

## **Prediction of Multi peptides Vaccination from VP10,VP21, VP51 against Reverse Transcriptase Human immunodeficiency Viruses Using Immuno- informatics Approach**

Safa Ahmad Almostafa <sup>1</sup>, Ienas Ibrahim mohmed <sup>2</sup>, Habab Abd elmoneim Siddig<sup>3</sup>, Sahar obi Abd albagi<sup>4</sup>, *Nadir Musa Khalil Abuzeid*<sup>5</sup>

<sup>1</sup>Department of microbiology, MSC of Medical Laboratory Sciences, National University , Sudan.

<sup>2</sup>Department of microbiology, MSC of Medical Laboratory Sciences, National University , Sudan.

<sup>3</sup> Department of microbiology, MSC of Medical Laboratory Sciences, National University , Sudan.

<sup>4</sup> Department of microbiology, faculty of Medical Laboratory Sciences, Alneelain University , Sudan

<sup>5</sup> (*Associate professor (Omdurman Islamic university) and first Consultant of medical Microbiology*)

### **Corresponding author:**

Safa Ahmad Almostafa, [safatayfour@outlook.com](mailto:safatayfour@outlook.com). +249 923207867

## **Abstract**

The human immunodeficiency virus-(HIV) is group of the genus Lentivirus within the family of Retroviridae, subfamily Ortho retrovirinae. Based on genetic characteristics and differences in the viral antigens, HIV is classified into the types 1 and 2 (HIV-1, HIV-2). HIV is identical single – stranded RNA molecule that are enclosed within the core of the virus particle proteins, the genome of the HIV Provirus, also known as DNA, is generated by the Protease against reverse transcriptase RNA genome into DNA, degradation of the RNA and integration of the double – stranded HIV DNA into the human genome. The aim of this study is to determine antigenic peptides from p10, p21, and p51 proteins that can be used for multiple peptide vaccine design using In-Silico study. A total of 73 sequences of three proteins were obtained from NCBI and subjected to multiple sequence alignments using CLUSTALW tool to determine conserved regions.

Immune Epitope Data Base tools were used to determine B cell epitopes, these tools are Bepipred Linear B cell epitopes prediction, surface accessibility and antigenicity prediction. Epitope binding to MHC class I and class II and their population coverage were also determined using IEDB software. The analysis results are as follow, for B cell binding from p10 (<sup>708</sup>**QGYSP**<sup>712</sup>), from p21 (<sup>704</sup>**QGYSP**<sup>708</sup>, <sup>73</sup>**CVPTDPNPQ**<sup>81</sup>) and (<sup>346</sup>**FKNL**<sup>349</sup>) from p51. All these peptides have high score in Linear B cell epitopes prediction, surface accessibility and antigenicity prediction. On another hand peptides that reacted to MHC class I were (<sup>47</sup>**EANTTLFCA**<sup>55</sup>, <sup>53</sup>**FCASDAKAY**<sup>61</sup>, <sup>55</sup>**ASDAKAYET**<sup>63</sup>) from p10, (<sup>38</sup>**YGVVPVWKE**<sup>46</sup>, <sup>10</sup>**PQEVFLVNV**<sup>18</sup> and <sup>29</sup>**AAGSTMGAA**<sup>37</sup>) from p21 and (<sup>63</sup>**EWEFVNTPP**<sup>71</sup>, <sup>70</sup>**PPLVKLWYQ**<sup>78</sup> and <sup>79</sup>**EKEPIVGA**<sup>87</sup>) from p51 protein respectively. It worth noting that the peptides(<sup>119</sup>**IISLWDQSL**<sup>127</sup>, <sup>108</sup>**CVKLTPLCV**<sup>116</sup>) from p10, (<sup>38</sup>**YGVVPVWKE**<sup>46</sup>, <sup>20</sup>**LLQYWSQEL**<sup>34</sup>, <sup>16</sup>**FNMWKNMNV**<sup>30</sup>) from p21 and (<sup>7</sup>**WKGSPAIFQ**<sup>21</sup>, <sup>11</sup>**WEFVNTPPL**<sup>25</sup> and <sup>58</sup>**FLWMGYELH**<sup>72</sup>) protein is also binds to MHC class II with high affinity. All T cell peptides had highest population coverage, and the combined coverage for all peptides in this study was found to be 100%. Using In-Silico studies will ensure less risk of virulence and side effects. Evaluation of antibodies response in animal models is needed to confirm efficacy of these epitopes in inducing protective immune response.

## 1.1.Introduction

The human immunodeficiency virus-(HIV) is group of the genus Lentivirus within the family of Retroviridae, subfamily Ortho retrovirinae (1). Based on genetic characteristics and differences in the viral antigens, HIV is classified into the types 1 and 2 (HIV-1, HIV-2). The immunodeficiency viruses of non-human primates (simian immunodeficiency virus, SIV) are also group of the genus Lentivirus. Epidemiologic and phylogenetic analyses currently available imply that HIV was introduced into the human population around 1920 to 1940. HIV-1 evolved from non-human primate immunodeficiency viruses from Central African chimpanzees (SIVcpz) and HIV-2 from West African sooty mangabeys (SIVsm) (2,3,4).

The HIV genome consists two identical single-stranded RNA molecules that are enclosed within the core of the virus particle. The genome of the HIV provirus, also known as proviral DNA, is generated by the reverse transcription of the viral RNA genome into DNA, degradation of the RNA and integration of the double-stranded HIV DNA into the human genome. The DNA genome is flanked at both ends by LTR (long terminal repeat) sequences. The 5' LTR region codes for the promotor for transcription of the viral genes. In the direction 5' to 3' the reading frame of the *gag* gene follows, encoding the proteins of the outer core membrane (MA, p17), the capsid protein (CA, p24), the nucleocapsid (NC, p7) and a smaller, nucleic acid-stabilizing protein. The *gag* reading frame is followed by the *pol* reading frame coding for the enzymes protease (PR, p12), reverse transcriptase (RT, p51) and RNase H (p15) or RT plus RNase H (together p66) and integrase (IN, p32). Adjacent to the *pol* gene, the *env* reading frame follows from which the two envelope glycoproteins gp120 (surface protein, SU) and gp41 (transmembrane protein, TM) are derived. In addition to the structural proteins, the HIV genome codes for several regulatory proteins: Tat (transactivator protein) and Rev (RNA splicing-regulator) are necessary for the initiation of HIV replication, while the other regulatory proteins Nef (negative regulating factor), Vif (viral infectivity factor), Vpr (virus protein r) and Vpu (virus protein unique) have an impact on viral replication, virus budding and pathogenesis(5,6).HIV-2 codes for Vpx (virus protein x) instead of Vpu, which is partially responsible for the reduced pathogenicity of HIV-2 (7). The genome structure of the



immunodeficiency viruses of chimpanzees (SIVcpand gorillas (SIVgor) is identical to that of HIV-1(8).

P10 Retroviral proteases are encoded by a part of the pol gene, for example in that of the human immunodeficiency virus (HIV). The protease gene is located between the gag gene (encoding structural proteins) and other enzymatic genes, such as reverse transcriptase and integrase. There are 93 sequences belonging to the retroviral protease family A2 of the aspartic peptidase clan AA at present, according to the Merops database, which provides information on viral as well as other proteases.

Inhibition of HIV-1 replication in monocyte-derived DCs (MDDCs) is associated with an increased expression of p21, a cell cycle regulator that is involved in the differentiation and maturation of DCs. Induction of p21 in MDDCs decreases the pool of dNTPs and increases the antiviral active.

HIV reverse transcriptase (RT) plays a multifunctional role in the transformation of viral RNA into dsDNA and represents a primary target for treatment of AIDS. Currently, all of the drugs in clinical use target the mature RT p66/p51 heterodimer, however, a single p66 peptide chain functions as the precursor for each subunit of the RT heterodimer, requiring a complex maturation process that includes subunit-selective elimination of a single ribonuclease H (RH) domain.

Based on their genetic make-up, HIV-1 viruses are divided into three groups (eg, M [main], N, and O group, ). These HIV-1 groups and HIV-2 probably result from distinct cross-species transmission events(9) Pandemic HIV-1 has diversified into at least nine subtypes and many circulating recombinant forms(10,11) which encode genetic structures from two or more subtypes (eg, A/E=CRF01; A/G=CRF02). The continuously evolving HIV-1 viral diversity poses an immense challenge to the development of any preventive or therapeutic intervention (10) . Infection with two or more genetically distinct viruses could lead to new recombinant viruses.

Recombination takes place at a higher rate than initially predicted,(11) and circulating recombinant forms account for as much as 20% of infections in some regions (eg, southeast Asia).(12)These findings are in agreement with the occurrence of co-infections with multiple distinct isolates in a close temporal context(13,14)Further, superinfections in which time points of virus acquisition are months to years apart have been described, although at a much lower frequency than co-infections(13,15,16) Collectively, these observations challenge the assumption that HIV-1 acquisition happens only once with a singular viral strain and that, thereafter, the infected individual is protected from subsequent infections(17) This lack of immunisation has



substantial implications for vaccine development. Emerging evidence suggests that clinical progression to AIDS might be more rapid in individuals with dual infections,(18) and encouraging safer sex practices in viraemic HIV-1-infected people might be appropriate to keep recurrent exposure to new viral strains to a minimum.

According to present knowledge, the spread of HIV started at the beginning of the 20th century (4,19) Zoonotic transmission of SIVcpz from chimpanzees (*Pan troglodytes troglodytes*) occurred for HIV-1 group M and group O around 1920 and for HIV-1 group N around 1960 (20,21) in West Central Africa. HIV-2 was transmitted zoonotically from sooty mangabey (*Cercocebus atys*) to human in West Africa around 1940 (22) Molecular genetic analyses suggest that HIV-1 was exported to Haiti probably in 1966 and arrived in North America approximately 2 years later (4,23). Since the mid-1980s the different HIV-1 M subtypes have been spreading, leading to a global pandemic. In contrast to HIV-1, HIV-2 initially remained restricted to West Africa because of its significantly lower infectivity.

