How can immunoinformatics be utilized to design a multi-epitope vaccine targeting the VP1 protein of Norovirus to enhance protection against gastroenteritis?

Summery

Norovirus, a leading cause of gastroenteritis, poses a significant public health challenge due to its high infectivity and genetic diversity. This study aimed to develop a multi-epitope vaccine targeting the VP1 protein of Norovirus using immunoinformatics methods. By analyzing the VP1 protein sequence, potential T-cell and B-cell epitopes were identified and evaluated for immunogenicity and conservation. A multi-epitope vaccine construct was designed, incorporating selected epitopes with appropriate linkers and an adjuvant. 1208 amino-acid vaccine sequence was obtained by manually joining all the epitopes in sequence-wise manner with the appropriate linkers, namely GPGPG Linker, EAAAK Linker, AAY Linker Additionally, Beta-defensin 3 Adjuvant was attached to increase the overall immune response towards the vaccine. As a result, enhanced overall protein stability, expression, immunostimulatory capabilities, and solubility of the designed construct were observed. Molecular dynamic simulations revealed the compactness and stability of the polypeptide construct. In silico analysis revealed the potential of the vaccine to induce a robust immune response against Norovirus. However, experimental validation is necessary to confirm the efficacy and safety of the vaccine. This research provides a foundation for further development of a Norovirus vaccine, offering potential solutions to the challenges associated with preventing this widespread disease.

1. Introduction

Gastroenteritis is an inflammation of the stomach and intestines that leads to symptoms like diarrhoea, vomiting, abdominal pain, and fever. It is often caused by infections with viruses, bacteria, or parasites. Among these, **Norovirus** is a leading cause, especially in cases of acute gastroenteritis worldwide. Norovirus is highly contagious and can spread through contaminated food, water, surfaces, and close contact with infected individuals.

Norovirus primarily infects humans by binding to specific carbohydrates on the surface of gut epithelial cells. Once inside, the virus disrupts the normal function of the digestive system, leading to rapid onset of symptoms such as vomiting and diarrhoea. The virus's high mutation rate also contributes to its ability to evade immune defences and its persistence in causing repeated outbreaks. The development of effective vaccines against Norovirus has been hindered by the virus's genetic diversity and high mutation rate, which contribute to the emergence of new strains that can evade the immune system. These challenges necessitate innovative approaches in vaccine design that can provide broad and durable protection against multiple strains of the virus. (Sell J, et al, 2018)

There are currently no specific antiviral treatments for norovirus-induced gastroenteritis. Management typically focuses on supportive care, including fluid replacement to prevent dehydration and symptomatic relief. However, because norovirus can cause severe dehydration and complications in vulnerable populations like young children, the elderly, and immunocompromised individuals, there is significant interest in developing targeted treatments and vaccines. (Zhang Y. J, et al, 2015)

Vaccine development is particularly important because of norovirus's global impact and its ability to cause recurrent outbreaks. Vaccines could help in preventing these infections, especially in high-risk populations, and reduce the burden of hospitalizations and deaths associated with severe gastroenteritis. Current research is focused on identifying the viral proteins (such as VP1) that are critical for eliciting an immune response, with the aim of designing multiepitope vaccines that can offer broad protection despite the virus's genetic diversity. (Glass, R. I et al, 2009)

This project focuses on the design of a multiepitope vaccine targeting Norovirus, specifically utilizing the VP1 protein**,** which is a major structural protein of the virus and plays a crucial role in its infectivity. Immunoinformatics tools are employed to identify and select conserved epitopes from the VP1 protein that can elicit a robust immune response, offering protection across different Norovirus strains. By integrating computational methods with immunological principles, this project aims to create a vaccine candidate that addresses the limitations of current vaccine strategies and contributes to the global effort in controlling gastroenteritis caused by Norovirus

1.1 Background

Immunoinformatics is a subfield of bioinformatics that applies computational tools and methods to immunology. It plays a pivotal role in understanding the immune system's response to pathogens and designing vaccines and therapies. The advantages of immunoinformatics include the ability to analyze vast amounts of immunological data quickly, predict epitopes that can elicit strong immune responses, and streamline the vaccine development process by reducing the need for extensive experimental trials. This field has become increasingly valuable in tackling complex infectious diseases like those caused by Norovirus, where traditional vaccine development faces numerous challenges.

Challenges in Vaccine Development for Gastroenteritis and the Role of Immunoinformatics

Developing vaccines for gastroenteritis, particularly against Norovirus, poses several challenges. Norovirus is characterized by its high genetic diversity, with multiple genotypes and strains co-circulating. This variability makes it difficult to create a single vaccine that provides broad protection. Additionally, Norovirus has a high mutation rate, leading to the emergence of new strains that can escape immune detection, further complicating vaccine design.

Another significant challenge is the lack of a robust, long-lasting immune response following natural infection with Norovirus. This means that even those who have recovered from the virus can be reinfected, underscoring the need for a vaccine that can induce more durable immunity.

Immunoinformatics offers potential solutions to these challenges by enabling the rational design of vaccines that target multiple epitopes from the virus's most conserved regions, such as the **VP1 protein**. By identifying epitopes that are less prone to mutation and can stimulate a strong and cross-reactive immune response, immunoinformatics can aid in the development of **multiepitope vaccines**. These vaccines can offer broader protection against diverse Norovirus strains and potentially reduce the impact of this pathogen on public health. Moreover, immunoinformatics tools can accelerate the vaccine development process, making it more cost-effective and efficient, which is crucial in responding to rapidly evolving pathogens like Norovirus. (Elsideeq E et al, 2020)

1.2 Objectives

- To identify and predict Immunogenic epitopes
- To design and construct a multi-epitope vaccine with Adjuvant and Linker molecules.
- To validate the predicted multi-epitope vaccine candidate by performing, structural analysis, molecular docking, and molecular dynamics simulations.
- To perform in silico immune simulations to produce effective immune response throughout life against Norovirus.

2. Methodology

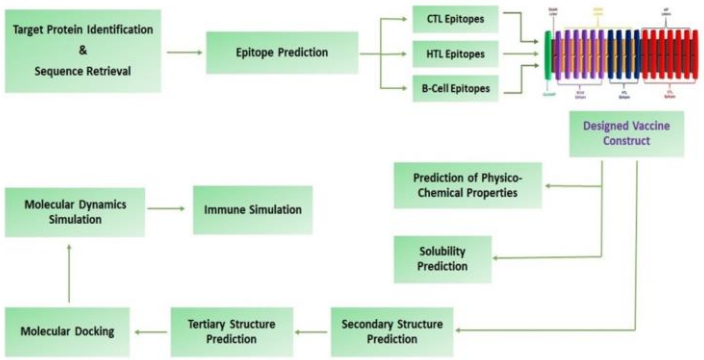


Figure 1 Workflow of Insilco Peptide based Vaccine Design (Martinelli, D. D. et al, (2022)

Step 1: Target Selection: VP1 protein FASTA sequence is retrieved from NCBI.

Disease: Gastroenteritis

Protein Selection: VP1 protein of Norovirus (GenBank: BAK68870.1)

Sequence: >BAK68870.1 VP1 protein [Norovirus sewage/GII/Influent/Nov2006/JPN]
MKMASNDASPSDGSAANLVPEVNNEVMALEPVVGAAIAAPVAGQQNVIDPWIRNNFVQAPGGEFTVSPRNAPGEILWS
APLGPDLNPYLSHLARMYNGYAGGFEVQVILAGNAFTAGKIIFAAVPPNFPTEGLSPSQVTMFPHIIVDVRQLEPVLIPLPDV
RNNFYHYNQSNDSTIKLIAMLYTPLRANNAGEDVFTVSCRVLTRPSPDFDFIFLVPPTVESRTKPFTVPILTVEEMTNSRFPIPL
EKLFTGPSSAFVVQPQNGRCTTDGVLLGTTQLSPVNICTFRGDVTHIAGSRNYTMNLASLNWNNYDPTEEIPAPLGTPDFV
GKIQGVLTQTTKGDGSTRGHKATVYTGSAPFTPKLGSVQFSTDTENDFETHQNTKFTPVGVIQDGSTTHRNEPQQWVLPS
YSGRNVHNVHLAPAVAPTFPGEQLLFFRSTMPGCSGYPNMDLDCLLPQEWVQHFYQEAAPAQSDVALLRFVNPDTGRVL
FECKLHKSGYVTVAHTGQHDLVIPPNGYFRFDSWVNQFYTLAPMGNGTGRRRAL

This amino acid sequence of the VP1 protein is used for further analysis.

Why VP1 Protein?

