European Journal of Biotechnology and Bioscience
ISSN: 2321-9122; Impact Factor: RJIF 5.44
Received: 21-06-2019; Accepted: 24-07-2019
www.biosciencejournals.com
Volume 7; Issue 5; September 2019; Page No. 51-59

T-cell epitope prediction and its interaction with MHC class i and class ii for ebola virus vaccine

Priyanka S Bharadwaj
BMS College of Engineering, Dept. of Biotechnology, Bangalore, Karnataka, India

Abstract
Vaccines are preventive pharmaceutical interventions that prepares the body for pathogens, making the immune system ready for future infections. It is highly cost-effective in comparison to secondary preventive measures. In pharmaceutical industry, product development and production of the same is normally very expensive and consumes a large time frame to achieve the same. Approaches focused on selection of appropriate antigenic structures, carriers and adjuvants are used to reduce the cost and time frame. For this matter, methodologies on bioinformatics can be utilized for vaccine development. Using bioinformatics with the knowledge of immunology gave rise to a new field known as immunoinformatics. This study involves the utilization of various bioinformatics tools and software, such as Jal View, IEDB Servers, and docking tools. The knowledge of reverse vaccinology is used to obtain a synthetic vaccine against EBOV nucleoprotein. Out of the 6 strains of Ebola virus present, only 4 are capable of infecting humans (Zaire ebola virus, Sudan ebola virus, Tai Forest ebola virus and Bundibugyo ebola virus). Their nucleoprotein sequences were retrieved from NCBI Protein database and their consensus sequence was obtained using Jal View software. Further, the binding MHC molecules and the epitopes were predicted. Using the tool Hex docking was done to fit the binding energy.

Keywords: vaccine, ebola virus, immunoinformatics, bioinformatics, binding MHC molecules

1. Introduction
Ebola virus is a class of virus that affects humans, non-human primates and pigs (Weingartl *et al*., 2012) [12]. It was first discovered at the banks of Ebola River, Zaire (now the Republic of Congo) and was hence named after the river. Since its first outbreak in 1976 and the recent 2013 to 2015 epidemic in West Africa, the fatality rate has been increasing at an alarming rate of 90% (Coltart *et al*., 2017) [3]. Because of its high mortality rate (up to 83-90%), EBOV is also listed as a select agent, World Health Organization Risk Group 4 Pathogen (requiring Biosafety level 4-equivalent containment), a U.S. National Institutes of Health / National Institute of Allergy and Infectious Diseases Category A Priority Pathogen, U.S. CDC Centers for Disease Control and Prevention Category A Bioterrorism agent, and listed as a Biological Agent for Export Control by the Australia Group (World Health Organization, 2014). Hence, developing a vaccine against this deadly disease should be of high priority. This fast rate of infection warrants further investigation of its vaccine for preventing further infections.

Vaccines are the pharmaceutical products that offer the best cost-benefit ratio in the prevention and treatment of diseases (Ehreth, 2003) [5]. Development of pharmaceutical products have been expensive and take years to accomplish the life cycle from development to production. A novel approach based on bioinformatics can be utilized to analyze and develop future vaccines (Scarselli *et al*., 2005) [9]. Using bioinformatics with the knowledge of immunology gives rise to a new field known as immunoinformatics. The birth of immunoinformatics lead to surge in in-silico vaccine and drug development and their design. This pre-process cuts short the usual 10 to 15 years of time period for vaccine development studies to a shorter period of 8 to 10 years (He *et al*., 2010) [6]. This work involves the application of comparative proteomics in the design and development of the vaccine against Ebola virus.

In this study, reverse vaccinology will be applied to create a potential formulation against Ebola virus using bioinformatics programs like NCBI Blast, JalView, IEDB servers, and docking programs for the synthetic vaccine. This study focuses only on the 4 out of the 6 strains of Ebola virus; these 4 are capable of infecting humans (Zaire ebola virus, Sudan ebola virus, Tai Forest ebola virus and Bundibugyo ebola virus). Most vaccines currently cultivated for Ebola virus have been developed based on B-cell immunity but this study mainly focused on T-cell immunity responses. The predicted epitopes are expected to give protective immunity against all four strains of Ebola virus that infect humans.

2. Material and method
A series of nucleoprotein sequencing, alignment and rendering programs will be used to generate the consensus sequence from the four known strains of ebola virus that are infective of humans. From these, the structures will be predicted and will include a model of protein structure expected to be produced as the main subunit for the Ebola vaccine.

2.1 National center for biotechnology information (NCBI)
The nucleoprotein sequences will be obtained from the National Center for Biotechnology Information (NCBI), which provides the biomedical and genomic information needed for this study. All the 4 strains of Ebola virus that are infective to humans are obtained here.

2.2 Clustal omega EMBL
Multiple sequence alignment is done with the Clustal

Omega, which utilizes seeded guide trees and HMM profile-profile techniques and then further renders the needed alignments for at least three sequences (Madeira *et al*, 2019). For this study the 4 sequences will be generated and the result is viewed in JalView.

2.3 JalView

The resulting sequences must be visualized and analyzed with the use of the Jal view program. The consensus sequence is obtained from the four Ebola virus nucleoprotein sequences. From this, the phylogenetic tree is calculated using Neighbor-Joining method (NJ method).

2.4 MHC class I binding predictions

The determination of the Class I MHC binding epitopes prediction was done using IEDB (Immune Epitope Database) MHC I prediction tool. For this prediction several HLA alleles were considered like HLA-A*03:01, HLA-A*68:01, HLA-B*40:01. The prediction was achieved by 'consensus' option. Among the obtained epitopes, the ones with rank percentile less than 0.2 were selected for further analysis.

2.5 MHC class II binding predictions

The determination of class II MHC binding epitopes was done using IEDB MHC Class II prediction tool. The reference set of alleles were HLA-DRB1*09:01, HLA-DPA1*01/DPB1*04:01, HLA-DRB5*01:01 etc. The prediction was done by 'consensus' method. All epitopes that bind to the alleles with rank percentile less than 0.2 were considered for further analysis. (Low percentile rank implies good binders)

2.6 Homology modelling

The complete 3-dimensional structure of the protein was obtained by Phyre2 tool, which uses homology modelling methods to predict a 3D structure of the protein molecule. To view the results, RasWin viewer was used. Two alleles of each MHC Class I and class II were selected. Their protein sequences were found using NCBI (protein). Using PDB, their structures were found and were viewed in RasWin.

2.7 Docking

The consensus protein was docked with each of the selected four HLA alleles (Table 1). Docking was carried out by a tool called Hex 8.0 and the model was visualized by RasWin. The E total, RMS, E Shape values were tabulated.

3. Results

Using the series of molecular sequencing and analysis programs, the needed nucleoprotein sequences and corresponding consensus sequence were obtained, 3D structure of the protein molecule was predicted and the obtained model has undergone docking. The sequences, structures and models are all documented here.