After HIV-2 was exported to Portugal and France probably during the mid-1960s, the spreading of HIV-2 with a low prevalence especially in Europe, South America and Asia is documented.

Globally, an estimated 35 million people were living with HIV in 2013 (24). Since 1999 the number of new infections has been decreasing continuously, and for 2013 a number of 1.9 million newly infected persons was estimated. About three quarters of HIV-infected persons live in Sub-Saharan Africa, and also about two-thirds of the reported new infections originate from this region. The countries most affected by HIV with a high prevalence among 15- to 49-year-olds are Swaziland (approximately 27%), Lesotho (approximately 23%) and Botswana (approximately 23%).

An estimated 38.6 (33.4–46.0) million people live with HIV-1 worldwide, while about 25 million have died already.(25)In 2005 alone, there were 4.1 million new HIV-1 infections and 2.8 million AIDS deaths.(25)These estimates mask the dynamic nature of this evolving epidemic in relation to temporal changes, geographic distribution, magnitude, viral diversity, and mode of transmission. Today, there is no region of the world untouched by this pandemic.

Sudan is bordered by countries with high rates of HIV infection. The first AIDS case was reported in 1986, and by 2011 the total number of cases reported had increased to 2218(26). In 2002, a



behavioral and epidemiological survey (BES) was carried out by Sudan National AIDS Program [9]. According to the survey, the prevalence among general population was 1.6% (26,27). Although this survey was the largest survey conducted in Sudan, it was inconclusive largely due to study limitations and mixing of samples that included both general and at-risk populations (27).

The spread of HIV is influenced by poverty and illiteracy, both of which are widespread in Sudan. The movement of people displaced by harsh environmental conditions has contributed to an increase in the number of HIV/AIDS cases. Refugees from other conflicts in the region also flee to Sudan. The war in Darfur resulted in the movement of large population fleeing from the war. This resulted in displacement of about 2.5 million people with many contracted the disease (28). The new war in Blue Nile and South Kordofan also contributed to unfavorable socioeconomic and health situations in Sudan (28). Today, there is no region of the world untouched by this pandemic.

The initial period following the contraction of HIV is called acute HIV, primary HIV or acute retroviral syndrome.(29,30) Many individuals develop an influenza-like illness or a mononucleosis-like illness 2–4 weeks after exposure while others have no significant symptoms.(31,32) Symptoms occur in 40–90% of cases and most commonly include fever, large tender lymph nodes, throat inflammation, a rash, headache, tiredness, and/or sores of the mouth and genitals.(29,32) The rash, which occurs in 20–50% of cases, presents itself on the trunk and is maculopapular, classically.(33) Some people also develop opportunistic infections at this stage.(29) Gastrointestinal symptoms, such as vomiting or diarrhea may occur.(32) Neurological symptoms of peripheral neuropathy or Guillain–Barré syndrome also occurs.(32) The duration of the symptoms varies, but is usually one or two weeks.(32)

Due to their nonspecific character, these symptoms are not often recognized as signs of HIV infection. Even cases that do get seen by a family doctor or a hospital are often misdiagnosed as one of the many common infectious diseases with overlapping symptoms. Thus, it is recommended that HIV be considered in people presenting with an unexplained fever who may have risk factors for the infection .(32)

The initial symptoms are followed by a stage called clinical latency, asymptomatic HIV, or chronic HIV.(29)Without treatment, this second stage of the natural history of HIV infection can last from



about three years(34) to over 20 years(35) on average, about eight years).(36)While typically there are few or no symptoms at first, near the end of this stage many people experience fever, weight loss, gastrointestinal problems and muscle pains. Between 50 and 70% of people also develop persistent generalized lymphadenopathy, characterized by unexplained, non-painful enlargement of more than one group of lymph nodes (other than in the groin) for over three to six months(37).

## Materials and Methods:

### In-Silico, Study Materials and Methods

A total of 10 HIV structural poly protein sequences were obtained from NCBI([www.ncbi.nlm.nih.gov/protein/?term=](http://www.ncbi.nlm.nih.gov/protein/?term=)) in FASTA format in Jun 2019 (Table1) For VP10.

Total of 6 HIV structural poly protein sequences were obtained from NCBI ([www.ncbi.nlm.nih.gov/protein/?term=](http://www.ncbi.nlm.nih.gov/protein/?term=)) in FASTA format in May 2019 For vp21. , reference sequences were identified from NCBI reference sequence.

Total of 57 reverse transcriptase protein (RT P51) sequences of HIV virus were retrieved from the National Center for Biotechnology Information database (NCBI) ([www.ncbi.nlm.nih.gov/protein/?term=](http://www.ncbi.nlm.nih.gov/protein/?term=)) in May 2019, reference sequences were identified from NCBI reference sequence.

All sequences of structural poly protein were subjected to multiple sequence alignments using CLUSTALW tool of BIOEDIT sequence alignment editor (version 7.2.5.0) in order to identify conserved regions between sequences. Then epitopes prediction and analysis of each protein was done using different tools of immune epitope data base IEDB software (<http://www.iedb.org>).

**Table: 1 HIV Countries and accession numbers of retrieved sequence**

Type of HIV p	Accession Code	Collection Year	Country
P10	ACN87192.1	2005	USA
	ACN87191.1	2005	USA
	ACN87186.1	2005	USA
	ACN87185.1	2005	USA
	ACN87183.1	2005	USA
	ACN87182.1	2005	USA
	ARA72305.1	NA	Switzerland
	AGW07432.1	2000	USA
	ACN87179.1	NA	USA
	ACN87176.1	2000	USA
P21	MH262621.1	2018	USA
	MG196385.1	2018	USA



	EU575977.1	2016	USA
	EU576083.1	2009	USA
	EU576177.1	2009	USA
	HQ217990.1	2011	USA
	HQ217935.1	2011	USA
P51	*NP_789739.1	2018	NA
	1TL3\B 158850	2018	NA
	1TL1\B <u>158849</u>	2018	NA
	1TKZ\B <u>158848</u>	2018	NA
	1TKT\B <u>158846</u>	2018	NA
	1TKX\B <u>158847</u>	2018	NA
	1SIV\B <u>28207</u>	2013	NA
	1S1U\B <u>28206</u>	2013	NA
	1S1W\B 28208	2013	NA
	1S1T\B 28205	2012	NA
	1JLQ \B 158768	2018	NA
	1LWC \B 20931	2013	NA
	1LWF \B 20933	2013	NA
	1LW0 \B 20928	2012	NA
	1LWE \B 20932	2013	NA
	1LW2 \B 20929	2012	NA
	1JLG \B 17237	2017	NA
	1JLF\B 17236	2012	NA
	1JLC \B 1734	2012	NA
	1JB \B 17233	2012	NA
	1JLE \B 17235	2012	NA
	1JLA \B 1732	2012	NA
	1JKH /B 1731	2012	NA
	1DTQ \B 12978	2014	NA
	1DTT \B 12888	2014	NA
	5J2Q \B 141564	2017	NA

### Determination of conserve sequence and phylogenetic analysis:

The retrieved sequences were aligned by multiple sequence alignment (MSA) using Bio Edit software (version 7.2.5.0) (38), to obtain the conserved regions. Which were considered candidate epitopes were analyzed by different prediction tools at the





Immune Epitope Database IEDB analysis resource (<http://www.iedb.org/>).

## **Prediction of B cell epitopes:**

Candidate epitopes were analyzed by several B-cell prediction methods which determine the antigenicity, hydrophilicity, flexibility and surface accessibility (39). The method for predicting B-cell epitopes is the hidden Markov model. The linear predicted epitopes were obtained by using BepiPred linear epitope prediction with a threshold value of -0.165 (40). The surface accessible epitopes were predicted from the conserved region with a threshold value of 1.000 using the Emini surface accessibility prediction tool (41). Using the kolaskar antigenicity we determined the antigenic sites with a threshold value of 1.000 (42).

## **MHC epitope prediction:**

### **MHC class I epitope prediction**

MHC I prediction tool has been used to analyze the referential (p10,p21,P51) RT protein of the HIV binding to MHC class I molecules (<http://tools.iedb.org/mhci/>). This prediction tool consists of 9 different prediction methods available in the Immune Epitope Database (IEDB) namely the covariance matrix SMMPMBEC, Artificial Neural Network (ANN), Stabilized Matrix Method (SMM), ComLib-Sidney2008, Consensus, NetMHCpan, PickPocket, NetMHCcons and IEDB Recommended.

Artificial neural network (ANN) was chosen as it is perfectly suitable in the incorporated such higher order connections when predicting the binding affinity (43,44). Analysis was done using HLA alleles (45). For all the alleles, peptide length was set to 9 amino acids prior to the prediction. The half maximal inhibitory concentration (IC<sub>50</sub>) values of the peptide's binding to MHC I molecules was calculated and the ones that had binding affinity (IC<sub>50</sub>) less than 100 nM were suggested to be promising candidate epitopes.

## **MHC Class II binding predictions**

The referential protein sequence was submitted in the (IEDB) MHC II binding prediction tool



(<http://tools.immuneepitope.org/mhcii/>). For the screening of promising epitopes, human allele reference set (HLA DR, DQ, DP) was used (46,47). In MHC II binding analysis at the (IEDB) there are seven prediction methods including IEDB recommended, Consensus, Net MHCII pan, NN-align, SMM-align, combinatorial library and Sturniolo methods. NN-align method was used with a half maximal inhibitory concentration (IC50) of 1000 nM. Peptides less or equal to the (IC50) value were proposed to be promising MHC II epitopes (48,49).

### **Population coverage:**

The estimation of the population coverage are based on the MHC binding with or without T cell restriction data, there for Nemours based tool was developed to predict population coverage of T cell epitope-based diagnostic and vaccines based on MHC binding with or without T cell restriction data (50). All alleles that interact with epitopes from P51 RT protein were subjected population coverage tool of IEDB ([http://tools.iedb.org/tools/population/iedb\\_input](http://tools.iedb.org/tools/population/iedb_input)) to calculate the whole world population coverage of MHC class I, MHC II and combined MHC I and II alleles for each protein (51).