Structural Importance: The VP1 protein is the major capsid protein of Norovirus. It forms the outer surface of the virus, making it a prime candidate for the immune system to recognize.
Immunogenicity: As a surface-exposed protein, VP1 is highly immunogenic. This means it can effectively trigger an immune response, making it a good target for vaccine development.
Conservation: VP1 is relatively conserved across different strains of Norovirus, which helps in designing a vaccine that could be effective against multiple strains. (Dar, M. A., et al, 2022)

Step 2: Epitope Selection:

The specific part of an antigen that is recognized by immune cells is called an epitope or antigenic determinant.

MHC-I Prediction:

- These epitopes are typically 8-11 amino acids long and are recognized by CD8+ cytotoxic T cells.
- Tools like IEDB's MHC-I prediction algorithms analyze the protein sequence to identify potential binding motifs that are likely to be presented by MHC-I molecules.
- The algorithm scores and ranks the peptides based on their binding affinity to various MHC-I alleles, predicting which ones are most likely to be recognized by cytotoxic T cells.
- The **ANN 4.0** (Artificial Neural Network) method is commonly used for MHC-I epitope prediction due to High Accuracy, Comprehensive Training Data, Flexibility and Adaptability, Integration with Other Tools, Proven Track Record.

MHC-II Prediction:

- MHC-II epitopes are generally longer, around 12-25 amino acids, and are presented to CD4+ helper T cells.
- IEDB's MHC-II prediction tools assess the protein sequence for potential binding to MHC-II molecules.
- These epitopes are critical for activating helper T cells, which are essential for the activation of other immune cells, including B cells.
- The **NN align** (NetMHCII 2.3) method is chosen for MHC-II epitope prediction because it provides a robust and accurate approach to predicting peptide binding, particularly for the variable-length peptides associated with MHC-II molecules. Its advanced neural network model, ability to handle flanking regions, high predictive accuracy, and broad allele coverage make it an ideal tool for this purpose.

B-Cell Epitope Prediction:

- B-cell epitopes are the regions of the protein recognized by antibodies.
- These epitopes can be linear (continuous sequence of amino acids) or conformational (spatial arrangement of amino acids in the 3D structure).
- Tools like IEDB predict B-cell epitopes by analyzing the protein sequence and structure for regions likely to be exposed on the surface and accessible to antibodies.

Step 3: Data Pre-processing:

Filtering Based on IC50 Value:

- Only include epitope candidates with a high binding affinity (IC50 < 100 nM) were considered for further analysis.
- The datasets are filtered in Excel, Using the filter function to select rows where the IC50 value meets the criterion (≤100 nM), the rows that do not meet this criterion are removed to focus on high-affinity epitopes.

Peptide Filtration According to Alphabetic Order:

- The peptide sequences are listed in alphabetical order for easier navigation and comparison.
- In Excel, the column containing the peptide sequences are selected, then sorted using "Sort A-Z" function to organize the data.

Step 4: Epitope Evaluation

These evaluation steps are performed to ensure that the selected epitopes are:

Safe: Non-allergenic and non-toxic.

Effective: High antigenicity to elicit a strong immune response.

Broadly Applicable: Recognized by a diverse range of MHC alleles, ensuring wide population coverage.

- **Allergenicity:** Evaluated using AllerTOP. Epitopes identified as allergens are rejected.
- **Antigenicity:** Evaluated using VaxiJen. Non-antigenic peptides are excluded.
- **Toxicity:** Evaluated using ToxinPred. Toxic peptides are excluded.
- **Conservation Analysis & Population Coverage:** Further analysis using IEDB tools to ensure that the selected epitopes are conserved and provide broad population coverage.
- **Conservation analysis**: helps in selecting regions of proteins that are conserved across different strains or species of a pathogen, ensuring that the vaccine remains effective against a broad range of viral virulent

Step 5: Linker and Adapter Selection

Linkers: Linkers are short peptide sequences used to connect multiple epitopes in a multi-epitope vaccine. They ensure that each epitope functions effectively and maintains its structural integrity.

Adapters: like the **Beta-defensin 3 (Q5U7J2)** are included in multi-epitope vaccines to enhance the immune response.

- **Gly-Pro-Gly-Pro-Gly (GPGPG)** is a peptide linker sequence. It consists of alternating glycine (Gly) and proline (Pro) residues, which provide Flexible Structure, Reduced Steric Hindrance, and Enhances Immunogenicity.
- **EAAAK Linker:** Provides rigidity to ensure structural stability. **EAAK**
- **AAY Linker:** Facilitates effective T-cell epitope processing. **AAY**
- **Beta-defensin 3 (Q5U7J2)**: is responsible for the antimicrobial and immunomodulatory functions of Beta-defensin 3, making it an effective molecule in enhancing immune responses when used as an adjuvant in vaccines.

MRIFYYLHFLCYVTFILPATCTLVNADRCTKRYGRCKRDCLESEKQIDICSLPRKICCTEKLYEEDDMF

Step 6: Vaccine Construction

The creation of a vaccine construct involves linking selected epitopes together using specific linkers and an adaptor like Beta-defensin 3. This process is done to develop a single, coherent vaccine that can effectively induce an immune response against multiple epitopes of a pathogen. The purpose is to ensure that the

immune system can recognize and respond to different parts of the pathogen, leading to a stronger and more comprehensive immune response.

MRIFYYLHFLCYVTFILPATCTLVNADRCTKRYGRCKRDCLESEKQIDICSLPRKICCTEK
LYEEDDMFEAAKAAYQSDALLIRYYLAHLSAMYTSSGDFLKYMQESEFSFYAVDWSG
TRYAAYFSGGFTPSYGPGPGPHVMCDVRALEPIQLGEKYYRTVASRVSKEITSILQAA
GTAFSIYCRRIDFLVYAESPVVKVYASLAAAAPLDLVFTAGKVVVALVPPYFMRMATP
SSASSVRNTLALAVRMGSQAAIKIPLDPAELRKCVGMTVGPGGEKYYRTVASRVSKE
GPVMCDVRALEPIQLGPGPGHVGNSISTGPGPGGGGLMGIIGLEPIQLPLLDGPGPG
MRMATPGPGPGVMCDVRALEPIQLPGPGPGPGPGITSILQAAGTAFSIYG

Step 7: Evaluation of the Vaccine Construct

After constructing a multi-epitope vaccine, it is crucial to evaluate its safety, efficacy, and stability before moving on to experimental validation. These evaluations ensure that the vaccine is not only effective in triggering an immune response but also safe for administration. The evaluation includes assessing the vaccine for

1. **Allergenicity:** To ensure that the vaccine does not cause allergic reactions in individuals. **(AllerTop)**
2. **Toxicity:** To confirm that the vaccine does not have toxic effects. **(ToxinPred)**
3. **Antigenicity:** To verify that the vaccine can effectively stimulate an immune response. **(VaxiJen)**
4. **Solubility:** To check whether the vaccine can be dissolved and administered effectively. **(SolPro/Soluprot)**
5. **Physicochemical Properties:**

   To assess various properties such as molecular weight, stability, and isoelectric point (pI), which are important for the vaccine's formulation and storage. Models to assess whether the epitopes in the vaccine can effectively stimulate an immune response.**(ProtParam)**

Allergenicity	Antigenicity	Toxicity	Solubility
Non-Allergic	Antigenic	Non-Toxic	0.576

Step 8: Vaccine Structure Validation

Vaccine structure validation is performed to ensure the correct folding and stability of the designed vaccine construct. This step is crucial because the structure of a vaccine is directly related to its efficacy and safety. Validating the structure ensures that the epitopes (antigenic regions) are properly presented in a way that will stimulate an effective immune response without causing unintended adverse effects.

1. **Psipred (2D Structure Prediction)**:

   Psipred is used to predict the secondary structure of proteins, which refers to the local folded structures such as alpha-helices, beta-sheets, and coils within the protein sequence.

2. **PHYRE (3D Structure Prediction)**:

   PHYRE (Protein Homology/analogY Recognition Engine) is used to predict the tertiary structure of a protein, which refers to the three-dimensional arrangement of the entire protein.

   Vaccine structure refining:

   In the vaccine structure validation process, the refining step is performed to improve the accuracy of the predicted 3D model. After the initial structure prediction (e.g., using tools like PHYRE), the predicted structure may have certain inaccuracies in atomic positioning, side-chain conformations, or backbone geometry. Refinement corrects these imperfections, bringing the model closer to the actual native structure of the protein. This step ensures that the vaccine construct will have proper stability, fold correctly, and interact efficiently with immune system components.

Tools for Refining Structure:

1. **Galaxy Refine (Web-based Tool)**: Galaxy Refine is used for refining predicted protein structures by performing iterative structural relaxation. It improves both local and global structure accuracy.
2. **Ramachandran Plot (Web-based Tool)**: The Ramachandran Plot is used to evaluate the stereochemical quality of the protein's backbone dihedral angles (phi and psi angles). It helps in identifying the regions of the structure that may have unrealistic or energetically unfavorable conformations.

