Construction of Vaccine Against Herpes Zoster Virus Glycoprotein E Using In-Silico Method

By

Jarin Tasnim
ID: 18346012

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (Hons.)

School of Pharmacy
Brac University
February 2023

Declaration

It is hereby declared that

1. The thesis submitted is my original work while completing a degree at BRAC University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. I have acknowledged all of the main sources of help.

Student's Full Name & Signature:

_____

Jarin Tasnim

Student ID: 18346012

Approval

The thesis/project titled "CONSTRUCTION OF VACCINE AGAINST HERPES ZOSTER VIRUS GLYCOPROTEIN E USING IN SILICO METHOD" submitted by Jarin Tasnim (18346012) has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelors of Pharmacy on [Date-of-Defense].

Supervised By:

Mohammad Kawsar Sharif Siam
Senior Lecturer
School of Pharmacy
Brac University

Approved By:

Assistant Dean &
Program Director:

Professor Dr. Hasina Yasmin
Program Director and Assistant Dean
School of Pharmacy
Brac University

Dean:

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
Brac University

Ethics Statement

While working on this thesis, no unethical practices are done. There are no clinical trials on living things during this study. No harm doing to any humans or animals during work. Very carefully maintained all applicable ethical norms.

Acknowledgments

I would want to begin by giving thanks to the Almighty, who is the source of all life, power, insight, grace, and kindness and is also our creator. All praise be to the Almighty who has given me the fortitude and bravery to do this task.

This project would not have been accomplished without the assistance of the individuals who are acknowledged here. I also want to express my deep gratitude to my supervisor, Mohammad Kawsar Sharif Siam, Senior Lecturer, School of Pharmacy, Brac University, for his professional guidance and advice as well as for always being there for me anytime I needed support, motivation, or encouragement.

Last but not least, I want to thank my parents and my husband for always being there for me and for inspiring me to work harder in all I do. Without their fervent prayers and constant support, I would not have progressed this far.

Jarin Tasnim

ID: 18146012

School of Pharmacy

Brac University

Abstract

Herpes zoster virus is a member of Alphaherpesvirinae subfamily and genus Varicellovirus of herpesviridae. Herpes zoster virus or Shingles is a viral disease. Recently, people are more suffering due to this virus. The varicella zoster virus (VZV), which also causes chickenpox, is what causes shingles. The initial virus infection in chickenpox, also known as varicella, often takes place during childhood or adolescence. There is only live attenuated vaccine for treating HZV and there are some antiviral drugs which are effective in early stage of this disease. But there is no promising in silico based vaccine for HZV. In this study, construction of herpes zoster vaccine through in silico method has been proposed against herpes zoster virus envelop glycoprotein E interferons. From UniProt database server Glycoprotein E sequence has been retrieved in FASTA file. Consequently, some online server are utilized to determine CTL, HTL and B-cell epitopes and other predictions. During study, a CTL , a HTL and a b-cell epitope shows best antigenic score and high inducing affinity with that herpes zoster glycoprotein E. The antigenicity rise happens from 0.51 to 0.53 when all antigenic epitopes are added in the construction. Z score, C-score ,GRAVY score are positive. The protein is hydrophilic which indicates it can be a good developed vaccine.

Keywords: In silico method, Toll-like receptor, Herpes Zoster Virus, Epitope based, Computational method.

Table of Content

List of Figures

List of Tables

List of Acronyms

1. VZV - Varicella Zoster Virus.
2. VHS - Virion Host Shutoff.
3. pOKA- parental strain Oka.
4. vOKA- vaccine strain Oka.
5. CTL- Cytotoxic T-Lymphocyte
6. IFN- Interferon.
7. MHC- Major histocompatibility complex.
8. HTL- Helper T-lymphocyte.
9. TLR- toll-like receptor.
10. IL4-Interleukin 4
11. GRAVY- Grand average of hydropathicity index.
12. IL10- Interleukin 10.
13. Inj- Injection.
14. NKC- Natural Killer Cell.

Chapter 1: Introduction

1.1 Structural information of Herpes Zoster Virus:

The varicella zoster virus (VZV) is one of many types of human herpesviruses that may infect humans. It is mainly a pathogenic human alpha-herpesvirus. During primary infection, VZV causes varicella known as chickenpox, but it can also go latent in the dorsal root ganglia and reawaken years later to produce herpes zoster which causes Shingles. Herpes zoster is a painful or pruritic cutaneous vesicular eruption that occurs in a typical dermatomal distribution. This may happen on its own or in response to any of a number of different triggers (Kennedy & Gershon et al,. 2018).

Shingles, a general illness characterized by a blistered, itchy rash on the skin. Usually, the rash appears as a single stripe across both sides of the face. There may be tingling or soreness there two to four days before the rash appears. Usually, the rash goes away in two to four weeks. Post herpetic neuralgia is the name for persistent nerve pain that can endure for months or years. People with weak immune systems may experience a widespread rash. Vision loss could happen if the rash affects the eyes. VZV reactivation within the body is what causes shingles. A VZV infection is the cause of the first case of chickenpox. The chickenpox virus may still be dormant in nerve cells after the symptoms have subsided. Older age, compromised immune system, and having had chickenpox before the age of 18 months are risk factors for reactivation.

Signs and symptoms are frequently used to make a diagnosis. In people between the ages of 50 and 80, the shingles vaccine reduces the risk of developing shingles by around half. Post herpetic neuralgia rates are also reduced by it. If taken within 72 hours of the onset of the rash, antiviral drugs such Acyclovir can lessen the severity and length of the illness. The acute pain

may be treated with NSAIDs or opioids. A third of people are predicted to experience shingles at some point in their lives. The illness can also affect children. Between 1.2 and 3.4 new cases per 1,000 people are reported annually (Chowdhury et al., 2016).

National seroprevalence data from the pre-vaccine era illustrates that less than 2% of individuals in the United States were vulnerable to varicella infection, and that more than 95% of people in the country contracted the disease before the age of 20 and varicella has attack rate of 90% with close contacts so it is highly communicable disease but people cannot get shingles or zoster from another person unlike chickenpox.

VZV belongs from same subfamily of HSV 1 and HSV 2 and this is a member of varicellovirus genus. The spherical icosahedral symmetry of the virus measures 159–200 nm and this virion contains the double-stranded DNA genome contains 125–240 kbp nucleotides, and 162 hollow hexagonal and pentagonal capsomeres make up the nucleocapsid, which coupled with the icosahedral protein capsid has an average diameter of 100 nm. A lipoprotein nature envelope surrounds the nucleocaspid. The particle inside the envelope has a diameter of 180 to 200 nm and is pleomorphic to spherical. Virus entrance is mediated by spikes of viral glycoproteins, which are 8 nm long and project from the trilaminar lipid host-derived envelope (Zerboni et al., 2014).

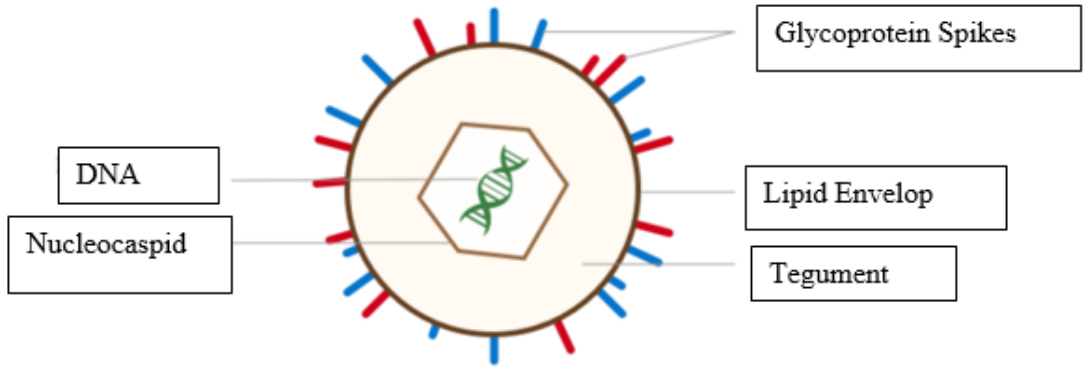

Figure 1: structure of varicella zoster virus. (Al-Anazi et al., 2019).

In Figure 1, structure of varicella zoster virus has been shown. The virus contains double stranded DNA covered by nucleocapsid. As this virus cell contains Glycoprotein spikes in its lipid envelope.

According to Creative Diagnostics (2019), Glycosaminoglycan, fusion, and initial attachment to host cell surfaces are all mediated by the proteins gB, gE, and gH, which are widely distributed. The tegument, an amorphous proteinaceous layer outside the capsid of mature virus particles, is encircled by a lipid envelope made of host cell membranes. The tegument is made up of enzymes like VP16, which is in charge of altering cellular proteins, enzymes involved in viral nucleic acid replication, and VHS (Virion Host Shutoff), which stops the creation of protein in the cytoplasm of the host cell (Zerboni et al., 2014)..

Moreover, the development of a live zoster vaccine was directly influenced by the availability of live attenuated varicella vaccine. It was necessary to employ a vaccine formulation 14 times stronger than varicella vaccine to enhance cellular protection in elderly patients who had chickenpox many years earlier. 50–60% of people over the age of 60 experienced protections against zoster after receiving this ZostavaxTM vaccination, which was created by Merck & Co. Unfortunately, this protection starts to diminish in some situations as early as the first year following inoculation and disappears completely within eight years. Boosters should not be used. Immunocompromised individuals are not guaranteed to be safe from this vaccine, as it may result in serious VZV infections. GlaxoSmithKline set out to create a vaccine, ShingrixTM that would be safe for vaccination in immunocompromised people while also providing superior protection for the elderly. The vaccination should be administered twice, with an interval of between two and six months between the two doses. Immunized healthy adults up to the age of 70 are protected from contracting the disease by roughly 97%. Immunocompromised patients are actively participating in immunogenicity and safety trials(Al-Anazi* et al., 2019). The most troublesome part of ShingrixTM is that it causes a lot

of adverse effects right after you get it injected, such as responses at the injection site, fever, and general malaise. However, major side effects are rare. Oral antiviral medication, such as acyclovir, famciclovir, or valacyclovir, should be given to patients as soon as possible in the event that they develop herpes zoster. Intravenous acyclovir may be given to patients suffering from severe herpes zoster, particularly those who are immunocompromised, particularly at the beginning of treatment .(Kennedy & Gershon et al,. 2018).

1.2 Genome structure of VZV:

VZV is accommodated by at least 70 genes. This virus consists of a double-stranded DNA genome, containing 125kb in length and this genome has a unique long region which is bounded by terminal long and internal long repeats. Additionally, a unique short region is there which is bounded by internal and terminal short repeats. At least three immediate-early (IE) proteins that are found in the tegument of virions and control viral transcription are encoded by VZV. Alpha herpesvirus, beta herpesvirus, and gamma herpesvirus are the three subfamilies of herpesviruses, and the VZV genome has approximately 41 "core genes" that are shared by all three. The VZV DNA polymerase, IE4, the single-stranded DNA-binding protein, helicase-primase components, ribonucleotide reductase, uracil-DNA glycosylase, dUTPase, DNase, ORF47 protein kinase, major capsid protein, protease, assembly protein, many tegument proteins, gB, gH, gL,gM and gN.

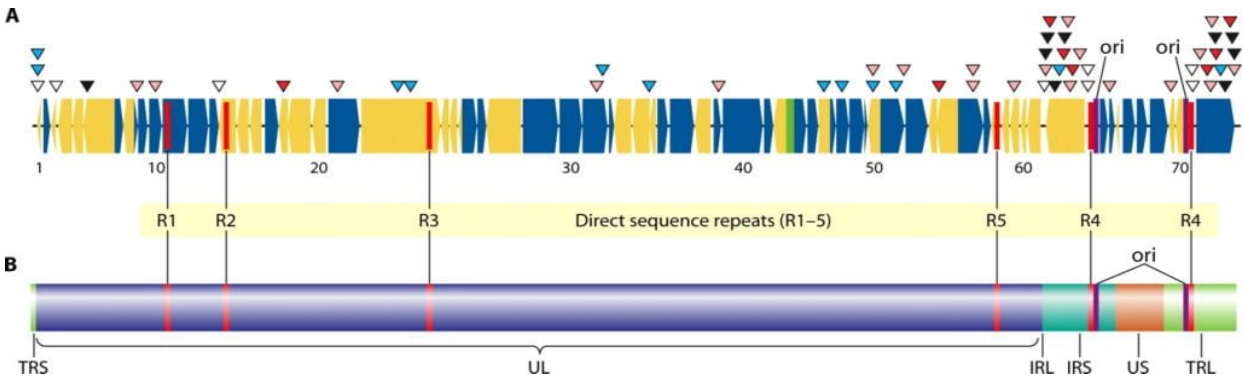

Figure 2: Genome of Varicella Zoster Virus. (Schmid & Jumaan et al,. 2010).