### **Homology modeling:-**

**Raptor X** protein structure prediction server available at (<http://raptorx.uchicago.edu/StructurePrediction/predict/>) was utilized for the creation of the **3D structure** and **Chimera 1.8** (52) was the tool used to visualize the selected epitopes belonging to the B cell, MHC class I and MHC class II.

## **Results**

### **Alignment and determination of conserved regions**

A total of 10 HIV p10, 6 of p21, and 57 HIV p51 sequences were obtained from NCBI ([www.ncbi.nlm.nih.gov/protein/?term=](http://www.ncbi.nlm.nih.gov/protein/?term=)) in FASTA format in May 2019. , reference sequences were identified from NCBI reference sequence .All sequences of HIV p10, p21 and P51 were subjected to multiple sequence alignments using CLUSTALW tool of BIOEDIT sequence alignment editor in order to



identify conserved regions between sequences. Then epitopes prediction and analysis of each protein was done using different tools of immune epitope data base IEDB software (<http://www.iedb.org>).

Figure 1: Multiple sequence alignment p10



Figure 2: Multiple sequence alignment;p21

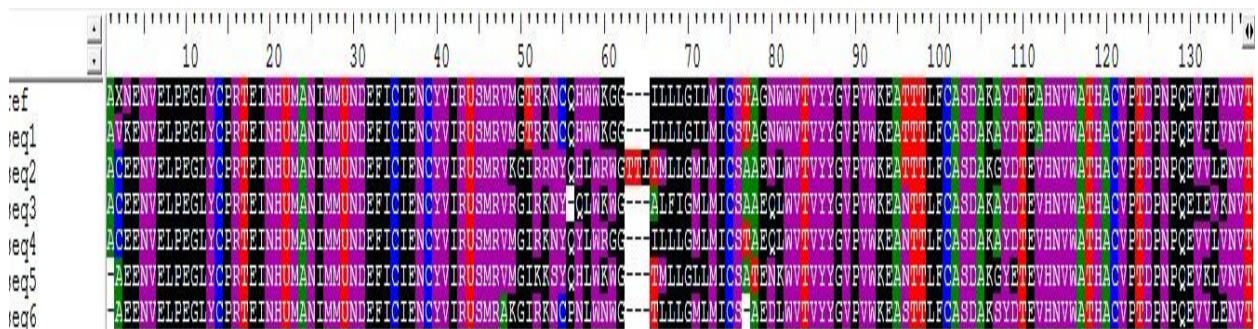
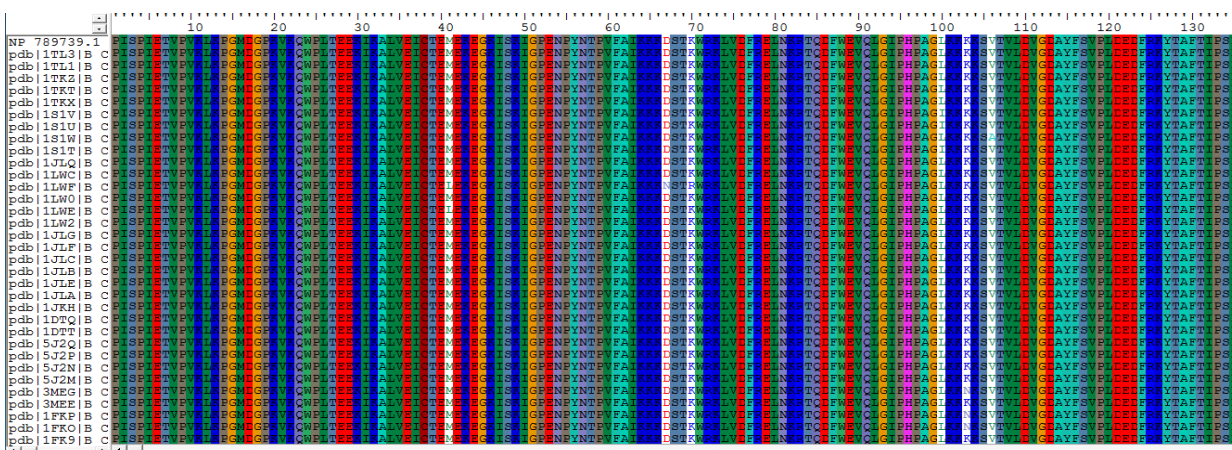


Figure 3: Multiple sequence alignment;p51





## Prediction of B cell epitopes:

Poly structural protein was tested using Bepipred linear epitope production, Emini surface accessibility and Kolaskar and Tongaonkar antigenicity tool of IEDB. All values equal or greater than the default threshold were considered as potential B-cell binders.

Regarding Emini surface accessibility prediction, the average binding score of poly structural protein was 1.00, and minimum and maximum values as displayed in Table2. All values equal or greater than the default threshold were predicted to have good surface accessibility. All proteins were subjected to Kolaskar and Tongaonkar antigenicity prediction tool of IEDB to predict peptides with probability of being antigenic, the average threshold value for poly structural protein was 1.01. Epitopes with values equal or greater than the average score are considered as antigenic peptides.

**Table 2: predicted B cell epitopes**

Type	NO	Start	End	Peptide	Length	Emini	Kolaskar
P10	1	708	712	QGYSP	5	2.511	1.055
P21	3	73	81	CVPTDPNPG	9	1.061	1.061
	8	204	207	KVSF	4	0.661	1.104
	23	521	525	GSTMG	5	0.559	0.899
	27	704	708	QGYSP	5	1.659	1.025
P51	1	1	7	PISPIET	7	0.521	1.029
	2	10	27	VKLKPGMDGPKVKQWP LT	18	0.697	1.024
	3	42	58	EKEGKISKIGPENPYNT	17	2.743	0.95
	8	121	125	DEDFR	5	1.578	0.909
	9	134	143	SINETPGIR	10	1.058	0.944
	10	151	159	QGWKGSPAI	9	0.172	0.986
	11	172	176	RKQNP	5	1.739	0.932
	15	234	241	LHPDKWTV	8	0.181	
	16	246	254	LPEKDSWTV	9	0.748	1.018
	20	315	324	HGVYYDPSKD	10	0.843	1.042
	21	330	337	QKQGQGQW	8	1.271	0.954
	22	346	349	FKNL	4	1.813	1.012
	24	359	367	GAHTNDVKQ	9	1.419	0.991
	27	395	400	KETWET	6	2.419	0.89
30	432	436	EPIVG	5	0.29	1.065	

The result of structural polyprotein showed that three peptides had passed the antigenicity prediction and surface accessibility prediction test, QGYSP from 704 to 708 was found to have greatest score among all other peptides in both tests in p10 and p21, and one peptide had passed



the antigenicity prediction and surface accessibility prediction test, **FKNL** from 346 to 349 was found to have greatest score among all other peptides in both tests in p51. The result is summarized in Table 2.

Figure .2 surface accessibility analyses using Emini surface accessibility scale

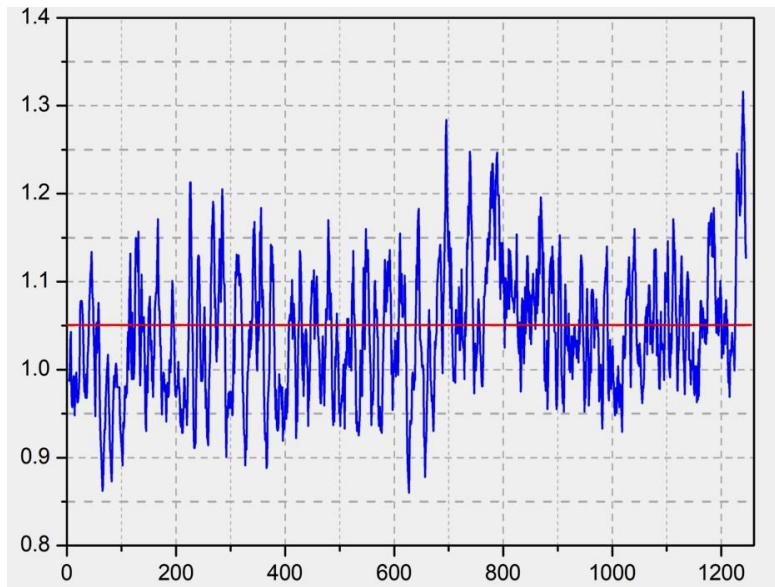
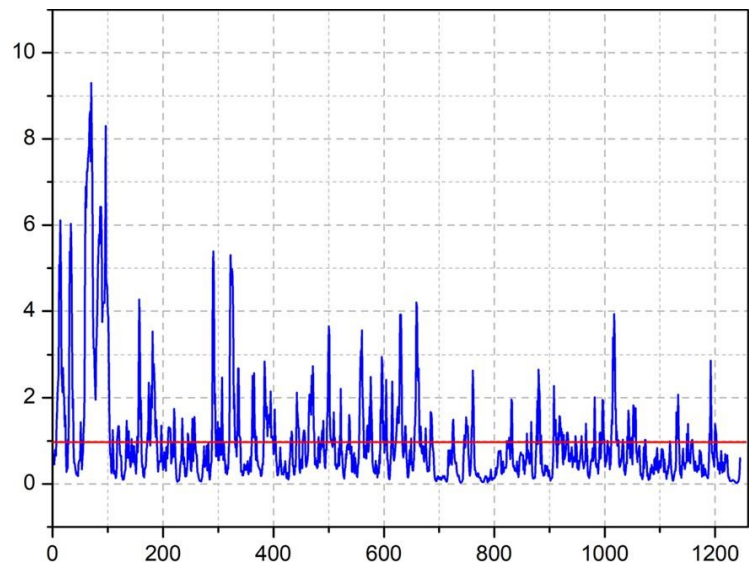
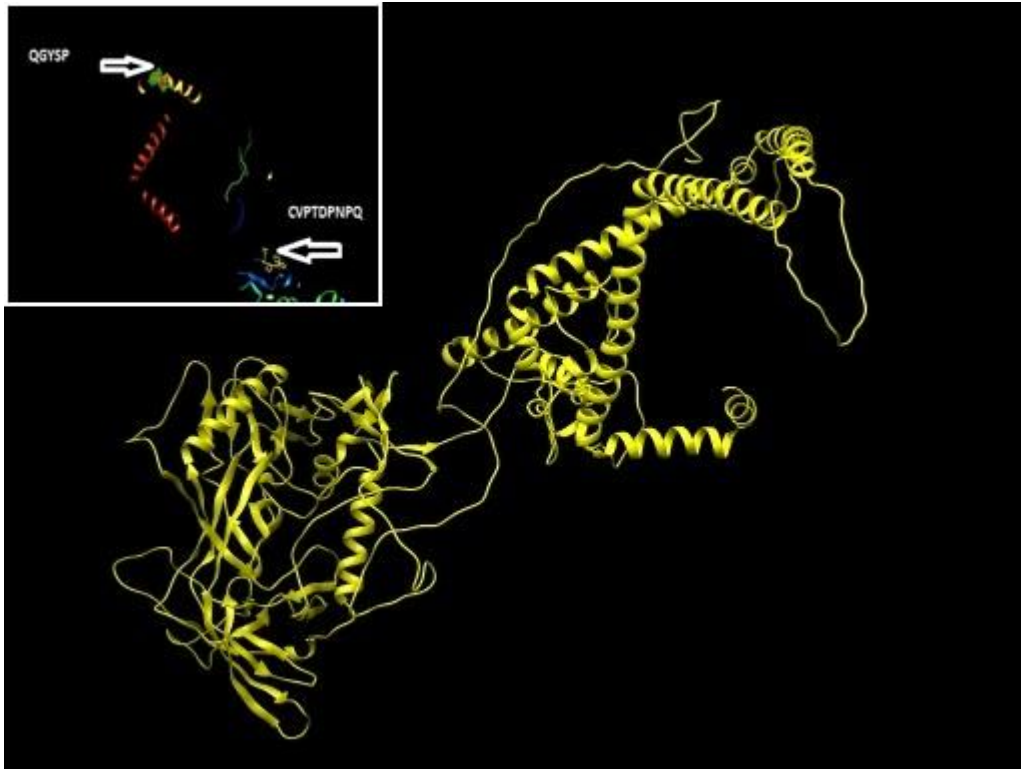