Step 9 Molecular Dynamics:

Molecular dynamics (MD) simulations are performed to ensure the stability and efficacy of the vaccine-receptor interactions. The vaccine construct needs to bind effectively to immune receptors such as TLR7 to elicit a strong immune response. Molecular dynamics provides insight into how the vaccine behaves when it interacts with the receptor over time, under physiological conditions. Molecular Dynamics Steps Are Performed to Ensuring Stability of Vaccine-Receptor Interaction**,** Analyzing Conformational Flexibility**,** Validating Docking Results**,** Predicting Real-World Behaviour**.**

Molecular dynamics steps, including docking analysis with ClusPro and simulation with iMod server, are crucial for validating the designed multiepitope vaccine against Norovirus. ClusPro predicts how well the vaccine binds to the immune receptor TLR7, and iMod server simulates the complex's behavior under physiological conditions. Together, these tools ensure that the vaccine construct is stable, maintains proper interaction with the receptor, and is likely to trigger a strong immune response, making the vaccine effective against gastroenteritis caused by Norovirus.

1. ClusPro (Docking Analysis)

ClusPro is used to predict how well the vaccine construct (designed from VP1 epitopes) docks to TLR7, the immune receptor. This docking process predicts the most favorable orientation and binding sites of the vaccine construct with TLR7.

2. TLR7 (Toll-Like Receptor 7)

TLR7 is an innate immune receptor that detects viral RNA and activates antiviral immune responses. By designing the vaccine to interact with TLR7, you can enhance the activation of the immune system against Norovirus.

The structure of TLR7 is retrieved from the PDB and used as the receptor in docking and molecular dynamics simulations. This receptor is chosen because activating TLR7 stimulates a strong antiviral immune response.

TLR7 is a key receptor involved in recognizing viral components. By designing a vaccine that targets TLR7, you ensure that the immune system is effectively alerted to the presence of Norovirus, increasing the likelihood of generating a protective immune response.

3. iMod Server (Molecular Dynamics Simulations)

iMod server is used for molecular dynamics simulations to analyze the dynamic behavior of the vaccine-receptor complex. It evaluates how stable the interaction between the vaccine construct and TLR7 is over time and assesses any conformational changes.

Step 10- Immune Stimulation:

The immune stimulation step using C-IMMSIM is performed to simulate how the multiepitope vaccine will interact with the immune system. It predicts the immune response, including activation of T-cells and B-cells, antibody production, and memory cell formation, providing insights into the vaccine's potential effectiveness and longevity. This step is essential for validating the immunogenicity and efficacy of the designed vaccine before proceeding to experimental validation

C-IMMSIM is a web-based in silico simulator that models the immune system's response to vaccine candidates. It simulates the interactions between antigens (vaccine epitopes) and immune cells, including T-cells, B-cells, macrophages, dendritic cells, and antibodies.

2.1 Data Sources

1. NCBI Protein (National Center for Biotechnology Information)

- NCBI Protein is a comprehensive database containing protein sequences derived from a variety of sources, including translations of coding sequences (CDS) from GenBank, RefSeq, and Swiss-Prot records.
- Usage in Vaccine Design**:** In the context of multi-epitope vaccine design, this resource is used for retrieving the protein sequences of viral proteins like the VP1 protein of Norovirus, which is critical for identifying potential epitopes (parts of antigens recognized by the immune system). These sequences are essential starting points for further epitope prediction and analysis.

2. IEDB (Immune Epitope Database)

- IEDB is a comprehensive repository of experimentally determined and predicted immune epitopes for a wide range of pathogens and diseases.

- Usage in Vaccine Design**:** IEDB is used for epitope prediction**,** specifically for B-cell, T-cell, and MHC binding predictions. It allows researchers to predict which segments of the viral proteins will likely trigger an immune response. This is key in multi-epitope vaccine design because only the most immunogenic epitopes are selected to create an effective vaccine.

3. VaxiJen

- VaxiJen is an online tool for the prediction of protective antigens based solely on their physicochemical properties without relying on sequence alignment.
- Usage in Vaccine Design**:** VaxiJen is used for antigenicity prediction. It helps determine whether a particular epitope or protein is likely to induce an immune response, which is crucial in the selection of candidates for vaccine development. For a multi-epitope vaccine, VaxiJen ensures that selected epitopes will be immunogenic.

4. AllerTop

- AllerTop is a machine-learning-based tool used to predict allergenicity of proteins.
- Usage in Vaccine Design**:** AllerTop is used to predict the allergenicity of selected epitopes. When designing a multi-epitope vaccine, it is important to ensure that the vaccine does not include epitopes that could trigger allergic reactions. AllerTop helps in screening epitopes to ensure safety.

5. ToxinPred

- ToxinPred is a web server designed for the prediction of toxicity of peptides and proteins.
- Usage in Vaccine Design**:** ToxinPred evaluates the potential toxicity of the chosen epitopes. This step is critical to ensure that the vaccine candidate does not contain toxic elements that could harm the patient. By excluding toxic epitopes, the safety of the vaccine is enhanced.

6. SolPro

- SolPro is a tool used to predict the solubility of a protein based on its amino acid sequence.
- Usage in Vaccine Design**:** SolPro predicts the solubility of the designed epitopes or proteins. In multi-epitope vaccine design, solubility is important for ensuring that the vaccine candidates are stable and can be expressed efficiently in host systems (e.g., during recombinant protein expression).

7. ProtParam

- ProtParam is a tool used to compute various physical and chemical parameters of a protein sequence, such as molecular weight, theoretical pI, amino acid composition, and more.
- Usage in Vaccine Design**:** ProtParam helps in analyzing the physicochemical properties of the selected epitopes or proteins. This analysis is important for understanding how the epitopes will behave in a biological system and aids in optimizing the vaccine design for stability and efficacy.

8. Psipred (2D Structure Prediction):

- Psipred uses neural networks trained on known protein structures to predict the likelihood of each amino acid in the sequence forming a specific type of secondary structure (helix, sheet, or coil).
- The input to Psipred is the amino acid sequence of the vaccine construct. It then analyzes the sequence and predicts the regions forming helices, sheets, and coils.
- This prediction helps you understand the overall folding pattern and how the epitopes are arranged spatially.

9. PHYRE (3D Structure Prediction):

- PHYRE works by comparing the input sequence to known protein structures in the Protein Data Bank (PDB) and predicts the 3D structure based on homology (sequence similarity to proteins with known structures).
- It uses threading and homology modeling techniques to generate a model that closely resembles the protein's true 3D shape.
- You provide the amino acid sequence of the vaccine construct to PHYRE, and it outputs a predicted 3D model. This model is essential for visualizing how the vaccine will interact with immune system components, such as antibodies.

10. Galaxy Refine (Web-based Tool):

- After receiving the initial 3D model (e.g., from PHYRE), the structure is uploaded to the Galaxy Refine web server.
- Galaxy Refine applies energy minimization techniques and molecular dynamics simulations to optimize the backbone and side-chain conformations.
- It performs repeated cycles of structure perturbation and relaxation, generating a set of refined models that are ranked based on energy scores (lower energy models are generally more stable).
- **Why Used**: Galaxy Refine improves the accuracy of the model by refining local geometric issues and improving hydrogen bonding and packing in the structure. It is commonly used because it effectively improves both the quality of the backbone and side-chain geometries.

11. Ramachandran Plot (Web-based Tool):

- After refining the model (e.g., with Galaxy Refine), the refined structure is analyzed using a Ramachandran Plot, which maps the dihedral angles of the protein's amino acids.
- The plot shows the phi (ϕ) and psi (ψ) angles of each amino acid and indicates whether these angles are in favorable, allowed, or disallowed regions of conformational space based on known protein structures.
- Ideally, the majority of the residues should fall within the favored and allowed regions of the plot. Residues in the disallowed regions indicate potential structural problems.
- **Why Used**: The Ramachandran Plot provides a clear visual representation of the quality of the protein's structure, specifically its backbone angles. It is a critical quality check for ensuring that the refined structure is stereochemically sound and that any conformational errors are corrected before proceeding further.

12. ClusPro (Docking Analysis)

- You input the structure of the vaccine construct and the receptor (TLR7) PDB file retrieved from the Protein Data Bank (PDB) into ClusPro.
- ClusPro generates several docking poses based on energy minimization and ranks them based on how well the vaccine construct fits with TLR7. This ensures that you identify the most favorable binding mode.
- **Why Used**: ClusPro helps identify the potential binding sites of the vaccine construct on TLR7. This is essential for ensuring that the vaccine construct can properly engage the receptor, which is a key step in activating the immune response.

