International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

An Innovative Approach towards Construction, Evaluation, Simulations and Cloning of a Multi-epitope Vaccine against Ebola Virus, through Immunoinformatics-based Study Design

Dr Paushali Pal

Assistant Professor, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India.

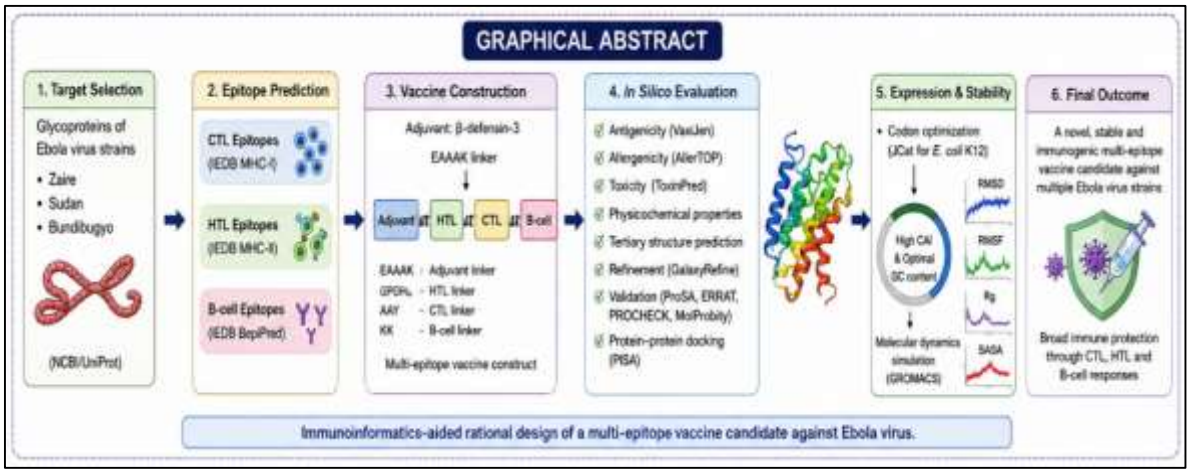


Figure 1: Graphical Abstract

HIGHLIGHTS

1. Multi-epitope vaccine construction incorporating B-cell and T-cell epitopes representing all 3 Ebola virus strains - Zaire, Sudan and Bundibugyo. .
2. Comprehensive Validation of the Vaccine construct by software like PROCHECK, Pro-SA, and MolProbity.
3. MD Simulations (GROMACS experiments) further validate the integrity and compactness of the vaccine structure.
4. Immunological potential of the vaccine construct is derived from C-IMMISIM results.
5. Codon Optimization and efficient cloning using SnapGene serve as a convenient E. coli K-12 expression system.

ABSTRACT

Introduction: Ebola virus has been preying on human populations for a long time. But recent outbreaks in Africa with symptoms such as severe haemorrhagic fever urge the immediate need for research on an

Ebola vaccine as an initial step. Hence, in this study, an in silico vaccine is designed against the Ebola virus, with epitopes representing all three strains: Zaire, Sudan, and Bundibugyo.

Methodologies: After initial sequence retrieval, they are prioritized based on physicochemical analyses and other suitable parameters. Epitopes are then predicted from selected sequences and subjected to stringent analyses – antigenicity, allergenicity, toxicity, conservation, solubility, among other parameters. An in silico vaccine is constructed using epitopes, incorporating Beta-defensin 3 at the C-terminus and a His6 Tag at the N-terminus. After a stringent set of analyses, the protein structure is subjected to Secondary and Tertiary structure prediction. After refinement of the structure, validation procedures are performed using PROCHECK, PRO-SA, and Molprobity. Simulation methods such as GROMACS and Immune Simulations, as implemented in C-IMMISIM, are used. Thereafter, docking with TLR4 (in PATCHDOCK software) , interaction analyses of the docking mechanism , by PISA software, Codon Optimization by JCAT and cloning into the pET28a(+) with SnapGene , are achieved.

Results: After applying stringent analytical methods, epitopes are selected for in silico vaccine construction against Ebola virus strains – Zaire, Sudan, and Bundibugyo. The constructed vaccine is subjected to similar analyses again, yielding a Vaxijen score of 0.6853. Secondary structure prediction by PSIPRED ensures reliable conformity, while tertiary structure prediction by Colab AlphaFold2 further reveals the vaccine structure. After refinement, the validation process reveals that the protein vaccine structure contains over 98% Ramachandran-favoured regions and an ERRAT score of 71%. Docking of the vaccine against the TLR4 immune receptor by PATCHDOCK is achieved, and PISA interaction analyses reveal a binding energy of -14.5kcal/mol. Again, Molecular Dynamics simulations reveal its structural integrity and conformity amid solvation analyses. Then, for C-IMMISIM results, both humoral and cell-mediated immunity are activated, along with memory cell propagation. Again, in JCAT, the CAI score is 0.9572, and a GC content of 52.95% is achieved, indicating high expression potential.

Conclusion: The immunoinformatics tools applied suggest its conformational integrity, compactness, and safety for further wet-lab research. World -class population coverage of 87.44% , along with approx 70% coverge for West, East and Central Africa , reflects the need for further labotory-based experiments with this vaccine candidate.

Keywords: Immunoinformatics, Vaccine design, Codon optimisation, In silico cloning, Molecular dynamics simulation, Epitope prediction.

1. Introduction

Ebola virus disease (EVD) poses a serious threat as it causes fatal haemorrhagic fever . The causative agent belongs to the genus *Ebolavirus* within the family *Filoviridae*. Its first recognized outbreak has been witnessed in Central Africa , in the year 1976. Since then, Ebola virus has remained a public health concern , because of its high fatality rate and rapid human-to-human transmission, thereby causing widespread epidemics [1]. Symptoms of EVD range from fever, fatigue, minor gastrointestinal malfunctions to severe haemorrhage and multi-organ failure , leading to fatality. In recent times , EVD has conquered regions like Western Africa , Central and Eastern Africa. The major causative agents include Zaire , Sudan and Bundibugyo strains of Ebola virus [2] .

Researchers are continuously striving to prepare effective vaccines against the Ebola virus , though technical challenges like limited viral representations, curb the development of broad-spectrum vaccines [3]. Again, vaccines with limited alleles representations , are not effective against a wide-range of

populace. The continuous emergence of new viral variants challenges the researchers to develop innovative vaccines , capable of cross-protective immunity against against multiple viral strains. Furthermore, conventional vaccine development is time-consuming , costly, and depend on extensive laboratory-based experimentation [16].

The viral glycoprotein (GP) plays a crucial role in host invasion process , such as host-cell attachment , membrane fusion, viral entry , making it an lucrative agent for vaccine development. Thus, the viral glycoproteins are easily recognized by the immune system and contain several immunodominant epitopes , capable of elliciting both humoral and cell-mediated immunity. Consequently, the glycoprotein-based vaccines have been an attractive strategy.

Advances in Bioinformatics-based software have revolutionized the vaccine-syntheses strategies by quick identification of antigenic epitopes , leading to the construction of multi-epitope vaccines , conferring protection to wide-range of population. These approaches integrate processes like epitope predictions, antigenic assessment, allergenicity screening, structural modelling, molecular docking, molecular docking simulations, immune simulations , to invent effective vaccines, within a suitable time frame at cheaper cost. In this context, multi-epitope vaccines are far more advantageous beacause they combine B-cell epitopes and T-cell epitopes , eliciting immune responses while minimizing unwanted antigenic components.

In this article , a multi-strain, multi-epitope vaccine has been devised with the sole aim of conferring protection to a wide- range of populations , recently affected by Ebola virus contamination. In order to cause an effective fruition, glycoprotein sequences from Zaire, Sudan and Bundibugyo strains of Ebola viruses are selected , stringently analyzed, epitopes-B-cells, CTL and HTL are predicted and from a pool among them, only a few are selected after stringent safety analyses , leading to the development of a suitable in-silico vaccine. Beta-defensin-3 adjuvant is incorporated at N-terminal of the vaccine while His6 Tags are included at the C-terminal . The designed vaccine is tested for safety parameters like antigenicity analyses, allergenicity screening, toxicity profiling , along with secondary structure prediction, tertiary structure prediction, , structural refinement, validation, docking with immune receptor, interaction analyses , molecular dynamics simulations, immune simulations, codon optimization and cloning into plasmid for suitable expression .

Hence, the objective of this study is to devise a promising and effective broad-spectrum multi-epitope vaccine candidate, capable of elliciting suitable immune responses against multiple Ebola virus strains.

2. Results and Discussion

2.1 Sequences Retrieval and Analyses

Zaire, Sudan, and Bundibugyo strain sequences were obtained from NCBI and UniProt. After that, Physicochemical analyses (ExPASy ProtParam results) of the retrieved sequences were performed. Table 1 below enlists the results of the Physicochemical properties.

Parameter	**Zaire**	**Sudan**	**Bundibugyo**
Sequence length (aa)	676	676	676
Molecular weight (kDa)	74.46	74.986	75.618
Theoretical pI	6.16	5.60	6.27
Extinction coefficient	99350	10783	97860
Estimated half-life (Mammalian)	30 h	30 h	30 h
Instability Index	38.36	41.60	36.61

Aliphatic Index	75.77	80.52	78.43
GRAVY	0.380	0.363	0.409

Table 1 : Enlists the results of Physico-chemical properties of Ebola virus strains' sequences (Zaire, Sudan, Bundibugyo) , as obtained from ExPasy Protparam.

As structural proteins of the Ebola virus are responsible for host invasion and entry, they constitute important antigenic determinants, crucial for vaccine construction concepts. Hence, before proceeding with the vaccine synthesis process, the antigenicity of the sequences is assessed using the Vaxijen score, with a threshold of 0.4, confirming that the sequences are antigenic. Again, AllerTOPv2.1 analyses are performed to ensure the sequences are safe for further analysis. Hence, all selected sequences are classified as 'non-allergenic'.

2.2 B-Cell Epitopes' Prediction and Selection

B-cell epitopes are predicted using BCPREDS, and each residue is selected based on prediction score, antigenicity, non-allergenic properties, conservancy (IEDB Tool), and solubility (Soluprot). Table 2 summarises B-cell epitopes and their essential characteristics[4]. Care is taken to represent all 3 strains in the B-cell epitopes. Hence, two epitopes from each of the 3 strains are considered.

Virus	**Pos.**	**Selected B-cell Epitope**	**Length**	**Vaxi Jen Score**	**Allergenicity**	**Toxicity**	**Conservancy**	**SoluProt Score**
Zaire	41–62	TLQVSDVDKLVCRDKLSSTNQL	22-mer	0.5271	Non-Allergenic	Non-Toxic	100%	0.927
Zaire	85–107	WGFRSGVPPKVVNYEAGEWAENC	23-mer	0.9821	Non-Allergenic	Non-Toxic	100%	0.927
Sudan	42–61	LEVTEIDQLVCKDHLASTDQ	20-mer	0.6702	Non-Allergenic	Non-Toxic	100%	0.900
Sudan	87–107	FRSGVPPQVVSYEAGEWAENC	21-mer	0.6541	Non-Allergenic	Non-Toxic	100%	0.916
Bundibugyo	105–121	ENCYNLDIKKADGSECL	17-mer	0.5510	Non-Allergenic	Non-Toxic	100%	0.837
Bundibugyo	295–316	NFTKTLSSEELSVILVPRAQDP	22-mer	0.5943	Non-Allergenic	Non-Toxic	100%	0.837

Table 2 : Depicts B-cell Epitopes from Zaire, Sudan and Bundibugyo strain sequences of Ebola virus.

2.3 CTL and HTL Epitopes Prediction and Selection

Cytotoxic- T-cell epitopes (CTL) or MHC-I epitopes and Helper T-cell epitopes (HTL) or MHC- II epitopes are predicted using IEDB (Immune Epitope Database) tool and the selection criteria is based on good IEDB score ,lower percentile rank, high conservancy, good binding affinity and size (9-mer for CTL and 15-mer for HTL). Precautionary measures adopted to represent all 3 strains of Ebola virus as 'epitopes'. Safety parameters adopted while selection is under process , such as high antigenicity, non-

allergenicity and non-toxicity. Table 3 and Table 4 depicts the CTL epitopes and HTL epitopes respectively [5-8].