3.1 Retrieving the nucleoprotein sequences of the 4 strains of human-infective ebola virus from NCBI protein database

An Ebola virus Database (EDB) was devised by the NCBI reference sequences for documentation of the actual sequence of the viral strains. This helps in determining the core genomes and necessary point within the gene for future researches. Here are the sequences of the four strains of Ebola virus that are infective of humans.

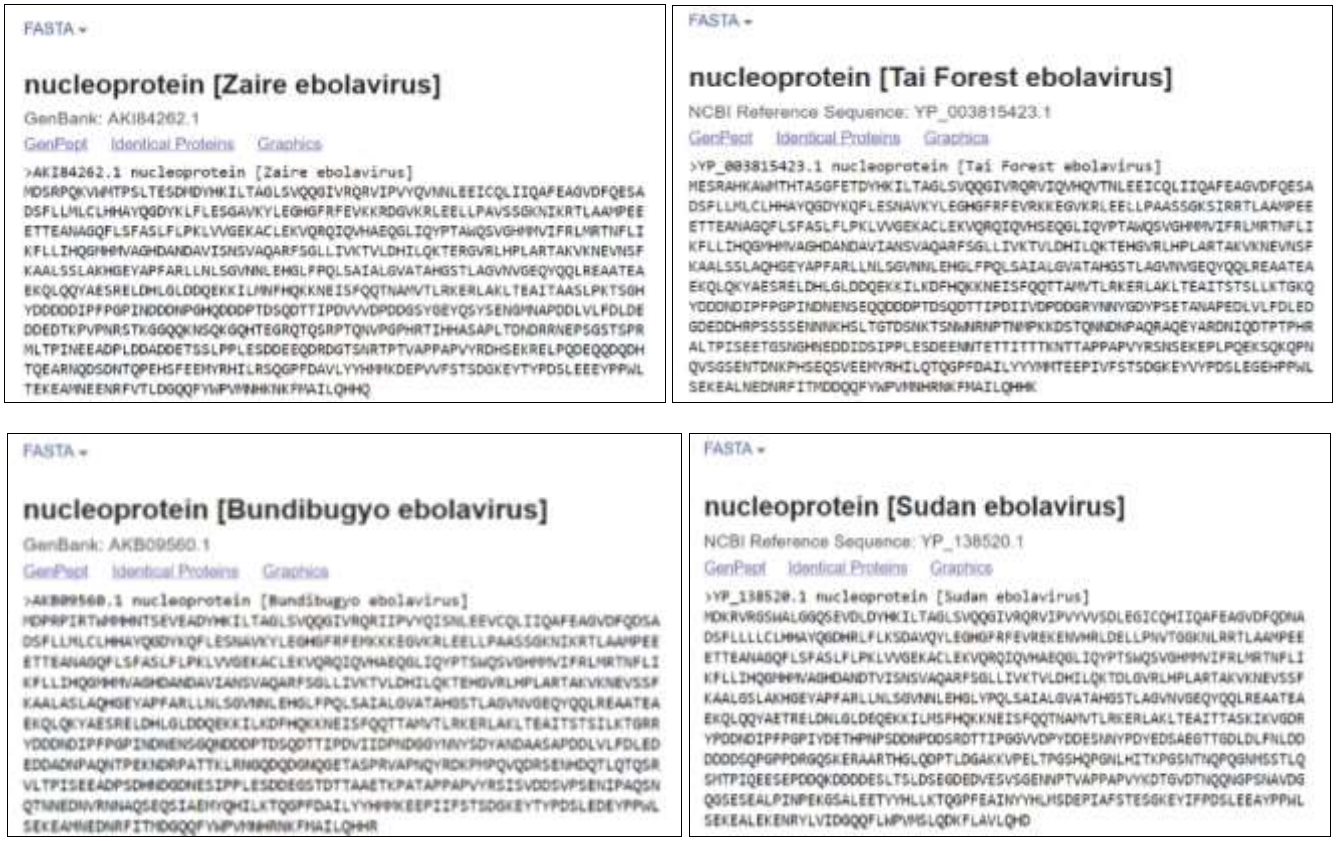


Fig 1: The retrieved sequences of all strains nucleoprotein sequences.

3.2 Alignment of the sequences using multiple sequence alignment tool

From the retrieved nucleoprotein sequences of ebola virus, an alignment has been done utilizing the Clustal Omega Multiple Sequence Alignment Tool.

Clustal omega

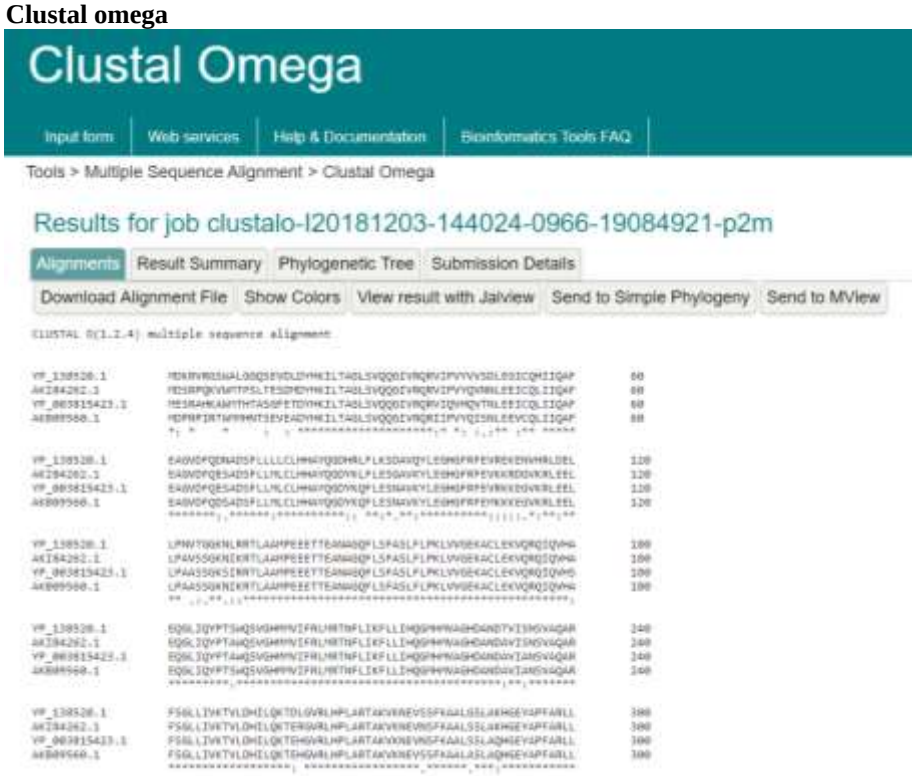


Fig 2: EMBL page for MSA

3.3 Alignment file in FASTA format is obtained and results viewed in JalView.

The obtained alignment is now downloaded in the text-based format of FASTA and is to be uploaded to the JalView software for visualization.