13. **iMod Server (Molecular Dynamics Simulations)**

- After docking, the vaccine-receptor complex is input into the iMod server for simulation. iMod applies normal mode analysis (NMA) to evaluate the flexibility, stability, and motion of the complex.
- The server outputs key metrics, such as B-factors (indicating flexibility), deformability, and energy scores, providing insight into the dynamic properties of the complex.
- **Why Used**: iMod server ensures that the docked complex remains stable and retains its interaction throughout the simulation. This is important for confirming that the vaccine construct will maintain its effectiveness in real biological systems.

14. **C-IMMSIM tool:**

- **Input Data**: The sequence of the vaccine construct (containing multiple epitopes) is uploaded into the C-IMMSIM tool.
- **Simulation Setup**: The simulation is configured to mimic various immune system parameters, including antigen presentation, immune cell activation, cytokine production, and memory formation.
- **Simulation Run**: C-IMMSIM then runs a series of simulations over multiple time steps, mimicking vaccine administration and the subsequent immune response over time.
- **Output**: The tool provides output data on the number and types of immune cells activated, levels of antibodies produced, and memory cell formation. This allows you to visualize how well the vaccine triggers an immune response and how long-lasting that response might be.

3. Results and Discussion

3.1 Protein collection: The amino acid sequences of capsid protein (GenBank: BAK68870.1) of Norovirus as well as Beta-defensin 3 (Q5U7J2), an adjuvant was retrieved from UniProtKB in FASTA format. These proteins were antigenic, non-allergenic, nontoxic in nature and were selected for the designing of a multi-epitope vaccine by immunoinformatics approach.

3.2 CTL (cytotoxic T lymphocytes) epitopes prediction:

NetCTL1.2 server predicted total of 51 CTL epitopes. In these epitopes, six non-allergenic epitopes were selected for vaccine designing based on a high binding affinity (IC50 < 100 nM) score. Based on the predicted scores, two epitopes were selected from A7YK09, A7YK10, and A7YK11each.

3.3 HTL (helper T lymphocytes) prediction:

MHC-II prediction module of IEDB was used for Helper T Lymphocytes (HTL) epitopes prediction forHLA-DRB101:01, HLA-DRB101:02, HLA-DRB101:03, HLA-DRB101:04, and HLA-DRB101:05 Human alleles. 9 HTL epitopes with the highest binding affinity were selected. The selected epitopes are situated at position145–159, 373–387, 117–131 (capsid protein),1–15,1631–1645,816–830,617–631,1089–1103and 193–207 (polyproteins).

4.4 B-cell epitope prediction:

B-cell epitope prediction was performed on IEDB Server, and six epitopes were selected.

4.5 Final multi-epitope vaccine construct

The final multi-epitope vaccine construct was composed of **6 CTL** and **B-cell epitopes** while **9 HTL** epitopes were selected based on high binding affinity scores. The topological representation of the final vaccine candidate is given in Figure2. AAY linkers were used to combine CTL epitopes, and GPGPG linkers joined HTL epitopes. In contrast, EAAAK linker was used for attachment of adjuvant to the N-terminal of vaccine, which amplifies its function.

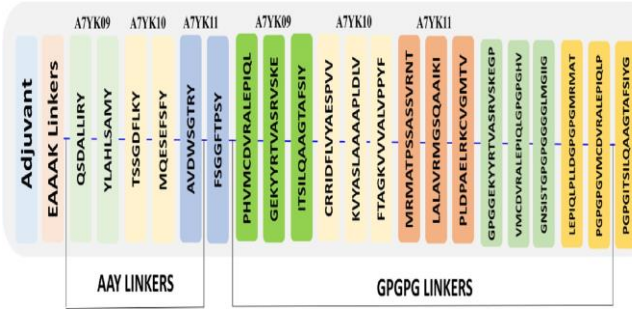


Figure 2: Vaccine Construct

4.6 Physiochemical properties:

The physiochemical propertis are predicted using ProtParam which provided the information about amino acid composition, molecular properties, and stability**.** The sequence has 1,239 amino acids with a molecular weight of 135,496.35 Daltons and a theoretical isoelectric point (pI) of 6.24, indicating the pH at which the molecule has no net charge. The extinction coefficient measures the protein's absorbance, useful for quantifying concentration. The instability index of 49.46 classifies the protein as stable, suggesting it may degrade quickly in cellular environments, while the aliphatic index (86.05) implies moderate thermostability. The GRAVY score of -0.056 indicates that the protein is slightly hydrophilic, potentially influencing its solubility in water. The

sequence's half-life estimates its stability across different organisms, suggesting moderate durability, especially in mammalian systems.

4.7 Prediction and validation of secondary structure:

Secondary structure of vaccine construct was predicted by PSIPRED server. Figure 3 shows the protein sequence alignment, a visual representation that compares the amino acid sequences of multiple proteins. Each row represents a different protein, and each column represents a specific amino acid position. The colours used in the image likely correspond to different features or properties of the proteins themselves.

Amino Acid Sequence: The horizontal axis represents the amino acid sequence of the vaccine construct.

Secondary Structure Predictions: The colored regions indicate the predicted secondary structure elements:

- **Yellow (Strand)**: Indicates beta-strand regions in the protein's secondary structure.
- **Pink (Helix)**: Indicates alpha-helices in the protein's structure.
- **Gray (Coil)**: Represents coil regions, which are neither helices nor strands.
- **Blue (Disordered)**: These regions lack a fixed structure, often highly flexible or unstructured.
- **Cyan (Transmembrane Helix)**: Denotes the presence of transmembrane helices, suggesting this protein might be membrane-associated.
- **Green (Re-entrant Helix)**: This typically refers to a helix that dips into the membrane but doesn't span it entirely.
- **Metal Binding**: Some residues may be involved in binding metal ions (though color and positioning are unspecified here).

Helices: 41 helices were predicted.

Strands: 44 strands were predicted.

Coils: 81 coils were predicted, making up approximately [Percentage of 10% of the total sequence.

Disordered Regions: 1 disordered regions were identified.

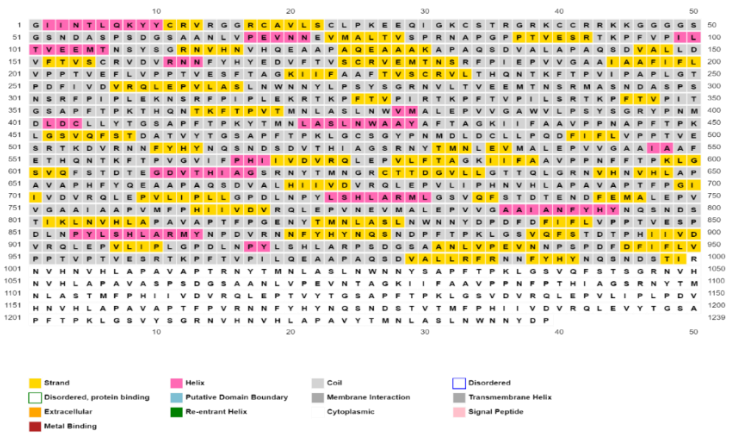


Figure 3 secondary structure and functional domain prediction

4.8 Prediction and validation of tertiary structure:

Tertiary structure of vaccine construct was obtained by Phyre2 Server.

64 residues (19% sequence) have been modelled with 96.9% confidence by the single highest scoring template. Out of these model with highest confidence of 96.9% and Coverage of 19% was chose as a best sequence identity suggests a reliable predicted model that likely reflects the actual structure of our vaccine.

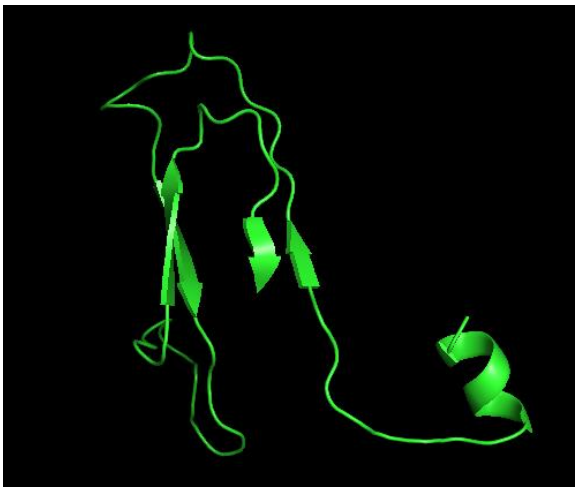


Figure 4 Tertiary Protein Structure.

4.9 Structure Refining: The Ramachandran plot is a graphical representation of the dihedral angles phi and psi in a protein structure. It helps visualize the allowed regions for amino acid conformations.

Key Observations:

- **High Percentage in Most Favoured Regions:** The plot indicates that a significant portion of the residues (81.5%) fall within the most favoured regions (A, B, L) of the Ramachandran plot. This suggests that the protein structure is generally well-structured and has a high degree of conformational stability.
- **Few Residues in Disallowed Regions:** No residues are found in the disallowed regions (XX), indicating that the protein structure is free from major steric clashes or unusual conformations.
- **Glycine and Proline Residues:** The presence of glycine and proline residues can sometimes lead to unusual phi and psi angles. The plot shows that these residues are distributed in a way that is consistent with their known conformational preferences.