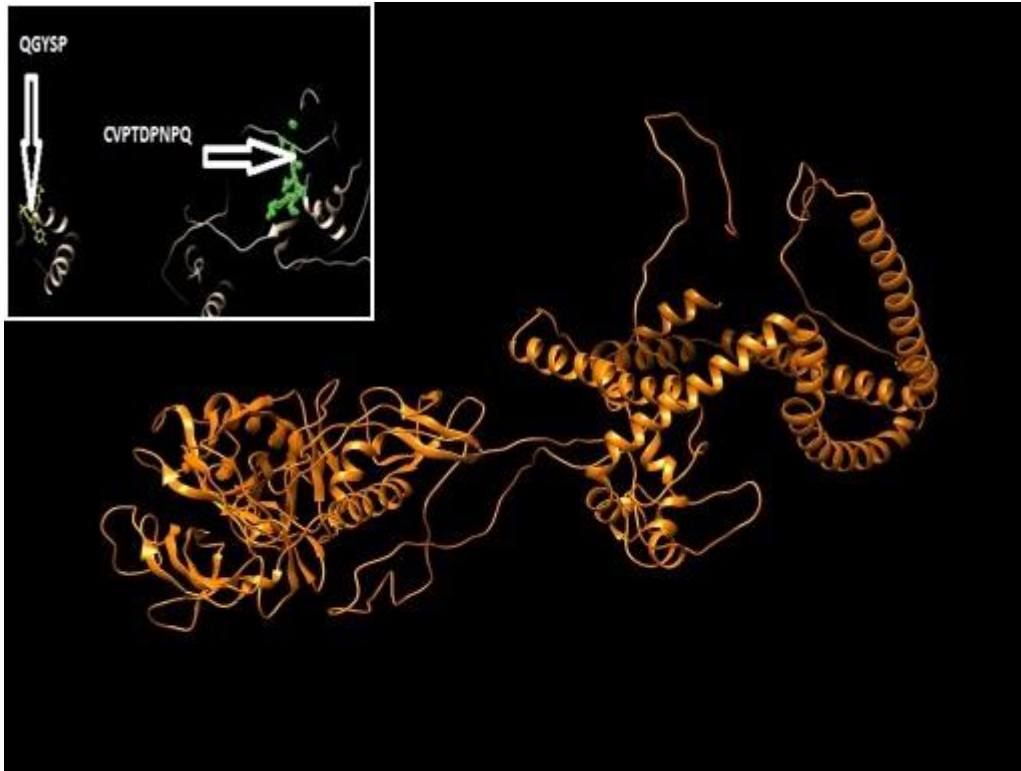
Figure 3.3 Kolaskar and Tongaonker antigenicity



Figuer4: prediction of Bcell epitopes p10 by chimera 1.8

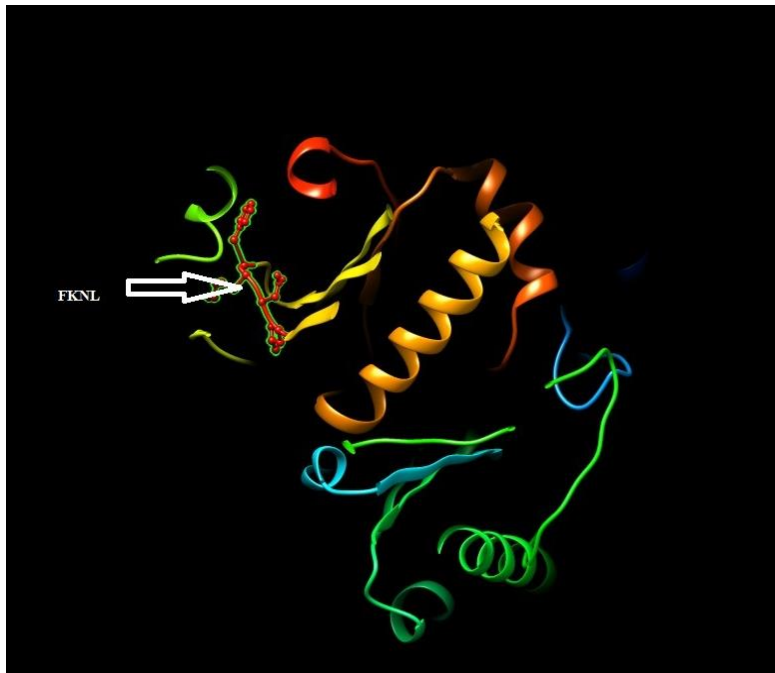


Figuer5: prediction of Bcell epitopes p21 by chimera 1.8





Figuer6: prediction of Bcell epitopes p51 by chimera 1.8



### Predictions of T cell Epitopes:-

IEDP server (<http://www.iedb.org>) was used through specific tools to determine MHC I and MHC II binding epitopes. This server uses specific scoring IC<sub>50</sub> (inhibitory concentration 50) to predict epitopes that bind to different MHC class I and MHC class II alleles.

### Prediction of T-cell epitopes and MHC class I interaction analysis

Epitopes sequences of structural poly protein subjected to MHC class I binding prediction tool of IEDP. T-cell epitope predicted to interact with different MHC class I alleles using ANN (artificial neural network) as prediction method and length of nine amino acids.

Using same score for all protein of 100 IC<sub>50</sub>, The peptide <sup>47</sup>EANTTLFCA<sup>55</sup>, <sup>53</sup>FCASDAKAY<sup>61</sup>, <sup>55</sup>ASDAKAYET<sup>63</sup> height affinity to interact with the largest number of (6) alleles for p10 ,12 peptide of **structural polyprotein** were found to interact with MHC class I. The peptide (<sup>38</sup> YYGVPVWKE<sup>46,10</sup> PQEVFLVNV<sup>18</sup> and <sup>29</sup>AAGSTMGAA<sup>37</sup>) had height affinity to interact with the largest number of alleles for p21.

These 10 sequenced epitopes were obtained from the IEDB MHC I binding epitope prediction tool using ANN method and the reference reverse transcripts protein (p51) as an input. There were 44 human MHC class-I alleles for which the epitopes were predicted the best 3 are <sup>63</sup>EWEFVNTPP<sup>71</sup>, <sup>70</sup>PPLVKLWYQ<sup>78</sup> and <sup>79</sup>EKEPIVGA<sup>87</sup> .

