required since SARS-CoV-2 is characterised by high infectivity and transmission speed. Immunoinformatics has helped to overcome some of the constraints of traditional vaccine design methodologies, resulting in a better knowledge of diagnostics, immune system response, and Reverse vaccinology [37]. An essential role in vaccine design is played by different tools applied for protein scaffolding and epitope prediction as this approach is of advantage since it is achieved by analysis of entire genome of pathogen as well as identification of proteins that can act as potential antigens.It has an advantage of flexible analysis over the traditional methods performed. This review highlights the immense work performed in identification and analysis of COVID-19 Viral epitopes and vaccine construct. The 28 research articles that we reviewed,used various immunoinformatics methodologies.

The objective of this review was to summarise the approaches that have been used in COVID-19 vaccine designing and here we also a draw a comparison between Different tools used in different studies by the researchers which gives an insight into the performance of the various tools. In general, the information gathered from the intense use of immunoinformatics tools for SARS-CoV-2 will pave the way for future investigations aiming at epitope discovery and vaccine development for a various other related viruses such as Bat COV RaTG13, pangolin coronavirus,mers-cov,H1N1 virus, and so on[38]. Finally, we report the important immunoinformatics tools required in vaccine development that pave the ultimate path towards overcoming the emerging COVID-19 Pandemic. On the basis of review of this study, future developments may focus on clinical trials of SARS-CoV-2 vaccine Development

.

REFERENCES

[1] Young, B. E., Ong, S. W. X., Kalimuddin, S., Low, J. G., Tan, S. Y., Loh, J., ... & Singapore 2019 Novel Coronavirus Outbreak Research Team. (2020). Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore. Jama, 323(15), 1488-1494.

[2] Hui, D. S., Azhar, E. I., Madani, T. A., Ntoumi, F., Kock, R., Dar, O., ... & Petersen, E. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. International journal of infectious diseases, 91, 264-266.

[3] CoV2020, GISAID EpifluDB. Archived from the original on 12 January 2020, 2020

[4] Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H. R., Becker, S., ... & Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. New England journal of medicine, 348(20), 1967-1976.

[5] Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., ... & SARS Working Group. (2003). A novel coronavirus associated with severe acute respiratory syndrome. New England journal of medicine, 348(20), 1953-1966.

[6] Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., Van Amerongen, G., Van Riel, D., ... & Osterhaus, A. D. (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. The Lancet, 362(9380), 263-270.

[7] De Groot, R. J., Baker, S. C., Baric, R. S., Brown, C. S., Drosten, C., Enjuanes, L., ... & Ziebuhr, J. (2013). Commentary: Middle east respiratory syndrome coronavirus (mers-cov): announcement of the coronavirus study group. Journal of virology, 87(14), 7790-7792.

[8] Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. New England Journal of Medicine, 367(19), 1814-1820.

[9] Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell, 181(2), 281-292.

[10] Hamming, I., Timens, W., Bulthuis, M. L. C., Lely, A. T., Navis, G. V., & van Goor, H. (2004).

Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 203(2), 631-637.

[11] Li, M. Y., Li, L., Zhang, Y., & Wang, X. S. (2020). Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. Infectious diseases of poverty, 9(1), 1-7.

[12] Hoffmann, M., Kleine-Weber, H., & Pöhlmann, S. (2020). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Molecular cell, 78(4), 779-784.

[13] Seib, K. L., Zhao, X., & Rappuoli, R. (2012). Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clinical Microbiology and Infection, 18, 109-116.

[14] Arya, H., & Bhatt, T. K. (2021). Role of Bioinformatics in Subunit Vaccine Design. In Molecular Docking for Computer-Aided Drug Design (pp. 425-439). Academic Press.

[15] Peng, M., Mo, Y., Wang, Y., Wu, P., Zhang, Y., Xiong, F., ... & Zeng, Z. (2019). Neoantigen vaccine: an emerging tumor immunotherapy. Molecular cancer, 18(1), 1-14.