Ramachandran Plot Statistics

- **High Percentage in Most Favored Regions:** The high percentage of residues in the most favored regions (81.5%) confirms the overall quality of the protein structure.
- **Low Percentage in Disallowed Regions:** The absence of residues in the disallowed regions reinforces the conclusion that the structure is free from major steric clashes.
- **G-Factors:** The average G-factors for various parameters are within acceptable ranges, suggesting that the structure is generally consistent with expected geometric properties.

Based on the Ramachandran plot and associated statistics, the protein structure appears to be of good quality. The high percentage of residues in the most favored regions and the absence of residues in the disallowed regions indicate a well-structured and stable protein. However, it's important to consider other factors like the resolution of the structure and any experimental validation to obtain a complete picture of the protein's quality.

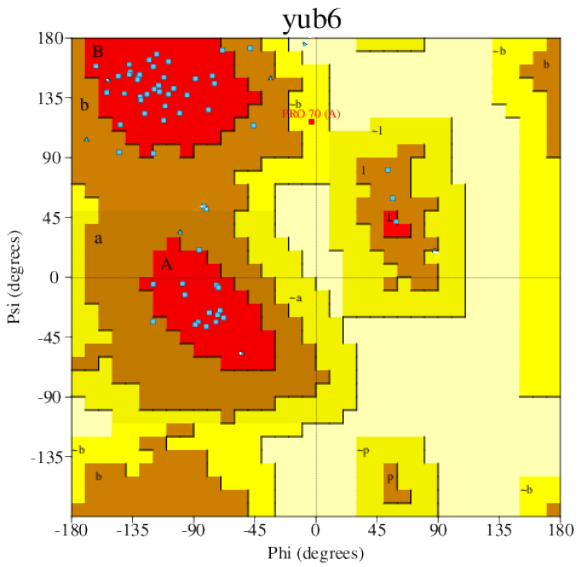


Figure 5 Ramachandran analysis residues

4.10 Docking Analysis: The docking score typically indicates how well the ligand (in this case, vaccine construct) binds to the target receptor (TLR7). The more negative the score, the stronger the predicted binding affinity. The unit is often in kcal/mol. The docking score was -1210.5 indicates a potential for the vaccine construct to effectively bind and elicit an immune response.

Conformational Stability: After docking, the conformation of the vaccine construct should remain stable when bound to the receptor, which indicates that the structure is likely to be immunogenic.

Binding Pocket and Surface Area: The vaccine construct ideally fit into the binding pocket of the receptor protein. The surface area involved in interaction is a good indicator of how much of the vaccine construct is exposed to immune recognition.

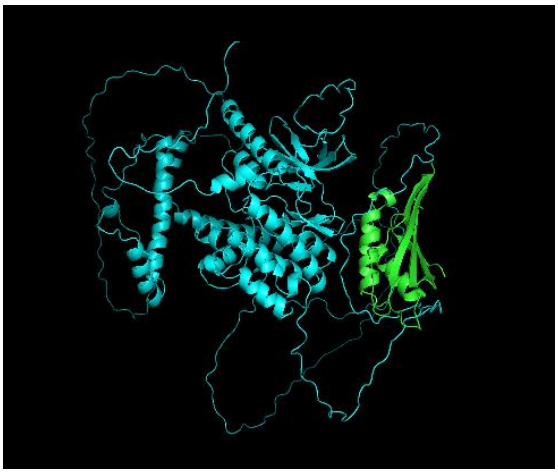


Figure 6 Docked Structure

The interaction pattern between TLR-7 (Chain A) and the vaccine construct (Chain B), based on data retrieved from the PDBsum server. It depicts the specific amino acids from each chain involved in the interaction.

Chain A likely represents the TLR-7 protein, while **Chain B** represents the vaccine construct. The amino acids in each chain are color-coded to visually represent their interactions. These lines indicate that the two amino acids are interacting with each other.

These amino acids represent the contact points or binding interactions between the TLR-7 receptor (Chain A) and the vaccine construct (Chain B). These interactions are critical in determining the binding affinity and the overall molecular dynamics between the two entities. Each color-coded oval represents a specific amino acid residue from either chain, with the lines between them indicating interactions such as hydrogen bonding, van der Waals forces, or electrostatic interactions.

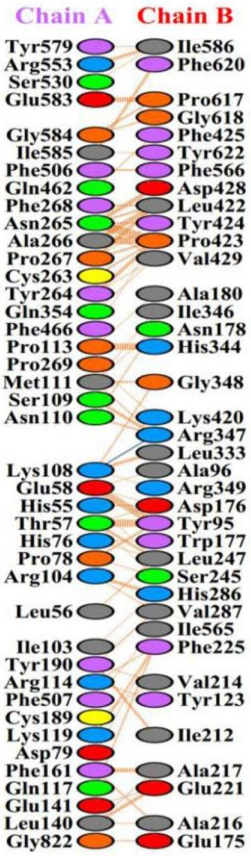


Figure 7 showing the interaction pattern of the TLR-7 and vaccine construct

4.11 Analysis of Protein Flexibility Using iMODS:

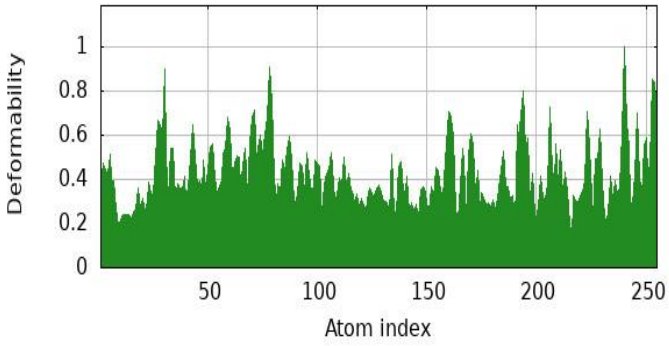


Fig. 7

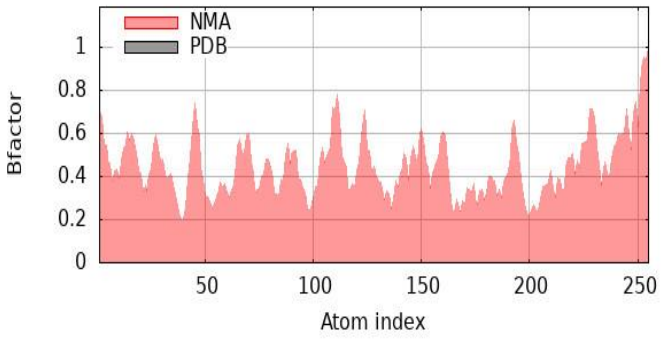


Fig. 8

Deformability: The deformability graph reveals regions within the protein structure that exhibit high flexibility, suggesting potential roles in protein-protein interactions, conformational changes, and ligand binding. By overlaying the deformability plot with the 3D protein structure, specific regions associated with these flexible properties can be identified, providing valuable insights into the protein's biological function. As shown in Fig. 7
B-Factor/Mobility Analysis: represents how much the main chain of the protein can deform at various residues. Regions with high deformability could indicate potential flexible or "hinge" regions that are crucial for protein function. High deformability might indicate important regions for protein dynamics, such as areas that allow the protein to undergo conformational changes, which can be important for binding or other functional activities. As shown in Fig.8

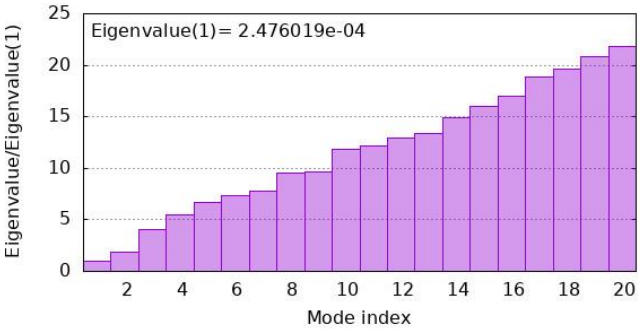


Fig. 9

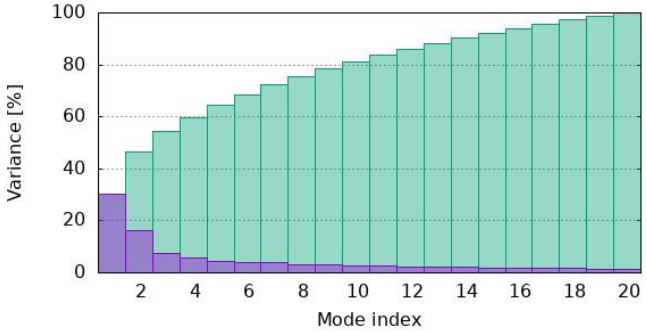


Fig.10

Eigenvalues: This indicates the stiffness of the motion related to the normal mode. Lower eigenvalues suggest that less energy is needed to deform the structure, which implies that the protein might undergo easier conformational changes. Low eigenvalues in specific modes could signal that the protein is flexible in those modes, which may be crucial for its function. As shown in Fig.9