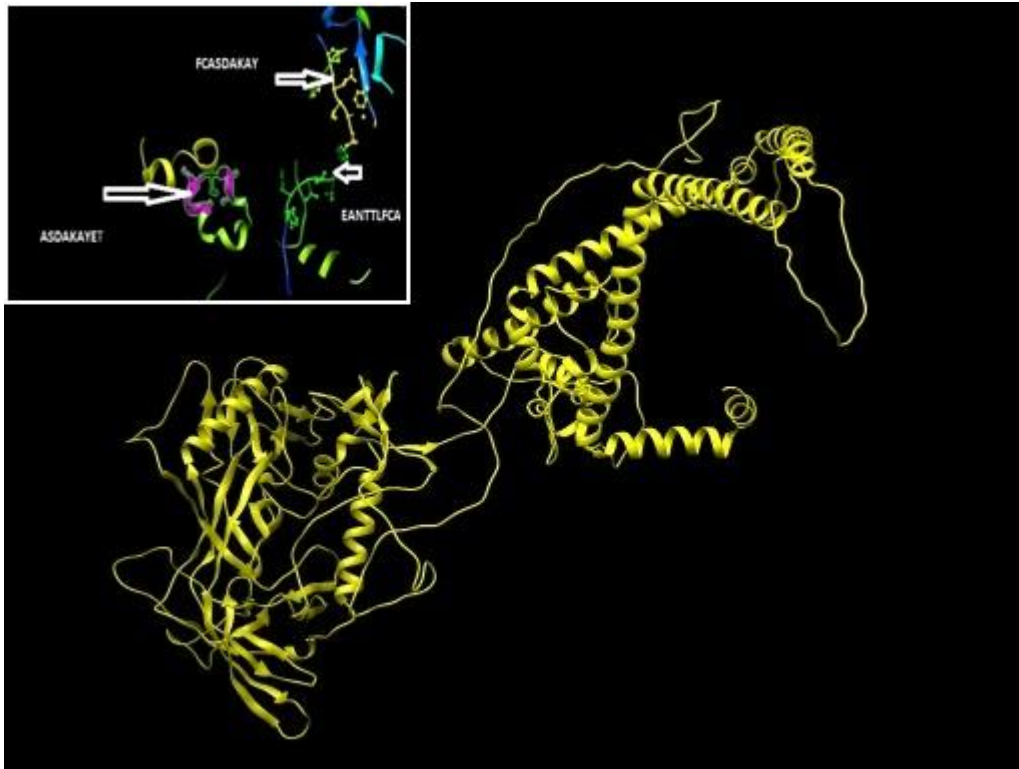
**Table 3: T cell MHC I**

Type	Allels	Start	End	Peptide	IC50
P10	HLA-A*01:01	1	9	ASDAKAYET	6550.91
	HLA-A*02:01	2	10		22617.1
	HLA-A*02:06	3	11		15291.4
	HLA-A*03:01	5	13		35840.5
	HLA-A*11:01	6	14		19071.5
	HLA-A*23:01	8	16		41574.3
	HLA-A*24:02	9	17		44720.3
	HLA-A*25:01	10	18		38771.4
	HLA-A*26:01	11	19		34380.7
	HLA-A*29:02	12	20		28157.9
	HLA-A*30:01	14	22		22186.4
	HLA-A*30:02	15	23		15192.9
	HLA-A*31:01	16	24		27697.4
	HLA-A*32:01	17	25		33368.1
	HLA-A*68:01	19	27		40802
	HLA-A*68:02	20	28		31124.6
	HLA-B*07:02	22	30		31386.7
	HLA-B*08:01	23	31		34477.2
	HLA-B*14:02	24	32		30692.2
	HLA-B*15:01	26	34		37307.6
	HLA-B*18:01	27	35		36324.9
	HLA-B*27:05	28	36		32121.2
	HLA-B*35:01	29	37		28876.8
	HLA-B*35:03	31	39		45802.1
	HLA-B*38:01	32	40		43091.6
	HLA-B*39:01	33	41		29277.9
	HLA-B*40:01	35	43		27886
	HLA-B*40:02	35	43		30107.1
	HLA-B*44:02	37	45		35715
	HLA-B*44:03	39	47		44467
	HLA-B*46:01	40	48		31376.8
	HLA-B*48:01	41	49		38134
	HLA-B*51:01	43	51		36904.5
	HLA-B*53:01	44	52		41255.6
	HLA-B*57:01	45	53		21268.8
	HLA-B*58:01	46	54		14381.8
	HLA-B*58:02	47	55		32581.8
	HLA-C*03:03	48	56		20664
	HLA-C*04:01	50	58		21804.9
	HLA-C*05:01	51	59		3197.57
HLA-C*06:02	53	61	35086.9		
HLA-C*07:01	54	62	34394.5		
HLA-C*07:02	56	64	42056.1		
HLA-C*08:02	56	64	3606.47		
HLA-C*12:03	58	66	31015		
HLA-C*14:02	60	68	37750.6		
HLA-C*15:02	60	68	3030.53		
HLA-E*01:01	61	69	33146.5		
P21	HLA-A*01:01	29	37	AAGSTMGAA	21669.68
	HLA-A*02:01	29	37		22489.96

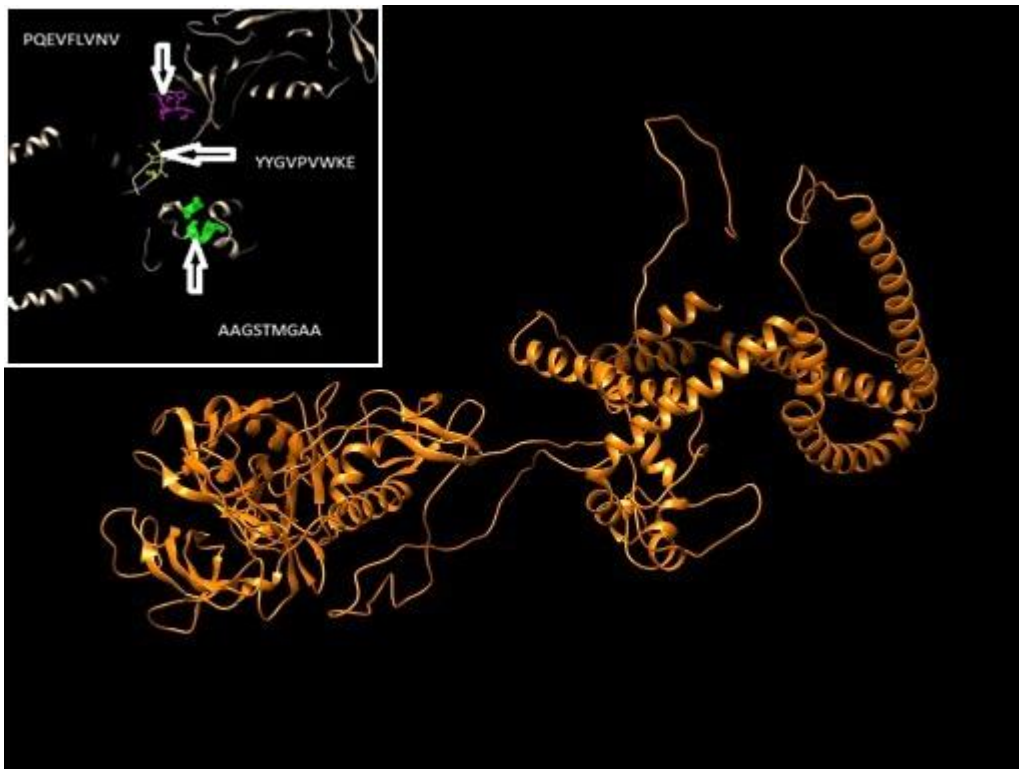
	HLA-A*02:06	29	37		7077.58
	HLA-A*03:01	29	37		27921.87
	HLA-A*11:01	29	37		25764.51
	HLA-A*23:01	29	37		43932.35
	HLA-A*24:02	29	37		45290.52
	HLA-A*25:01	29	37		28076.67
	HLA-A*26:01	29	37		22191.44
	HLA-A*29:02	29	37		31607.49
	HLA-A*30:01	29	37		4975.53
	HLA-A*30:02	29	37		4627.86
	HLA-A*31:01	29	37		33250.28
	HLA-A*32:01	29	37		44890.48
	HLA-A*68:01	29	37		43386.27
	HLA-A*68:02	29	37		23074.13
	HLA-B*07:02	29	37		3462.62
	HLA-B*08:01	29	37		26171.06
	HLA-B*14:02	29	37		6865.34
	HLA-B*15:01	29	37		10540.55
	HLA-B*15:02	29	37		32196.69
	HLA-B*18:01	29	37		34263.76
	HLA-B*27:05	29	37		31952.03
	HLA-B*35:01	29	37		13316.33
	HLA-B*35:03	29	37		46211.71
	HLA-B*39:01	29	37		29455.85
	HLA-B*40:01	29	37		25910.71
	HLA-B*40:02	29	37		35477.38
	HLA-B*44:02	29	37		30584.78
	HLA-B*44:03	29	37		43716.59
	HLA-B*46:01	29	37		18806.18
	HLA-B*48:01	29	37		39003.64
	HLA-B*51:01	29	37		31125.25
	HLA-B*53:01	29	37		37891.77
	HLA-B*57:01	29	37		19363.68
	HLA-B*58:01	29	37		19822.41
	HLA-B*58:02	29	37		39589.09
	HLA-C*03:03	29	37		7969.87
	HLA-C*04:01	29	37		23049.17
	HLA-C*05:01	29	37		32721.72
	HLA-C*06:02	29	37		43038.42
	HLA-C*07:01	29	37		39252.57
	HLA-C*07:02	29	37		41778.98
	HLA-C*08:02	29	37		25734.7
	HLA-C*12:03	29	37		19312.84
	HLA-C*14:02	29	37		26701.35
	HLA-C*15:02	29	37		31794.78
	HLA-E*01:01	29	37		44165.89
P51	HLAA*01:01	70	78	PPLVKLWYQ	28659.53
	HLAA*02:01	70	78		42240.3
	HLA-A*02:06	70	78		40383.8
	HLA-A*03:01	70	78		39480.9
	HLA-A*11:01	70	78		37117.9
	HLA-A*23:01	70	78		43669.8
	HLA-A*24:02	70	78		45103.7