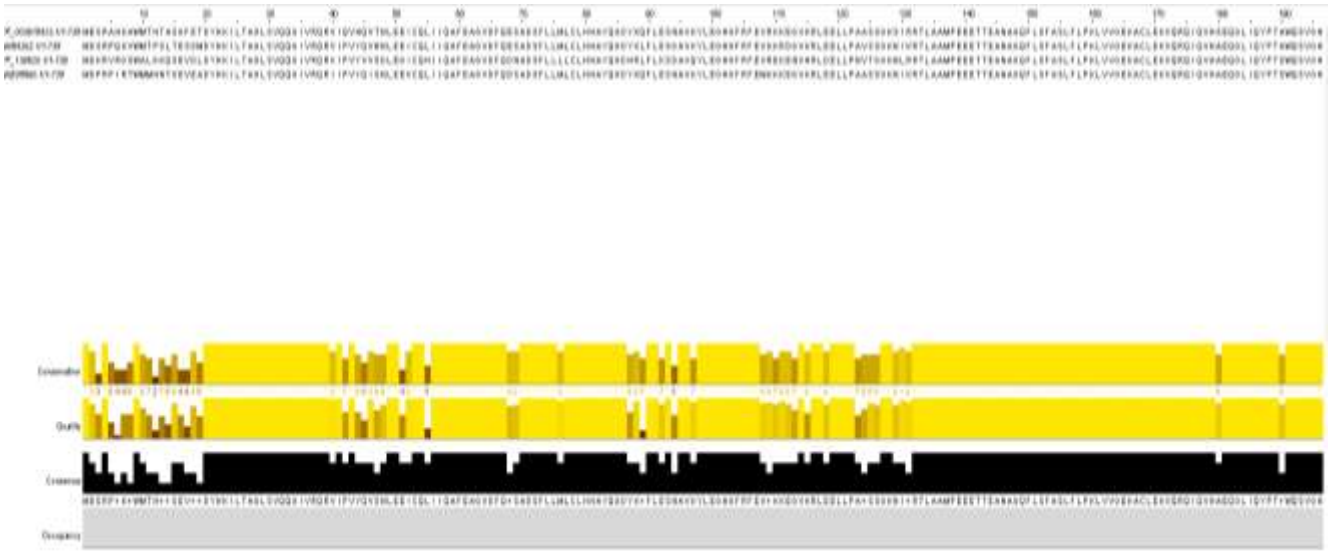


Fig 3: Output of Jal View

3.4 Consensus sequence obtained by taking percent identity 80.

Through the alignments processed, the sequences are processed to generate the consensus sequence to depict the most frequent nucleotide and amino acid residues that appear within the sequence.

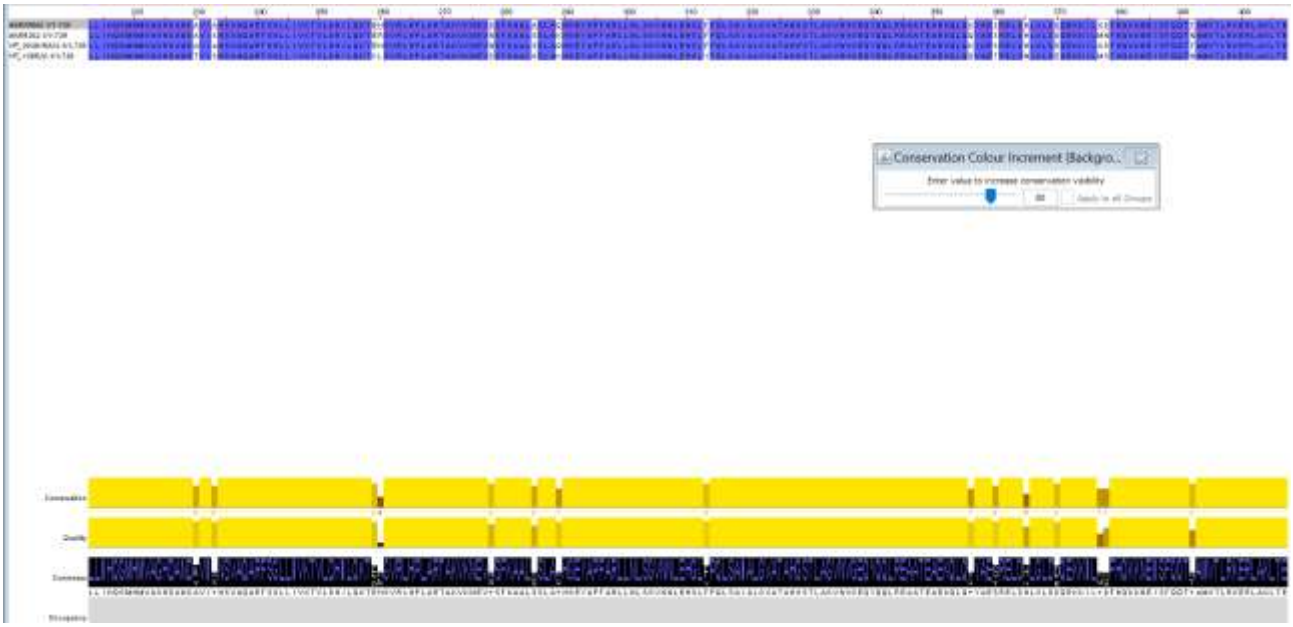


Fig 4: Consensus sequence

The consensus sequence obtained is encoded as follows:
MDSRPHKAWMTHGASEVDADYHKILTAGLSVQQGI
VRQRVIPVYQVSNLEEICQLIIQAFEAGVDFQDSADSF
LLMLCLHHAYQGDYKLFLESNAVKYLEGHGFRFEVK
KKEGVKRLEELLPAASSGKNIKRTLAAMPEEETTEAN
AGQFLSFASLFLPKLVVGEKACLEKVQRQIQVHAEQ
GLIQYPTAWQSVGHMMVIFRLMRTNFLIKFLLIHQG
MHMVAGHDANDAVIANSVAQARFSGLLIVKTVLDHI
LQKTEHGVRLHPLARTAKVKNEVNSFKAALSSLAKH
GEYAPFARLLNLSGVNNLEHGLFPQLSAIALGVATAH
GSTLAGVNVGEQYQQLREAATEAEKQLQKYAESREL
DHLGLDDQEKKILKDFHQKKNEISFQQTNAMVTLRK
ERLAKLTEAITSASLLKTGDRYDDDNDIPFPGPINDNE
NPGQQDDDPTDSQDTTIPDVIVDPDDGEYNNYGDYA
DDAANAPDDLVLFDLDDDDDDDHPAPNDSEKNDEH
AATGLDHGQDDDGNQGATAMPEKAPTQNNDGPHRT
IHHAAAKAQDNARDNDQDGSTPHRALTPISEEADPD
DHNDDDDESIPPLESDDEEDTETTAAETKPTTAPPAPV
YRDHSEDDELPQEESQAQDNQGSASENDDNNPQSEQ
SFEEMYRHILKTQGPFDAILYYHMMKDEPIVFSTSDG
KEYTYPDSLEEEYPPWLSEKEALNEDNRFITMDGQQF
YWPVMNHRNKFMAILQHHK

3.5 Phylogenetic Tree was obtained via JalView calculation NJ using BLOSUM62

From the consensus sequence obtained, a phylogenetic tree was generated using the Neighbor-Joining method of the JalView software. The expected phylogenetic tree is calculated with the substitution matrix BLOSUM62 to identify the evolutionarily similar protein sequences within the set of sequences.