Figure 2 illustrates genome of varicella zoster virus. (A) indicates Model of the VZV genome illustrating the single-base differences between pOka (parental strain Oka) and vOka (vaccine strain Oka). White triangles, SNP located in noncoding region; black triangles, SNP present uniformly among vaccine strains leading to amino acid substitutions; red triangles, SNP present uniformly among vaccine strains that does not lead to amino acid substitutions; pink triangles, SNP present as a mixed base in vaccine preparations and vaccine-associated SNP that leads to amino acid substitutions; blue triangles, SNP present as a mixed base in vaccine preparation that does not lead to amino acid substitutions. Dark blue ORFs are transcribed from the sense strand; yellow ORFs are transcribed from the antisense strand. (B) Display of VZV genome architecture. TRS, terminal repeat short; TLR, terminal repeat long; UL, unique long; US, unique short; IRL, internal repeat long; IRS, internal repeat short; ori, origin of replication.

(B) Displays of VZV genome architecture. TRS, terminal repeat short; TLR, terminal repeat long; UL, unique long; US, unique short; IRL, internal repeat long; IRS, internal repeat short; ori, origin of replication.

Additionally, according to NCBI, There are at least 70 genes encoded by VZV, and three of them (ORF62, 63, 64) are located in both the long and short repeat regions. At least three tegument-localized, transcriptionally regulatory IE proteins are encoded by VZV. Promoter regions that are intermediate, late, or early can all be activated by IE4 and IE62. Multiple VZV promoters are repressed by IE63, and IE63 also reduces interferon-alpha activity by binding to anti-silencing protein 1. Even though it has not been established that ORF61 is an IE gene, the ORF61 protein activates IE, early, and late viral promoters (Cohen et al,. 2010.).

1.3 Life cycle and replication of VZV:

Once VZV particles have made it to mucosal epithelial points of entry, they can infect the human host. Whenever VZV has replicated locally, it travels to the tonsils and other nearby

lymphoid tissues, where it infects T cells by replicating there. The T cells become infected and transport the virus to the skin where it can replicate. After that VZV reaches the neuronal nucleus via axons, it becomes latent in the sensory ganglia. A second phase of replication in skin is possible after reactivation from latency, and this usually results in lesions in the dermatome innervated by the afflicted sensory ganglion (Zerboni et al., 2014).

Replication:

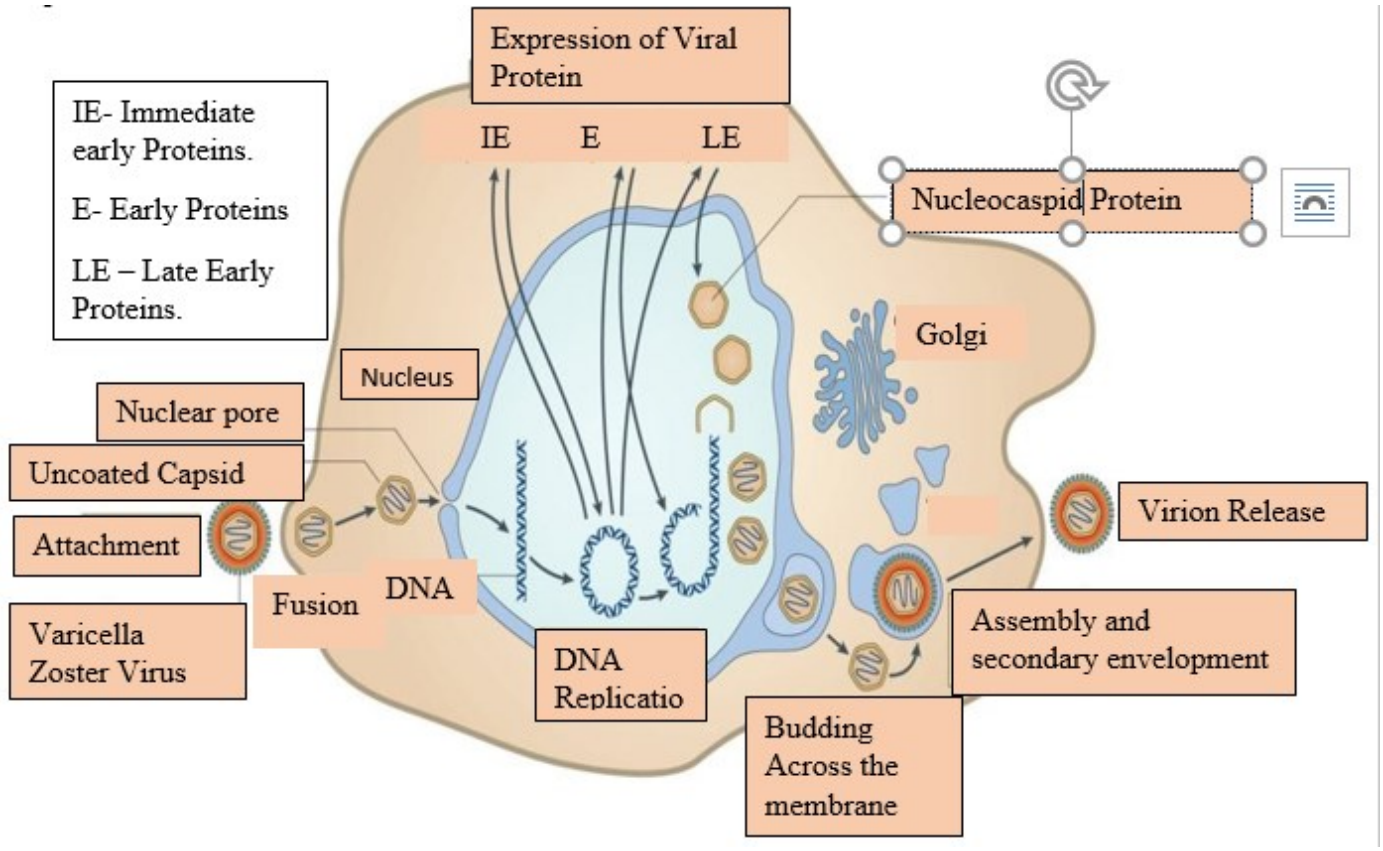

Figure 3: Replication of Varicella Zoster Virus (Zerboni et al., 2014).

In figure 3, replication of varicella zoster virus has been shown the details has been shown in a flow chart below in figure 4.

Replication Steps:

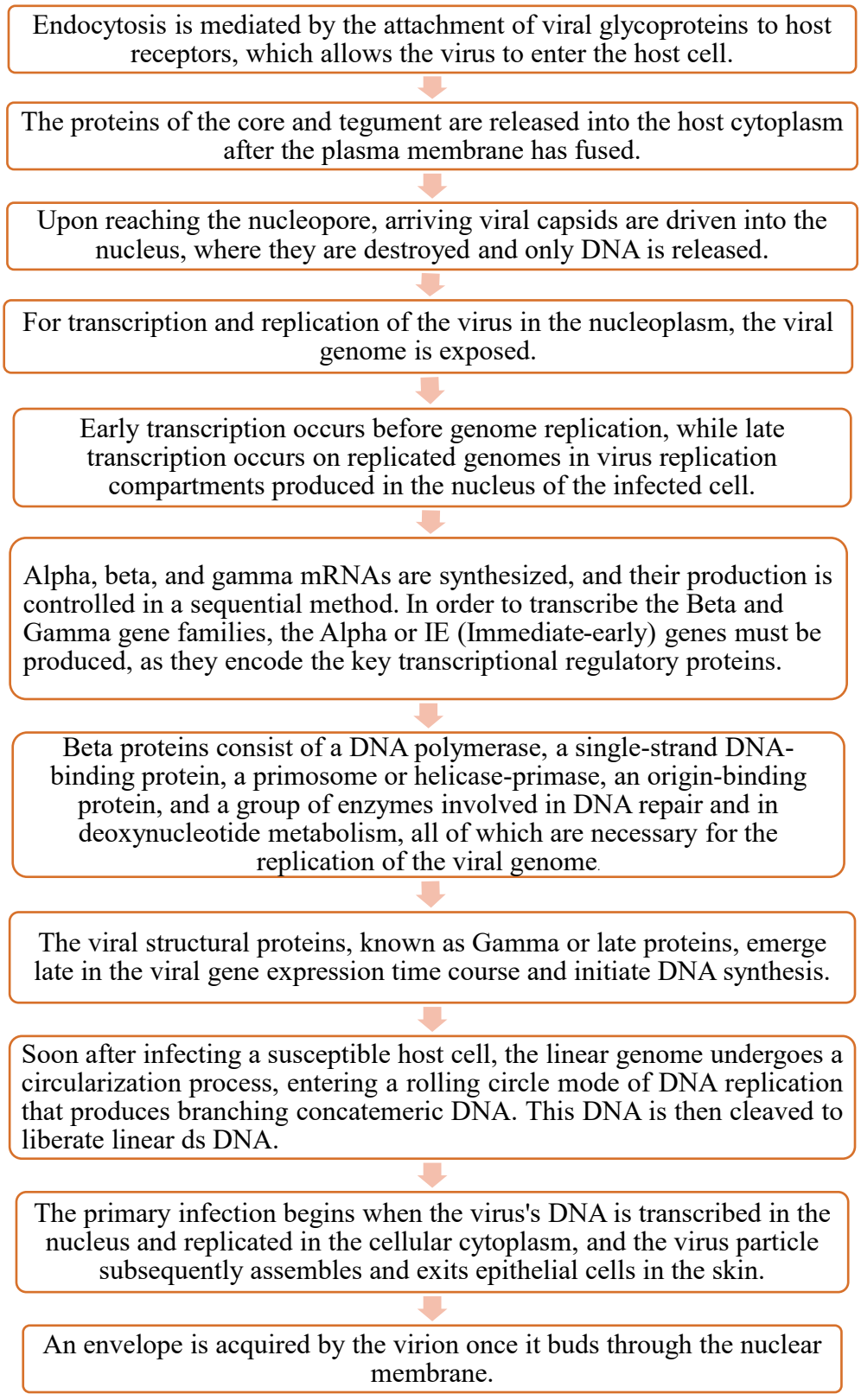

Figure 4: Replication of VZV.

1.4 Pathogenesis of VZV:

The pathogenesis of herpes zoster virus (HZV) involves several stages, starting from initial infection to the development of shingles and complications.

1. Initial infection: The virus enters the body through inhalation of respiratory droplets or direct contact with an infected person. It initially infects the upper respiratory tract and then spreads to regional lymph nodes.
2. Primary viremia: The virus then enters the bloodstream and infects organs such as the skin, mucous membrane, liver, and spleen. This stage is characterized by the appearance of rashes on the skin.
3. Latency: After the initial infection has been resolved, the virus remains latent in sensory ganglia in the body.
4. Reactivation: If the immune system of the infected individual becomes compromised, the virus can reactivate and spread to other cells in the body, leading to the development of shingles.
5. Rash development: The reactivated virus infects cells in the skin, leading to the development of a painful skin rash characterized by fluid-filled blisters.
6. Complications: In some cases, the virus can cause complications such as postherpetic neuralgia (PHN), which is a persistent pain that can persist long after the rash has disappeared. Other complications include vision loss, hearing loss, and bacterial infections of the skin.