[16] Chukwudozie, O. S., Gray, C. M., Fagbayi, T. A., Chukwuanukwu, R. C., Oyebanji, V. O., Bankole, T. T., ... & Daniel, E. M. (2021). Immuno-informatics design of a multimeric epitope peptide based vaccine targeting SARS-CoV-2 spike glycoprotein. Plos one, 16(3), e0248061.

[17] Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., ... & Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. Journal of medical virology, 92(6), 618-631.

[18] Dong, R., Chu, Z., Yu, F., & Zha, Y. (2020). Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches. Frontiers in immunology, 11, 1784.

[19] Bhatt, P., Sharma, M., & Sharma, S. (2021). Prediction and identification of T cell epitopes of COVID-19 with balanced cytokine response for the development of peptide based vaccines. In silico pharmacology, 9(1), 1-17.

[20] Jakhar, R., & Gakhar, S. K. (2020). An immunoinformatics study to predict epitopes in the envelope protein of SARS-CoV-2. Canadian Journal of Infectious Diseases and Medical Microbiology, 2020.

[21] Shehata, M. M., Mahmoud, S. H., Tarek, M., Al-Karmalawy, A. A., Mahmoud, A., Mostafa, A., ... & Ali, M. A. (2021). In Silico and In Vivo Evaluation of SARS-CoV-2 Predicted Epitopes-Based Candidate Vaccine. Molecules, 26(20), 6182.

[22] Anand, R., Biswal, S., Bhatt, R., & Tiwary, B. N. (2020). Computational perspectives revealed prospective vaccine candidates from five structural proteins of novel SARS corona virus 2019 (SARS-CoV-2). PeerJ, 8, e9855.

[23] Abdelmageed, M. I., Abdelmoneim, A. H., Mustafa, M. I., Elfadol, N. M., Murshed, N. S., Shantier, S. W., & Makhawi, A. M. (2020). Design of a multiepitope-based peptide vaccine against the E protein of human COVID-19: an immunoinformatics approach. BioMed research international, 2020.

[24] Tahir ul Qamar, M., Rehman, A., Tusleem, K., Ashfaq, U. A., Qasim, M., Zhu, X., ... & Chen, L. L. (2020). Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches. PloS one, 15(12), e0244176.

[25] Ayyagari, V. S., TC, V., & Srirama, K. (2020). Design of a multi-epitope-based vaccine targeting M-protein of SARS-COV-2: an immunoinformatics approach. Journal of Biomolecular Structure and Dynamics, 1-15.

[26] Safavi, A., Kefayat, A., Mahdevar, E., Abiri, A., & Ghahremani, F. (2020). Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches. Vaccine, 38(48), 7612-7628.

[27] Mishra, S. (2020). T cell epitope-based vaccine design for pandemic novel coronavirus 2019-nCoV.

[28] Quiros-Fernandez, I., Poorebrahim, M., Fakhr, E., & Cid-Arregui, A. (2021). Immunogenic T cell epitopes of SARS-CoV-2 are recognized by circulating memory and naïve CD8 T cells of unexposed individuals. EBioMedicine, 72, 103610.

[29] Bukhari, S. N. H., Jain, A., Haq, E., Mehbodniya, A., & Webber, J. (2021). Ensemble Machine Learning Model to Predict SARS-CoV-2 T-Cell Epitopes as Potential Vaccine Targets. Diagnostics, 11(11), 1990.

[30] Joshi, A., Joshi, B. C., Mannan, M. A. U., & Kaushik, V. (2020). Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. Informatics in medicine unlocked, 19, 100338.

[31] Yashvardhini, N., Kumar, A., & Jha, D. K. (2021). Immunoinformatics Identification of B- and T-Cell Epitopes in the RNA-Dependent RNA Polymerase of SARS-CoV-2. Canadian Journal of Infectious Diseases and Medical Microbiology, 2021.

[32] Dar, H. A., Waheed, Y., Najmi, M. H., Ismail, S., Hetta, H. F., Ali, A., & Muhammad, K. (2020). Multiepitope subunit vaccine design against covid-19 based on the spike protein of sars-cov-2: An in silico analysis. Journal of immunology research, 2020.