Virus Strain	Selected CTL Epitope	HLA Allele	Percentile Rank	VaxiJen Score	Allergenicity	Toxicity	Conservancy
Zaire	RTFSILNRK	HLA-A*11:01	0.01	1.8122	Non-allergen	Non-toxic	100%
Zaire	TTIGEWAFW	HLA-B*58:01	0.04	0.9650	Non-allergen	Non-toxic	100%
Zaire	TYVQLESRF	HLA-A*24:02	0.02	1.4723	Non-allergen	Non-toxic	100%
Zaire	NTIAGVAGL	HLA-A*68:02	0.02	0.8147	Non-allergen	Non-toxic	100%
Zaire	GFRSGVPPK	HLA-A*30:01	0.03	0.8887	Non-allergen	Non-toxic	100%
Sudan	RTYTILNRK	HLA-A*03:01	0.01	0.9918	Non-allergen	Non-toxic	100%
Sudan	EVTEIDQLV	HLA-A*68:02	0.01	0.4981	Non-allergen	Non-toxic	100%
Sudan	RISDRATRK	HLA-A*03:01	0.02	1.3024	Non-allergen	Non-toxic	100%
Bundibugyo	SPRPVPTST	HLA-B*07:02	0.02	0.9705	Non-allergen	Non-toxic	100%
Bundibugyo	KVFPIPLGV	HLA-A*02:01	0.02	1.5020	Non-allergen	Non-toxic	100%
Bundibugyo	TYVQLEPRF	HLA-A*24:02	0.01	2.2638	Non-allergen	Non-toxic	100%
Bundibugyo	FPIPLGVVH	HLA-B*35:01	0.03	1.4099	Non-allergen	Non-toxic	100%

Table 3 : Depicts a list of Cytotoxic T-cell epitopes (CTL) , representing all three strains of Ebola virus - Zaire, Sudan and Bundibugyo , with distinct alleles representations.

Virus Strain	Selected HTL Epitope	HLA Allele	Percentile Rank	VaxiJen Score	Allergenicity	Toxicity	Conservancy
Zaire	WVIILFQRTFSIPLG	HLA-DRB1*07:01	0.44	0.8741	Non-allergen	Non-toxic	100%

Virus Strain	Selected HTL Epitope	HLA Allele	Percentile Rank	VaxiJen Score	Allergenicity	Toxicity	Conservancy
Zaire	PDGIRGFPRCRYVHK	HLA-DRB1*13:02	0.47	0.4066	Non-allergen	Non-toxic	100%
Sudan	QETITETTATIIGTN	HLA-DRB1*07:01	0.08	0.5472	Non-allergen	Non-toxic	100%

Table 4 : Shows a list of Helper T-cell epitopes (HTL) from Zaire and Sudan strains of Ebola virus, after stringent analyses.

2.3.1 Immune Alleles

In the case of CTL epitopes multiple alleles representation is undertaken to represent a wide variety of population. Eight varieties of alleles are included (Table 3) along with two different alleles for HTL epitopes (Table 4).

2.4 Vaccine Synthesis (In Silico)

A single multi-epitope vaccine is constructed employing the selected epitopes , aiming to be effective against all 3 strains of Ebola virus - Zaire, Sudan and Bundibugyo. The adjuvant used in this case is Beta-Defensin-3 , as this enhances the immune responses by activating the dentritic cells and macrophages of the immune system by accurately representing the protein bodies on receptors like TLR1 and TLR2 , improving both cell-mediated and humoral immunity.One of the prominent Linker - EAAAK is included as it compartmentalizes the adjuvant from epitope region , so that each functions with their own integrity. Again, its rigid helical structure restricts from unwanted interactions , enabling the vaccine components to function properly. Thereafter, the CTL epitopes comprising as one of the main components of the vaccine , activates cell mediated immunity by presenting the MHC-I epitopes on CD8+ T lymphocytes , directly eliminating the Ebola virus upon entry. For HTL epitopes , the representation of MHC-II epitopes on CD4+ T-lymphocytes , again activating cell mediated immunity , eliminating the virus upon entry. Other important linkers such as AAY -linker (CTL-Linker) facitilaes accurate MHC-I epitopes representation by its appropriate cleavage and TAP transport. Likewise , the GPGPG linker elicit the immune responses by enabling appropriate representation of MHC-II on CD4+T-lymphocytes. B-cell epitopes , on the otherhand enhances humoral immunity by activating B-cell lymphocytes producing sufficient antibodies, counteracting and eliminating pathogens. KK-linker or B-cell linker is included to enhance humoral immunity and separates the B-cells from other portions of the vaccine , enabling them to function properly. Lastly, the His-6 tag added at the C-terminal facilitates in protein purification by Affinity Chromatography. Figure 2 depicts the vaccine structure.

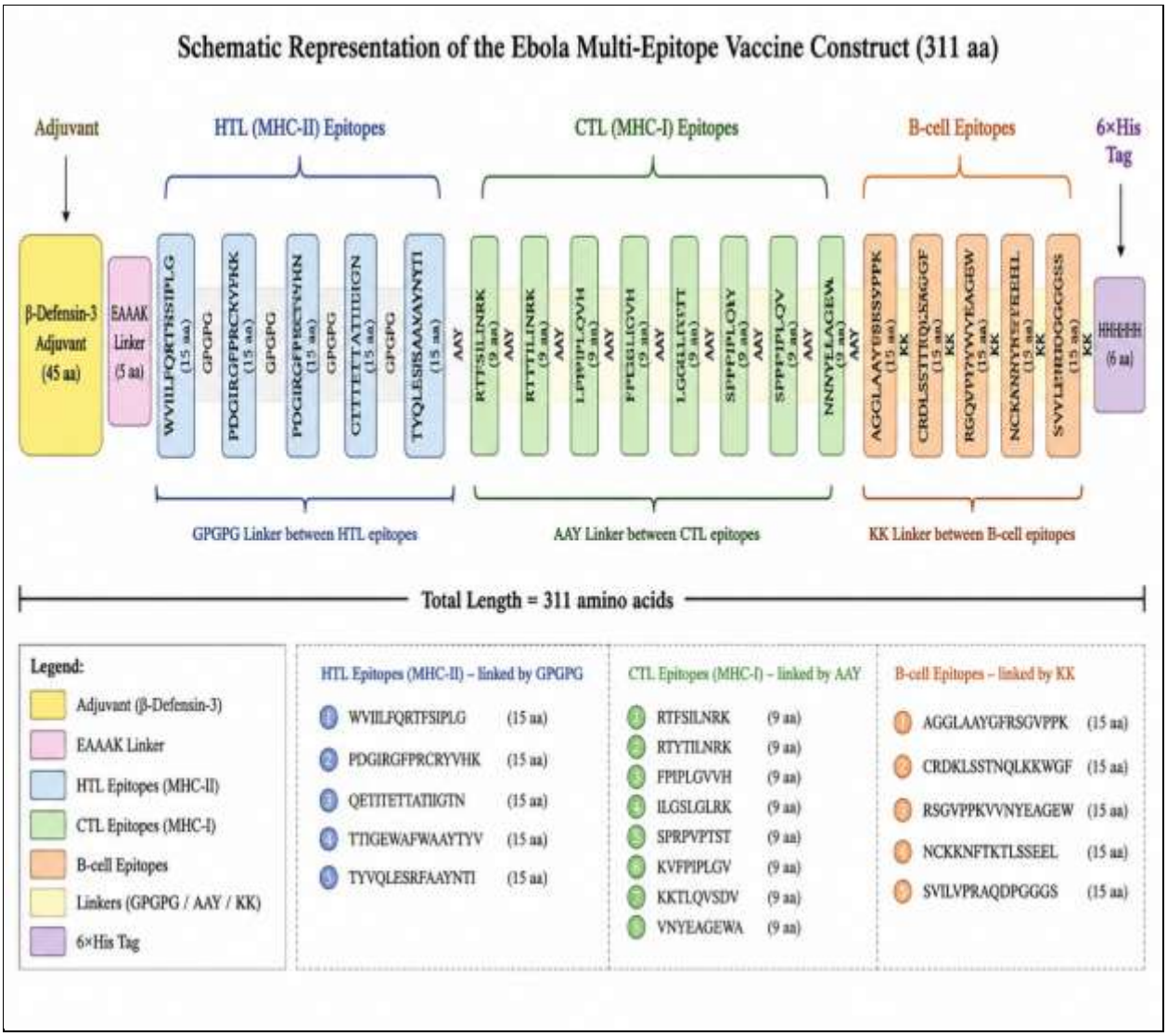


Figure 2 : Schematic diagram of Ebola vaccine construct

2.5 Analyses of Vaccine Characteristics

The physicochemical properties of the constructed vaccine structure is evaluated employing ExPASY Protparam . Parameters such as total amino acid count, molecular weight in terms of kDa , theoretical pI, extinction coefficient , half-life , instability index (I.I), aliphatic index (A.I)and hydropathicity (GRAVY) is analyzed (Table 5) .

Safety parameters judged stringently , include : high antigenic score in Vaxijen , non-allergenicity in Allertop tool , non-toxicity as depiced by Toxinpred and solubilty assay by Soluprot (Table 6).

2.6 Secondary Structure Interpretation of Protein Vaccine

In order to evaluate the secondary structure of the proteinaceous structure of the vaccine , PSIPRED tool is employed for the purpose. The structure comprises of strands, helices , coils with no deformabilities.The same is depicted in Figure 3. Mostly the secondary structure depicts high-confidence level (C), ensuring the integrity of the structural conformation of the vaccine structure. (Figure 4).

Table 5 & 6 : Represents Physico-chemical characteristics of Ebola vaccine and other parameters' characteristics of the vaccine construct, respectively.

Parameter	Result
Number of amino acids	311
Molecular Weight	34.125 kDa
Theoretical pI	9.94
Total Negatively Charged Residues (Asp + Glu)	17
Total Positively Charged Residues (Arg + Lys)	45
Molecular Formula	C1545H2422N438O419S9
Total Number of Atoms	4833
Extinction Coefficient (Cystines formed)	53,330 M^{-1} cm^{-1}
Extinction Coefficient (Reduced Cys)	52,830 M^{-1} cm^{-1}
Estimated Half-life (Mammalian reticulocytes, in vitro)	30 hours
Estimated Half-life (Yeast, in vivo)	>20 hours
Estimated Half-life (Escherichia coli, in vivo)	>10 hours
Instability Index (II)	27.60
Stability Prediction	Stable
Aliphatic Index	78.17
GRAVY (Grand Average of Hydropathicity)	−0.248

Tool	**Parameter**	**Result**	**Interpretation**
AllerTOP v2.0	Allergenicity Prediction	Probable NON-ALLERGEN	Vaccine construct is predicted to be non-allergenic
VaxiJen v2.0	Antigenicity Score	0.6853 (Threshold =0.4)	Score exceeds threshold (0.40)
VaxiJen v2.0	Overall Prediction	Probable ANTIGEN	Vaccine construct is predicted to be antigenic
ToxinPred	Toxicity Prediction	Non-toxic	No toxic regions predicted for the vaccine construct
SoluProt	Solubility Probability	0.9192	High probability of soluble expression upon overexpression

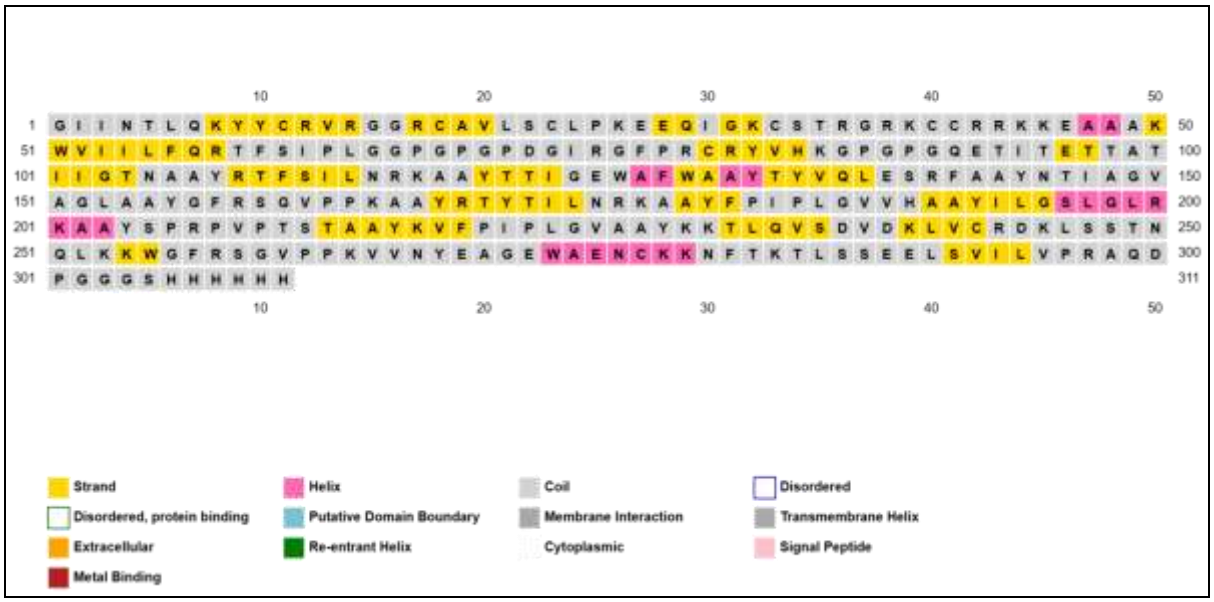


Figure 3 : Depicts Secondary structure prediction of Ebola vaccine construct by PSIPRED