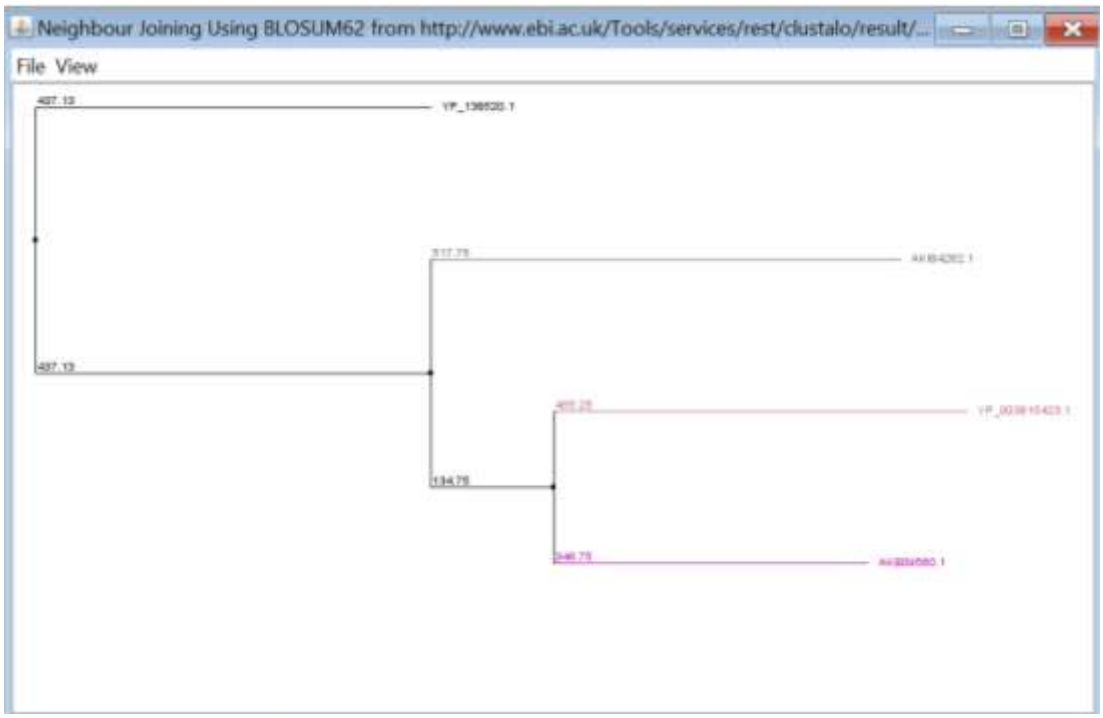


Fig 5: Phylogenetic Tree using NJ method (using BLOSUM62)

Where,
YP_138520.1 nucleoprotein [Sudan ebola virus]
AKB09560.1 nucleoprotein [Bundibugyo ebola virus]
AKI84262.1 nucleoprotein [Zaire ebola virus]
YP_003815423.1 nucleoprotein [Tai Forest ebola virus]

3.6 The structure of the consensus sequence was predicted using Phyre2 with 52% coverage and 100% confidence.

The protein structure prediction was done using the Protein Homology/Analogy Recognition Engine (Phyre) 2, to visualize the shape of the protein and generate a model of it.

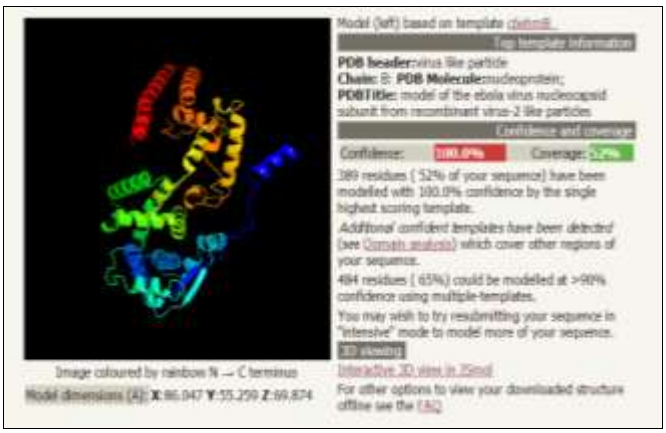


Fig 6: Homology modelling of consensus sequence using Phyre2

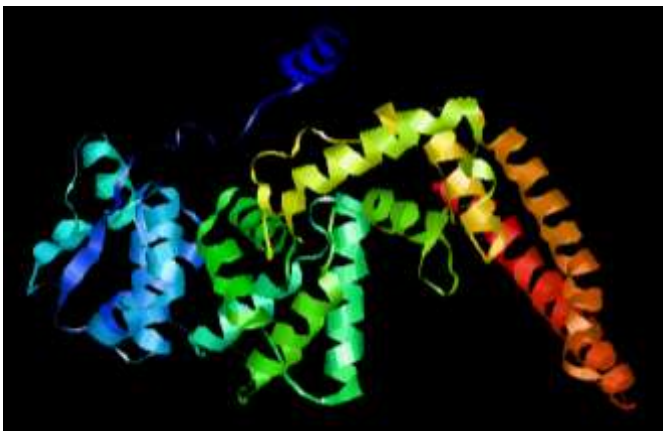
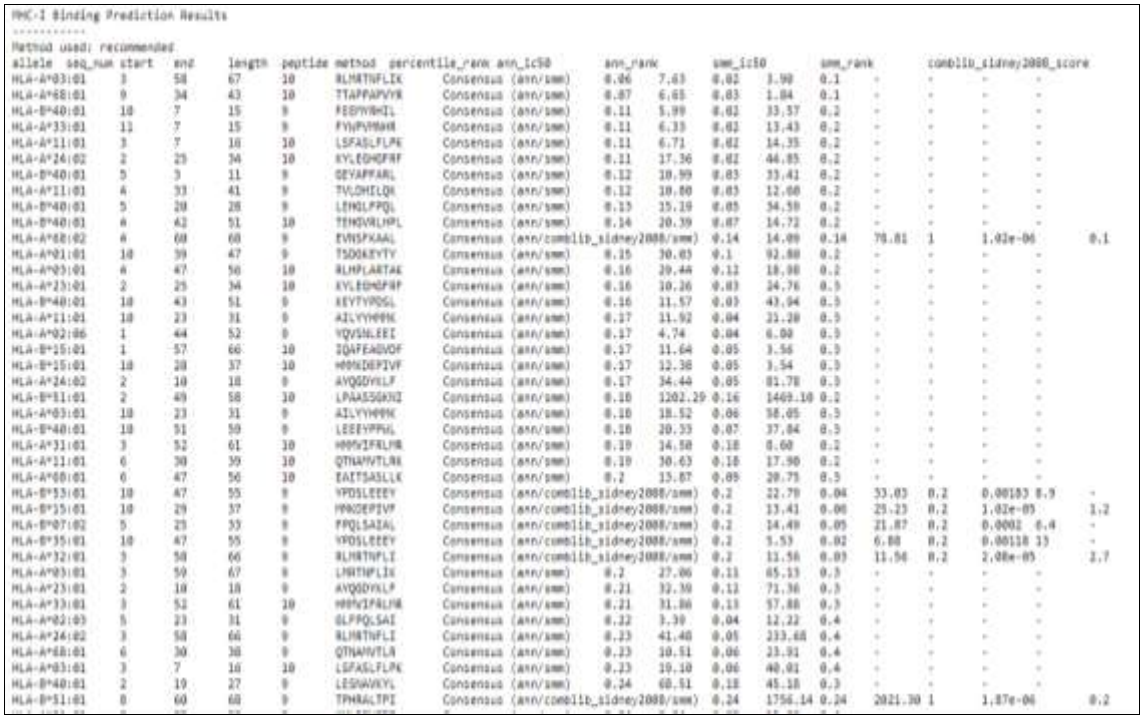