[33] Khairkhah, N., Aghasadeghi, M. R., Namvar, A., & Bolhassani, A. (2020). Design of novel multiepitope constructs-based peptide vaccine against the structural S, N and M proteins of human COVID-19 using immunoinformatics analysis. Plos one, 15(10), e0240577.Mitra, D., Pandey, J., Jain, A., & Swaroop, S. (2020). In silico design of multi-epitope-based peptide vaccine against SARS-CoV-2 using its spike protein. Journal of Biomolecular Structure and Dynamics, 1-14.

[34] Mitra, D., Pandey, J., Jain, A., & Swaroop, S. (2020). In silico design of multi-epitope-based peptide vaccine against SARS-CoV-2 using its spike protein. Journal of Biomolecular Structure and Dynamics, 1-14.

[35] Bahrami, A. A., Payandeh, Z., Khalili, S., Zakeri, A., & Bandehpour, M. (2019). Immunoinformatics: in Silico approaches and computational design of a multi-epitope, immunogenic protein. International reviews of immunology, 38(6), 307-322.

[36] Dhama, K., Sharun, K., Tiwari, R., Dadar, M., Malik, Y. S., Singh, K. P., & Chaicumpa, W. (2020). COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. Human vaccines & immunotherapeutics, 16(6), 1232-1238.

[37] Sunita, Sajid, A., Singh, Y., & Shukla, P. (2020). Computational tools for modern vaccine development. Human vaccines & immunotherapeutics, 16(3), 723-735.

[38] Abdelrahman, Z., Li, M., & Wang, X. (2020). Comparative review of SARS-CoV-2, SARS-CoV, MERS-CoV, and influenza a respiratory viruses. Frontiers in immunology, 11, 2309.

PAPER NAME

n turnitin.pdf

WORD COUNT

10714 Words

CHARACTER COUNT

61413 Characters

PAGE COUNT

53 Pages

FILE SIZE

879.6KB

SUBMISSION DATE

May 2, 2022 12:51 PM GMT+5:30

REPORT DATE

May 2, 2022 12:52 PM GMT+5:30

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- Crossref database
- 2% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Bibliographic material

Sandeep

5% Overall Similarity

Top sources found in the following databases:

- 3% Internet database
- Crossref database
- 2% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	link.springer.com Internet	1%
2	cttjournal.com Internet	<1%
3	"Immunoinformatics", Springer Science and Business Media LLC, 2014 Crossref	<1%
4	Rochester Institute of Technology on 2007-04-09 Submitted works	<1%
5	Namrata Tomar. "Immunoinformatics: an integrated scenario : Immun... Crossref	<1%
6	ncbi.nlm.nih.gov Internet	<1%
7	"Coronavirus Disease - COVID-19", Springer Science and Business Med... Crossref	<1%
8	Bulbuli Khanikor. "Comparative mode of action of some terpene comp... Crossref	<1%

9	Sunway Education Group on 2020-07-24 Submitted works	<1%
10	CSU, Los Angeles on 2022-02-21 Submitted works	<1%
11	Central Queensland University on 2020-04-20 Submitted works	<1%
12	University of Keele on 2021-09-28 Submitted works	<1%
13	WWW.medicalnewstoday.com Internet	<1%
14	coek.info Internet	<1%
15	hdl.handle.net Internet	<1%
16	National University of Ireland, Galway on 2020-12-10 Submitted works	<1%
17	www2.oakland.edu Internet	<1%
18	Gabb, H.A.. "Modelling protein docking using shape complementarity, ... Crossref	<1%
19	Jun Zhang, Weichun Tang, Hailong Gao, Christy L. Lavine et al. "Structu... Crossref posted content	<1%
20	University of Northumbria at Newcastle on 2022-04-23 Submitted works	<1%