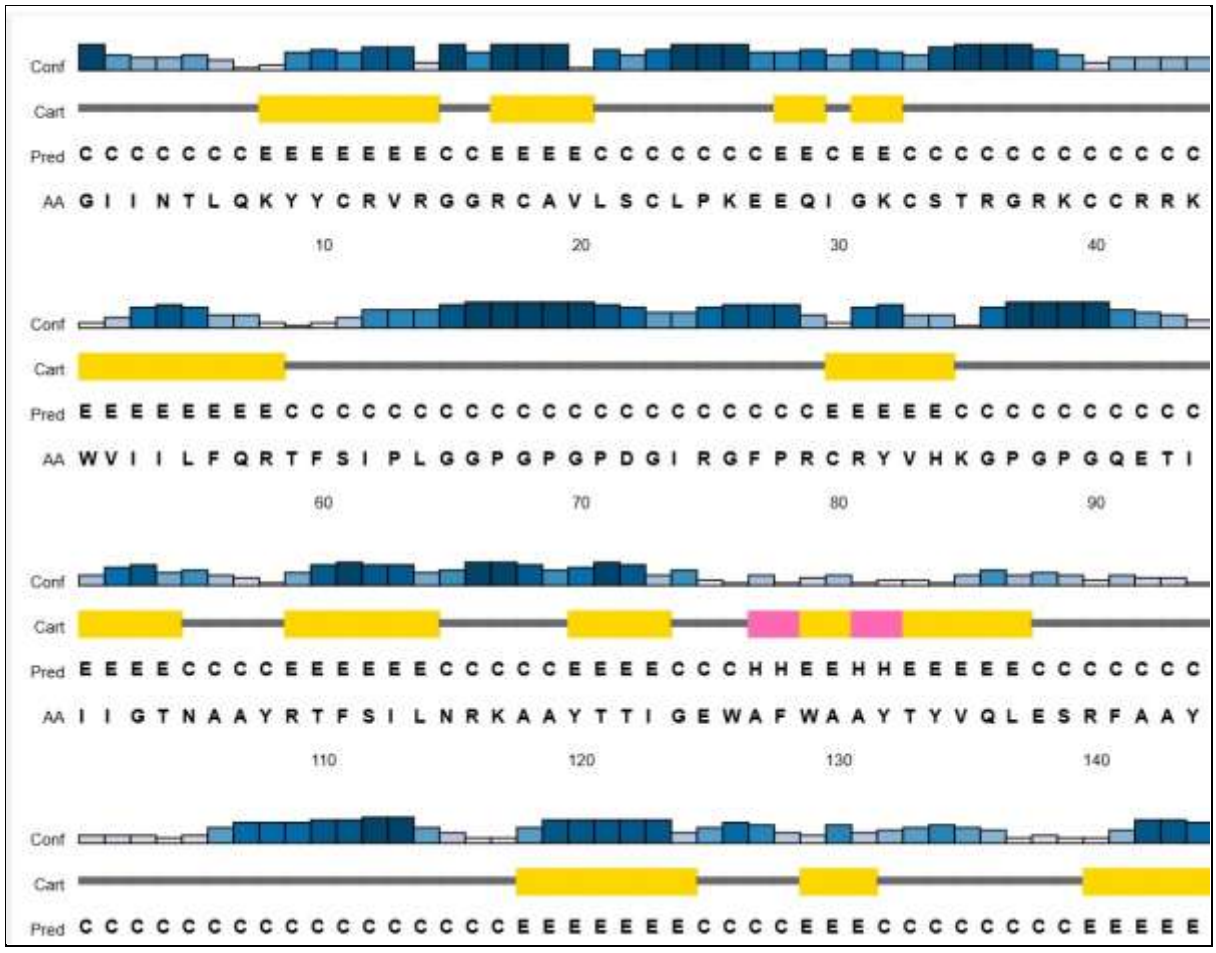


Figure 4 : Depicts high order of Confidence level (C) , signifying good structural integrity of the protein , Ebola vaccine structure.

2.7 Tertiary Structure Interpretation of Protein Vaccine

The three-diamensional structure interpretation of the vaccine protein is done with ColabFold (AlphaFold2)[10-11]. The results obtained after the analyses, generates 5-structural models with rankings based on their prediction confidence (pLDDT). Amongst the 5, rank 1 structure is considered for further analyses such as structural validation, molecular docking and molecular dynamics simulations. This rank 1 structure is also analyzed in terms of sequence coverage, Predicted Alignment Error (PAE) and Predicted Local Distance Difference Test (pLDDT) analyses. Figure 5 below shows the 3d structure of the protein vaccine , with Rank1 , as obtained from the visualization tool , Chimera1.19.

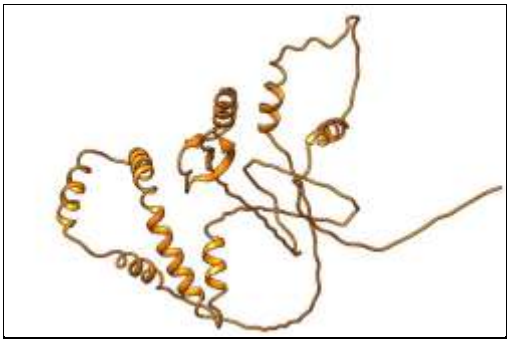


Figure 5 : Tertiary protein structure of the Ebola vaccine construct, as obtained from ColabFold (AlphaFold2)

2.7.1 Sequence Coverages

The sequence coverage plot is constructed after multiple sequence alignment coverage as against the query sequence. The regions of the protein vaccine close to the N-Terminal represents higher coverage being the conserved regions , as compared to the other linker -rich regions . These variations are highly expected as the vaccine construct is composed of heterogenous epitopes , each joined by linkers , completing the in silico process of synthesis. Figure 6 represents the Sequence coverage plot .

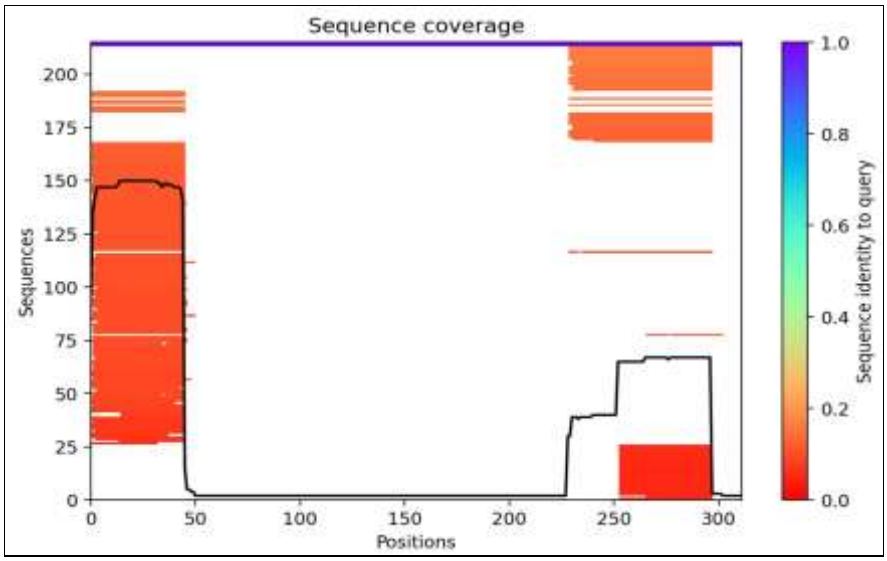


Figure 6 : Depicts the Sequences Coverage plot for Ebola vaccine construct , as obtained from ColabFold (AlphaFold2). Red= conserved regions , traversed near the N-terminal of the vaccine construct.

2.7.2 Prediction Alignment Error (PAE)

The PAE analysis reveals low alignment error indicated by blue diagonal line , interpreting the reliability of the three-diamensional structure of the protein , Ebola vaccine. High level of consistency is observed between the rigid distant regions and the flexible portions of the vaccine , in the presence of the linkers. Figure 7 depicts the PAE analyses.

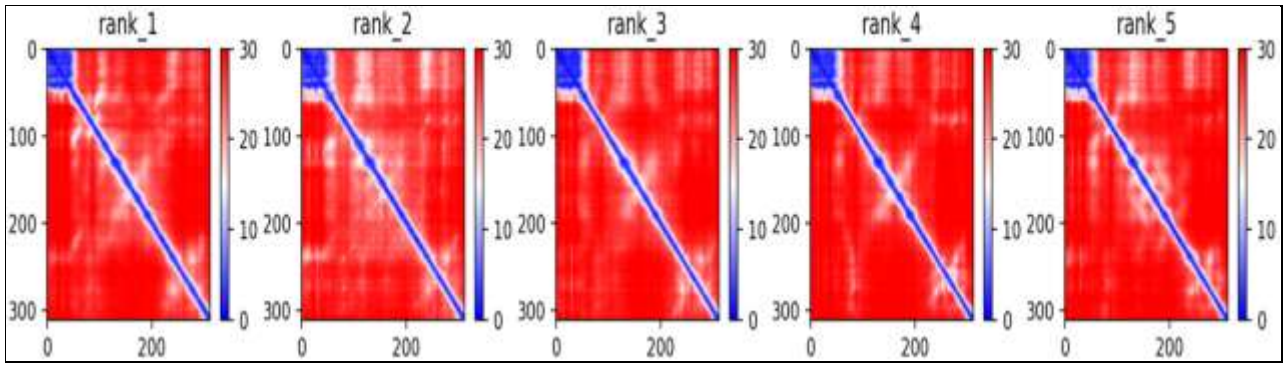


Figure 7 : Depicts the PAE analysis results of the Ebola vaccine construct from ColabFold (AlphaFold2)

2.7.3 Predicted Local Distance Difference Test (pLDDT)

The pLDDT profile depicts high confidence level upto 95 , whereas low confidence is observed near the central and C-terminal portion. The reduced confidence level pertains to the presence of adjuvant (human Beta-Defensin -3), linkers and exposed epitope segments, which are the common characteristics of a multi-epitope protein vaccine . Figure 8 depicts the pLDDT plot for the best ranked 3d structure of the protein vaccine. This structure is thus considered for further refinement, validation and other downstream analyses.

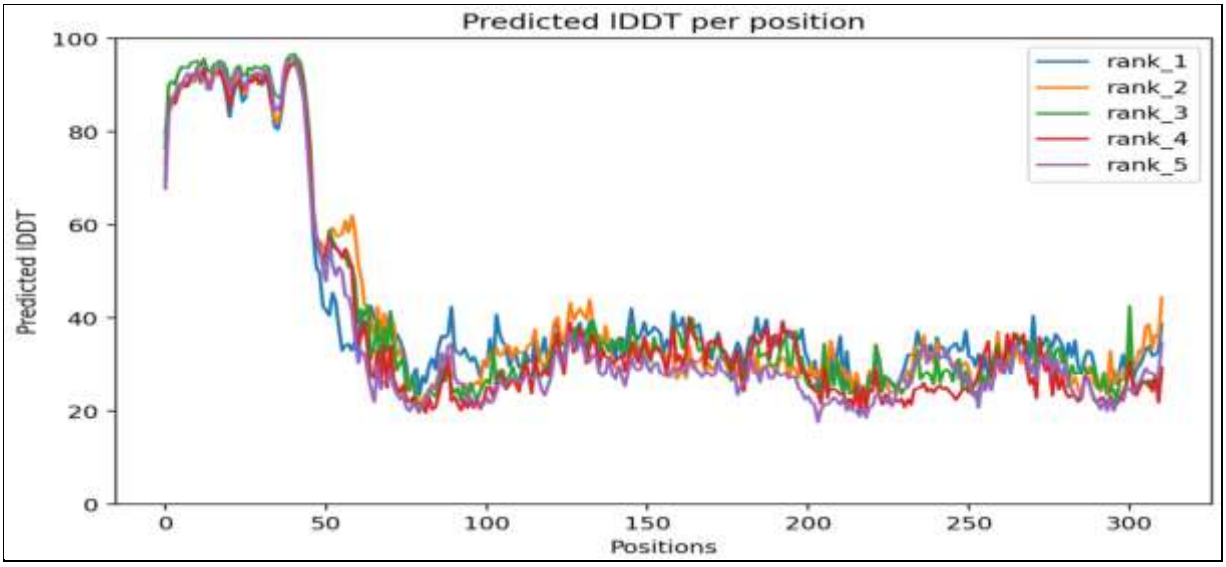


Figure 8 : Depicts the pLDDT profile analysis results of the Ebola vaccine construct from ColabFold (AlphaFold2)

2.8 Protein Vaccine Structure Refinement

The vaccine structure is refined using Galaxyrefine server [9] , resulting in five generated Models. Amongst these, Model1 and Model3 are chosen initially. Steric clashes (Figure 10 and Figure9,obtained respectively from Model3 and Model1 using Chimera1.19) are more intense in case of Model1 , as compared to Model3. Hence, Model3 is selected. Thus, Model3 includes 99% Ramachandran favoured regions , indicating high order of stereochemical protein structure. Table 7 summarizes the Models' parameters while Figure11 shows refined protein vaccine structure, obtained from the visualization tool - Chimera1.19.