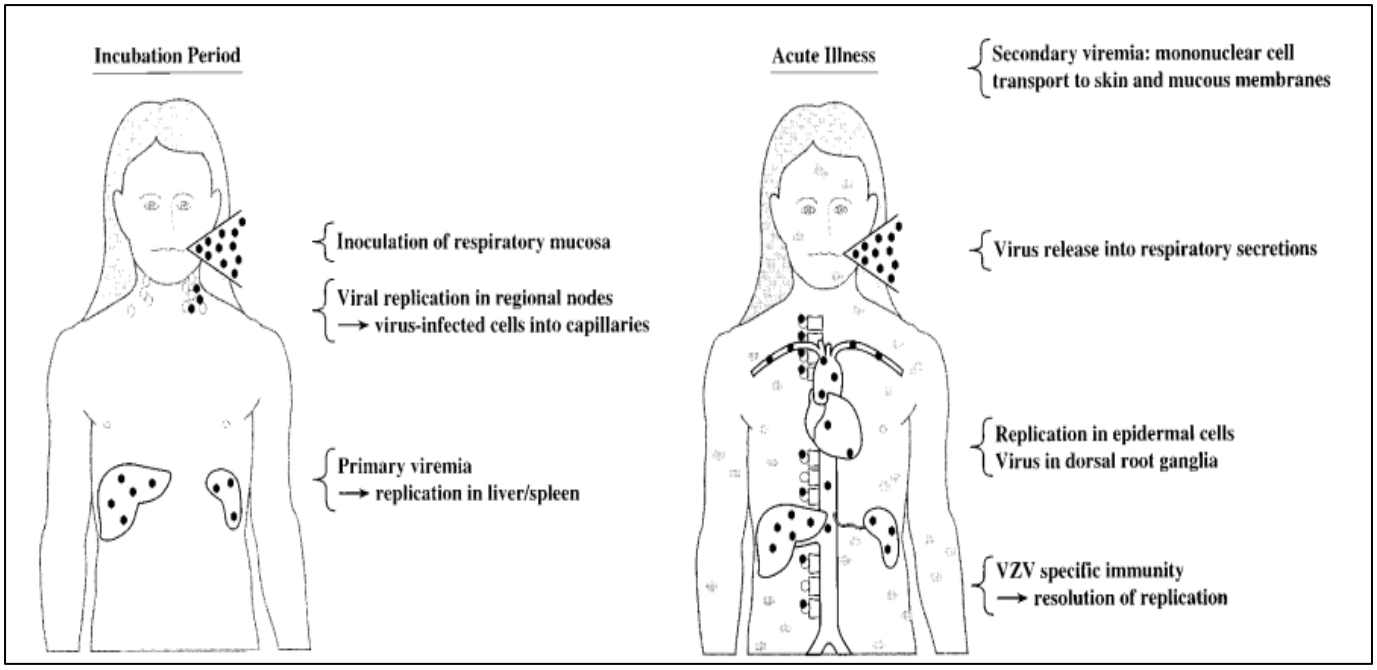

Figure 5: Pathogenesis of VZV (Arvin et al,. 1996).

Figure 5 explains pathogenesis of varicella zoster virus. According to A. N. N. M. Arvin, (1996),

- **Virus entry:** through inhalation of respiratory droplets or contact with infected persons
- **Initial infection:** in the upper respiratory tract
- **Replication and proliferation:** in regional lymph nodes of the upper respiratory tract
- **Primary viremia:** spread of the virus in the bloodstream
- **Localization in organs:** skin, mucous membrane, liver, and spleen
- **Secondary viremic stage:** manifesting in characteristic rashes
- **Development of rashes:** from macules to fluid-filled vesicles to pustules and scabs
- **Clearance of virus-infected host cells**: through T-cell-mediated immune mechanisms and antibody-dependent cell-mediated cytotoxicity

- **Penetration of nerve endings:** transport to dorsal root ganglia
- **Latency:** virus remains dormant in cranial-nerve and dorsal-root ganglia
- **Reactivation:** can trigger dermatomal outbreaks or widespread "shingles" in immunocompromised individuals.

Chapter 2: Methodology

In the realm of vaccine design, computational vaccination is gaining traction as a formidable strategy. The creation of potent vaccines for an array of viruses entails a multifaceted analysis of a multitude of factors such as the forecasting of T cell and B cell epitopes, scrutiny of antigen processing, antigenicity and population coverage, conservation analysis, allergenicity evaluation, toxicity prediction, and protein-peptide docking. These analyses can be facilitated by an assortment of bioinformatics tools and web-based servers that have been designed for this purpose (Kardani et al., 2020).

A multitude of in silico methods exist to explore linear B-cell epitopes, helper T lymphocytes (HTL), and cytotoxic T lymphocytes (CTL) epitopes. The use of state-of-the-art bioinformatics techniques allows for the evaluation of the antigenicity, human population coverage, physicochemical properties, toxicity, allergenicity, and secondary structure of the vaccine to guarantee its efficacy. Moreover, there are in silico techniques available for forecasting, enhancing, and verifying the three-dimensional (3D) structures of the vaccine candidates. Epitope discovery experiments can be expedited with the aid of in silico models. These models use semi-automated techniques to identify potential epitopes and employ high-throughput experimental tests to detect MHC-peptide binding affinities. This saves both time and resources while maintaining high accuracy in classifying peptide binding and identifying immunogenic peptides. However, the results of these in silico predictions should be verified by a competent immunologist to ensure their reliability and relevance.

During in silico vaccine design of VZV,the chosen method and tools have been mentioned in figure 06.

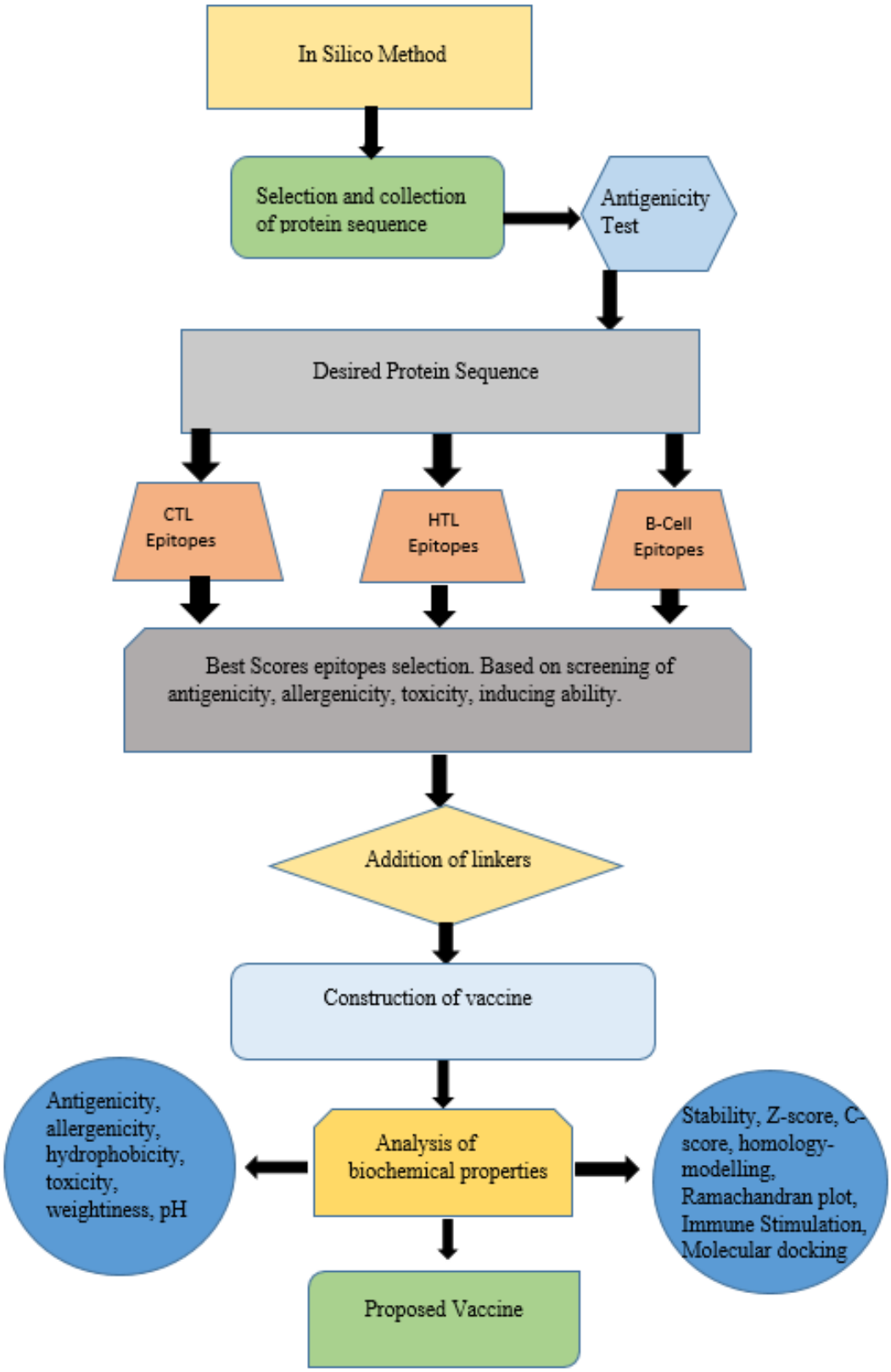

Figure 6: Methodology of constructed Vaccine.

2.1 Selection and Collection of Protein Sequence of Herpes Zoster Virus:

UniProt Knowledgebase was chosen to select protein sequence of herpes zoster vaccine. The UniProt Knowledgebase strives to offer a broad, and openly available collection of protein sequences that are annotated with functional details, in order to provide its users with comprehensive knowledge (Consortium et al,.2021). The most effective application of UniProtKB is to annotate binding sites in for biologically relevant (cognate) ligands using the chemical ontology ChEBI, which is both comprehensive and computationally tractable (Chemical Entities of Biological Interest) and additionally, to aid in the investigation and prediction of functionally significant interactions between protein sequences and structures and small molecule ligands.. (Coudert et al., 2023). (Website: www.uniprot.org)

After entering the website, there is a search-bar where desired protein sequences can be found by entering disease name and numerous of results will be there as outcome.

So, the way information was gotten from UniProtKB:

Chrome Search Bar → UniProtKB → Search Engine → Herpes Zoster Virus → Table format was chosen→ Desire protein was selected → all information about the protein sequence was taken → sequence in FASTA format was downloaded .

2.2 Identifying Antigenicity of Selected Protein Sequence:

The activation of the immune response by B- and T-cells is dependent on antigenic determinants. According to Huang and Lai (2021), the host's antibody response to a virus is a measure of its antigenicity. Each virus causes the host to produce a unique set of antibodies, which originate from different parts of the germline and are typically honed through somatic

mutation to produce antibodies with a particularly strong affinity for the antigen. Antigenic sites are the specific locations on proteins that are identified by specific antibodies. (Website: http://www.jenner.ac.uk/VaxiJen.)

Vaxijen 2.0 server was utilized for the prediction of antigenicity of the protein sequence. To date, VaxiJen is the only server that can forecast protective antigens against bacteria, viruses, and tumors without requiring an alignment. Models constructed from ACC preprocessed amino acid characteristics can be found on the server. Internal leave-one-out cross-validation on training sets and external validation on test sets were used to assess the models' prediction performance. All three models have internal and external validation accuracies between 70% and 89%. Combinations of the positive set and five separate negative sets revealed outstanding stability in the models. Therefore, VaxiJen is an accurate and consistent method for determining which protective antigens will be produced. In reverse vaccinology, it can be employed independently or in tandem with other bioinformatics methods (Doytchinova & Flower et al., 2007b).

In the Vaxijen 2.0 server, Virus was selected as target organism and the threshold was kept 0.5. Since the majority of the models performed best at an accuracy criterion of 0.5 threshold, this value was selected across the board (Doytchinova & Flower, 2007).

Chrome search bar → Vaxijen 2.0 → protein sequence in fasta format → virus (target organism) → threshold: 0.5 → Submit.

2.3 Determination of Cytotoxic T-lymphocyte (CTL) Epitopes:

Rational vaccine development relies on accurate predictions of Cytotoxic T lymphocyte (CTL) epitopes. Mostly, they can reduce the time and money spent on experiments to find epitopes. For determining the presence of human CTL epitopes in a protein of interest, NetCTL 1.2 is

used during constructing final vaccine. It's mainly a web-based program developed for this purpose. This is achieved by combining TAP transport efficiency, MHC class I affinity, and predictions of proteasomal cleavage (Larsen et al., 2007).

The URL: https://services.healthtech.dtu.dk/service.php?NetCTL-1.2 .

While determining CTL epitopes, there are some parameters which are considered such as binding capacities, hydrophobicity, toxicity, antigenicity. So the CTL epitopes are chosen through the following steps

NetCTL 1.2 server → Services and products → Submission → Selected protein sequence is inputted in FASTA format→ Submit

Result shows the location of peptide sequence which has effective binding with Major histocompatibility complex (MHC).

2.4 Determination of MHC class I Alleles:

After getting peptides from the result chart of NetCTL 1.2, NetMHC Pan 4.1 server has been utilized to determine which peptides shows strong bonds.

The URL: https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1

NetMHCpan 4.1→ services and products → Submission → peptides (Selected from NetCTL 1.2 server) are inputted in FASTA format → peptide length: 9mer peptides →HLA all → submit.