Variance: Variance is inversely related to eigenvalues, with higher variance suggesting greater flexibility in that mode of motion. By analyzing the variance, one can determine which modes contribute significantly to the overall flexibility of the protein. High variance in a particular mode could suggest an important functional movement. As shown in Fig.10

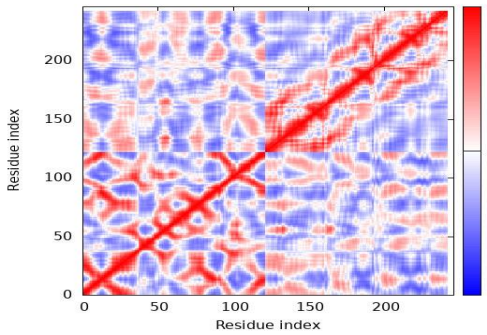


Fig. 11

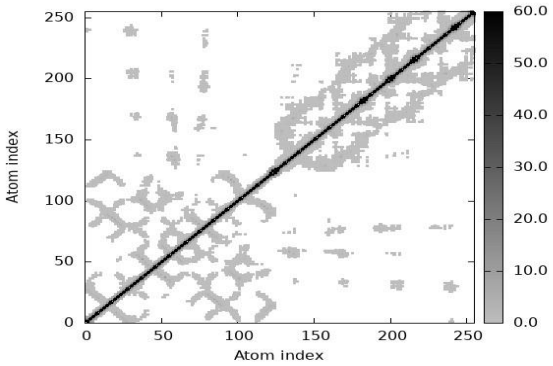


Fig.12

Residue Index: This map reveals how different residues in the protein structure are coupled, either positively (correlated motions) or negatively (anti-correlated motions). Red indicates correlated motions, blue indicates anti-correlated motions, and white indicates no correlation. Strong correlations between residues can indicate structural regions that move together during protein function, which may be important for maintaining structural integrity or facilitating conformational changes during interactions. As shown in Fig. 11

Elastic Network: Stiffness Representation**,** This describes the network of "springs" between atoms. Darker dots indicate stiffer springs, representing stronger connections between atoms. The elastic network helps identify the stability of the structure. Regions with stiffer springs are more resistant to deformation, suggesting structural stability, whereas regions with less stiffness might be more flexible and capable of conformational changes. As shown in Fig.12

4.12 **Immune simulation:**

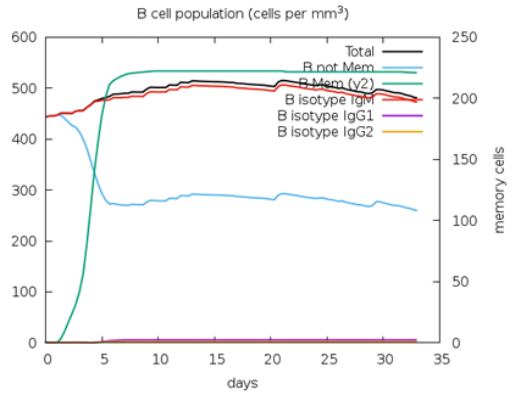


Fig. 13

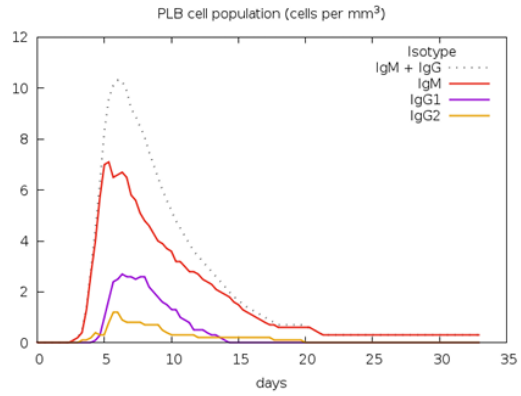


Fig.14

Fig.13 illustrates the dynamics of B cell populations over time, including total B cells, memory cells, and sub-divided isotypes (IgM, IgG1, IgG2). The total B cell population (black line) undergoes a rapid expansion following exposure, indicative of a robust immune response. A significant population of memory cells (green line) is generated, suggesting the development of long-lasting immunity. A shift from IgM production (red line) to IgG production (blue and purple lines) occurs, indicating class switching and the maturation of the immune response. The predominance of IgG1 (blue line) over IgG2 (purple line) suggests a typical pattern of antibody class switching in response to this antigen.

Fig. 14 depicts the dynamics of plasma B lymphocyte populations sub-divided by isotype: IgM, IgG1, and IgG2. These isotypes represent different classes of antibodies that play distinct roles in the immune response. A sharp peak in IgM levels is observed during the early stages of the immune response, indicating a rapid activation of B cells and production of IgM antibodies. Over time, a decline in IgM levels and a corresponding increase in IgG1 and IgG2 levels are seen, suggesting class switching from IgM to IgG subclasses. IgG1 becomes the predominant isotype, indicating a mature and effective immune response. IgG levels remain elevated, suggesting the generation of long-lived plasma cells and the development of a memory response.

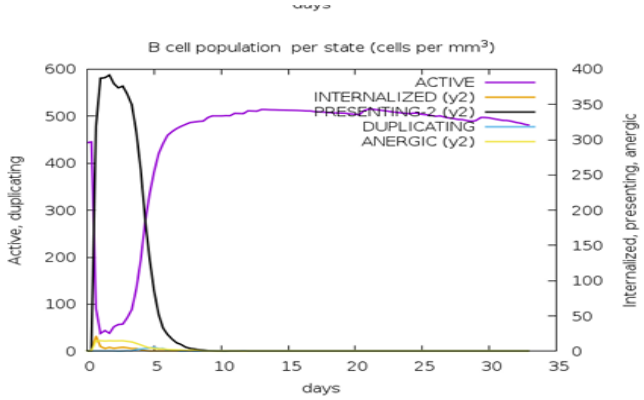


Fig. 15

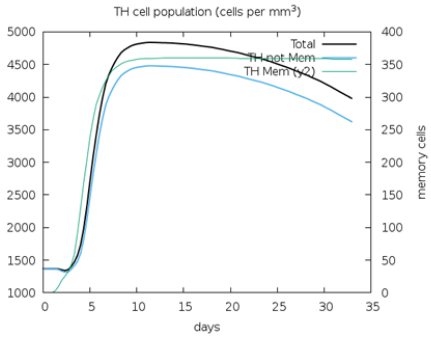
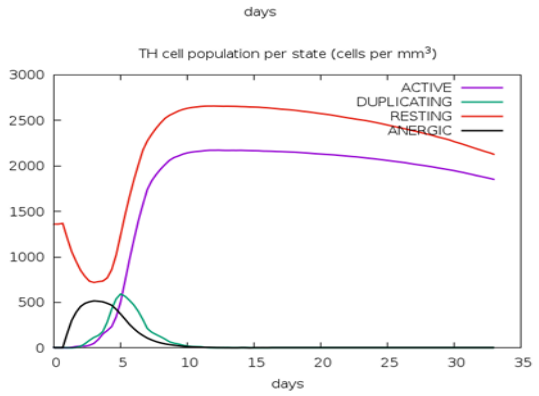


Fig. 16

Fig. 15 depicts the dynamics of B cell populations across different states: Active, Internalized, Presenting-2, Duplicating, Anergic. A rapid increase in the active B cell population is observed, indicating a strong initial response to the antigen. A subsequent increase in internalized and presenting B cells suggests efficient antigen processing and presentation to T cells. A significant expansion of the duplicating B cell population occurs, leading to the generation of a large number of effector and memory cells. A portion of the B cell population becomes anergic, potentially as a mechanism to regulate the immune response and prevent excessive activation.

Fig. 16 depicts the dynamics of CD4 T-helper lymphocyte populations over time, including total T cells and memory cells. The total T cell population (black line) undergoes a significant expansion following exposure,

indicative of a robust immune response. A substantial population of memory cells (green line) is generated, suggesting the development of long-lasting immunity. The total T cell population reaches a peak and then declines, reflecting the natural dynamics of the immune response.