Figure 9: Steric clashes of Model 3 of Ebola vaccine refined structure

Figure10: Steric clashes of Model 1 of Ebola vaccine refined structure

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.338	13.0	12.5	65.0
MODEL 1	0.8601	0.657	0.940	1.8	0.0	99.0
MODEL 2	0.8617	0.665	0.838	1.2	0.0	99.0
MODEL 3	0.8674	0.670	0.908	1.6	0.0	99.4
MODEL 4	0.8641	0.667	0.940	1.8	0.0	99.4
MODEL 5	0.8553	0.666	0.969	2.0	0.0	98.4

Table 7: Depicts the parameters of the Models generated by GalaxyRefine tool

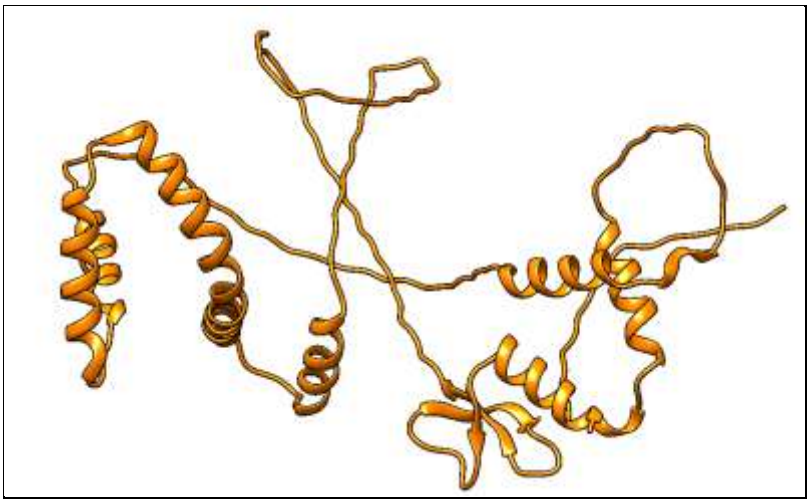


Figure11: Protein structure of the Ebola vaccine after refinement (Model 3).

2.9 Vaccine Protein Structure Validation

Before undertaking docking analyses , the protein structure of the vaccine(Model3) is validated using three different servers , such as Molprobity , Procheck and ProSA. Both the Procheck server as well as MolProbity [12] depicts that the Ramachandran favoured regions of the protein construct is above 98% (Figure 12 and Figure13) , indicating high order of protein stereochemical strucure of the vaccine. Again, the ProSA analysis, reveals the Z-score to be -0.5 as the protein vaccine structure consists of linker-rich regions , prepared in-silico(Figure14).

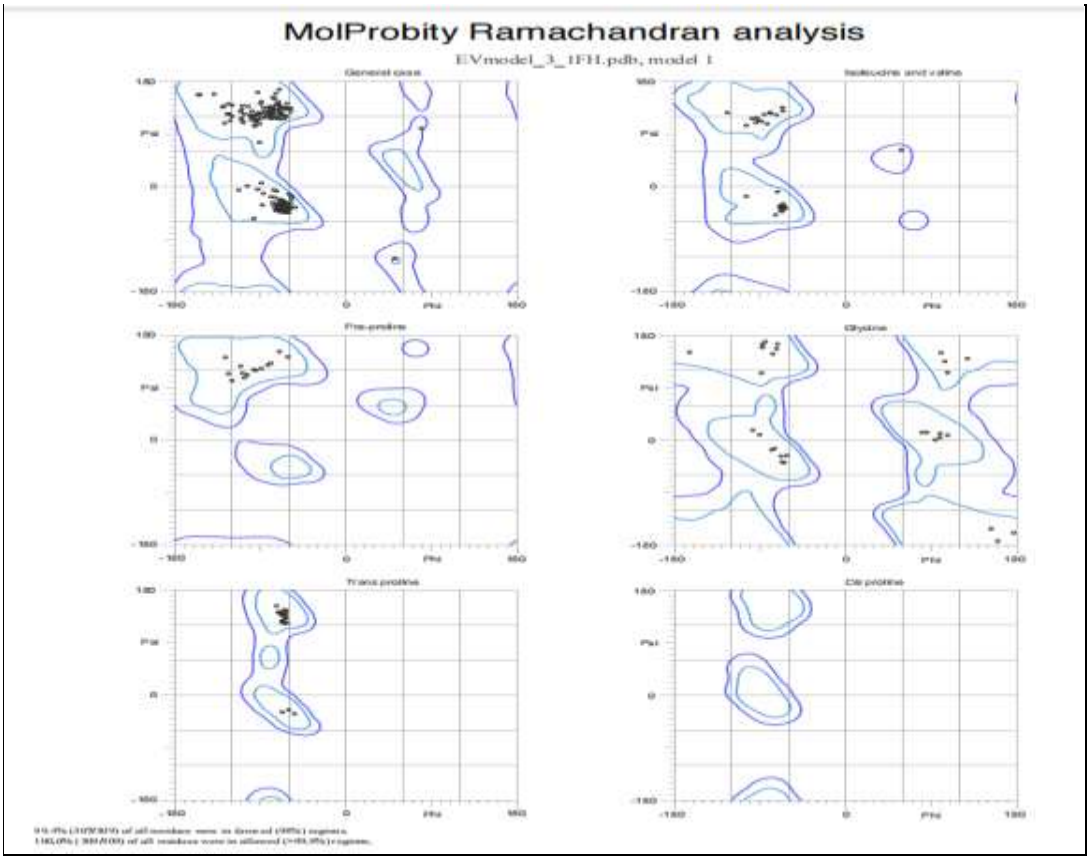


Figure12 : Depicts the Ramachandran plot of the protein , Ebola vaccine with above 98% as Ramachandran favoured regions .

```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
|                                                                            |
| /var/www/SAVES/Jobs/546679/saves.pdb   1.5                  311 residues  |
|                                                                            |
+| Ramachandran plot:   98.1% core    1.2% allow    0.8% gener    0.0% disall |
|                                                                            |
+| All Ramachandrans:    3 labelled residues (out of 309)                     |
| Chi1-chi2 plots:      0 labelled residues (out of 163)                     |
| Side-chain params:    5 better     0 inside      0 worse                   |
|                                                                            |
+| Residue properties: Max.deviation:     8.9             Bad contacts:    0  |
+|                     Bond len/angle:    4.3    Morris et al class:  1  1  3  |
|                                                                            |
| G-factors           Dihedrals:   0.42  Covalent:  -0.17    Overall:   0.20  |
|                                                                            |
| Planar groups:   100.0% within limits   0.0% highlighted                   |
|                                                                            |
+----------------------------------------------------------------------------+
  + May be worth investigating further.  * Worth investigating further.
```

Figure 13 : Shows the results of PROCHECK validation for the protein Ebola vaccine structure

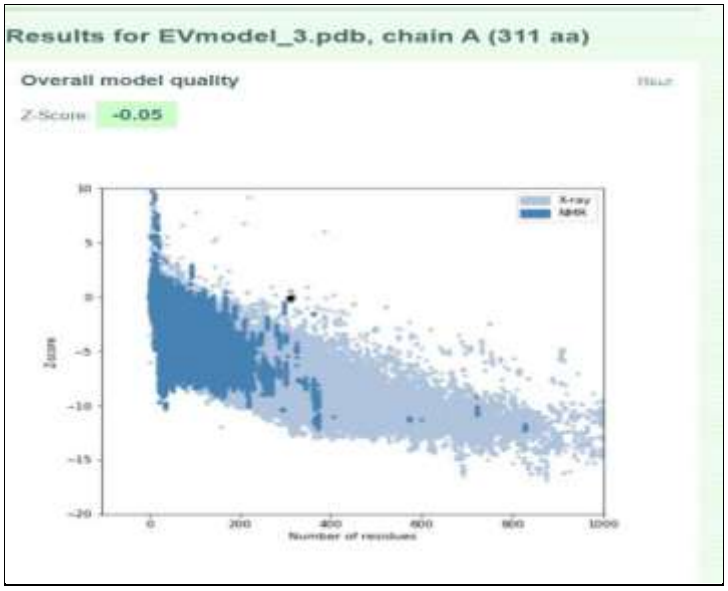


Figure 14 : Validation result , Z-score obtained from Pro-SA web server tool for the Ebola vaccine construct

2.10 Docking Analyses of the Protein Vaccine Structure Against Immune Receptors

The docking interactions are undertaken between the immune receptor, TLR4 with PDB ID 2Z63 (Figure15) and the refined , validated protein vaccine structure, Model3. This docking analyses is performed using PATCHDOCK server. The H-bonds (red colour) and the Vanderwaals contacts (black pseudobonds within 0.4A)are depicted in Figure17. Again, Figure 18 depicts the results of PATCHDOCK server. Upon comparison of the score as well as other parameters , it is seen that the Model1 , scores the best. Hence for all further downstream analyses , such as the Molecular Dynamics Simulations , are done considering Model1 from PATCHDOCK.

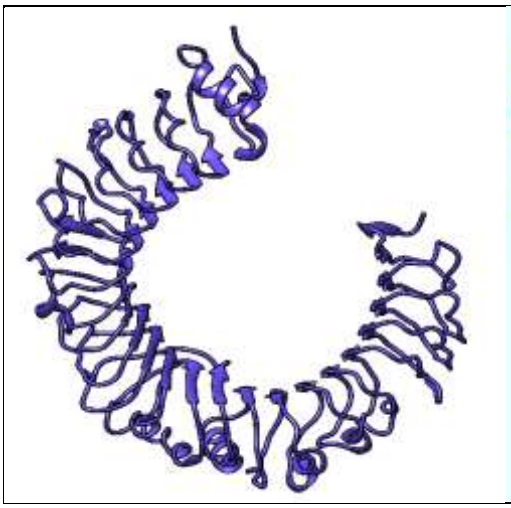


Figure15 : Immune Receptor TLR4 (PDB ID 2Z63)

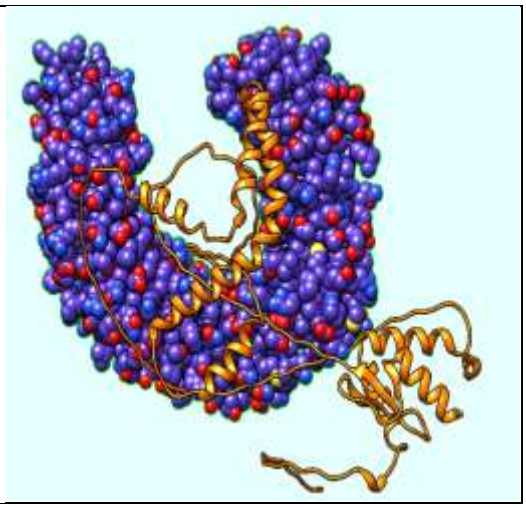


Figure16 : Docked structure. Blue = TLR4, Orange = Refined Ebola Vaccine structure