These 9mer peptides are filtered by toxicity test by using Toxinpred server. This server shows which peptides are toxic and which are nontoxic.

The URL: (http://crdd.osdd.net/ raghava/toxinpred/).

ToxinPred is the first of its type in silico technology that can accurately forecast the toxicity of peptides. It will be helpful in finding harmful areas in proteins and developing the least hazardous peptides. Peptide-based medication research will receive a boost from ToxinPred's development (Gupta et al., 2013).

Prediction of toxic peptides done by following steps:

ToxinPred → Batch Submission → selected peptide sequences in Fasta format → Run analysis.

The result comes with a data table which shows the toxic and non-toxic peptides separately.

2.5 Determination of Helper T Lymphocyte (HTL) Epitopes:

Peptide binding prediction to MHC is a potent tool for estimating the likely specificity of a T-cell immune response. Tools used to predict MHC-ligand interactions are typically trained on binding affinities or ligands eluted by mass spectrometry (Reynisson et al., 2020).

For emerging HTL epitopes, NetMHC II 4.0 is used at the very first place. Steps that are followed-

NetMHCIIpan-4.0 →service and products → Submission → Protein sequences in FASTA Format → peptide length: 15 → A1 Alleles: 20 at a time →Submit→SB / WB collect

Collecting SB (Strong Bindings) and WB (Weak Bindings) is very important. "Strong binding" refers to the predicted peptides that are most likely to bind to the MHC class I molecules with high affinity. It is indicating that they are more likely to elicit a T-cell immune response compared to the peptides with lower predicted binding affinities.

2.6 Assessment Based on Cytokine Stimulating Ability of Strong Binding HTL Epitopes:

IFN epitope is filtered by utilizing this server (http://crdd.osdd.net/raghava/ifnepitope/) -

IFN epitope → all strong binder HTL epitope using FASTA format → Submit

A result box comes with IFN positive and negative epitopes. All positive IFN epitopes are collected.

After that, IL4 Pred server is used to find out il4 inducer peptide. The URL: http://crdd.osdd.net/raghava/ifnepitope/

IL4 server → virtual screening of IL4 peptides → IFN positive peptides in FASTA format → threshold- 0.2 → virtual screening

Peptides which are IL4 inducer and non-inducer comes in a result chart. Inducer Il4 peptides are collected.

IL10 server is applied to find out if the collected peptides are IL10 inducer or not. Server:

https://webs.iiitd.edu.in/raghava/il10pred/index.html

IL10 server → predict → IL4 inducer peptides in FASTA format → default settings → Run Analysis

Results is showing non inducer and inducer peptides. Inducer peptides are taken.

So, eventually, peptides which are collected at the end by IL10 pred, these are HTL peptides candidate for final vaccines as they are IFN positive, IL4 inducer and IL10 inducer.

More specific, all the candidate peptides are screened through Vaxigen 2.0 to check out antigenicity.

2.7 Determination of B-cell Epitopes

A critical first step in the creation of epitope-based vaccines, antibody production, illness prevention, and disease detection is the identification of B-cell epitopes (Shepherd & Lee et al., 2018). To be able to combine with the B-cell, B-cell epitopes must be found in the solvent-exposed area of the antigen. Predicting the structural protein sequence's surface accessibility is significant (Z. Yang et al., 2021).

To figure out B-cell epitopes for final vaccine, IEDB server (http://tools.iedb.org/main/bcell/) is used. The Immune Epitope Database, or IEDB for short, is a particular kind of internet interface that gives access to several epitope-related technologies that have been carefully defined and confirmed (Vita et al., 2015).

IEDB server → B-cell tools → prediction of linear epitopes from protein sequence → Home → Antigenicity texted protein sequence in plain format → Bepi linear epitope prediction 2.0 → submit.

A new webpage appears which shows b-cell epitopes in different length and graph.

An antigenicity test is done via Vaxigen 2.0 server using threshold 0.5 to find of best b-cell epitopes for final vaccine.

2.8 Determination of Linkers:

For the final vaccine construction, EAAAK, GPGPG and KK linkers are selected.

The vaccine construct's N-terminal end had an EAAAK linker added in the current investigation. EAAAK is a peptide linker with a closed-packed backbone that forms a rigid -helix and intramolecular hydrogen bonds. Comparing rigid linkers versus flexible linkers, there

are a number of benefits. By maintaining a set distance and little interference between the epitopes, EAAAK linkers effectively separate the functional domains while maintaining the specific functional characteristics of each epitope (Manuscript, 2014). This facilitates the efficient domain separation in a bifunctional fusion protein (Arai et al., 2001).

It has been shown that GPGPG linkers can stimulate TH lymphocyte (HTL) responses, which is essential for a multi-epitope vaccination. Moreover, the GPGPG linker is a useful tool for removing junctional immunogenicity, which results in the restoration of the individual epitopes' immunogenicity (Livingston et al., 2002).

Now a days KK linker was used to link b-cell epitopes (Sarkar et al., 2020). Cathepsin B, a lysosomal protease involved in the digestion of the antigenic peptides for their presentation on the cell surface in an MHC-II limited antigen presentation, has as its target the Lysine linker. Also, by preventing the development of antibodies for the peptide sequence that arises when individual epitopes are connected linearly, it is essential in lowering the junctional immunogenicity (Yano et al., 2005). Studies show KK linkers increase immunogenicity (Paper et al., 2016).

2.9 Assembly of peptides: vaccine construction

For final vaccine, assembly of all peptides and linkers are done in corresponding way.

At first, selected protein sequence (PS) is applied then linker is added as a bridge between PS and CTL. Secondly, CTL is added and another linker is linked which is a bridge between CTL and HTL. Finally HTL and B-cell is added through another linker.

Table 1: Process of vaccine construction.

Protein Sequence	Linker EAAAK	CTL	Linker GPGPG	HTL	Linker KK	B-cell

2.10 Biochemical Analysis of Vaccine candidate:

To check vaccine's stability, Expasy Protparam tool is relied on. The URL: https://web.expasy.org/protparam/

Many physicochemical qualities that can be inferred from a protein sequence are computed by ProtParam. The protein under examination doesn't need to be described in any further detail. Either a Swiss-Prot/TrEMBL accession number or ID, or a raw sequence, can be used to identify a protein. Numbers and white space are disregarded (Gasteiger et al., n.d.).

Expasy ProtParam tool → vaccine sequence → compute parameters

It shows, molecular weight, hydrophilicity or hydrophobicity, number of Amino acids, theoretical pI, atomic composition, formula, and number of atoms and so on. It makes easier to analyze.

For vaccine construction, Frame for molecular weight is under 100kDa is chose. More than that is considering heavy and rejected.

2.11 Assessment of Antigenicity and Toxicity and Allergenicity:

Vaxigen 2.0 server is applied to check vaccine antigenicity using threshold 0.5 and also this time it must be observed that antigenicity is increased or decreased in compare to plain protein sequence.

T3DB server is used to determine vaccine toxicity. The Toxin-Toxin Target Database (T3DB - www.t3db.ca). About 2900 commonly occurring hazardous compounds were listed in the T3DB, along with comprehensive details on their chemical characteristics, descriptions,

targets, toxic effects, toxicity thresholds, sequences (for both targets and toxins), processes, and references (Wishart et al., 2014).

T3DB → sequence search → Vaccine → submit

If result page shows no result it indicates the vaccine is nontoxic and non-harmful.

After that , AllergenFP v.1.0 is applied to confirm the vaccine's allergenicity.

The URL : https://ddg-pharmfac.net/AllergenFP/index.html

AllergenFP v.1.0 → home → vaccine sequence → get the result

This server shows if the sequence provides allergic action or not.

2.12 Homology Modeling and C-score:

I-tasser server is used to determine homology modelling. Website: https://zhanggroup.org/I-TASSER/

I-TASSER was initially developed for simulations of iterative threading and assembly used in modeling protein structure. By comparing structure predictions with recognized functional templates, it was recently expanded for structure-based function annotation. Here, the I-TASSER Suite is provided, a standalone application that implements the pipelines for modeling protein structure and function based on I-TASSER. Although the community has long used the online I-TASSER server, the lack of adequate computer power from a single laboratory has hindered widespread implementations of these techniques. There is an anticipation that the creation of the standalone package will eliminate the restrictions on computational resources, enabling the benchmarking of novel structural and functional modeling techniques (J. Yang et al., 2014).

I-tasser → vaccine sequence → Email / password → Run i-tasser

I-tasser sends the result to the email which is used during application of server. It may take hours to send the result. The mail contains pdb file of model for further analysis.

A confidence score, or C-score, is used by I-TASSER to gauge the caliber of predicted models. It is determined using the importance of threading template alignments and the convergence parameters from simulations of the structure construction. A model with a high confidence level has a higher C-score, and vice versa. The C-score normally falls between [-5,2].

2.13 Evaluating Ramachandran Plots:

The Ramachandran plot is crucial in figuring out how stereochemically accurate the protein model is. The conformation angles (and), also known as the Ramachandran angles, govern the polypeptide chain's backbone in a protein molecule. The Ramachandran Plot, which is a very helpful indicator for structural biologists and protein modelers, plots these angles against one another. Understanding structural motifs comprising these residues is aided by the identification of the precise interactions that impact the backbone of glycine and preproline, which also demonstrates how the interactions enhance the performance of the existing force fields. High resolution, non-homologous protein crystal structures have been used to study the distribution of amino acid residues in the forbidden area of the Ramachandran plot. Yet, statistical research demonstrates that the distribution of amino acid residues along the polypeptide chain is not random and that each residue has a unique inclination while retaining the secondary structure (Gopalakrishnan et al., 2007).

For Ramachandran plot the server which has been applied is Ramachandran SWISS model. (https://swissmodel.expasy.org/assess/help) this server shows, mobility, Ramachandran outliers, Ramachandran favored, bad bindings etc.

The working process –

Ramachandran SWISS model → Tools →Structure Assessment → upload pdb file → start assessment

2.14 Structure analysis and Z-score Evaluation of Vaccine

For structure analysis and z-score, Prosa-web is utilized, this server provides overall model quality. (https://prosa.services.came.sbg.ac.at/prosa.php)

The energy difference between a protein's native fold and the average of a group of misfolds, expressed in units of the group's standard deviation, is known as a protein's Z-score. The knowledge-based potentials' capacity to distinguish the native fold from other options is frequently assessed using the Z-score. The Z-score range that should exist if one had a proper potential is unknown, though. Here, an estimate for Z-scores derived from protein calorimetric observations are provided. The energy discovered through these experimental data are contrasted with those discovered through computer simulations of a protein lattice. In comparison to the experimental values, it is recommended that the Z-scores estimated from various knowledge-based potentials are typically too low (Zhang & Skolnick, 1998).

The identification of mistakes in experimental and theoretical models of protein structures is a significant issue in structural biology. The ProSA program (Protein Structure Analysis) is a well-known tool with a sizable user base that is frequently used in structure prediction and modeling as well as in the improvement and validation of experimental protein structures. Protein structural analysis is typically a challenging and time-consuming task. The new service described here is a clear-cut and user-friendly addition to the traditional ProSA software that takes advantage of the benefits of interactive web-based applications for the presentation of scores and energy plots that draw attention to potential issues in protein structures. A 3D molecule viewer is used to illustrate and highlight problematic areas of a structure, and the

quality scores of a protein are displayed in relation to all known protein structures. The service focuses on the requirements encountered in the verification of protein structures discovered using theoretical calculations, NMR spectroscopy, and X-ray analysis (Wiederstein & Sippl et al., 2007). In ProSA-web, a negative score indicates that the energy of the protein structure being analyzed is within the range of energies expected for native, well-folded proteins. A lower negative score typically indicates a better-quality protein structure. On the other hand, a positive score indicates that the energy of the protein is outside of this range and may suggest that the protein structure is misfolded or contains structural problems. In short, a negative score in ProSA-web is generally considered a good sign for the quality of a protein structure, while a positive score may indicate potential problems or errors in the structure (Sippl et al., 1993).

The application process –

Prosa-Web → choose file → pdb file → analyze

Results comes with Z-score and a graph. Graph shows a blue gray whitish area where z-score pointed as black dot. That black zot should be in the gray-blue area. If it is in white area, the model should be canceled or modified.

2.15 Molecular Docking of Vaccine:

For molecular docking of vaccine with human receptors, Patchdock server is used.