Fig 7: Rasmol view of the consensus protein

3.7 T- cell epitope prediction and its interaction with MHC Class I

Based on 'consensus' prediction many epitopes were obtained. Among these, the epitopes with percentile rank less than 0.2 were selected. For further analysis in this project, two epitopes were selected. (Table1).



MHC-I Binding Prediction Results

Method used: recommended

allele	seq_num	start	end	length	peptide	method	percentile_rank	ann_ic50	ann_rank	smm_ic50	smm_rank	comblib_sidney2008_score	
HLA-A*03:01	3	58	67	10	RLMRTNFLIK	Consensus (ann/smm)	0.06	7.63	0.02	3.90	0.1	-	-
HLA-A*68:01	9	34	43	10	TTAPPAPVYR	Consensus (ann/smm)	0.07	6.65	0.03	1.84	0.1	-	-
HLA-B*40:01	10	7	15	9	FEDYVHKIL	Consensus (ann/smm)	0.11	5.99	0.01	33.57	0.2	-	-
HLA-A*33:01	11	7	15	9	FYLPVMHHR	Consensus (ann/smm)	0.11	6.33	0.01	13.43	0.2	-	-
HLA-A*11:01	3	7	16	10	LSFASLFLPK	Consensus (ann/smm)	0.11	6.71	0.02	14.35	0.2	-	-
HLA-A*24:02	2	25	34	10	KYLEQHDFRF	Consensus (ann/smm)	0.11	17.36	0.02	44.85	0.2	-	-
HLA-B*40:01	5	3	11	9	GEVAPFARL	Consensus (ann/smm)	0.12	10.99	0.03	33.41	0.2	-	-
HLA-A*11:01	4	33	41	9	TVLDHILQK	Consensus (ann/smm)	0.12	10.80	0.03	12.08	0.2	-	-
HLA-B*40:01	5	29	28	9	LENGLFPQL	Consensus (ann/smm)	0.13	15.19	0.05	54.59	0.2	-	-
HLA-B*40:01	6	42	51	10	TFNDVRLHPL	Consensus (ann/smm)	0.14	20.39	0.07	14.72	0.2	-	-
HLA-A*68:02	8	60	68	9	EVNSFKAAL	Consensus (ann/comblib_sidney2008/smm)	0.14	14.00	0.14	78.81	1	1.03e-06	0.1
HLA-A*01:01	10	39	47	9	TSDDAEYTY	Consensus (ann/smm)	0.15	30.03	0.1	92.80	0.2	-	-
HLA-A*03:01	6	47	56	10	RLMPLARTAK	Consensus (ann/smm)	0.16	20.44	0.12	18.98	0.2	-	-
HLA-A*23:01	2	25	34	10	KYLEQHDFRF	Consensus (ann/smm)	0.16	10.26	0.03	24.76	0.3	-	-
HLA-B*40:01	10	43	51	9	AEYTYPDSL	Consensus (ann/smm)	0.16	11.57	0.03	43.94	0.3	-	-
HLA-A*11:01	10	23	31	9	AILVYHHPK	Consensus (ann/smm)	0.17	11.92	0.04	21.28	0.3	-	-
HLA-A*02:06	1	44	52	9	VQVSNLEEI	Consensus (ann/smm)	0.17	4.74	0.04	6.00	0.3	-	-
HLA-B*15:01	1	57	66	10	IQAFEAGVQF	Consensus (ann/smm)	0.17	11.64	0.05	3.56	0.3	-	-
HLA-B*15:01	10	28	37	10	HHMKIDEPIVF	Consensus (ann/smm)	0.17	12.38	0.05	3.54	0.3	-	-
HLA-A*24:02	2	10	18	9	AYQGDYKLF	Consensus (ann/smm)	0.17	34.44	0.05	81.78	0.3	-	-
HLA-B*51:01	2	49	58	10	LPAASSGKHI	Consensus (ann/smm)	0.18	1202.29	0.16	1469.10	0.2	-	-
HLA-A*03:01	10	23	31	9	AILVYHHPK	Consensus (ann/smm)	0.18	18.52	0.06	58.05	0.3	-	-
HLA-B*40:01	10	51	59	9	LEEEYPFGL	Consensus (ann/smm)	0.18	20.33	0.07	37.84	0.3	-	-
HLA-A*31:01	3	52	61	10	HHMVIFRLMR	Consensus (ann/smm)	0.19	14.50	0.18	0.60	0.2	-	-
HLA-A*11:01	6	30	39	10	QTNAMVTLRK	Consensus (ann/smm)	0.19	30.63	0.16	17.90	0.2	-	-
HLA-A*68:01	6	47	56	10	EAITSASLLK	Consensus (ann/smm)	0.2	15.87	0.05	20.75	0.3	-	-
HLA-B*53:01	10	47	55	9	YPDSLEEEY	Consensus (ann/comblib_sidney2008/smm)	0.2	22.70	0.04	53.03	0.2	0.00103 0.9	-
HLA-B*15:01	10	29	37	9	HMKDEPIVF	Consensus (ann/comblib_sidney2008/smm)	0.2	13.41	0.06	25.23	0.2	1.02e-05	1.2
HLA-B*07:02	5	25	33	9	FPQLSAIAL	Consensus (ann/comblib_sidney2008/smm)	0.2	14.49	0.05	21.87	0.2	0.0002 0.4	-
HLA-B*35:01	10	47	55	9	YPDSLEEEY	Consensus (ann/comblib_sidney2008/smm)	0.2	5.53	0.02	6.88	0.2	0.00118 13	-
HLA-A*32:01	3	58	66	9	RLMRTNFLI	Consensus (ann/comblib_sidney2008/smm)	0.2	11.56	0.09	11.56	0.2	2.08e-05	2.7
HLA-A*03:01	3	59	67	9	LMRTNFLIK	Consensus (ann/smm)	0.2	27.06	0.11	65.13	0.3	-	-
HLA-A*23:01	2	10	18	9	AYQGDYKLF	Consensus (ann/smm)	0.21	32.30	0.11	71.36	0.3	-	-
HLA-A*33:01	3	52	61	10	HHMVIFRLMR	Consensus (ann/smm)	0.21	31.86	0.13	57.88	0.3	-	-
HLA-A*02:03	5	23	31	9	GLFPQLSAI	Consensus (ann/smm)	0.22	3.30	0.04	12.22	0.4	-	-
HLA-A*24:02	3	58	66	9	RLMRTNFLI	Consensus (ann/smm)	0.23	41.40	0.05	233.68	0.4	-	-
HLA-A*68:01	6	30	38	9	QTNAMVTLR	Consensus (ann/smm)	0.23	10.51	0.06	23.91	0.4	-	-
HLA-A*03:01	3	7	16	10	LSFASLFLPK	Consensus (ann/smm)	0.23	19.10	0.06	40.01	0.4	-	-
HLA-B*40:01	2	19	27	9	LESWAKYL	Consensus (ann/smm)	0.24	68.51	0.18	45.18	0.3	-	-
HLA-B*51:01	8	60	68	9	TPNRALTPI	Consensus (ann/comblib_sidney2008/smm)	0.24	1756.14	0.24	2021.30	1	1.87e-06	0.2