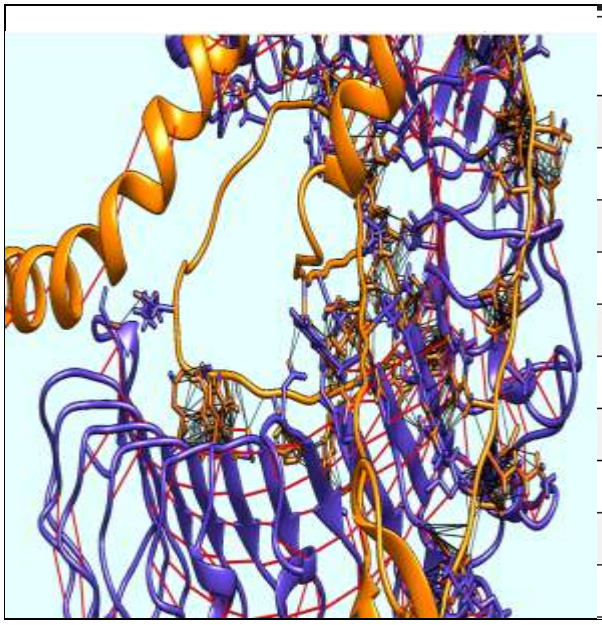


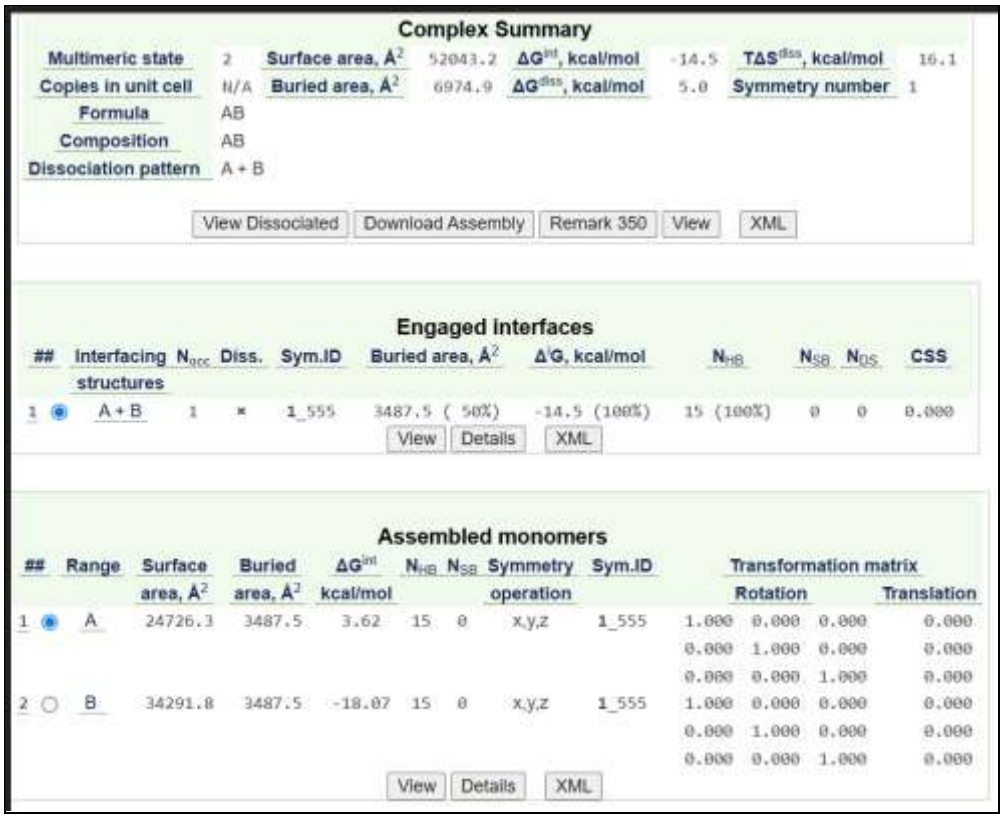
Figure17: Depicts bond interactions within 0.4A

Solution number	Score	Area	ACE
1	18278	4052.3	142.7
2	17236	2395.8	282.78
3	17152	3107.7	226.68
4	17120	2761.5	304.55
5	16750	2584.1	340.58
6	16714	2528.3	63.72
7	16570	2593.1	-236.5
8	16280	2440.7	-21.14
9	16198	2482.7	200.59
10	16136	3228.7	486.96

Figure 18: Shows the results of Docking mechanism , using PATCHDOCK software

2.11 Evaluation of Interaction Analyses

Employing the PISA server , the docking interaction analyses between the vaccine and TLR-4 , is accomplished [13]. This interaction analyses reveal the binding energy to be -14.5 kcal/mol , amongst many other metrics as shown in Figure 19 and Figure 20.



Complex Summary

Multimeric state	2	Surface area, Å²	52043.2	ΔG^int, kcal/mol	-14.5	TΔS^diss, kcal/mol	16.1
Copies in unit cell	N/A	Buried area, Å²	6974.9	ΔG^diss, kcal/mol	5.0	Symmetry number	1
Formula	AB						
Composition	AB						
Dissociation pattern	A + B						

View Dissociated | Download Assembly | Remark 350 | View | XML

Engaged Interfaces

##	Interfacing structures	N_occ	Diss.	Sym.ID	Buried area, Å²	ΔG, kcal/mol	N_HB	N_SB	N_DS	CSS
1	A + B	1	*	1_555	3487.5 (50%)	-14.5 (100%)	15 (100%)	0	0	0.000

View | Details | XML

Assembled monomers

##	Range	Surface area, Å²	Buried area, Å²	ΔG^int kcal/mol	N_HB	N_SB	Symmetry operation	Sym.ID	Transformation matrix: Rotation			Translation
1	A	24726.3	3487.5	3.62	15	0	x,y,z	1_555	1.000	0.000	0.000	0.000
									0.000	1.000	0.000	0.000
									0.000	0.000	1.000	0.000
2	B	34291.8	3487.5	-18.07	15	0	x,y,z	1_555	1.000	0.000	0.000	0.000
									0.000	1.000	0.000	0.000
									0.000	0.000	1.000	0.000

View | Details | XML

Figure 19 : Shows the results of docking interactions using PISA software

Interface Summary XML

	Structure 1		Structure 2	
Selection range	A		B	
class	Protein		Protein	
symmetry operation	x,y,z		,,	
symmetry ID	1_555		0_555	
Number of atoms				
interface	389	8.6%	356	11.9%
surface	2290	50.5%	2389	79.9%
total	4536	100.0%	2989	100.0%
Number of residues				
interface	122	21.4%	74	23.8%
surface	505	88.6%	309	99.4%
total	570	100.0%	311	100.0%
Solvent-accessible area, Å				
interface	3230.6	13.1%	3744.3	10.9%
total	24726.3	100.0%	34291.8	100.0%
Solvation energy, kcal/mol				
isolated structure	-596.7	100.0%	-216.1	100.0%
gain on complex formation	3.6	-0.6%	-18.1	8.4%
average gain	-3.3	0.5%	-24.2	11.2%
P-value	0.865		0.845	

Figure 20 : Depicts the results of docking interactions from PISA software

2.12 Molecular Dynamics Simulations (MD Simulations)

10 ns Md Simulations are performed at 300^{o}K by employing the GROMACS server , in order to ascertain

the structural stability and behavioural conformity of the designed protein vaccine [15] . The simulation confirms thermodynamic stability, with consistent temperature , pressure , density and energy profiles throughout the experimental run. The overall structural intergrity of the multi-epitope vaccine with minor fluctuations, is confirmed by the structural analyses such as , RMSD, RMSF , Radius of Gyration plot, Hydrogen bond occupancy ,SASA and potential energy. Hence, these results confirm favourable dynamic stability in the midst of stimulated physiological conditions.

Figure 21-27 represent the results of GROMACS, MD Simulations . The lattice structure of the Solvation experiment is depicted in Figure 21. As for the interpretation of the results, RMSD plot (Figure 22) shows an increased trend from 0 to 0.8nm , then stable and the last 2ns (0.8-1nm) witnessed small fluctuation. Hence, the overall RMSD plot confirms stability .Again RMSF plot (Figure 23)for per-residue analysis reveals limited fluctuations (<0.3nm) across mot of the backbone , with sharp peaks around residues 100-120, 150-200 and near 260 indicating flexible regions such as surface-exposed loops or termini.For Radius of Gyration (rGyr) , Figure 24 analysis reveals that the rGyr completes its stretch from 3.4nm to 3.6 nm , representing compactness of the vaccine protein. SASA plot (Figure 25) or the Solvent Accessible Surface area plot indicates a distinct decreasing trend ~540 nm^2 to ~495 nm^2 over a fixed time period of upto 6000ps , revealing its hydrophobic tendency and the progressive compaction of the protein multi-epitope based vaccine. The Hydrogen-bonds' occupancy graph (Figure 26) indicate a strong network of hydrogen bonding within its core structure as very small amount of fluctuation (~490–500)is observed throughout the trajectory. The potential energy of the system, as observed from the plot (Figure 27) , stabilized within the first ~10 ps and remains stable around -4.79×10^6 kJ/mol for the remainder of the trajectory, confirming energetic equilibration.

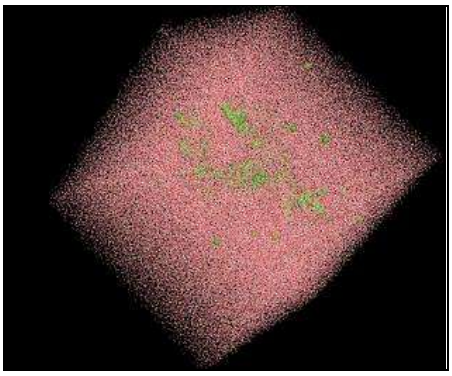


Figure 21 : Depicts the lattice structure in which the vaccine structure is embedded , during the Solvation experiment of MD Simulations

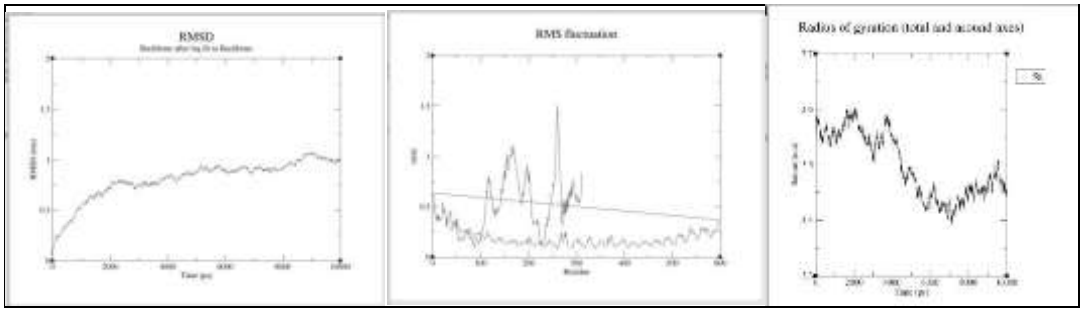


Figure 22 : RMSD Plot Figure 23 : RMSF Plot Figure 24 :Plot showing Radius of Gyration

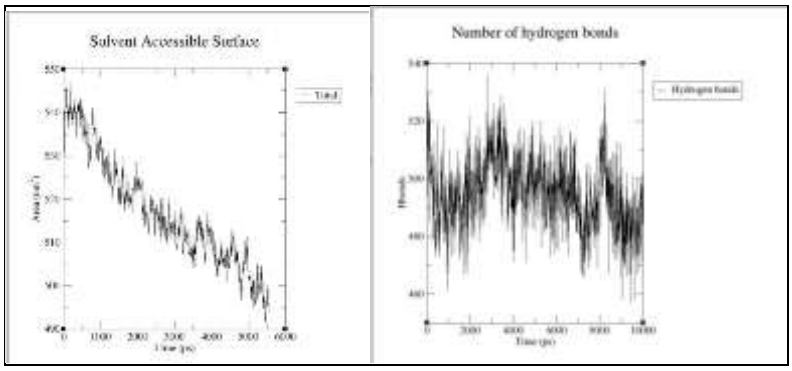


Figure 25 : SASA Plot **Figure 26 : Plot showing H-bonds concentration within vaccine structure**

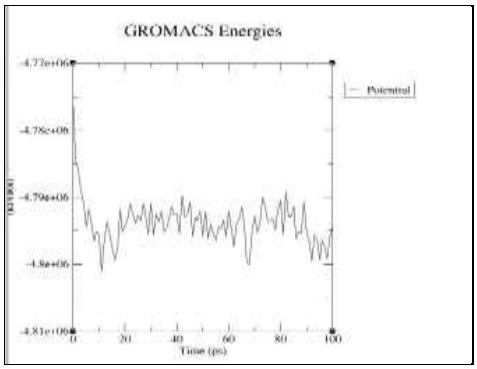


Figure 27 :Potential Energies plot during GROMACS experiments

2.13 Immunology of the Designed Vaccine

C-IMMISIM results have witnessed a robust and coordinated response of the immune system against the predicted antigen, Ebola virus. The results have evidenced the antigen clearance within the first 5 days (Figure 28) , following which the IgM and IgG1 antibody titers have raised-up prominently with their peak volume around 12-15 days , before declining , and simultaneously establishing memory B cells (Figure 29). T-Helper cells' populations increased steadily , with a peak of ~6000 cells/mm³ by day 10–12. On the otherhand, Cytoxicity-Tcells' populations as well as memory TC -cells'populations shows consistent enhancement (Figure 30). Again, the antigen-presenting cells (APCs) , especially the macrophages and dendritic cells represents sound functioning ,including persistent presentation activity throughout the simulation, as it is revealed from the C-IMMISIM results. Cytokine profiling reveals that the IFN-γevelates to maximum level at day 8-10, accompanied by IL-2, TGF-β, and IL-10 elevations, consistent with a Th1-skewed inflammatory response (Figure 31).Many immunodominant epitope candidates are observed using Parker's hydrophilicity scale and MHC class I/II binding prediction , for example, STRGRKCCRRKKEA for B cells; YVQLESRFA and YNTIAGVAG for MHC class II, thereby supporting the epitope's potency and utlity for for successful downstream vaccine design(Figure 32). Again, Figure 33 , shows the molecular mechanisms of immune stimulation by Ebola vaccine injections , as per C-IMMISIM results.