URL: http://bioinfo3d.cs.tau.ac.il/ PatchDock/help.html

Scoring functions are used in molecular docking systems to calculate the binding energies of anticipated ligand-receptor complexes. The binding constant (Kd) and the Gibbs free energy (GL) determine the energy variation brought on by the development of the ligand-receptor structure. The most significant physical-chemical processes involved in ligand-receptor

binding, including as intermolecular interactions, desolvation, and entropic effects, are assessed in order to predict the binding energy (Ferreira et al., 2015).

A geometry-based molecular docking algorithm is PatchDock. Its goal is to identify docking modifications that produce favorable complementarity of molecule shape. When these transformations are used, they result in both large interface areas and minor steric collisions. The attached molecules' matched local features that have complementary properties are included in a wide interface. . The transformation values are used to describe the spatial relationship and orientation of the two protein structures in the predicted complex. Specifically, the transformation values describe the translation (movement) and rotation of one protein structure relative to the other in order to form the predicted complex. The Connolly dot surface representation of the molecules is divided into concave, convex, and flat patches by the PatchDock algorithm. Then, candidate transformations are created by matching complementary patches. (Schneidman-duhovny et al., 2005)

Patchdock → receptor: 3W3G → ligand: pdb file from selected model → submit form

Result is send via email to the applier.

2.16 Determination of immune response with corresponding stimulations

C-Immsim online tool is chosen for determination of immune response stimulations. An agent-based model like C-ImmSim offer the chance to simulate deterministic behavior, avoiding the mean field approximation and taking into account the effect of the spatial distribution. The versatility of this model can be exploited to make a virtual laboratory, where experiments can be made prior to the design of a real in vitro or in vivo experiment (Castiglione & Bernaschi, 2004).

The C-ImmSim model simulates the humoral response and cytokine generation of the immune system of a mammalian organism to the presence of antigens at the cellular level (such as viruses, bacteria, and so on) (Schneidman-duhovny et al., 2005).

While designing Herpes varicella virus vaccine, steps for stimulation is chosen 546 (Samuel & Castiglione et al., 2023).

The application process:

C-Immsim online → home –login → stimulation → simulation steps (600) → inj no 1→ time step of inj (1) → vaccine sequence in fasta → add injection → inj no 2 →time step of inj (273)→ vaccine sequence in fasta → add injection → inj no 3 → time step of inj 546 → vaccine sequence in fasta → submit job

After submitting, there will be result webpage including 14 graphs. Each graph shows different information. But it can be find out clearly how improved vaccine is after it is administered.

As 1 stimulation step=8hrs, the expectations for this constructed vaccine is

(273*8)/24=91days (3 months - second dose schedule)

(546*8)/24=182 days (6 months –third dose schedule)

Chapter 3: Results

3.1 Selection of Protein Sequence:

At the very first, gE was chosen for the target glycoprotein. VZV glycoprotein E has a unique N-terminal region that is responsible for replication and causing herpes zoster virus from her simplex virus (Berarducci et al., 2010). While searching in uniprot, most of protein sequence could not meet up the criteria of antigenicity based on threshold 0.5 and some do not show virus host on humans. Among them Q9J3M8|GE_VZVO shows antigenicity based on expectations and also infects human. Figure 7 shows,

General information of selected protein,

Gene: Envelope glycoprotein E

ORF: ORF68

OS=Varicella-zoster virus

OX=341980

Virus host: Homo sapiens

Protein Sequence:

MGTVNKPVVGVLMGFGIITGTLRITNPVRASVLRYDDFHIDEDKLDTNSVYEPYYHS
DHAESSWVNRGESSRKAYDHNSPYIWPRNDYDGFLENAHEHHGVYNQGRGIDSGE
RLMQPTQMSAQEDLGDDTGIHVIPTLNGDDRHKIVNVDQRQYGDVFKGDLNPKPQG
QRLIEVSVEENHPFTLRAPIQRIYGVRYTETWSFLPSLTCTGDAAPAIQHICLKHTTCF
QDVVVDVDCAENTKEDQLAEISYRFQGKKEADQPWIVVNTSTLFDELELDPPEIEPG
VLKVLRTEKQYLGVYIWNMRGSDGTSTYATFLVTWKGDEKTRNPTPAVTPQPRGA
EFHMWNYHSHVFSVGDTFSLAMHLQYKIHEAPFDLLLEWLYVPIDPTCQPMRLYST
CLYHPNAPQCLSHMNSGCTFTSPHLAQRVASTVYQNCEHADNYTAYCLGISHMEPS
FGLILHDGGTTLKFVDTPESLSGLYVFVVYFNGHVEAVAYTVVSTVDHFVNAIEERG
FPPTAGQPPATTKPKEITPVNPGTSPLLRYAAWTGGLAAVVLLCLVIFLICTAKRMRV
KAYRVDKSPYNQSMYYAGLPVDDFEDSESTDTEEEFGNAIGGSHGGSSYTVYIDKTR

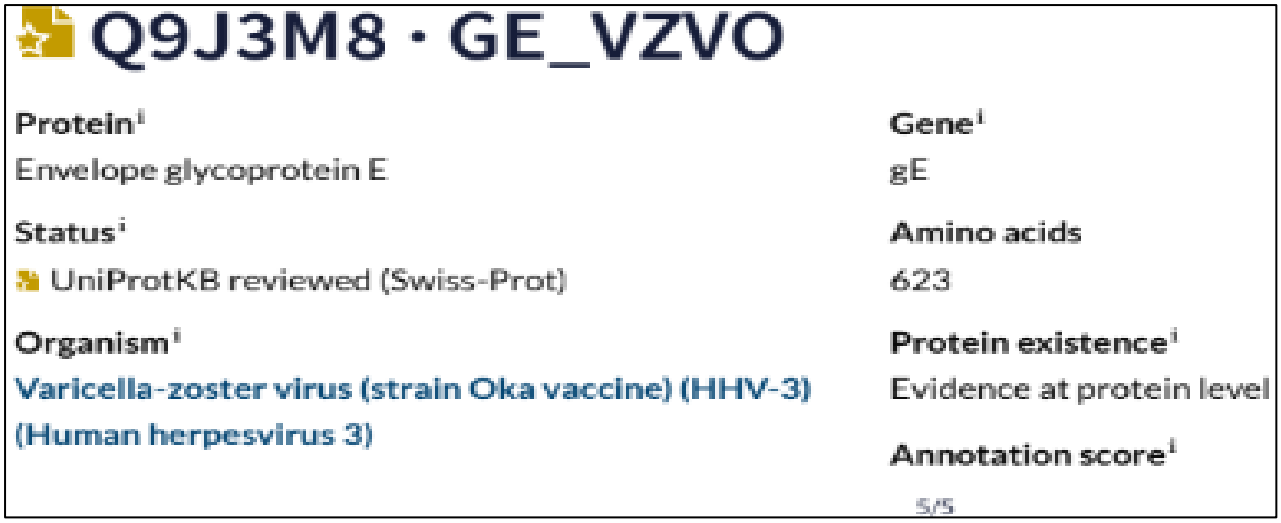

Figure 7: Selected Protein from UniProt (Consortium et al., 2021).

3.2 Antigenicity of Selected Protein:

From the figure it can be aid that Antigenicity score for the selected protein is 0.5165 which indicates it is antigenic.

Figure 8: Antigenicity check through Vaxijen *2.0 of expected protein (Doytchinova & Flower et al*,. 2007).

3.3 Determination CTL Epitopes:

From Netctl 1.2 server, 17 CTL epitopes are collected as these are showing efficient binding ability with MHC I alleles.

The Epitopes are shown in the table 2.

Table 2: CTL Epitopes with their corresponding combined score (Larsen et al., 2007).

CTL Epitopes	Combined Scores
PVRASVLRY	2.7730
DTNSVYEPY	2.5410
HSDHAESSW	0.6060
KAYDHNSPY	3.4030
NAHEHHGVY	2.8770
YTETWSFLP	0.0510
NTSTLFDEL	1.0130
RTEKQYLGV	0.4410
GTSTYATFL	0.6110
TSTYATFLV	0.2680
GAEFHMWNY	2.7860
FSLAMHLQY	2.8980
PTCQPMRLY	2.2930
HMEPSFGLI	0.4750
TPESLSGLY	2.3160
TVDHFVNAI	0.5300
KSPYNQSMY	2.9180

3.4 Determination of MHC Class I Alleles Specific to CTL Epitopes

From NetMHC Pan 4.1 server, 9 peptide sequences are determined which show significant binding capacity with MHC I alleles.

Filtered CTL Epitopes:

DTNSVYEPY,GAEFHMWNY,FSLAMHLQY,TPESLSGLY,KSPYNQSMY,KAYDHNSPY,DTNSVYEPY,NAHEHHGVY,HSDHAESSW

3.5 Prediction of Antigenicity and Toxicity of CTL Epitopes:

After filtering the CTL epitopes through NetMHC pan 4.1 server, epitopes are tested with antigenicity and toxicity to predict best score of result.

Table 3: Antigenicity and toxicity test of CTL epitopes. (Larsen et al., 2007) (Doytchinova & Flower et al., 2007)

CTL epitopes	Antigenicity	Toxicity
DTNSVYEPY	Protective Antigen = **0.4393** **NON-ANTIGEN**	Non-Toxic
GAEFHMWNY	Protective Antigen = **0.5301** **ANTIGEN**	Non-Toxic
FSLAMHLQY	Protective Antigen = **1.0683** **ANTIGEN**	Non-Toxic
TPESLSGLY	Protective Antigen = **0.5240** **ANTIGEN**	Non-Toxic
KSPYNQSMY	Protective Antigen = **0.3825** **NON-ANTIGEN**	Non-Toxic
KAYDHNSPY	Protective Antigen = **0.6851** **ANTIGEN**	Non-Toxic
DTNSVYEPY	Protective Antigen = **0.4393** **NON-ANTIGEN**	Non-Toxic
NAHEHHGVY	Protective Antigen = **0.3996** **NON-ANTIGEN**	Non-Toxic
HSDHAESSW	Protective Antigen = **0.6318** **ANTIGEN**	Non-Toxic

From the table 3, most antigenic and nontoxic epitope sequence are detected which is "FSLAMHLQY" and it is taken as a CTL epitope for further vaccine construction.

3.6 Determination of Strong Binding HTL Epitopes

From NetMHC II pan 4.0 server, 178 peptide sequences are found with strong binding ability. There are many which have same core sequences. Now, they are remove and then further filtered with other tools.

3.7 HTL Epitopes based on Cytokine Stimulating Ability and their Antigenicity:

After determination of strong binding Epitopes, further filtration is done with IFN positivity, IL4 and IL10 inducing ability.

After IL10 inducing prediction, there are only three epitope sequence which show IFN positivity and IL4 and IL10 inducing ability.

For better vaccine construction, antigenicity tests are done of each sequences to predict their curriculum in the construction. Expecting non antigenic epitopes or epitopes with poor antigenicity do not provide better result in final vaccine antigenicity.

Table 4: HTL Epitopes with their inducing ability and corresponding antigenicity. (Jurtz et al., 2017)

HTL Epitopes	IFN	IL4	IL10	Antigenicity
GLYVFVVYFNGHVEA	POSITIVE	Inducer	Inducer	-0.0882(NON-ANTIGEN)
IEPGVLKVLRTEKQY	POSITIVE	Inducer	Inducer	0.5673(ANTIGEN)
PGVLKVLRTEKQYLG	POSITIVE	Inducer	Inducer	0.3209 (NON-ANTIGEN

From table 4, it can be said only one HTL epitope shows antigenic ability where other two are completely non antigenic with desired threshold. So, the HTL epitope which is selected for further constructing vaccine is “IEPGVLKVLRTEKQY”.

3.8 Determination of B-cell Epitopes with their antigenicity and toxicity:

19 peptides are found by the IEDB analysis where length of 1-9 and 101 are not chosen due to their non-antigenicity and high and low weight which causes instability.

Table 5: B- cell epitopes including their antigenicity and toxicity result (Kardani et al., 2020).