HLA-A*25:01	70	78	47522
HLA-A*26:01	70	78	45153.5
HLA-A*29:02	70	78	36471.4
HLA-A*30:01	70	78	35786.6
HLA-A*30:02	70	78	38541.8
HLA-A*31:01	70	78	30897.4
HLA-A*32:01	70	78	44932.7
HLA-A*33:03	70	78	30362.5
HLA-A*68:01	70	78	37656.8
HLA-A*68:02	70	78	44876.9
HLA-A*74:01	70	78	41444
HLA-B*07:02	70	78	39124.5
HLA-B*08:01	70	78	27134.71
HLA-B*14:02	70	78	30915.81
HLA-B*15:01	70	78	40519.97
HLA-B*15:02	70	78	41754.11
HLA-B*18:01	70	78	31651.65
HLA-B*27:05	70	78	40488.86
HLA-B*35:01	70	78	25955.04
HLA-B*35:03	70	78	29749.87
HLA-B*38:01	70	78	45755.97
HLA-B*39:01	70	78	35098.63
HLA-B*40:01	70	78	33807.12
HLA-B*40:02	70	78	44794.41
HLA-B*44:02	70	78	42183.21
HLA-B*44:03	70	78	47158.29
HLA-B*48:01	70	78	36140.28
HLA-B*46:01	70	78	30652.35
HLA-B*51:01	70	78	24535.34
HLA-B*53:01	70	78	16808.18
HLA-B*57:01	70	78	30827.97
HLA-B*58:01	70	78	28839.33
HLA-B*58:02	70	78	37189.89
HLA-C*03:03	70	78	44548.88
HLA-C*04:01	70	78	28983.22
HLA-C*05:01	70	78	47401.79
HLA-C*06:02	70	78	40127.73
HLA-C*07:01	70	78	37536.75
HLA-C*07:02	70	78	41852.27
HLA-C*08:02	70	78	43381.12
HLA-C*12:03	70	78	46617.48
HLA-C*14:02	70	78	44622.68
HLA-C*15:02	70	78	38660.8
HLA-E*01:01	70	78	43346.39

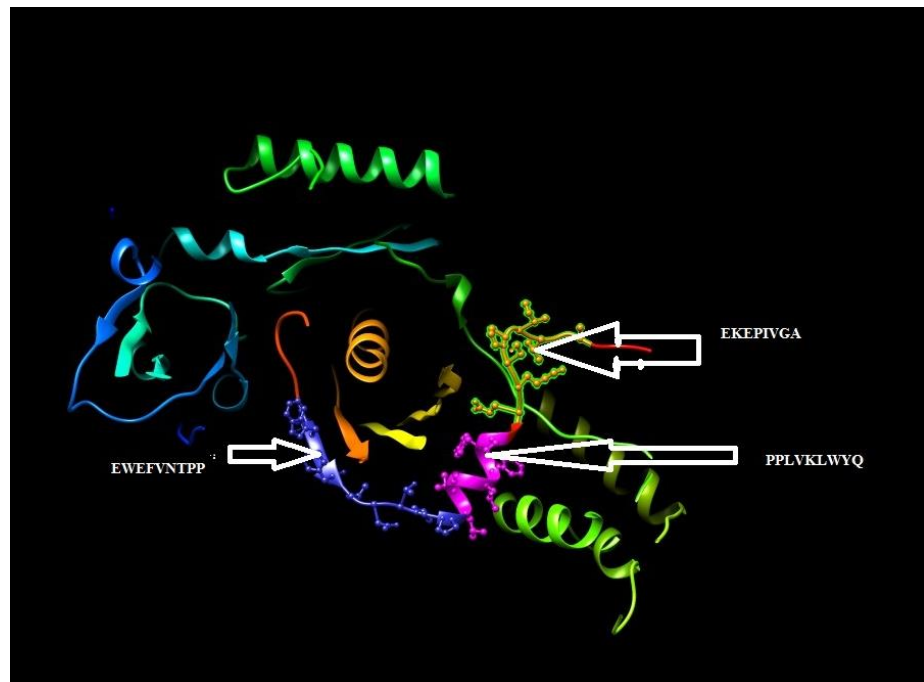
Figuer7: prediction peptides with MHC I of p10 by chimera 1.8



Figuer8: prediction peptides with MHC I of p21 by chimera 1.8



Figuer9: prediction peptides with MHC I of p51 by chimera 1.8



## Prediction of T Helper cell epitopes and interaction with MHC class II

Using MHC class II binding prediction method of IEDB and based on NN align and IC50 of 500 for structural polyprotein. From all predicted epitopes of **structural polyprotein** that bind to MHC class II, the core sequence<sup>119</sup>**IISLWDQSL**<sup>127, 108</sup>**CVKLTPLCV**<sup>116</sup> bind to (25) alleles. which is largest number of alleles in contrast to other predicted, core sequences in p10. the core sequence <sup>38</sup>**YGVVPVWKE**<sup>52</sup>, <sup>17</sup>**FNMWKNNMV**<sup>31</sup>, <sup>20</sup>**LLQYWSQEL**<sup>34</sup> bind to 43 alleles. which is largest number of alleles in contrast to other predicted, core sequences in p21. the core sequence <sup>7</sup>**WKGSPAIFQ**<sup>21</sup> that showed affinity for 27 MHC- II molecules, <sup>11</sup>**FLWMGYEL**<sup>25</sup> that showed affinity for 43 MHC- II molecules ,and <sup>58</sup>**WEFVNTPPL**<sup>72</sup> that showed affinity for 37 MHC- II molecules .

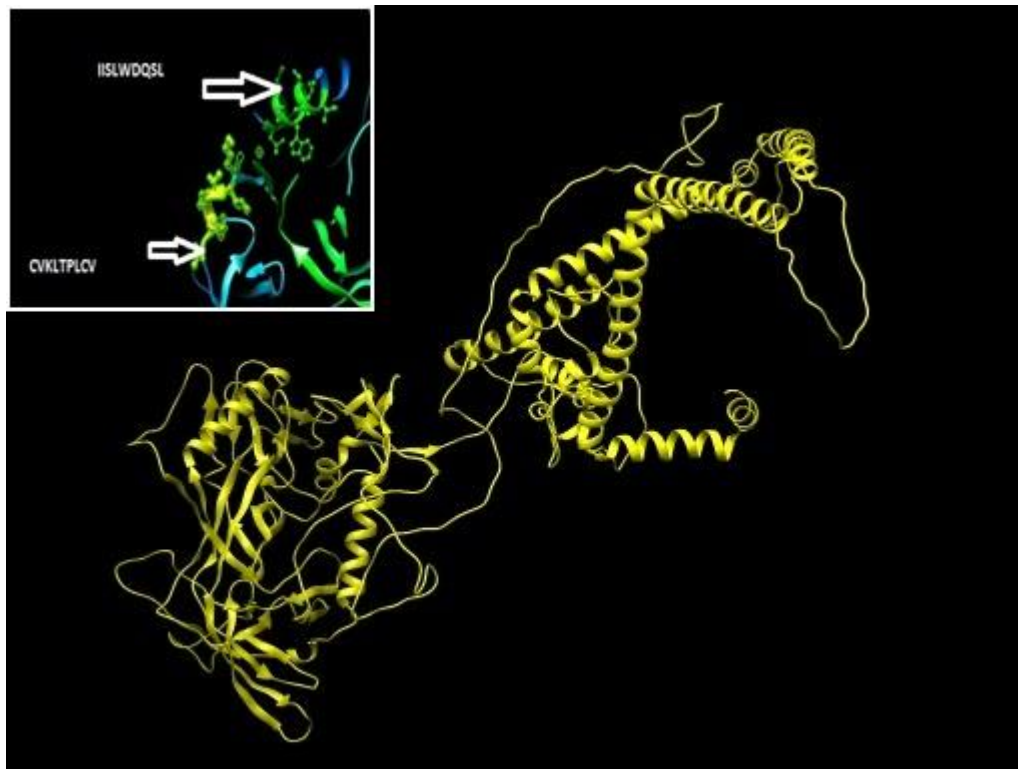
Table 4: T cell MHC II of HIV

Type	Allels	Start	End	Core	Peptide	IC50
P10	HLADRB1*01:01	46	60	CVKLTPLCV	LKPCVKLTPLCVTLN	21.9
	HLADRB1*01:01	45	59		SLKPCVKLTPLCVTL	25
	HLADRB1*01:01	44	58		QSLKPCVKLTPLCVT	33.8
	HLADRB1*01:01	47	61		KPCVKLTPLCVTLNC	36
	HLADRB1*01:01	43	57		DQSLKPCVKLTPLCV	44.5
	HL-DRB1*01:01	48	62		PCVKLTPLCVTLNCT	64.7
	HLADRB1*01:01	49	63		CVKLTPLCVTLNCTD	88
	HLADRB1*07:01	43	57		DQSLKPCVKLTPLCV	40.3
	HLADRB1*07:01	44	58		QSLKPCVKLTPLCVT	52.7
	HLADRB1*07:01	45	59		SLKPCVKLTPLCVTL	58.2