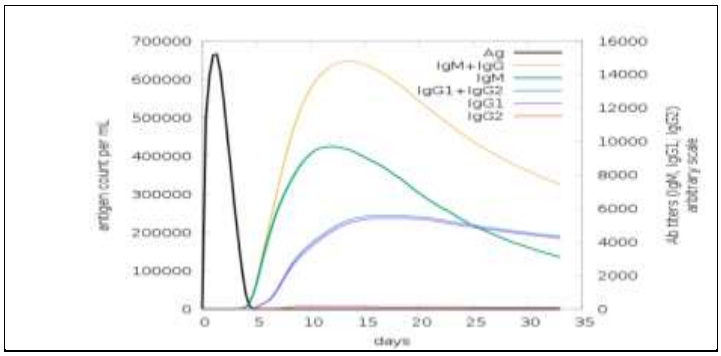


Figure 28: Showing antigen clearance within 5 days

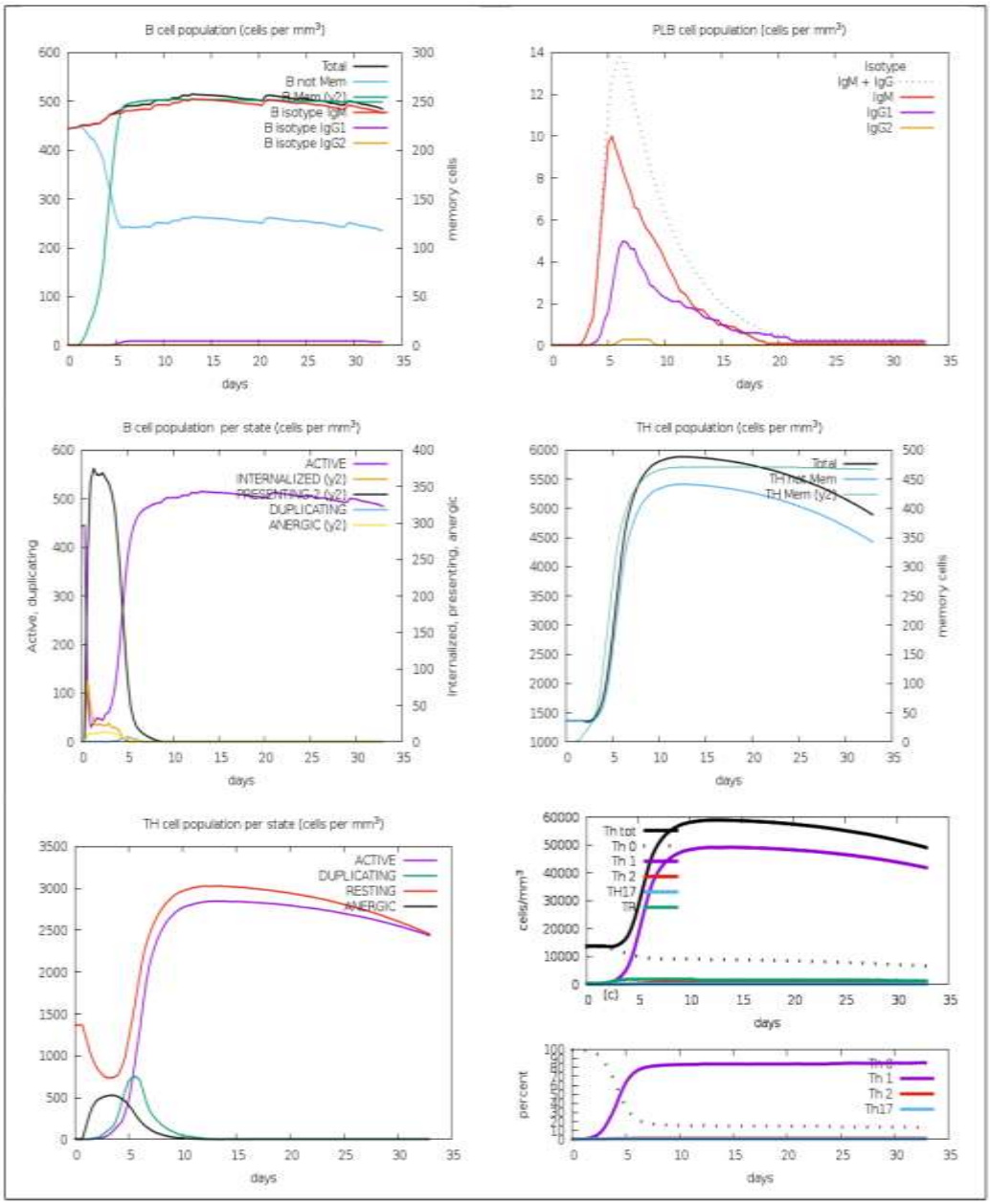


Figure 29: C-IMMISIM results showing B-cells' activation and memory B-cells' proliferation

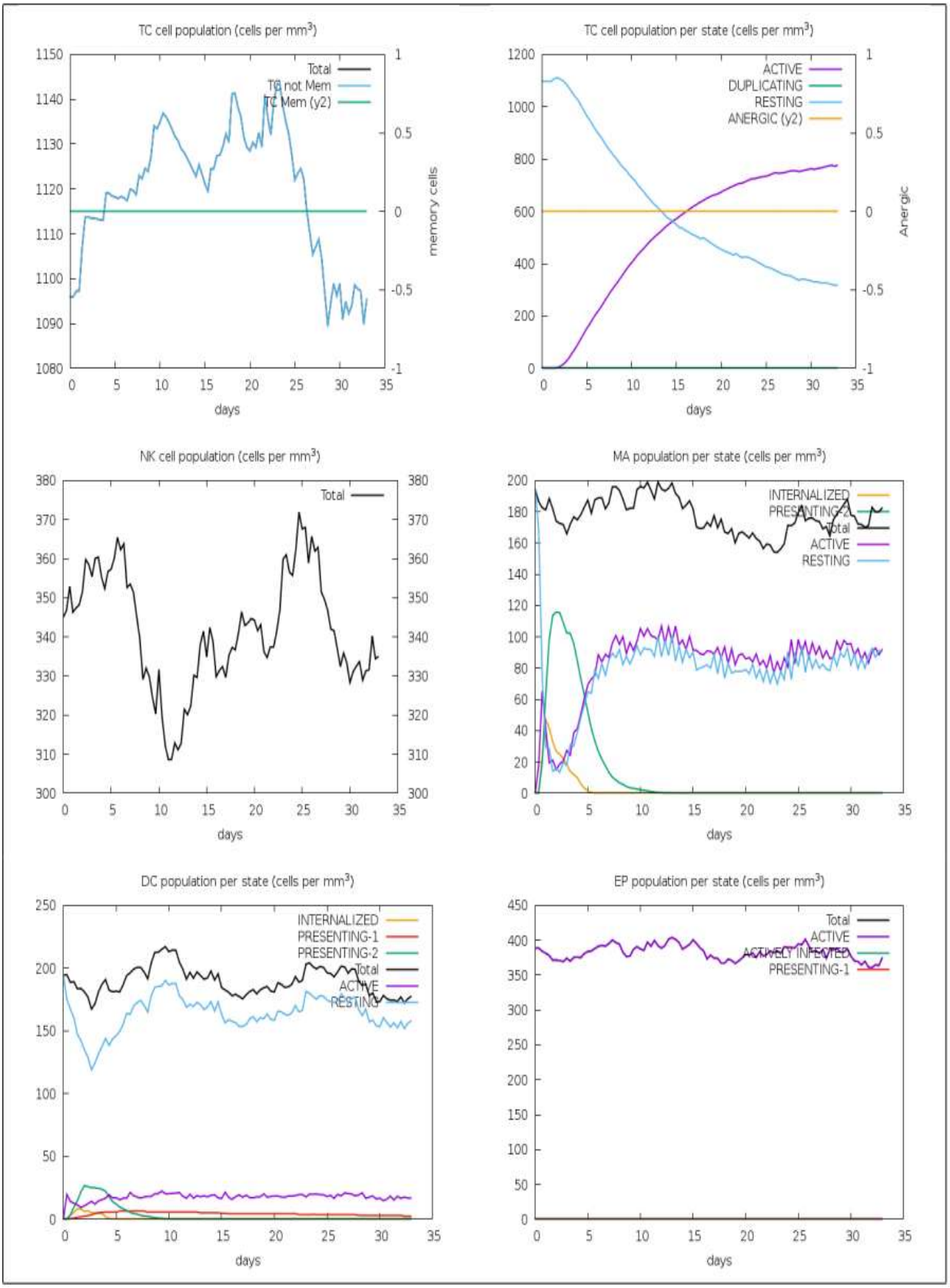


Figure 30 : C-IMMISIM results showing T-cells activation and memory T-cells proliferation

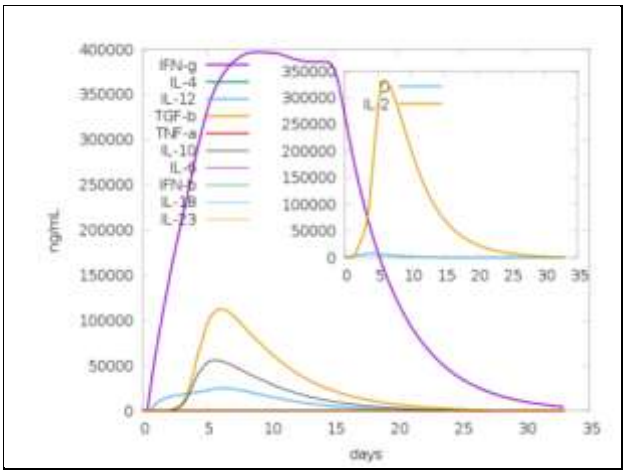


Figure 31 : Cytokine & Inerleukin Profile

```
Parker_B_scale_epitope_find
Use Parker's propensity scale, takes an antigen block as input,
and creates a list of residues that are possible epitopes.


---------------------------------
Antigen sequence file: /opt/lampp/htdocs/C-IMMSIM/Jobs/input/42642_20260708-052241_5_ruDxQECvH2BN.FSA_1_001
---------------------------------


GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAKWVIILFQRTFSIPLGGPGPGPDGIRGFPRC
RYVHKGPGPGQETITETTATIIGTNAAYRTFSILNRKAAYTTIGEWAFWAAYTYVQLESRFAAYNTIAGVAGLAAYGFRS
GVPPKAAYRTYTILNRKAAYFPIPLGVVHAAYILGSLGLRKAAYSPRPVPTSTAAYKVFPIPLGVAAYKKTLQVSDVDKL
VCRDKLSSTNQLKKWGFRSGVPPKVVNYEAGEWAENCKKNFTKTLSSEELSVILVPRAQDPGGGSHHHHHH

00000000000000000000000000011110011111111111111000000000000000000011100100000000
00000011110011100000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000010000
00000111011000000000000000000000000011001000110000000000001111111100000
Epitopes of protein 0 ------------
1]        pos=27 len=4        EQIG
2]        pos=33 len=14        STRGRKCCRRKKEA
3]        pos=86 len=4        PGPG
4]        pos=298 len=8        QDPGGGSH
```

Figure 32 : Showing list of epitopes that can block Th1-skewed inflammatory response