B-cell Epitopes	Antigenicity	Toxicity
LFDELELDPPEIEPG	ANTIGEN (0.99)	NON-TOXIC
DEKTRNPTPAVTPQPRG	ANTIGEN (1.07)	NON-TOXIC
NAPQCLSHMNS	NON ANTIGEN (0.2)	NON-TOXIC
QNCEHADNYTAYCLGISHMEPSFGLILHDGG	ANTIGEN (0.8932)	NON-TOXIC
GDDRHKIVNVDQRQYGDVFKGDLNPKPQGQRL	ANTIGEN (0.5234)	NON-TOXIC
EERGFPPTAGQPPATTKPKEITPVNPGTSP	ANTIGEN(1.0788)	NON-TOXIC
YRVDKSPYNQSM	NON ANTIGEN (0.2135)	NON-TOXIC
VDDFEDSESTDTEEEFGNAIGGSHGG	NON-ANTIGEN (0.3480)	NON-TOXIC

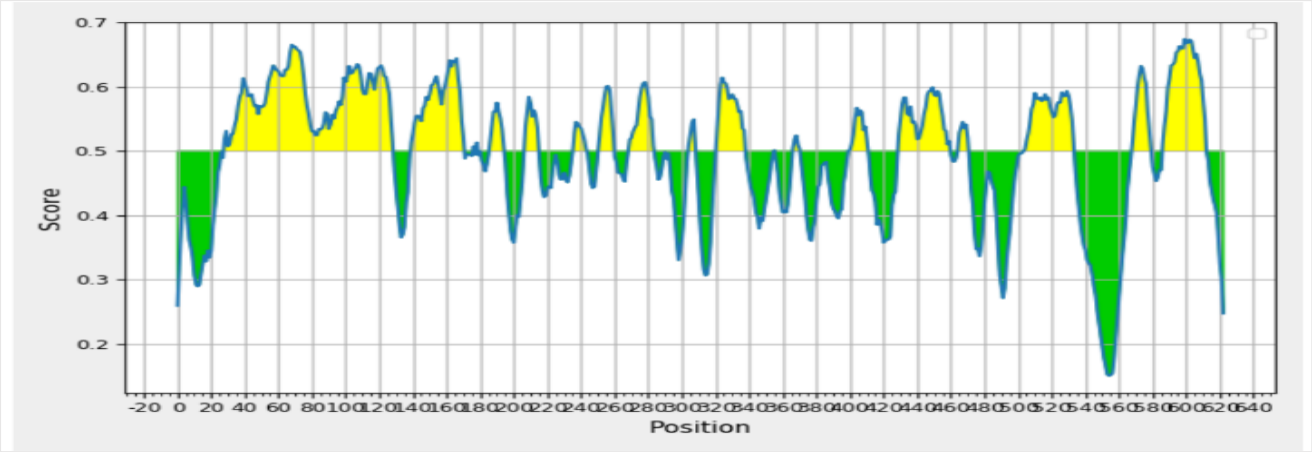

Figure 9: B -cell result graph of protein from IEDB server (Vita et al., 2015).

Figure 9 illustrates, the yellow colored region shows the peptide sequences score above 0.5 and green colored region shows peptide sequence score below 0.5.

3.9 Assembly of peptides and vaccine construction:

Combination of plain protein sequence with all selected CTL, HTL and B cell epitopes using linkers.

- Protein sequence – EAAAK – CTL -GPGPG- HTL - KK -B cell

FINAL VACCINE:

MGTVNKPVVGVLMGFGIITGTLRITNPVRASVLRYDDFHIDEDKLDTNSVYEPYYHS
DHAESSWVNRGESSRKAYDHNSPYIWPRNDYDGFLENAHEHHGVYNQGRGIDSGE
RLMQPTQMSAQEDLGDDTGIHVIPTLNGDDRHKIVNVDQRQYGDVFKGDLNPKPQG
QRLIEVSVEENHPFTLRAPIQRIYGVRYTETWSFLPSLTCTGDAAPAIQHICLKHTTCF
QDVVVDVDCAENTKEDQLAEISYRFQGKKEADQPWIVVNTSTLFDELELDPPEIEPG
VLKVLRTEKQYLGVYIWNMRGSDGTSTYATFLVTWKGDEKTRNPTPAVTPQPRGA
EFHMWNYHSHVFSVGDTFSLAMHLQYKIHEAPFDLLLEWLYVPIDPTCQPMRLYST
CLYHPNAPQCLSHMNSGCTFTSPHLAQRVASTVYQNCEHADNYTAYCLGISHMEPS
FGLILHDGGTTLKFVDTPESLSGLYVFVVYFNGHVEAVAYTVVSTVDHFVNAIEERG

FPPTAGQPPATTKPKEITPVNPGTSPLLRYAAWTGGLAAVVLLCLVIFLICTAKRMRV

KAYRVDKSPYNQSMYYAGLPVDDFEDSESTDTEEEFGNAIGGSHGGSSYTVYIDKTR

EAAAKFSLAMHLQYGPGPGIEPGVLKVLRTEKQYKKEERGFPPTAGQPPATTKPKEI

TPVNPGTSP

3.10 Biochemical Analysis of Predicted Vaccine:

Number of amino acids: 689

Molecular weight: 76991.52

Theoretical pI: 5.46

Formula: $C_{3443}H_{5239}N_{921}O_{1041}S_{25}$
Total number of atoms: 10669

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 102400
Abs 0.1% (=1 g/l) 1.330, assuming all pairs of Cys residues form cystines

Ext. coefficient 101650
Abs 0.1% (=1 g/l) 1.320, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 38.34
This classifies the protein as stable.

Aliphatic index: 73.54

Figure 10: Biochemical Analysis through Expasy ProtParam tool of construed vaccine (Gasteiger et al., 2010).

Figure 10 shows biochemical analysis , Expasy Protparam provides molecular weight 76991.52 da or 76.991 kDa (<100kDa) and indicates the sequence as stable vaccine. Number of Amino acid 689 where only protein sequence was 623.

Theoritical pI is 5.46 which indicates vaccine is slightly acidic in nature and instability index is 38.34 where index more than 40 indicates unstable molecule.

Aliphapic index is 73.54 which means it is stable in different temparature.

Half ife is 30 hrs (mamalian reticulocutes, in vitro > 20 hrs (yeast in vivo > 10 hrs (E.coli, invivo)

3.10 Antigenicity, Allergenicity and Toxicity Prediction of Constructed Vaccine:

Antigenicity:

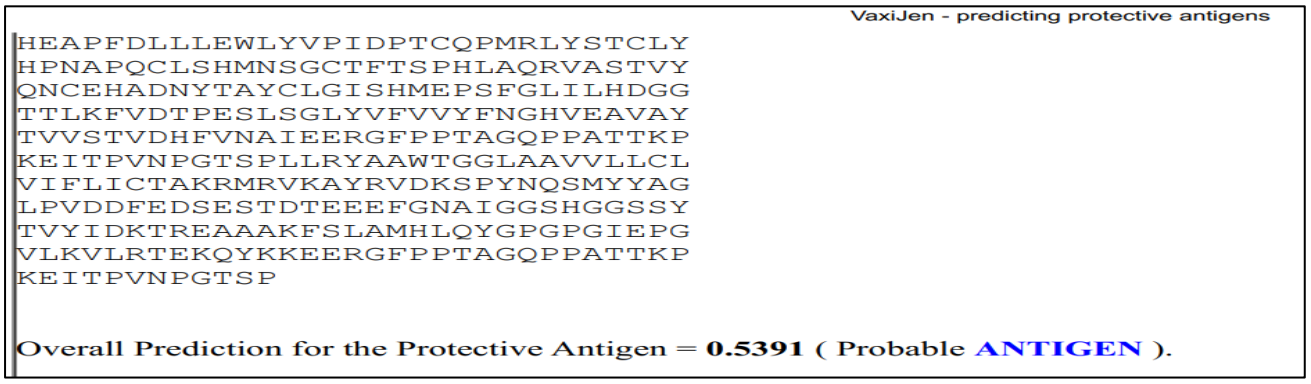

Figure 11: antigenicity of vaccine through Vaxijen 2.0 of contructed vaccine (Doytchinova & Flower, 2007b).

According to figure 11, the antigenicity of the designed vaccine is 0.5391 where the plain protein sequence antigenicity is 0.5165.

But when non-antigenic and antigenic CTL/HTL and B cell are added to the construction , the total antigenicity is decreased that's why it is decided to construct vaccine with only best score antigenic provider peptides so that total antigen increase.

Allergenicity:

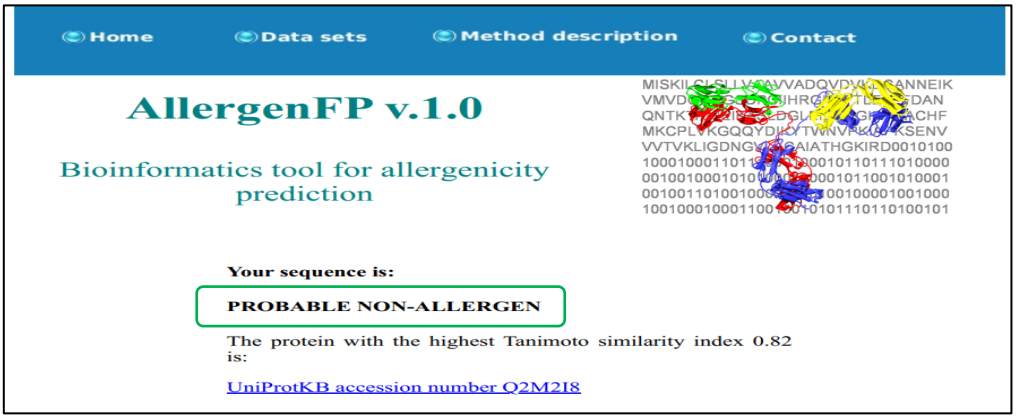

Figure 12: Allergenicity test of vaccine candidate (Dimitrov et al., 2014).

Figure 12 presents allergenicity test. AllergenFP v.1.0 shows that the designed vaccine is non-allergen. This server was used during construction of herpes vaccines.(Kumar et al., 2020) As Herpes Zoster virus is a sub class of Herpes family. It was followed to check allergenicity.

Toxicity:

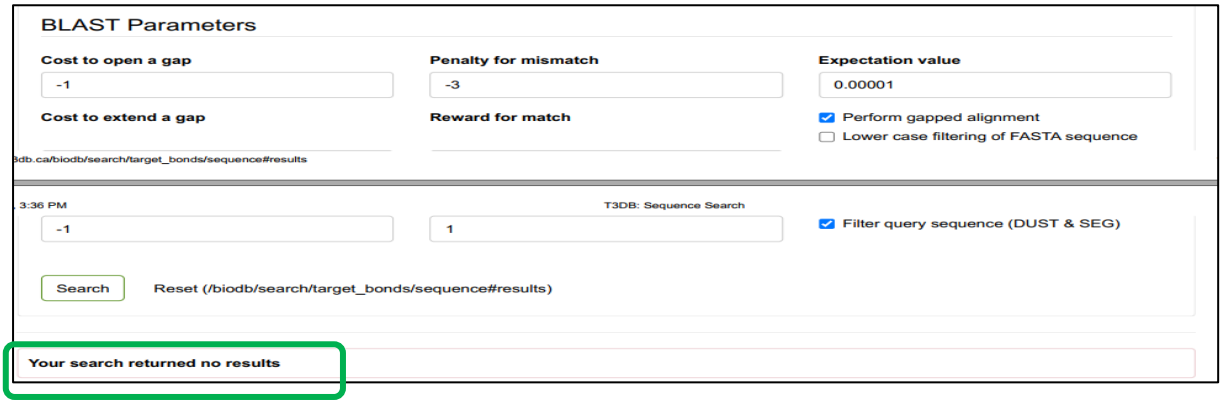

Figure 13 Toxicity result (Lim et al., 2010)

According to figure 13, the bar which shows search retuned no results which indicates there is no toxicity in the designed vaccine.

3.11 Homology modeling and C-score:

I-Tasser:

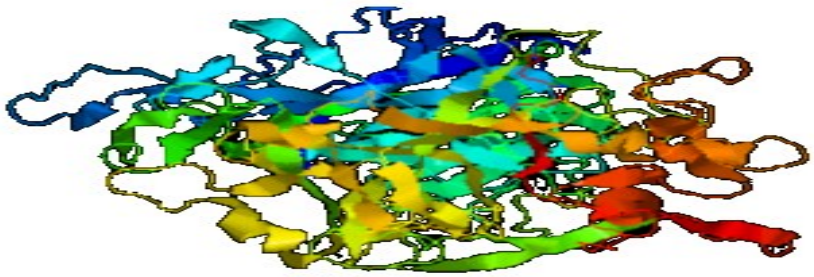

Figure 14 : I-tasser (model 1) (J. Yang et al., 2015).

Figure 14 is the result which is got from I-tasser mail. This is the model 1. This shows very prominent score so that would be chosen as vaccine model. C-score=-0.78. C-score basically calculated based on significant threading template and prove by coverage parameters of the structure. The acceptable parameters [-5, 2], the vaccine model come up with -0.78 which is acceptable according to the parameters.