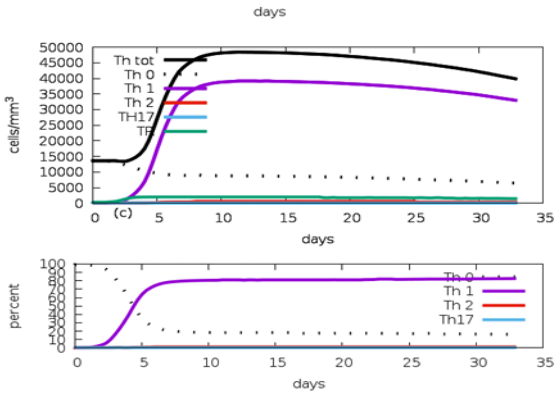
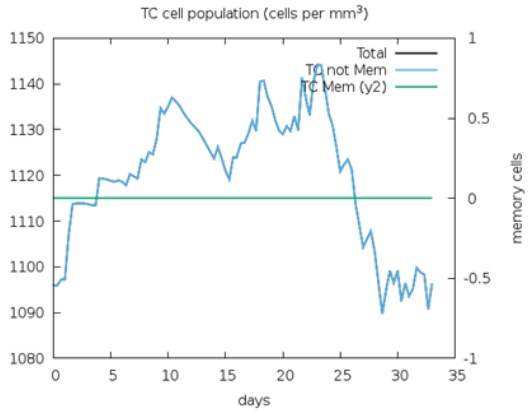


Fig. 17

Fig.18

Fig. 17 and Fig.18 depicts the dynamics of CD4 T-helper lymphocyte populations across different states: Active, Duplicating, Resting, Anergic. A rapid increase in the active T cell population is observed, indicating a strong initial response to the antigen. A significant expansion of the duplicating T cell population occurs, leading to the generation of a large number of effector and memory cells. A portion of the T cells transition to a resting state, suggesting the establishment of a memory population. A small population of T cells becomes anergic, potentially as a mechanism to regulate the immune response.



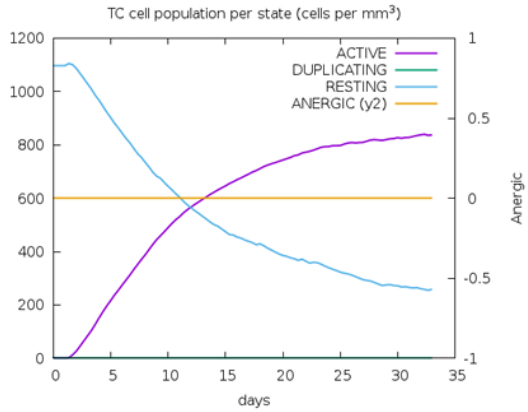


Fig. 19

Fig.20

Fig. 19 depicts the dynamics of CD8 T-cytotoxic lymphocyte populations over time, including total T cells and memory cells. The total T cell population (black line) undergoes a rapid expansion, followed by a decline and subsequent fluctuations. A smaller population of memory cells (green line) is generated compared to the CD4 T-helper lymphocytes. The total T cell population exhibits f luctuations throughout the simulation period, suggesting a dynamic response.

Fig. 20 depicts the dynamics of CD8 T-cytotoxic lymphocyte populations across different states: ACTIVE, DUPLICATING, RESTING, ANERGIC. A rapid increase in the active T cell population (purple line) is observed, followed by a decline. A transient increase in the duplicating T cell population (green line) suggests an initial burst of proliferation. A significant portion of T cells transition to a resting state (blue line), indicating the establishment of a memory population. A small population of T cells becomes anergic (yellow line), potentially as a mechanism to regulate the immune response.

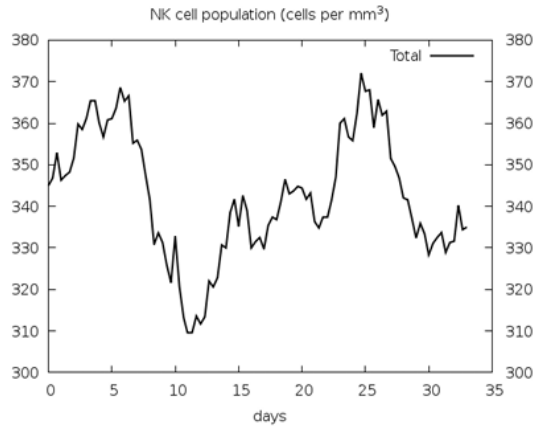


Fig.21

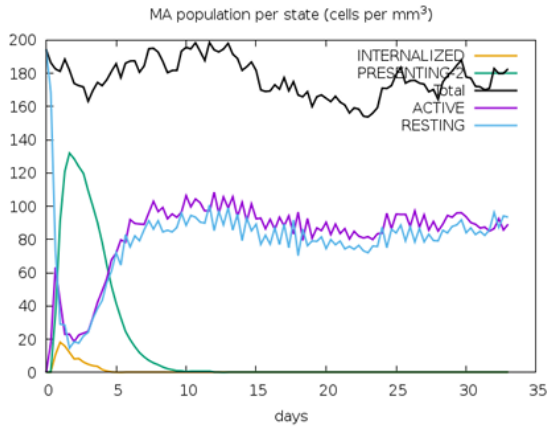


Fig. 22

Fig.21 depicts the dynamics of the natural killer (NK) cell population over time. The y-axis represents the cell count per cubic millimeter, and the x-axis represents the time in days. The NK cell population shows initial fluctuations during the early stages of the simulation. This might reflect the initial response to the antigen or other factors influencing NK cell activation. After the initial f fluctuations, the NK cell population stabilizes at a relatively constant level. This suggests that the immune system has reached a steady state in terms of NK cell numbers.

Fig. 22 depicts the dynamics of macrophage populations across different states: Total, INTERNALIZED, PRESENTING-2, ACTIVE, RESTING. A rapid increase in the total macrophage population (black line) and the active macrophage population (purple line) is observed, indicating an early response to the antigen. The internalized macrophage population (yellow line) rises, reflecting the uptake of antigens for processing. The presenting-2 macrophage population (green line) increases, indicating the presentation of antigens to CD4 T cells. The active macrophage population declines, while the resting macrophage population increases, suggesting a transition from an activated state to a more quiescent state.

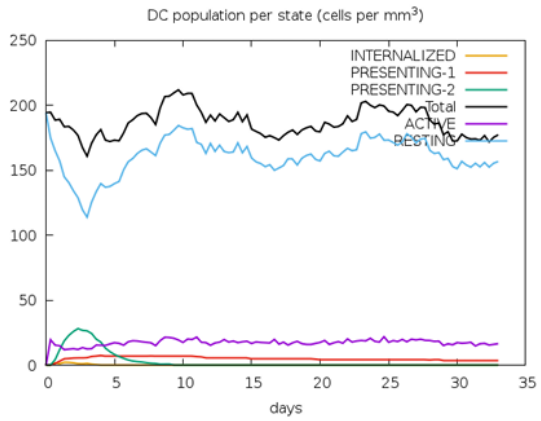


Fig. 23

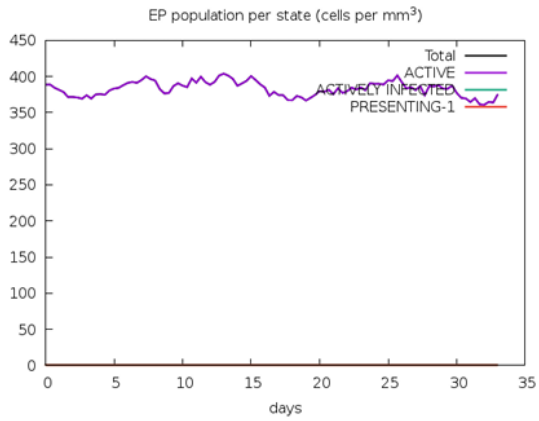


Fig. 24

Fig.23 depicts the dynamics of dendritic cell (DC) populations across different states: Total, INTERNALIZED, PRESENTING-1, PRESENTING-2, ACTIVE, RESTING. A rapid increase in the total DC population (black line) and the active DC population (purple line) is observed, indicating an early response to the antigen. The internalized DC population (yellow line) rises, reflecting the uptake of antigens for processing. The presenting-1 (red line) and presenting-2 (green line) DC population's increase, indicating the presentation of antigens to both CD8 and CD4 T cells. The active DC population declines, while the resting DC population increases, suggesting a transition from an activated state to a more quiescent state.

Fig. 24 depicts the dynamics of epithelial cell populations across different states: Total, ACTIVE, ACTIVELY INFECTED, PRESENTING-1. A rise in the actively infected epithelial cell population (green line) is observed, indicating the spread of the virus. An increase in the total epithelial cell population (black line) and the active epithelial cell population (purple line) suggests a response to the infection, potentially involving cell proliferation and activation. The presenting-1 epithelial cell population (red line) increases, indicating the presentation of viral antigens to CD8 T cells. The decline in the actively infected population and the

stabilization of the total epithelial cell population suggest a successful immune response and control of the infection.