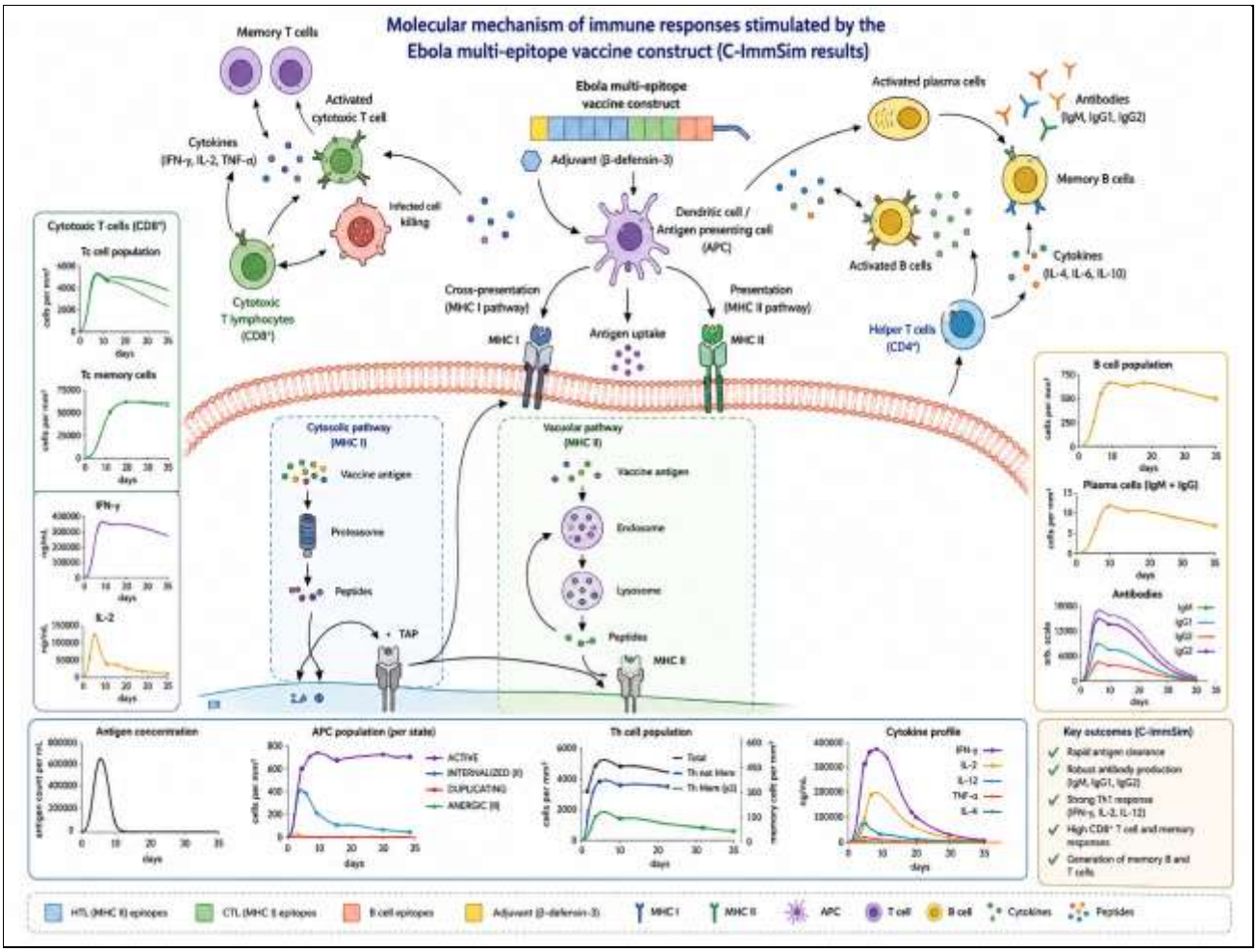


Figure 33 : Depicts the molecular Mechanism of immune responses as stimulated by the Ebola vaccine , as per C-IMMISIM results

2.14 Codon Optimization for the designed Vaccine

Careful strategy is devised not only for vaccine construction , but also for the expression system , applicable for human injections. Hence, for this purpose, JCAT, Codon Optimization tool is employed to obtain optimized sequence for *Escherichia coli* (strain ATCC 27325 / DSM 5911 / W3110 / K12). A score (CAI) of 0.957, highly favourable for E.coli expression is thus obtained. GC content of the codon is 52.95% , closely matching to GC content of the E.coli (50.80%), minimizing tanslational inefficiencies.

2.15 Plasmid Construction

The SnapGene software is employed for the purpose and the gene of interest containing the vaccine construct(Figure 35) is inserted into the plasmid, pET28a(+) ,as in Figure 34, using restriction enzymes such as NdeI and XhoI . The final construct (chimeric molecule) resulted in 6099bps (Figure 36). The vaccine insert is present downstream of T7promoter and His-tag,enabling IPTG-inducible protein expression with C-terminal purification capability.

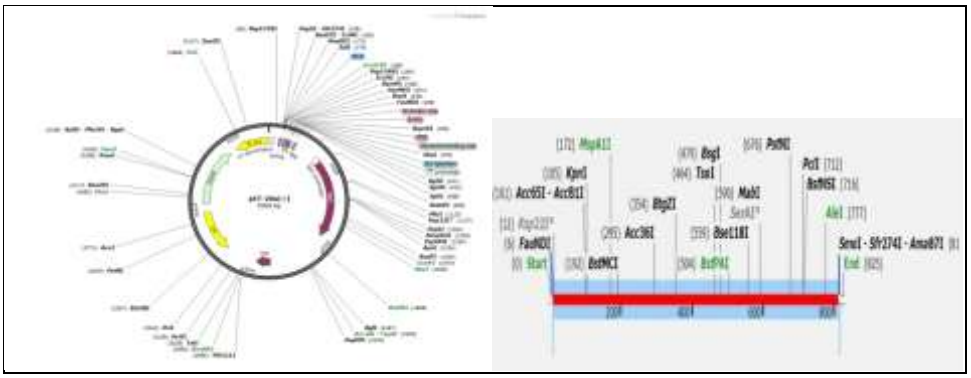


Figure 34 : pET28a(+) plasmid used for cloning

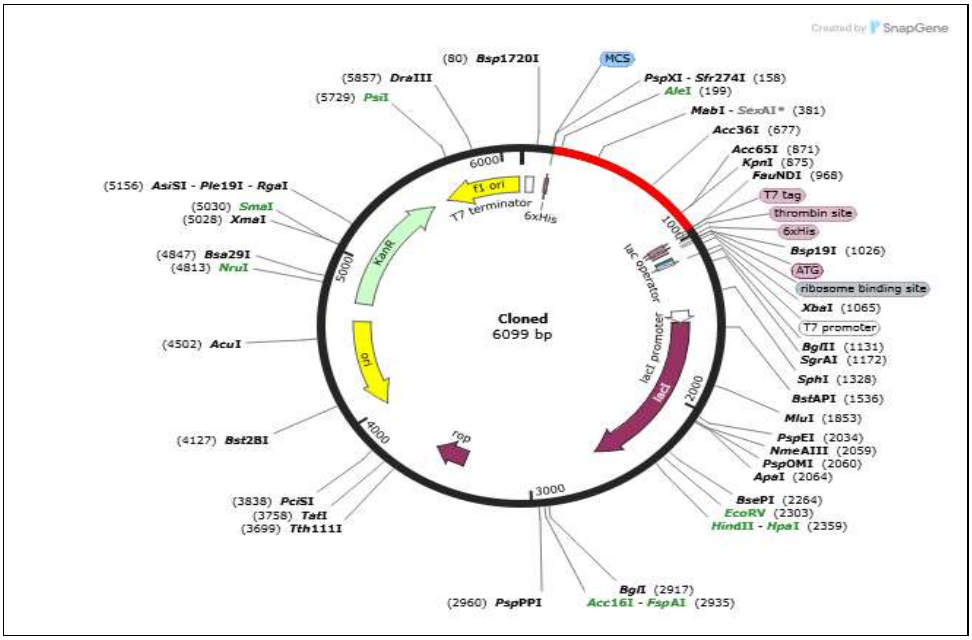


Figure 35 : Gene of interest with restriction sites , containing the vaccine structure

Figure 36: Cloned Plasmid with gene of interest (red colour) , containing the vaccine structure and His6 tag as easy protein-expression system.

3. Methodologies

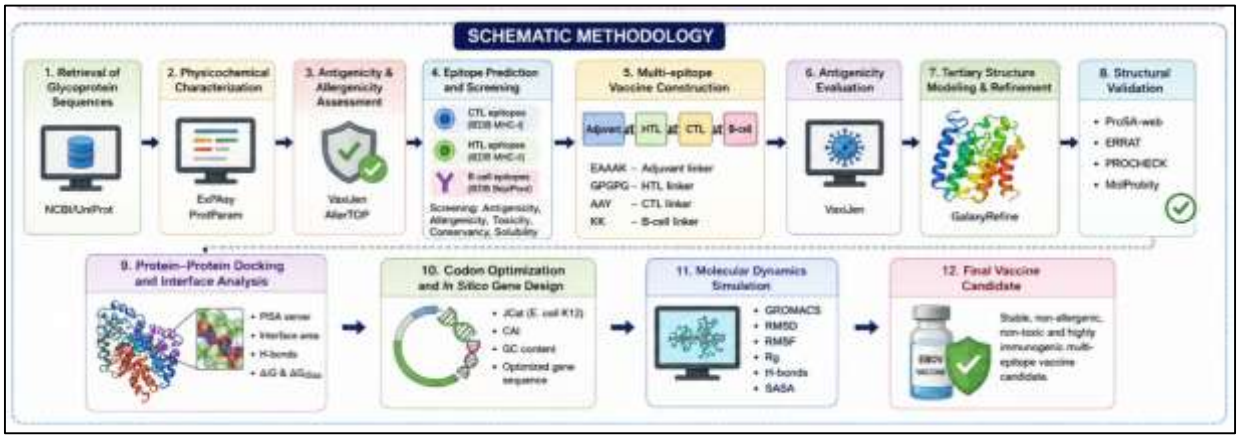


Figure 37 : Schematic diagram illustrating the processes involved in Methodologies

3.1 Sequences' Retrieval and Prioritization

As a solution to combat Ebola virus , this study aims to construct and validate a multi-epitope vaccine , with its epitopes representing all three viral strains - Zaire, Sudan and Bundibugyo.

Initially, the protein sequences of these strains , are derived from NCBI/UniProt alongwith their Accession numbers in FASTA format. These sequences are then analyzed for physico-chemical properties (ExPasy Protparam), allergenicity assessment (Allertop), antigenicity (Vaxijen) ,toxicity (Toxinpred) and solubility (Soluprot).

3.2 Epitopes' Selection for Vaccine Preparation

Protein epitopes are selected with the help of IEDB tool. Three sets of epitopes are obtained ,namely , B-cell epitopes, Cytotoxic -Tcell epitopes (CTL) and Helper T-cell epitopes (HTL) . Again, a big list of epitopes from each of the three categories , the epitopes are funnelized to a handful epitopes , based on stringent analyses - physico-chemical properties (threshold 0.4) , allergenicity , antigenicity ,toxicity, conservancy and solubility. In the case of CTL, 9mer sized-epitopes are constructed and prioritized while for HTL the size are 15mer .

3.3 Multi-epitope Ebola Vaccine Constrution

Ebola vaccine construct is processed by incorporating all the selected B-cell , CTL and HTL epitopes and linking them with suitable linker sequences - GPGPG linkers for HTL, AAY linkers for CTL and KK linker for B-cell epitopes. This constitutes the main body of the vaccine construct. Then, at the N-terminal , Beta-defensin-3 sequence is added while at the C- terminal His6 tag are added. The final construct counts to 311 amino acids.

Subsequent to the vaccine contruction, the vaccine construct sequence is subjected to the same set of analyses- physico-chemical properties (ExPasy Protparam), allergenicity assessment (Allertop), antigenicity (Vaxijen) ,toxicity (Toxinpred) and solubility (Soluprot).

3.4 Secondary and Tertiary Structure Prediction of the Vaccine Construct

The Secondary structure of the vaccine construct , predicted with the Psipred tool , reveals the regions of strandedness (high confidence) ,helices and coils . The Tertiary structure is assessed using ColabFold (AlphaFold2). Subsequent to these predictions, the vaccine protein structure is further refined and all other downstream analyses are done.

3.5 Refinement of Vaccine construct

The protein vaccine structure is further refined using GalaxyRefine server. The refinement is performed through repeated side-chain rebuildingand molecular dynamics relaxation procedures. Multiple refined models are generated and the quality is assessed using Global Distance Test High Accuracy (GDT-HA), Root Mean Squre Diviation (RMSD) , MolProbity score , Clash score percentage of poor rotamers and most importantly Ramachandran favoured regions. Based on these criteria , a particular Model is chosen for all further downstream analyses.

3.6 Validation of the Vaccine Construct

The refined vaccine model is validated using PROCHECK, MolProbity and ProSA web server tools. The overall stereochemical assessment is done using PROCHECK, Ramachandran plot is obtained using MolProbity and through Pro-SA tool, suitable Z-score is obtained , which is one of the quality parameters. Thus, the refined model structure with high order of quality checks like high percentage of Ramachandran favoured regions , are considered for further analyses.