3.12 Evaluation of Z-Score and Ramachandran Plots:

Z-score: -4.97

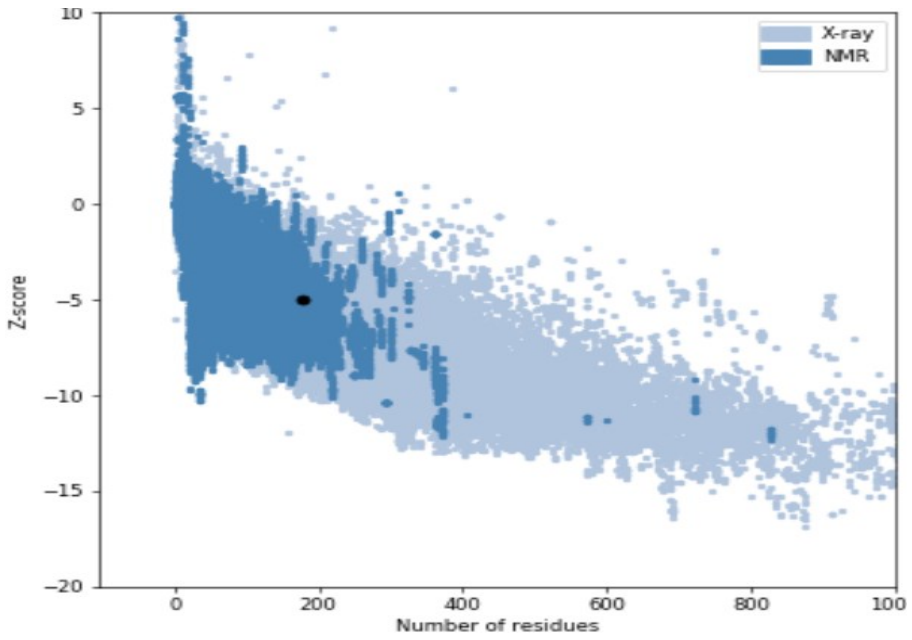

Figure 15 Z- score (Wiederstein & Sippl et al., 2007).

Figure 15 indicates Z-score. A positive outcome is the black spot is located on the blue gray area. And score -4.97 which indicates it is active. A z-score of -4.97 in ProSA-web indicates that the protein structure being analyzed has an energy or stability that is significantly lower than the average energy of the reference set. A negative z-score of this magnitude suggests that the protein structure is highly stable and energetically favorable.

Ramachandran Plot:

Figure 16

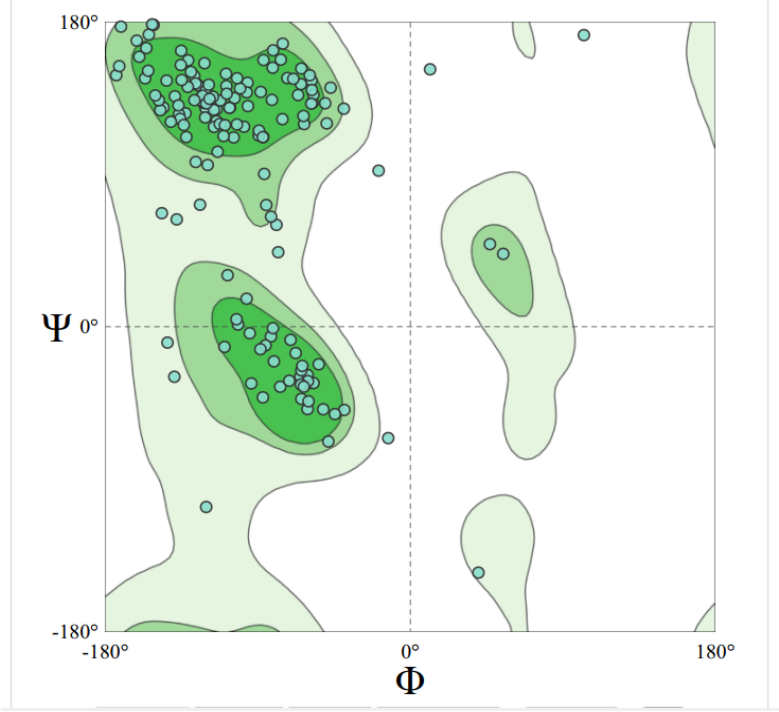

Figure 16 : Ramachandran plot (Gopalakrishnan et al., 2007)

In figure 16 , Ramachandran plot is identified where Ramachandran favored score is 87.57% which indicates good assessment and therefore, a Ramachandran favored score of 87.57% suggests that the protein structure is likely to be well-folded and stable, with the majority of its residues adopting conformations that are consistent with the known structures of well-folded proteins. Ramachandran outlier is 5.65% which is not in parameters favor which makes this as a limitation for the designed vaccine. For this reason, designed vaccine should have some modification.

3.14 Molecular Docking of Final Vaccine:

Receptor	Ligand	Complex Type	Clustering RMSD	User e-mail	Receptor Site	Ligand Site	Distance Constraints
3w3g	vaccine_final.pdb	Default	4.0	jtasnim50@gmail.com	-	-	-

Solution No	Score	Area	ACE	Transformation	PDB file of the complex
1	19702	2889.10	342.10	-2.41 -0.91 2.49 -27.43 13.06 39.32	result.1.pdb

Figure 17 Molecular docking (Schneidman-duhovny et al., 2005).

TLR8 receptor binder and the pdb format of protein are used to determine molecular docking.

Outcome shows highest score solution no 1 which has binding encompassed 2899.10 square meters and total 342.10 ACE and its transformation -2.41,-0.91,2.49,-27.43,13.06,39.32. The best combination between TLR8 and suggested vaccine got the maximum score of 19702 with a transformation mentioned above.

3.15 Immune response with corresponding Simulations:

Immunogenicity and immunological response to the vaccination were in silico modelled by the C-IMMSIM server. The effects of the suggested vaccine on the immune system and the immunological response can be anticipated using c-immsim. Monitoring antibody or immunoglobulin levels in response to vaccination timing and dosage is necessary to evaluate the body's immunological response. Monitoring the growth and decline of immune cells in response to vaccine doses is also crucial.

As the server has been used to check at 600 stimulations step following addition of 3 injections.

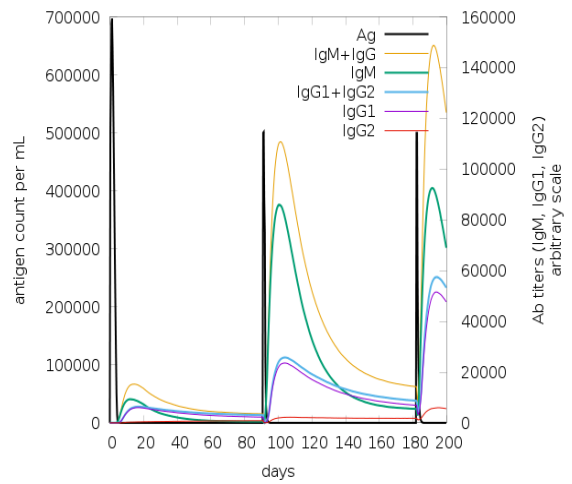

Figure 18(a): Antigen and immunoglobulins

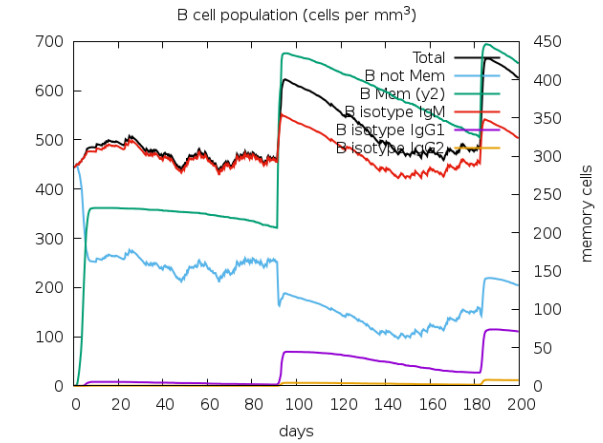

Figure 18(b): B cell population

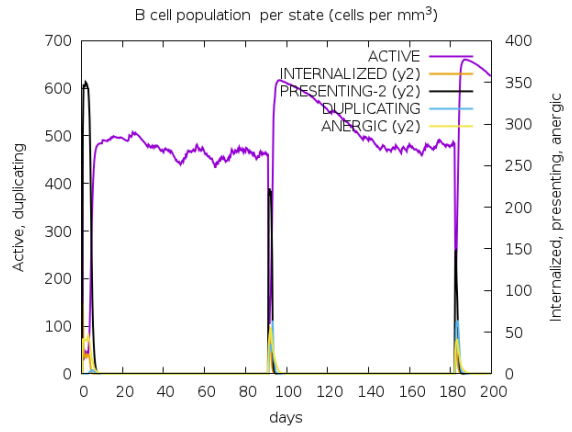

Figure 18 (C): B lymphocytes population per entity-state.

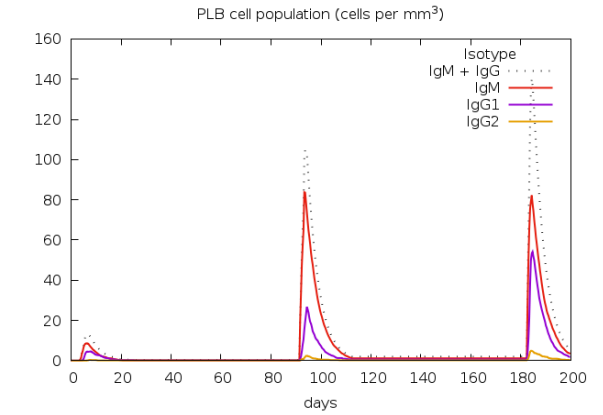

Figure 18 (d): PLB cell population

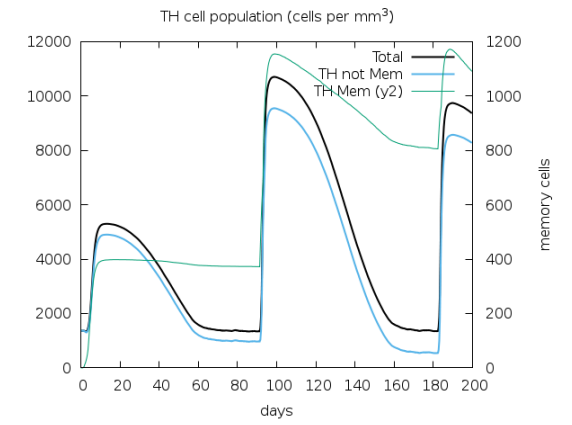

Figure 18 (e): T cell population

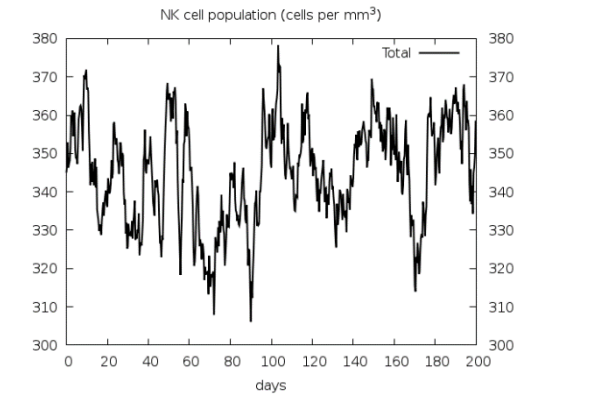

Figure 18 (f): Natural killer cell total

Figure 18 (a-f) C-immsim results of constructed vaccine (Castiglione & Bernaschi et al,. 2004).

From figure 18 (a), it appears that the graph represents the dynamics of antigen (Ag) and immunoglobulin (IgM and IgG) levels over a span of 180 days.

Here's a breakdown of the information

Antigen (Ag) levels:

Day 1: 700,000

Day 90: 500,000

Day 180: 500,000

Antigen (Ag) levels: The initial antigen level on day 1 is 700,000. Over the course of 90 days and 180 days, the antigen levels remain consistent at 500,000. This indicates a relatively stable or constant presence of the antigen throughout the observed time period.