	HLADRB1*07:01	46	60		LKPCVKLTPLCVTLN	85.5
	HLADRB1*07:01	47	61		KPCVKLTPLCVTLNC	123.3
	HLADRB1*07:01	48	62		PCVKLTPLCVTLNCT	191.6
	HLADRB1*07:01	49	63		CVKLTPLCVTLNCTD	210.9
	HLADRB1*09:01	46	60		LKPCVKLTPLCVTLN	308.2
	HLADRB1*09:01	47	61		KPCVKLTPLCVTLNC	319.3
	HLADRB1*09:01	45	59		SLKPCVKLTPLCVTL	391
	HLADRB1*09:01	48	62		PCVKLTPLCVTLNCT	399.8
	HLADRB1*09:01	44	58		QSLKPCVKLTPLCVT	470.2
	HLADRB1*09:01	49	63		CVKLTPLCVTLNCTD	479.1
	HLADRB1*11:01	46	60		LKPCVKLTPLCVTLN	289
	HLADRB1*15:01	46	60		LKPCVKLTPLCVTLN	307.5
	HLADRB1*15:01	47	61		KPCVKLTPLCVTLNC	308
	HLADRB1*15:01	45	59		SLKPCVKLTPLCVTL	363.6
	HLADRB1*15:01	48	62		PCVKLTPLCVTLNCT	391.9
P21	HLADRB1*01:01	16	30	FNMWKNMNV	VNVTENFNMWKNMNV	61
	HLADRB1*04:01	16	30		VNVTENFNMWKNMNV	374.9
	HLADRB1*04:04	16	30		VNVTENFNMWKNMNV	60.4
	HLADRB1*04:05	16	30		VNVTENFNMWKNMNV	136.7
	HLADRB1*07:01	16	30		VNVTENFNMWKNMNV	19.2
	HLADRB1*015:01	16	30		VNVTENFNMWKNMNV	217.1
	HLADRB1*01:01	17	31		NVTENFNMWKNMNV	60.7
	HLADRB1*04:01	17	31		NVTENFNMWKNMNV	490.9
	HLADRB1*04:04	17	31		NVTENFNMWKNMNV	28.9
	HLADRB1*04:05	17	31		NVTENFNMWKNMNV	153.9
	HLADRB1*07:01	17	31		NVTENFNMWKNMNV	25.5
	HLADRB1*15:01	17	31		NVTENFNMWKNMNV	233.3
	HLADRB1*01:01	18	32		VTENFNMWKNMNV	38.5
	HLADRB1*04:01	18	32		VTENFNMWKNMNV	419.6
	HLADRB1*04:04	18	32		VTENFNMWKNMNV	38.4
	HLADRB1*04:05	18	32		VTENFNMWKNMNV	170.9
	HLADRB1*07:01	18	32		VTENFNMWKNMNV	34.1
	HLADRB1*15:01	18	32		VTENFNMWKNMNV	229.4
	HLADRB1*01:01	19	33		TENFNMWKNMNV	415.1
	HLADRB1*04:01	19	33		TENFNMWKNMNV	20.1
	HLA-DRB1*4:04	19	33		TENFNMWKNMNV	316.8
	HLA-DRB1*4:05	19	33		TENFNMWKNMNV	26.3
	HLA-DRB1*7:01	19	33		TENFNMWKNMNV	152.9
	HLADRB1*02:01\DPB1*01:01	19	33		TENFNMWKNMNV	36.1
	HLADRB1*01:01	19	33		TENFNMWKNMNV	340
P51	HLADRB1*01:01	7	21	WKGSPAIFQ	NVLPQGWKGSPAIFQ	277
	HLADRB1*07:01	7	21		NVLPQGWKGSPAIFQ	380
	HLADRB1*09:01	7	21		NVLPQGWKGSPAIFQ	249.8
	HLADRB1*01:01	8	22		VLPQGWKGSPAIFQS	162.1
	HLADRB1*07:01	8	22		VLPQGWKGSPAIFQS	424.6
	HLADRB1*09:01	8	22		VLPQGWKGSPAIFQS	193.5
	HLADQA:01/DQB1*03:1	9	23		LPQGWKGSPAIFQSS	42.3
	HLADRB1*01:01	9	23		LPQGWKGSPAIFQSS	86.5
	HLADRB1*07:01	9	23		LPQGWKGSPAIFQSS	464.5
	HLADRB1*09:01	9	23		LPQGWKGSPAIFQSS	154.7
	HLADQA1*01:02/DQB1*06:02	10	24		PQGWKGSPAIFQSSM	184.8
	HLADRB1*01:01	10	24		PQGWKGSPAIFQSSM	41.2
	HLADRB1*04:01	10	24		PQGWKGSPAIFQSSM	471.4

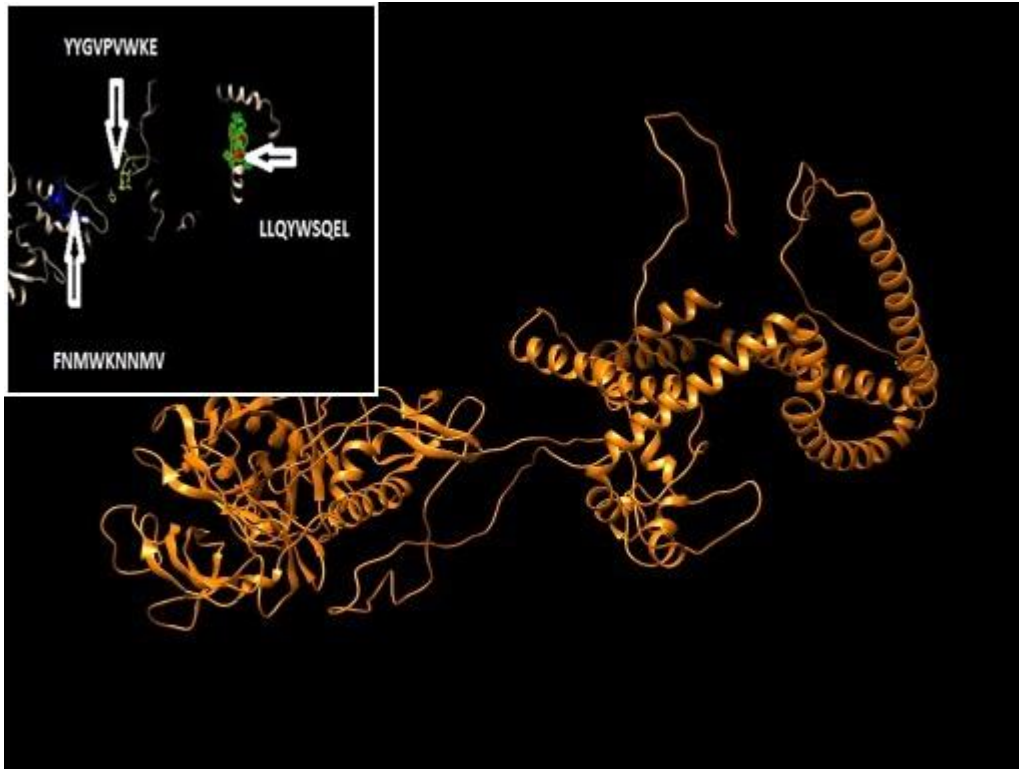
HLADRB1*07:01	10	24	PQGWKGSPAIFQSSM	438.7
HLADRB1*09:01	10	24	PQGWKGSPAIFQSSM	135.2
HLADRB1*11:01	10	24	PQGWKGSPAIFQSSM	395.6
HLADRB5*01:01	10	24	PQGWKGSPAIFQSSM	472.5
HLADQA1*05:01/DQB1*03:01	11	25	QGWKGSPAIFQSSMT	66.3
HLADRB1*01:01	11	25	QGWKGSPAIFQSSMT	59.9
HLADRB1*07:01	11	25	QGWKGSPAIFQSSMT	491.4
HLADRB1*09:01	11	25	QGWKGSPAIFQSSMT	220
HLADRB1*01:01	12	26	GWKGSPAIFQSSMTK	115.9
HLADRB1*04:01	12	26	GWKGSPAIFQSSMTK	359.7
HLADRB1*07:01	12	26	GWKGSPAIFQSSMTK	486
HLADRB1*09:01	12	26	GWKGSPAIFQSSMTK	268.8
HLADRB1*01:01	13	27	WKGSPAIFQSSMTKI	152.5

Figuer10: prediction peptides with MHC II of p10 by chimera 1.8

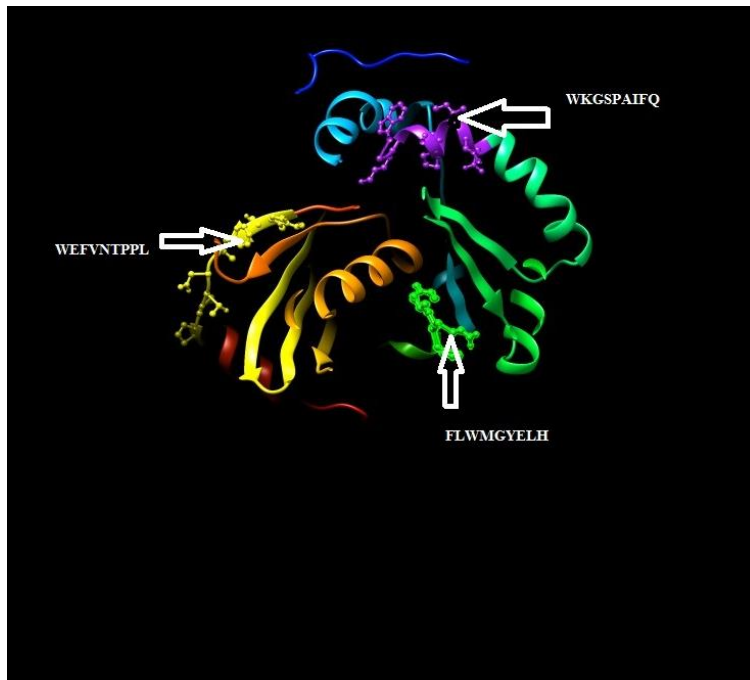




Figuer11: prediction peptides with MHC II of p21 by chimera 1.8



Figuer12: prediction peptides with MHC II of p51 by chimera 1.8



## population coverage:

Analysis of population coverage all epitopes from the reverse transcriptase proteins which selected as epitopes with high affinity to interact with MHC class I and MHC class II, subjected to IEDB population coverage set against the whole world population. For MHC class I, three Epitopes with highest population coverage resulted in epitope set of 88.87% . In MHC class II, towepitopes that interact with class II alleles. resulted in epitope set of 97.24% .for combined population coverage was also calculated for combined MHC class I and class II using proposed epitopes from reserve transcriptase protein which resulted in epitope set of 99.86%.

In p21 For MHC class I, three Epitopes with highest population coverage were . YYGVPVWKE,PQEVFLNVN and AAGSTMGAA resulted in epitope set of 97.55%. In MHC class II, three epitopes that interact with class II alleles YYGVPVWKE, FNMMWKNMNV , LLQYWSOEL resulted in epitope set of 95.94% for combined population coverage The data demonstrated in Table7 Figure 4.3 Prediction of population coverage was also calculated for combined MHC class I and class II using proposed epitopes from structural polyprotein which resulted in epitope set of 99.62%.