21	repub.eur.nl Internet	<1%
22	hindawi.com Internet	<1%
23	"Immunoinformatics", Springer Science and Business Media LLC, 2020 Crossref	<1%
24	"Systems Medicine", Springer Science and Business Media LLC, 2016 Crossref	<1%
25	Mahzad Akbarpour, Laleh Sharifi, Amir Reza Safdarian, Pooya Farhang... Crossref	<1%
26	Stephen N. Crooke, Inna G. Ovsyannikova, Richard B. Kennedy, Gregory... Crossref posted content	<1%
27	Tuoyu Zhou. "SARS-CoV-2 triggered excessive inflammation and abnor... Crossref posted content	<1%
28	Yeungnam University on 2019-08-20 Submitted works	<1%
29	journals.plos.org Internet	<1%
30	neurosciencenews.com Internet	<1%
31	Ain Shams University on 2020-07-15 Submitted works	<1%
32	CSU, Los Angeles on 2017-05-30 Submitted works	<1%

33	Namrata Tomar, Rajat K. De. "Chapter 3 Immunoinformatics: A Brief Re... Crossref	<1%
34	National University of Singapore on 2011-04-06 Submitted works	<1%

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi - 110042

CANDIDATE'S DECLARATION

I Sandeep kumar Roll Number: 2K20/MSCBIO/25, student of M.Sc. Biotechnology, hereby declare that the work which is presented in the Major Project entitled –" **Covid-19 vaccine development through immunoinformatics guided multiple epitoping and demonstration of Drug Repurposing through molecular docking"** ‖ in the fulfillment of the requirement for the award of the degree of Master of Science in Biotechnology and submitted to the Department of Biotechnology, Delhi Technological University, Delhi, is an authentic record of my own carried out during the period from January- May 2022, under the supervision of Dr. Yasha Hasija.

The matter presented in this report has not been submitted by me for the award for any other degree of this or any other Institute/University. The work has been accepted in SCI/SCI expanded /SSCI/Scopus Indexed Journal OR peer reviewed Scopus Index Conference with the following details:

Title of the Paper: Immunoinformatics Tools: A Boon in Vaccine Development Against Covid-19
Author Names: Kumar, Sandeep and Hasija, Yasha
Name of Conference: Delhi Section International Conference on Electrical, Electronics and Computer Engineering (DELCON-2022)– IEEE Conference
Conference Date and Venue: 11th-13th Feb 2022 at Netaji Subhas University of Technology, New Delhi, India
Registration: Done
Status of Paper: Published
Date of Paper Communication: 31 December 2021
Date of Paper Acceptance: 16th January 2022
Date of Paper Publication: 21 April 2022

Sandeep

Sandeep kumar

Date: 6/05/2022

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College Of Engineering)
Bhawana Road, Delhi-110042

Certificate

I hereby certify that the Project Dissertation titled "**Covid-19 vaccine development through immunoinformatics guided multiple epitoping and demonstration of Drug Repurposing through molecular docking**" which is submitted by **Sandeep Kumar (2K20/MSCBIO/25)**, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Science is recorded for the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any degree or any diploma to this university or elsewhere.

Place: Delhi
Date : 06/5/22

06/05/22

Prof. Yasha Hasija

(Supervisor)

Professor

Department of Biotechnology

Delhi Technological University

06/05/2022

Prof. Pravir Kumar

Head of Department

Department of Biotechnology

Delhi Technological University

Acknowledgement

I would like to express my gratitude towards my supervisor, Prof. Yasha Hasija, for giving me the opportunity to do research and providing invaluable guidance throughout this research. Her dynamism, vision, sincerity and motivation have deeply inspired me. She has motivated to carry out the research and to present my work works as clearly as possible. It was a great privilege and honor to work and study under her guidance. I am extremely grateful for what he has offered me. Her insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future.

I would also like to thank to all PhD scholars of complex systems and genome informatics laboratory Raj Kumar Chakraborty, Jaishree meena, Neha kumari, Priya rai without them I wouldn't be able to complete my thesis.

A special thanks goes to my best friends Deepak kumar , Neha Yadav, Anjali Mishra, Shraddha Mohanty and Shubham arman jha for their moral support, tolerance and help from the beginning to the end.

I would also like the institution Delhi Technological University, Delhi for giving me the opportunities throughout the tenure of study.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly

Sandeep

Sandeep kumar