3.7 Molecular Dynamics Simulations (MD Simulations)

The structural stability and compactness of the vaccine structure is assessed using GROMCS, tool for

Molecular Dynamics Simulations. Subsequent to energy minimization , equilibration is carried out under constant volume (NVT) and constant pressure (NPT) conditions. The production MD is then carried out in explicit solvent environment (cubic simulation box with diamensions14.59 × 14.59 × 14.59 nm^3) . The simulations are carried for 5,000,000 steps (10ns) with trajectories over 50,001 frames, under 300 K (300.09 K average) and pressure at ~1 bar using appropriate thermostat/barostat coupling. The simulation is run on 1 MPI rank with 12 OpenMP threads, achieving a performance of 1.181 ns/day, with a total wall-clock time of ~8.5 days (731,729.9 s).

As results of the MD Simulations Root Mean Square Deviation (RMSD) plot is obtained revealing the structural stability of the vaccine construct , Root Mean Square Fluctuations(RMSF) reveals the residue-wise flexibility , Radius of Gyration (rGyr) plot signifies compactness of the vaccine construct , Solvent Accessibility Surface Area (SASA) reveals its hydrophobicity nature , important for the structural integrity of the vaccine construct.

3.8 Docking and Interaction Analyses

Protein-protein docking mechanism is performed with Immune receptor, TLR4 (PDB ID-2Z63) and the refined , evaluated vaccine construct as ligand , using the tool PATCHDOCK. The best docked complex based on binding affinity scores , is considered for revealing the interaction characteristics with the tool named PISA.

Thus, PISA is employed to reveal its interaction area, buried surface area, hydrogen bond interactions, and Gibbs' free energy associations.

3.9 Immune Simmulations

The constructed Ebola vaccine is subjected to Immune simulations by C-IMMISIM, which reveals the potency of the vaccine to arouse humoral and cell mediated immunity upon Ebola virus exposure. In this simulation, the role of the vaccine in activating B-cells, CTL , HTL , macrophages, dendritic cells, NK cells, epithelial cells , antibody titers , memory cells propagation , antigen clearance kinetics, and cytokine/interleukin profiles (IFN-γ, IL-2, IL-4, IL-10, IL-12, TNF-α, TGF-β, among others) , are revealed.

As a means of input to start the experiments, three injections at time of delivery -1, 84, 168 days respectively and Simulation steps=1100, are prioritized as important criteria for starting the experiment.

3.10 Codon Optimization and Cloning

Initially , with the vaccine construct as input , the JCAT tool is employed for codon -optimization and the desired gene of interest is achieved . The CAI (Codon Adaptation Index) score obtained is also desirable. The gene of interest is then cloned into excised pET28a(+) plasmid , with the restriction enzymes -NdeI and XhoI , as per literature study. The resulting chimeric molecule is of 6099bps , achieved by the SnapGene software.

4. Conclusion

In this study, immunoinformatics-based software are employed for the development of potential vaccine [16-25]. The vaccine candidate achieves a global population coverage of 87.44% (Figure 36), when the analyses are done using IEDB tool for MHCI and MHCII combined. Again, when population coverage analyses are performed for West Africa , East Africa and Central Africa . approximately 70% coverage is achieved (Figures 38-41) . As the prognosis is maximum in these regions , emphasis is given for the effectiveness of the vaccine candidate , for these parts of the world. The hits range from 1.32 to 1.39 and PC90 values of 0.33, are achieved.

As these results are software-based, in-vivo and in-vitro lab experiments are the ‘need of the hour’ , for a desirable fruition.

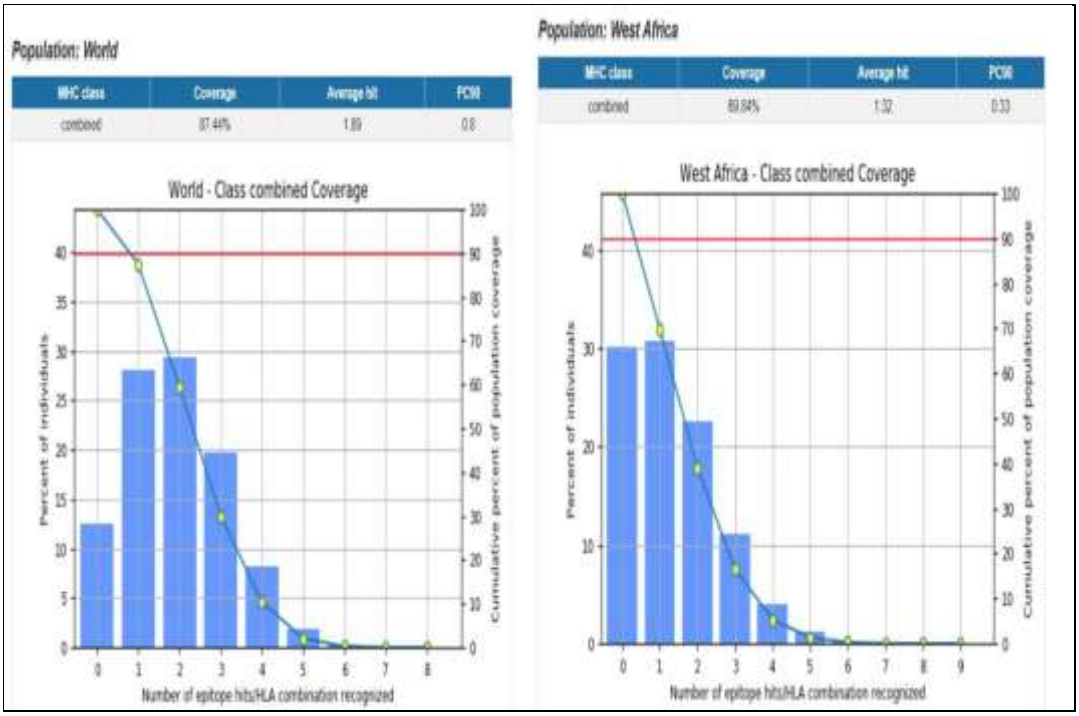


Figure 38 : World Population Coverage

Figure 39: Population Coverage (West Africa)

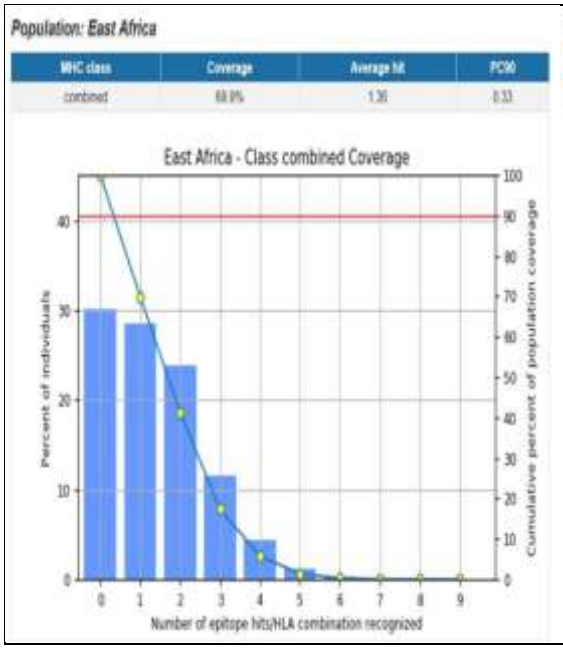


Figure 40: Population Coverage (East Africa)

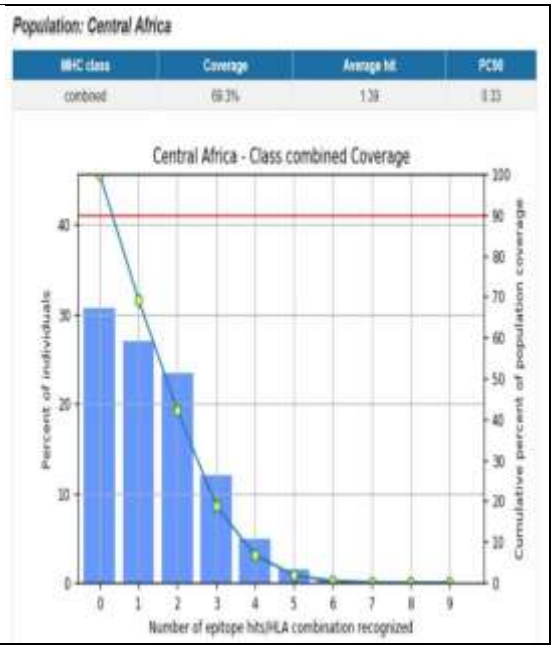


Figure 41: Population Coverage (Central Africa)

Acknowledgement

I am grateful to Prof Dr Hammad Saleem (Karachi, Pakistan) , for guiding me to complete this project. I am also grateful to my institution for allowing me to complete this work . Moreover, the Blessings of my parents and The Almighty gave me the right inspiration to complete this project.

5. References

1. Feldmann H, Geisbert TW. Ebola haemorrhagic fever. *Lancet*. 2011;377(9768):849–862.
2. Kuhn JH, Amarasinghe GK, Perry DL, et al. ICTV Virus Taxonomy Profile: Filoviridae. *J Gen Virol*. 2019;100(6):911–912.
3. Henao-Restrepo AM, Camacho A, Longini IM, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease. *Lancet*. 2017;389(10068):505–518.
4. Vita R, Mahajan S, Overton JA, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47(D1):D339–D343.
5. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*. 2007;8:4.
6. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2—a server for in silico prediction of allergens. *J Mol Model*. 2014;20(6):2278.
7. Gupta S, Kapoor P, Chaudhary K, et al. In silico approach for predicting toxicity of peptides and proteins. *PLoS One*. 2013;8(9):e73957.
8. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*. Humana Press; 2005:571–607.
9. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res*. 2012;40(Web Server issue):W294–W297.
10. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35(Web Server issue):W407–W410.
11. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26(2):283–291.
12. Chen VB, Arendall WB III, Headd JJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 1):12–21.
13. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774–797.
14. Grote A, Hiller K, Scheer M, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*. 2005;33(Web Server issue):W526–W531.
15. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1–2:19–25.
16. Malik S, Gupta N, Shukla R, et al. Ebola Virus Disease Vaccines: Development, Current Perspectives and Challenges. *Vaccines*. 2023;11(2):268.
17. Alizadeh M, Amini E, Khoei SG, et al. Designing a novel multi-epitope vaccine against Ebola virus using immunoinformatics and reverse vaccinology approaches. *Scientific Reports*. 2022;12:7602.
18. Khan MA, Hossain MU, Rakib-Uz-Zaman SM, Morshed MN. Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study. *F1000Research*. 2015;4:365.
19. Henao-Restrepo AM, Longini IM, Egger M, et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination trial. *Lancet*. 2017;389:505–518.
20. Regules JA, Beigel JH, Paolino KM, et al. A recombinant vesicular stomatitis virus Ebola vaccine. *New England Journal of Medicine*. 2017;376:330–341.

21. Kennedy SB, Bolay F, Kieh M, et al. Phase 2 placebo-controlled trial of two vaccines to prevent Ebola in Liberia. *New England Journal of Medicine*. 2017;377:1438–1447.
22. Marzi A, Feldmann H. Ebola virus vaccines: an overview of current approaches. *Expert Review of Vaccines*. 2014;13(4):521–531.
23. Basmenj ER, Pajhouh SR, Fallah AE, et al. Computational epitope-based vaccine design with bioinformatics approaches: a review. *Heliyon*. 2025.
24. De Groot AS, Rappuoli R. Genome-derived vaccines. *Expert Review of Vaccines*. 2004;3(1):59–76.
25. Flower DR. Immunoinformatics and the in silico prediction of immunogenicity. *Current Pharmaceutical Design*. 2008;14(26):2679–2683.
26. Tomar N, De RK. Immunoinformatics: an integrated scenario. *Immunology*. 2010;131(2):153–168.
27. He Y, Xiang Z. Bioinformatics analysis of vaccine targets and vaccine design. *Journal of Biomedical Biotechnology*. 2013;2013:864806.
28. Rappuoli R. Reverse vaccinology: a genome-based approach to vaccine development. *Vaccine*. 2001;19(17–19):2688–2691.
29. Sette A, Rappuoli R. Reverse vaccinology: developing vaccines in the era of genomics. *Immunity*. 2010;33(4):530–541.
30. Poland GA, Ovsyannikova IG, Kennedy RB. Personalized vaccinology: a review. *Vaccine*. 2018;36(36):5350–5357.