Immunoglobulin (IgM+IgG) levels:

Day 1: Below 100,000

Day 90: 450,000

Day 180: 650,000

Immunoglobulin (IgM+IgG) levels: The initial level of IgM+IgG on day 1 is below 100,000. Over time, the levels increase to 450,000 on day 90 and further to 650,000 on day 180. This suggests an increase in the production or accumulation of IgM and IgG antibodies in response to the antigen presence. The trend observed in the immunoglobulin levels indicates that the immune response to the antigen is gradually developing and maturing over time. The increase in IgM+IgG levels suggests an ongoing immune reaction and antibody production, possibly indicating a mounting immune response against the antigen.

From figure 18(b) shows B cell population per mm^3. The graph depicting the B cell population per mm3 typically portrays the number of B cells on the vertical axis and time on the horizontal axis. It illustrates distinct categories of B cells, such as naïve B cells, memory B cells, or plasma cells, using separate lines or data points are activated and increased by time. The B cell population graph offers insights into various facets of the immune response, such as the scale and kinetics of B cell activation, the development of B cell memory, and the potential for sustained production of antibodies in the long term.

Figure 18(c) illustrates, B cell population per state. The graph clearly shows it's activeness through purple lines which indicates after each doses the activation of B cell eventually increase. Graph shows B cells concentrations are generated over time.

Figure 18 (d) explains number of plasma B cell according to their isotopes. Here, the production, activation and growth of various B cell isotopes are observed and also memory formation is witnessed.

Figure 18 (e) shows number of T cells. The graph indicates the activation of T cells. A great increment can be observed over time and doses. T memory cell increment is very high after all three doses.

Figure 18 (f) justifies a well and excellent level of natural killer cells during stimulation. The expansion of natural killer cell over time may suggest the activation and proliferation of NK cell indicating ongoing immune response. NK cells has great cytotoxic capabilities so activation of NK shows a strong immune repons during stimulation.

Chapter 4: Discussion

Recently, there are vaccines for herpes zoster virus but they are live attuned. In silico method based herpes zoster vaccine is now a running interest. Many paper focus on the Glycoprotein so did this study too. As glycoprotein's replication is the reason behind that simplex zoster virus reactive and turns into herpes zoster virus. By using computational technology, this study find antigenic vaccine using in silico method which has very low odd result. In this case, the chosen protein sequence has a strong binding affinity with one CTL, one HTL and one B cell based on their antigenicity. The antigenicity of the designed vaccine sequence enhanced as a result due to antigenic epitopes.

The vaccine has antigenicity score 0.5391 and it shows non allergen properties. The vaccine has not shown any toxicity as it has been constructed by by choosing best score of CTL, HTL epitopes and B cell epitopes.

The designed vaccine has no harmful or toxic consequences and also rise I the body's antibody levels. The model shows -0.7 in the fixed parameters of [-5, 2]. Which indicates it is acceptable designed.

The vaccine shows -4.97 as z-score. A z-score of -4.97 indicates that the energy of the protein structure is approximately 4.97 standard deviations below the mean energy of the reference set. Such a low z-score suggests that the structure is likely to be well-folded and have a favorable energy landscape, indicating a potentially reliable and stable protein conformation.

In the Ramachandran assessment, favored score is 87.57%. A Ramachandran favored score of 87.57% indicates the percentage of residues in a protein structure that fall within the favored regions of the Ramachandran plot. Furthermore, A Ramachandran outlier score of 5.65% suggests that a small portion of residues in the protein structure have torsional angles that fall

within the outlier regions of the Ramachandran plot. Outliers represent conformations that are rarely observed in well-folded and stable protein structures.

This vaccine has great immune response while checking for its stimulation for 546 time steps which indicates 6 months. Three doses has been utilized to boost the immunity and over time the great immune response has been observe. While checking graph activation, increment of number of B cell, T cell and NK cell has been observed. Over time and doses antibodies increases and number of antigen decreases.

This vaccine has better weight, Z score, c-score, stability which indicates it as a nearly good vaccine. So in the immune stimulation tests, this vaccine shows increasing antibody, antigenicity. But Ramachandran outlier are not in best score where a development can be done. The structure coverage can be increased in further development. By practicing that, development of new, better, complete, safer, vaccine should be prepared and it is in capable range. There is a need of more effective vaccine design and for herpes zoster, better and effective in silico vaccine is now must.

After all the discussion and interpretations of every results , it can be said that this vaccine has a good future and further research and development of this vaccine can clearly take place.

Chapter 5: Conclusion

Varicella-zoster virus (VZV), which often causes chickenpox in childhood, reactivates to create herpes zoster (HZ), also known as shingles. HZ manifests as a vesicular, painful rash that spreads in a dermatomal pattern along the ganglia of the dorsal root or cranial nerve. Physical, emotional, and social functioning are all hampered as a result of this excruciating rash, which has a detrimental impact on patients' quality of life. More than 90% of individuals have this virus, which puts them at risk of reactivation. Immuno-senescence causes reactivation to happen more frequently in persons over 50, but it can happen at any age, especially in people with impaired immune systems. Post herpetic neuralgia, which affects 10–20% of patients and gets worse with age, is one of the complications of HZ. (Patel et al., 2019)

Recently, Bangladeshi people are suffering from herpes zoster without knowing about the disease. It is increasing day by day due to environment, food habit, and lack of nutrients so on. Now this very concern to human health. Day by day necessities of multidisciplinary strategy for curing the disease is very important.

Corona like disease can weaken immune system where it is very easy to reactivation of HZV. To save from this kind of disease which can even cause come and death, better prevention curriculums should be taken. To better this situation, HZV vaccination is must and modeling or designing this kind of vaccine can be a good approach for future steps. The advantage is quick detection patterns without applying it in vivo.

Though the vaccine has higher range of Ramachandran outliers. While a lower Ramachandran outlier score is generally desirable, it is important to note that the presence of a few outliers does not necessarily imply that the entire protein structure is incorrect or biologically non-functional. It is common for protein structures to have a small percentage of outliers due to

specific local conformational requirements or experimental limitation. Without these it can be said that designed vaccine is almost good vaccine.

For the better understanding, the designed vaccine is in now at its primary stage. There are need of many improvement, development, modifications of the model to get advanced support. After advanced modelling, other steps will be required such as clinical trials in vitro and in vivo testing.

Reference:

Al-Anazi*, K. A., WK, A.-A., & AM, A.-J. (2019). The beneficial effects of varicella zoster virus. In *Journal of Hematology and Clinical Research* (Vol. 3, Issue 1). https://doi.org/10.29328/journal.jhcr.1001010

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., & Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering Design & Selection*,*14*(8),529–532. https://doi.org/10.1093/protein/14.8.529

Arvin, A. M. (1996). Varicella-zoster virus. *Clinical Microbiology Reviews*, *9*(3), 361–381. https://doi.org/10.1128/cmr.9.3.361

Berarducci, B., Rajamani, J., Zerboni, L., Che, X., Sommer, M., & Arvin, A. M. (2010). Functions of the unique N-terminal region of glycoprotein E in the pathogenesis of varicella-zoster virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 282–287. https://doi.org/10.1073/pnas.0912373107

Bernaschi, M., & Castiglione, F. (2001). Design and implementation of an immune system simulator. *Computers in Biology and Medicine*, *31*(5), 303–331. https://doi.org/10.1016/s0010-4825(01)00011-7

Bateman, A., Martin, M., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L. G., . . . Masson, P. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

Coudert, E., Gehant, S., De Castro, E., Pozzato, M., Baratin, D., Neto, T., Sigrist, C. J. A., Redaschi, N., & Bridge, A. (2022). Annotation of biologically relevant ligands in UniProtKB using ChEBI. *Bioinformatics*, *39*(1). https://doi.org/10.1093/bioinformatics/btac793

Doytchinova, I., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1). https://doi.org/10.1186/1471-2105-8-4

Ferreira, L. F., Santos, R. R. D., Oliva, G., & Andricopulo, A. D. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, *20*(7), 13384–13421. https://doi.org/10.3390/molecules200713384

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. M. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *Humana Press eBooks* (pp. 571–607). https://doi.org/10.1385/1-59259-890-0:571

Gopalakrishnan, K., Sowmiya, G., Sheik, S. S., & Sekar, K. (2007). Ramachandran Plot on The Web (2.0). *Protein and Peptide Letters*, *14*(7), 669–671. https://doi.org/10.2174/092986607781483912

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. S. (2013). In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLOS ONE*, *8*(9), e73957. https://doi.org/10.1371/journal.pone.0073957

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology*, *199*(9), 3360–3368. https://doi.org/10.4049/jimmunol.1700893

Kardani, K., Bolhassani, A., & Namvar, A. (2020). An overview of *in silico* vaccine design against different pathogens and cancer. *Expert Review of Vaccines*, *19*(8), 699–726. https://doi.org/10.1080/14760584.2020.1794832

Kennedy, P. G. E., & Gershon, A. A. (2018). Clinical Features of Varicella-Zoster Virus Infection. *Viruses*, *10*(11), 609. https://doi.org/10.3390/v10110609

Kumar, N., Sood, D., & Chandra, R. (2020). Vaccine Formulation and Optimization for Human Herpes Virus-5 through an Immunoinformatics Framework. *ACS Pharmacology and Translational Science*, *3*(6), 1318–1329. https://doi.org/10.1021/acsptsci.0c00139

Larsen, M. V, Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). *Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction*. *12*, 1–12. https://doi.org/10.1186/1471-2105-8-424

Livingston, B. D., Crimi, C., Newman, M. F., Higashimoto, Y., Appella, E., Sidney, J., & Sette, A. (2002). A Rational Strategy to Design Multiepitope Immunogens Based on Multiple Th Lymphocyte Epitopes. *Journal of Immunology*, *168*(11), 5499–5506. https://doi.org/10.4049/jimmunol.168.11.5499

Chen, X., Zaro, J. L., & Shen, W. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, *65*(10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039

Li, X., Guo, L., Kong, M., Su, X., Yang, D., Zou, M., Liu, Y., & Lu, L. (2015). Design and Evaluation of a Multi-Epitope Peptide of Human Metapneumovirus. *Intervirology*, *58*(6), 403–412. https://doi.org/10.1159/000445059

Shah, R. A., Limmer, A. L., Nwannunu, C. E., Patel, R. M., Mui, U. N., & Tyring, S. K. (2019). Shingrix for Herpes Zoster: A Review. *PubMed*, *24*(4), 5–7. https://pubmed.ncbi.nlm.nih.gov/31339679

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., & Nielsen, M. (2020). *Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry* . https://doi.org/10.1021/acs.jproteome.9b00874

Ullah, A., Johora, F., 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. https://doi.org/10.1016/j.imbio.2020.151955

Schneidman-duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). *PatchDock and SymmDock : servers for rigid and symmetric docking*. *33*, 363–367. https://doi.org/10.1093/nar/gki481

Shepherd, A. J., & Lee, G. (2018). *iBCe-eL : A New ensemble Learning Framework for Improved Linear B-Cell epitope Prediction*. *9*(July). https://doi.org/10.3389/fimmu.2018.01695

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, D., Cantrell, J. R., Wheeler, D. K., Gabbard, J. L., Hix, D., Sette, A., & Peters, B. (2015). *The immune epitope database (IEDB) 3 . 0. 43*(October 2014), 405–412. https://doi.org/10.1093/nar/gku938

Wiederstein, M., & Sippl, M. J. (2007). *ProSA-web : interactive web service for the recognition of errors in three-dimensional structures of proteins*. *35*, 407–410. https://doi.org/10.1093/nar/gkm290

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., Liu, Y., Goldansaz, S. A., & Rappaport, S. M. (2014). *T3DB : The toxic exposome database T3DB : the toxic exposome database*. *November*. https://doi.org/10.1093/nar/gku1004

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213

Yang, Z., Bogdan, P., & Nazarian, S. (2021). An in silico deep learning approach to multi - epitope vaccine design : a SARS - CoV - 2 case study. *Scientific Reports*, 1–21. https://doi.org/10.1038/s41598-021-81749-9

Yano, A., Onozuka, A., Asahi-ozaki, Y., Imai, S., Hanada, N., Miwa, Y., & Nisizawa, T. (2005). *An ingenious design for peptide vaccines*. *23*, 2322–2326. https://doi.org/10.1016/j.vaccine.2005.01.031

Zerboni, L., Sen, N., Oliver, S. L., & Arvin, A. M. (2014). Molecular mechanisms of varicella zoster virus pathogenesis. *Nature Reviews Microbiology*, *12*(3), 197–210. https://doi.org/10.1038/nrmicro3215

Zhang, L., & Skolnick, J. (1998). What should the Z-score of native protein structures be? *Protein Science*, *7*(5), 1201–1207. https://doi.org/10.1002/pro.5560070515