Fig 8: Result page of IEDB MHC I binding prediction

3.8 T-cell epitope prediction and its interaction with MHC Class II

Based on 'consensus' prediction many epitopes were obtained. Among these, the epitopes with percentile rank less than 0.2 were selected. For further analysis in this project, two epitopes were selected. (Table2)

The 3-dimensional structures of the alleles viewed using RasMol is given in figure 10 (a, b, c, d).

Allele	#	Start	End	Peptide	Method used	Percentile rank
HLA-DRB1*09:01	1	276	290	EVNSFKAALSSLAKH	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB1*09:01	1	277	291	VNSFKAALSSLAKHG	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	671	685	AILYYHMMKDEPIVF	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	672	686	ILYYHMMKDEPIVFS	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	673	687	LYYHMMKDEPIVFST	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	674	688	YYHMMKDEPIVFSTS	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	675	689	YHMMKDEPIVFSTSD	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	676	690	HMMKDEPIVFSTSDG	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	677	691	MMKDEPIVFSTSDGK	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	714	728	NEDNRFITMDGQQFY	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	715	729	EDNRFITMDGQQFYW	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	716	730	DNRFITMDGQQFYWP	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	717	731	NRFITMDGQQFYWPV	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	718	732	RFITMDGQQFYWPVM	Consensus (comb.lib./smm/nn)	0.01
HLA-DPA1*01:03/DPB1*02:01	1	201	215	FRLMRTNFLIKFLLI	Consensus (comb.lib./smm/nn)	0.01
HLA-DPA1*01:03/DPB1*02:01	1	202	216	RLMRTNFLIKFLLIH	Consensus (comb.lib./smm/nn)	0.01
HLA-DPA1*01:03/DPB1*02:01	1	203	217	LMRTNFLIKFLLIHQ	Consensus (comb.lib./smm/nn)	0.01
HLA-DPA1*01:03/DPB1*02:01	1	204	218	MRTNFLIKFLLIHQG	Consensus (comb.lib./smm/nn)	0.01
HLA-DPA1*01:03/DPB1*02:01	1	205	219	RTNFLIKFLLIHQGM	Consensus (comb.lib./smm/nn)	0.01
HLA-DPA1*01/DPB1*04:01	1	201	215	FRLMRTNFLIKFLLI	Consensus (comb.lib./smm/nn)	0.01
HLA-DRB3*01:01	1	466	480	YNNYGDYADDAANAP	Consensus (comb.lib./smm/nn)	0.02
HLA-DRB3*01:01	1	467	481	NNYGDYADDAANAPD	Consensus (comb.lib./smm/nn)	0.02
HLA-DRB3*01:01	1	468	482	NYGDYADDAANAPDD	Consensus (comb.lib./smm/nn)	0.02
HLA-DRB3*01:01	1	469	483	YGDYADDAANAPDDL	Consensus (comb.lib./smm/nn)	0.02
HLA-DRB3*01:01	1	470	484	GDYADDAANAPDDLV	Consensus (comb.lib./smm/nn)	0.02
HLA-DRB3*01:01	1	472	486	YADDAANAPDDLVLF	Consensus (comb.lib./smm/nn)	0.02
HLA-DPA1*01/DPB1*04:01	1	202	216	RLMRTNFLIKFLLIH	Consensus (comb.lib./smm/nn)	0.02

Fig 9: Result page of IEDB Class II MHC binding prediction tool

Table 1: Epitope data

Type of MHC	Allele	Epitope	Percentile rank	Start	End	Length
Class I	HLA-A*23:01	KYLEGHGFRF	0.11	25	34	10
Class I	HLA-B*40:01	LEHGLFPQL	0.13	20	28	9
Class II	HLA-DRB1*09:01	EVNSFKAALSSLAKH	0.01	276	290	15
Class II	HLA-DRB5*01:01	RLMRTNFLIKFLLIH	0.01	202	216	15