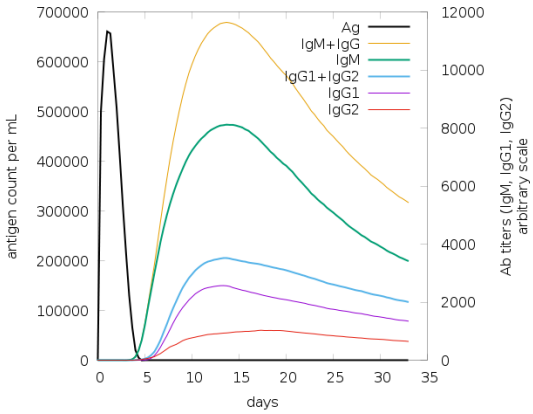


Fig. 24

Fig. 24 illustrates the dynamics of antigen levels and antibody responses over time. Here are the key observations: The antigen levels (black line) rapidly decrease following initial exposure, suggesting effective clearance by the immune system. The initial rise in IgM levels (green line) indicates a rapid primary immune response. The subsequent increase in IgG levels (blue, purple, and red lines) represents the development of a secondary immune response, characterized by higher antibody titers and a more sustained response. The presence of IgG1 and IgG2 subclasses suggests class switching, which is a hallmark of a mature and effective immune response. Concentration of cytokines and interleukins. Inset plot shows danger signal together with leukocyte growth factor IL-2

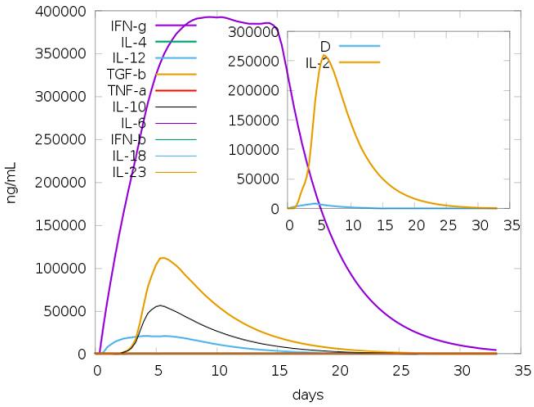


Fig. 25

Fig. 25 depicts the dynamics of various cytokines and interleukins, which are signaling molecules involved in immune responses. The inset plot shows the concentration of the danger signal (D) over time. A rapid increase in the levels of several cytokines, including IFN-gamma, IL-12, TNF alpha, and IL-6, is observed during the early stages of the immune response. The elevated levels of IFN-gamma and IL-12 suggest a predominant Th1-type immune response, while the increased levels of IL-4 and IL-10 suggest a potential Th2-type response. The production of IL-10 and TGF beta, which are anti-inflammatory cytokines, might help regulate the immune response and prevent excessive inflammation. The presence of the danger signal (D) suggests the activation of innate immune cells and the release of inflammatory mediators. The overall cytokine profile indicates a balanced immune response with both pro-inflammatory and anti-inflammatory cytokines being produced.

Discussion

Norwalk viruses are considered as a leading etiology of epidemic acute gastroenteritis as well as an important cause of sporadic cases of acute gastroenteritis (Peng, J. et al, 2011). Currently, there is no vaccine to prevent human norovirus infection, and there is no specific therapy available to treat it (Chen Z, et al, 2018). The norovirus genome has three open reading frames (ORFs) of which ORF2 and ORF3 encode the major capsid protein (VP1) that determines the antigenicity of the virus, as well as the minor capsid protein (VP2). The majority of the studies performed to design vaccine for norovirus used VP1 as a vaccine construct. For instance vaccine using recombinant adenovirus expressing the norovirus major capsid protein VP1was developed. The vaccine measured the cross-reactive neutralizing antibody responses which are required for a successful norovirus vaccine (Ettayebi K, et al, 2018). A study by Tucker et al (2008) developed a currently human clinical trials vaccine that employs a recombinant adenovirus expressing the norovirus GI.1 major capsid protein (VP1) in an oral tablet formulation developed by Vaxart, Inc. Moreover recombinant adenovirus vaccine expressing the norovirus GII.4 major capsid protein VP1 (Guo L,et al, 2008) and multiple VLP vaccine developed from VP1 protein were developed by the Chinese center for disease control and prevention. Both studies demonstrated the antigenicity of VP1 as a vaccine candidate. Therefore this study aimed to propose multiple epitopes vaccine candidates from the capsid protein (VP1) to elicit B and T lymphocytes and act as a vaccine candidate using immune-informatics tools. In the current study the B cell epitopes were predicted from the capsid protein VP1 to find the potential epitopes that would interact with B lymphocytes and initiate immune response. For the vaccine to be recognized by the B cell antibodies it must be linear and located on the surface of the antigen protein to be easily accessible. In addition to that the candidate vaccine should demonstrate greater antigenicity to elicit antibodies production. Therefore several tools from IEDB analysis resources were used to identify B cell epitopes such as Bepipred linear epitope prediction analysis Emini surface accessibility prediction

and Kolaskar and Tongaonkar antigenicity scale (Jespersen,et al, 2017). These tools provided multiple epitopes that were linear, on the protein surface and antigenic. However, for the epitopes to be proposed as a vaccine candidate it should be nonallergic and nontoxic to the host cells (Azim KF, et al, 2017). Thus the predicted epitopes further subjected to Vaxigen antigenicity, allergenicity and toxicity investigations. **6 CTL** and **B-cell epitopes** while **9 HTL** epitopes were selected based on high binding affinity scores. Since the immune response of T cell is long lasting response compared to B cell, where the antigen can easily escape the antibody memory response. This considered that CD8+T and CD4+T cells response play a major role in antiviral immunity [48]. Cytotoxic CD8+T lymphocytes are considered as an important parameter in recognizing and killing infected cells or producing specific cytokines that prevent the infection in the body (Garcia,et al, 1999). Accordingly these epitopes were strongly recommended as promising epitopes vaccine candidates against T lymphocyte cells. Recently a study by Azim et al (2019) used multiple bioinformatics tools to predict epitopes from the VP1 and VP2 proteins of the noroviruse. However none of their predicted epitopes for B and T lymphocytes were corroborated to our proposed epitopes. This might be attributed to the differences in the software used in both studies to predict vaccine candidates.

Reference:

1. Zhang, Y. J., Li, S., Gan, R. Y., Zhou, T., Xu, D. P., & Li, H. B. (2015). Impacts of gut bacteria on human health and diseases. *International journal of molecular sciences*, *16*(4), 7493–7519. https://doi.org/10.3390/ijms16047493
2. Glass, R. I., Parashar, U. D., & Estes, M. K. (2009). Norovirus gastroenteritis. *The New England journal of medicine*, *361*(18), 1776–1785. https://doi.org/10.1056/NEJMra0804575
3. Sell J, Dolan B. Common Gastrointestinal Infections. Prim Care. 2018 Sep;45(3):519-532.
4. Dar, M. A., Kumar, P., Kumar, P., Shrivastava, A., Dar, M. A., Chauhan, R., ... & Dhingra, S. (2022). Designing of peptide based multi-epitope vaccine construct against gallbladder cancer using immunoinformatics and computational approaches. *Vaccines*, *10*(11), 1850.
5. Peng, J. and J. Xu, A multiple-template approach to protein threading. Proteins: Structure, Function, and Bioinformatics. 2011. 79(6): p. 1930-1939.
6. Chen Z, Sosnovtsev SV, Bok K, Parra GI, Makiya M, Agulto L, et al. Development of Norwalk virus-specific monoclonal antibodies with therapeutic potential for the treatment of Norwalk virus gastroenteritis. Journal of virology. 2013; 87(17): 9547-57.
7. Ettayebi K, Crawford SE, Murakami K, et al. Replication of human noroviruses in stem cell-derived human enteroids. Science. 2016; 353(6306): 1387-1393
8. Tucker SN, Tingley DW, Scallan CD. Oral adenoviral-based vaccines: historical perspective and future opportunity. Exp Rev Vaccines. 2008; 7: 25-31.
9. Guo L, Wang J, Zhou H, et al. Intranasal administration of a recombinant adenovirus expressing the norovirus capsid protein stimulates specific humoral, mucosal, and cellular immune responses in mice. Vaccine. 2008; 26; 460-468.
10. Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P. BepiPred 2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017; 45, W24-W29.
11. Azim KF, Mahmudul H, Hossain Md N, Somana SR et al. Immunoinformatics approaches for designing a novel multi epitope peptide vaccine against human norovirus (Norwalk virus). Infection, Genetics and Evolution. 2019; 74 (103936)
12. Garcia, K.C., Teyton, L., Wilson, I.A. Structural basis of T cell recognition. Annu. Rev. Immunol. 1999; 17, 369-397.