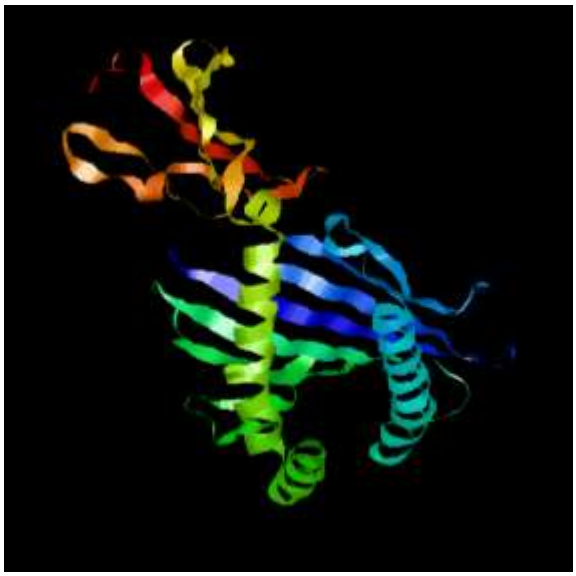


Fig 10a: 3D structure of HLA-A

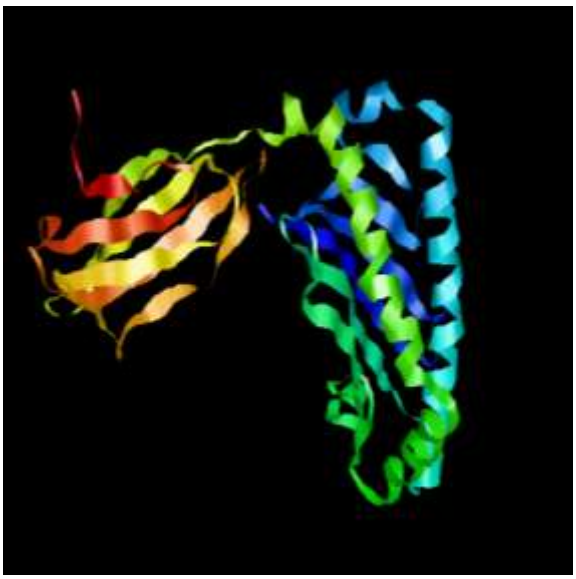


Fig 10b: 3D structure of HLA-B

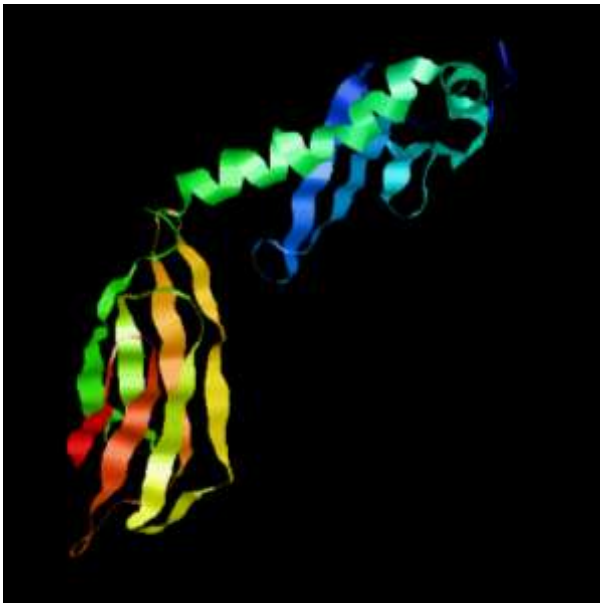


Fig 10c: 3D structure of HLA-DRB1

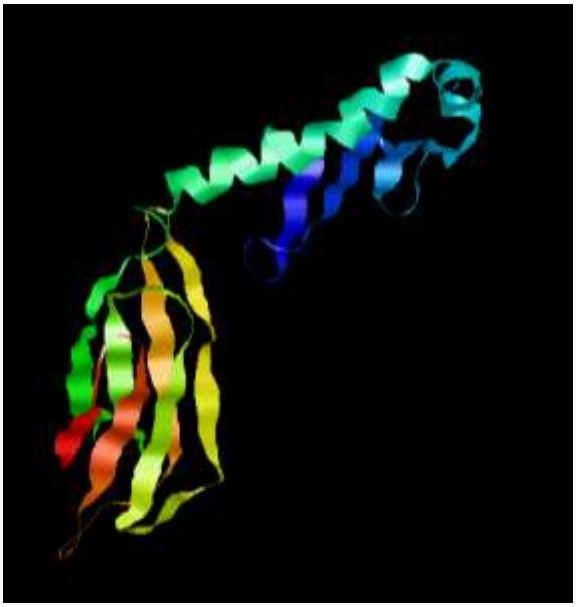


Fig 10d: 3D structure of HLA-DRB5

Fig 10: (a, b, c, d)

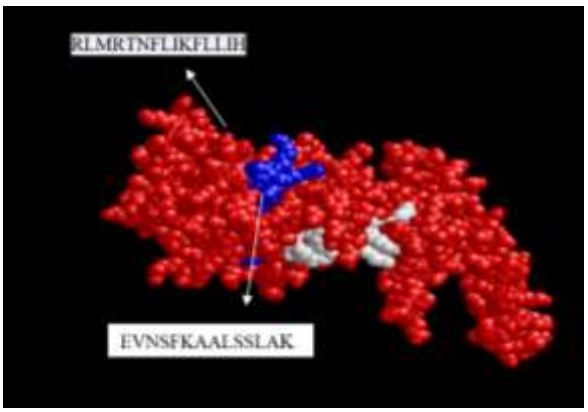


Fig 11: Two Class II MHC molecules' epitopes are indicated

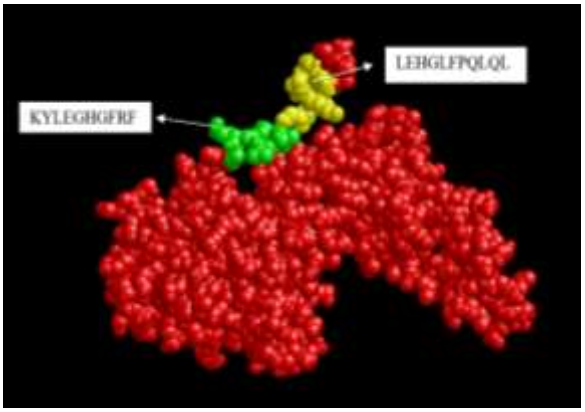


Fig 12: Two Class I MHC molecules' epitopes are indicated

3.9 Docking

The selected alleles shown in Table 1 were docked with the protein using Hex 8.0. The docking results are shown in figure 15. The energy and RMS values are shown in Table 2.

Based on the binding energy in kcal/mol unit, the lowest binding energy was selected to obtain best binding and to predict real T cell epitopes as possible.

Table 2: Showing the docking output

Type of MHC	Allele	Epitope	E (Total)	E (Shape)	RMS
Class I	HLA-A	KYLEGHGFRF	-762.12	-762.18	-1.00
Class I	HLA-B	LEHGLFPQL	-729.47	-729.47	-1.00
Class II	HLA-DRB1	EVNSFKAALSSLAKH	-906.48	-906.48	-1.00
Class II	HLA-DRB5	RLMRTNFLIKFLLIH	-832.68	-832.68	-1.00

EVNSFKAALSSLAKH- HLA-DRB1 had the highest minimum total energy. Hence it might be considered as the best binder among all the four epitopes.

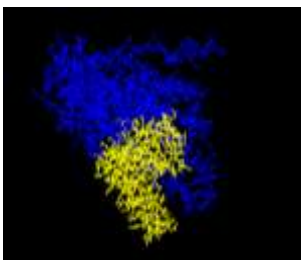


Fig 13a: HLA-DRB5

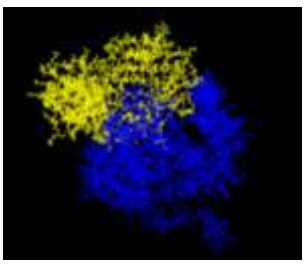


Fig 13b: HLA-A

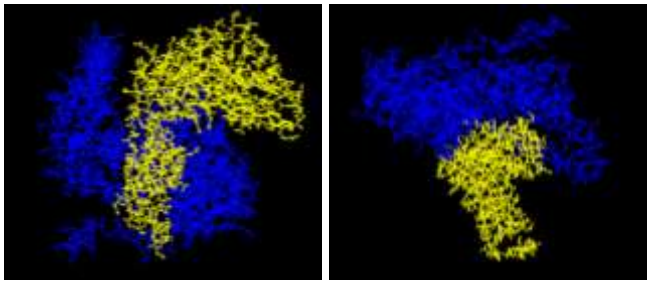


Fig 13c: HLA-DRB1 **Fig 13d:** HLA-B

Fig 13: The docked images of the respective MHC molecules with the protein

4. Discussion

Using the JalView, the relationship between the Ebola virus strains was predicted through the Neighbor-Joining method and with the substitution matrix BLOSUM62, the generated phylogenetic tree clarifies the relationship between the Bundibugyo BDBV and Tai Forest TAFV as sister groups within the clade. Their common ancestor is more closely related to the Zaire Ebola virus making the Sudan SUDV an out group from the clade. From the previous Ebola virus outbreak in Guinea, Bundibugyo BDBV, Reston RESTV, Sudan SUDV, Tai Forest TAFV and Zaire Ebola virus EBOV species were evaluated to identify which strain actually started the outbreak (Dudas, & Rambaut, 2014) [4]. The Zaire viral Ebola virus was suspected to be the primary strain, along with its divergent lineage, to have caused the outbreak on the Democratic Republic of Congo, Gabon and on the recent outbreak in Guinea. (Baize *et al*., 2014) [1]. From the obtained phylogenetic tree, it shows that the Zaire Ebola virus is the middle ground from the Sudan SUDV, Bundibugyo BDBV and Tai Forest TAFV, thus the formulated vaccine for this study can indeed deal with the four strains.

Through the Immune Epitope Database (IED), chosen epitopes with the percentile rank less than 0.2 were singled out to guarantee the most immunogenic potential. Mainly focused on the T-cell immunity responses, predicting T-cell epitope is relatively more reliable than that of the B-cell (Sanchez-Trincando *et al.*, 2017) [8]. Some vaccines currently under development are based on B-cell immunity, which takes advantage of the cell's capacity to linger within the body for up to years for an extended immunogenic potential. But predicting the B-cell epitope, both in linear and conformational forms, are unreliable. Also, linear epitopes have a tendency to activate antibodies that do not bind to intrinsic antigens within the circulation and conformational B-cell epitopes are difficult to isolate in their protein form.

Using proteomics approach for reverse vaccinology, it has proved to be a highly efficient way to obtain vaccines for target diseases. This approach is advantageous over traditional vaccinology as it can (i) reduce time and cost in vaccine development; (ii) refine the number of proteins to be studied and facilitates the selection process; (iii) identify antigens present in small amounts or expressed only at certain stages; and (iv) allow for the study of non-cultivable or risky microorganisms.

5. Conclusion

From the predicted epitopes in this study, the resulting vaccine can be made and is expected to give protective immunity against all four strains of Ebola virus that infect humans. Vaccination has proven to be an effective method of prevention and treatment of various fatal infectious diseases. As what is usually administered nowadays, live-attenuated or inactivated forms of microbial pathogens, e.g. viruses, bacteria, have been used for induction of immune responses that would potentially protect the host against the subsequent infections-but with the risk of allergic responses. This is where the central role of the EBOV nucleoprotein comes in as it makes a suitable antigenic target, compared to the other antigens used, for the potential vaccine of this viral hemorrhagic disease. Also, current vaccines have been developed based on B-cell immunity, while this study mainly focused on T-cell immunity responses further innovating the discoveries on Ebola virus vaccines. From here, EBOV vaccines can be concocted and subjected to trials to be finally used in prophylaxis for anticipated viral infections, providing more accessible vaccines and save many lives.

6. Further work

The potential Ebola virus vaccine as a product which is proposed in this study need to undergo clinical testing to assess its safety, efficacy and presence of any possible adverse effects. The trials need to focus on whether the antigen predicted would not elicit an allergic response to the host. Along with this, the vaccine must also initiate an immune response significant enough to protect people from anticipated infection. The totality of the vaccine must be cleared of any adverse effect and document any possible contraindications, so that the administration of the vaccine to individuals can be regulated. The World Health Organization has already supported ongoing clinical trials for EBOV vaccines ever since the outbreak in 2014 (Calain, 2018) [2]. The quality of potential EBOV vaccines in the near future can be increased with improved formulations.

The critical role of antigenic epitopes in eliciting robust immune responses has led to the development of synthetic vaccines. Epitopes represent the relevant part of the antigen recognized by T and/or B cells, mediating adaptive immunity. The risk of causing pathogenic or off-target responses with epitope vaccines is much lower compared with conventional vaccines and are therefore safer. Epitope-ensemble vaccines have great potential in terms of variety as it can be prepared with several subunits. Synthetic peptides are acclaimed for easier production and are cost-effective (Skwarczynski, & Toth, 2016) [10]. With the current efforts of the scientific community, it is recommended to research more on synthetic vaccines, which could potentially increase rate of production to offset the supply-demand and eradicate prevalent diseases in the increasing global population.

7. References

1. Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, Gunther S. Emergence of Zaire Ebola Virus Disease in Guinea-Preliminary Report. N Engl J Med. 2014 Apr 16. PubMed PMID: 24738640. https://doi.org/10.1056/nejmoa1404505, 2014.
2. Calain P. The Ebola clinical trials: a precedent for research ethics in disasters, Journal of Medical Ethics. 2018; 44:3-8. https://doi.org/10.1136/medethics-2016-103474.
3. Coltart CE, Lindsey B, Ghinai I, Johnson AM, Heymann DL. The Ebola outbreak, 2013-2016: old lessons for new epidemics. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 2017, 372(1721), 20160297. https://doi.org/10.1098/rstb.2016.0297
4. Dudas G, Rambaut A. Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. PLOS Currents Outbreaks. May 2, Edition 1 https://doi.org/10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d, 2014.
5. Ehreth J. The global value of vaccination, Vaccine. 2003; 21 (7-8):596-600. https://doi.org/10.1016/ S0264-410X(02)00623-0, 2016.
6. He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. Journal of Biomedicine and Biotechnology. 2010, 218-590. http://dx.doi.org/10.1155/2010/218590.
7. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, *et al*. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acid Res. https://doi.org/10.1093/nar/gkz268, 2019.
8. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and Methods for T- and B-Cell Epitope Prediction, Journal of Immunology Research, vol. 2017, Article ID 2680160, 14, 2017. https://doi.org/10.1155/2017/2680160.
9. Scarselli M, Giuliani MM, Adu-Bobie J, Pizza M, Rappuoli R. The impact of genomics on vaccine design. Trends Biotechnol. 2005; 23(2):84-91. https://doi.org/10.1016/j.tibtech.2004.12.008
10. Skwarczynski M, Toth I. Peptide-based synthetic vaccines. Chemical science. 2016; 7(2):842-854. https://doi.org/10.1039/c5sc03892h.
11. Stein RA. What is Ebola? Int J Clin Pract. 2015; 69(1):49-58. Epub 2014 Dec 11. https://doi.org/10.1111/ijcp.12593
12. Weingartl HM, Embury-Hyatt C, Nfon C, Leung A, Smith G, Kobinger G. Transmission of Ebola virus from pigs to non-human primates. Scientific reports, 2, 811, 2012. https://doi.org/10.1038/srep00811.
13. World Health Organization, Ebola virus disease Fact sheet N°103. Retrieved from, 2014, https://www.who.int/en/news-room/fact-sheets/detail/ ebola-virus-disease on 12 April 2014.