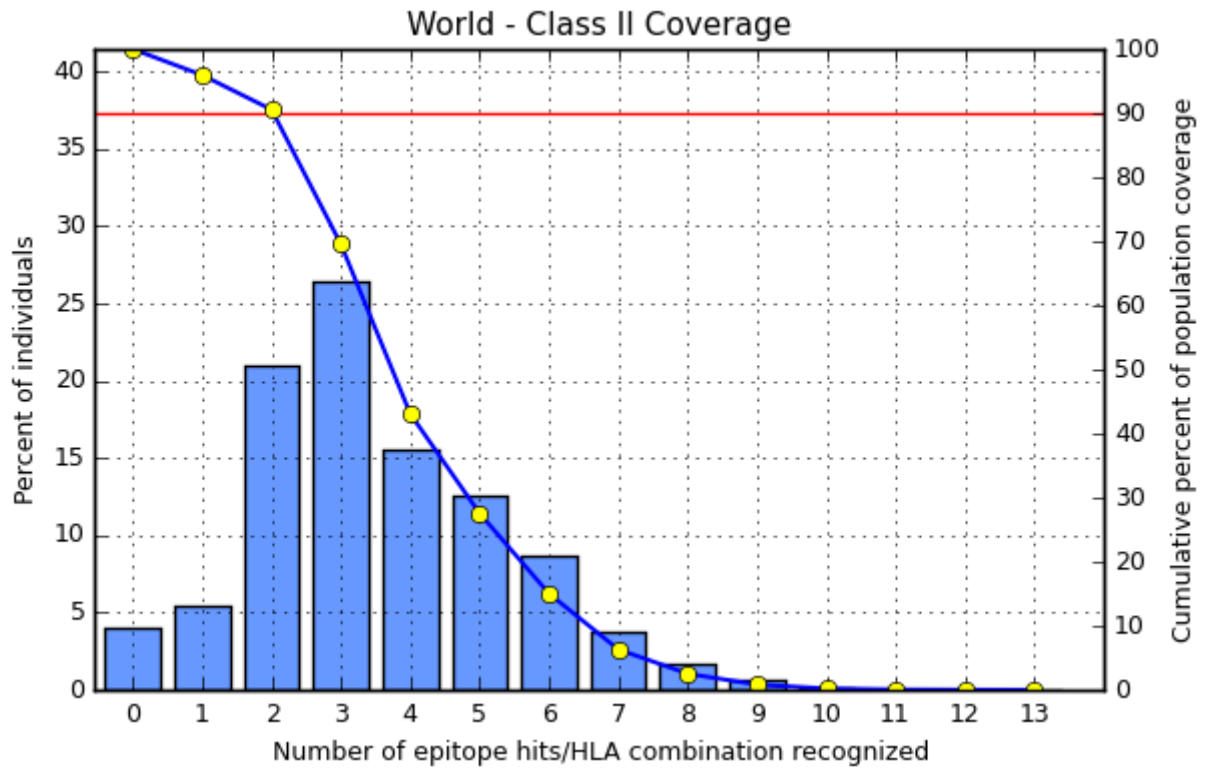
In p51 For MHC class I, five Epitopes with highest population coverage resulted in epitope set of 94.45%. In MHC class II, five epitopes that interact with class II alleles. resulted in epitope set of 93.6%.for combined population coverage was also calculated for combined MHC class I and class II using proposed epitopes from reserve transcriptase protein which resulted in epitope set of 99.76%.

**Table 5: population coverage**

Population \area	Class I			Class II			Class combined		
	coverage	average	Pc90	Coverage	average	Pc90	Coverage	Average	Pc90
World	88.87%	4.66	0.9	97.24%	3.73	2.1	99.68%	4.39	2.66
P10	97.55%	5.75	2.05	95.94%	3.52	2.03	99.62%	4.93	2.47
P21	94.45%	5.23	1.52	93.6%	3.21	1.19	99.76	5.83	3.75

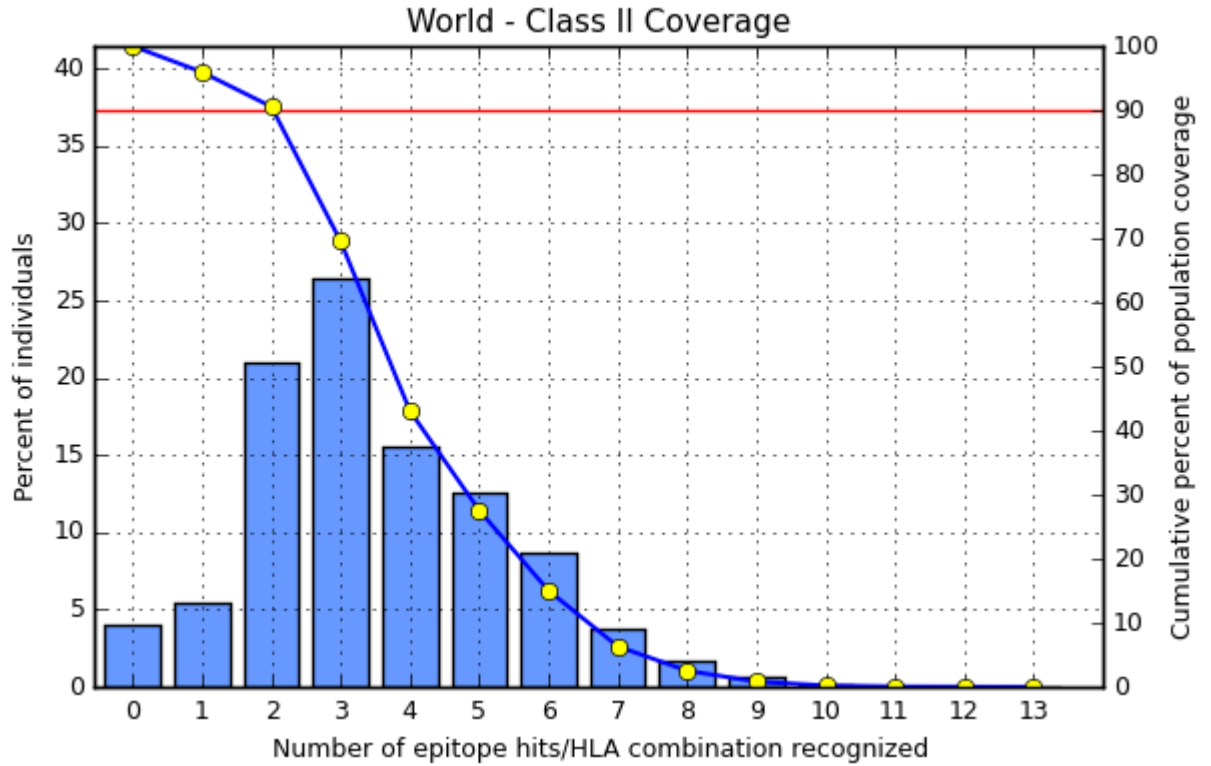


Figuer13: population coverage MHC class II of p10 HIV

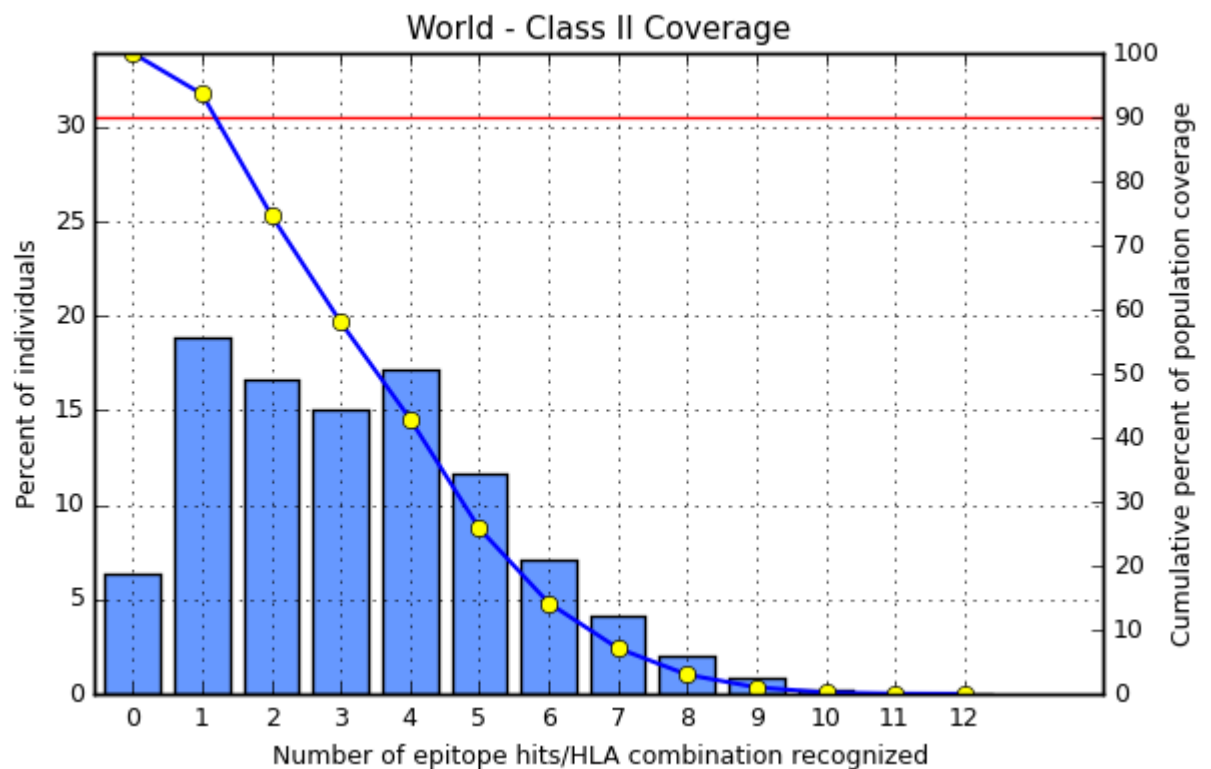




Figuer14: population coverage MHC class II of p21 HIV



Figuer15: population coverage MHC class II of p51 HIV

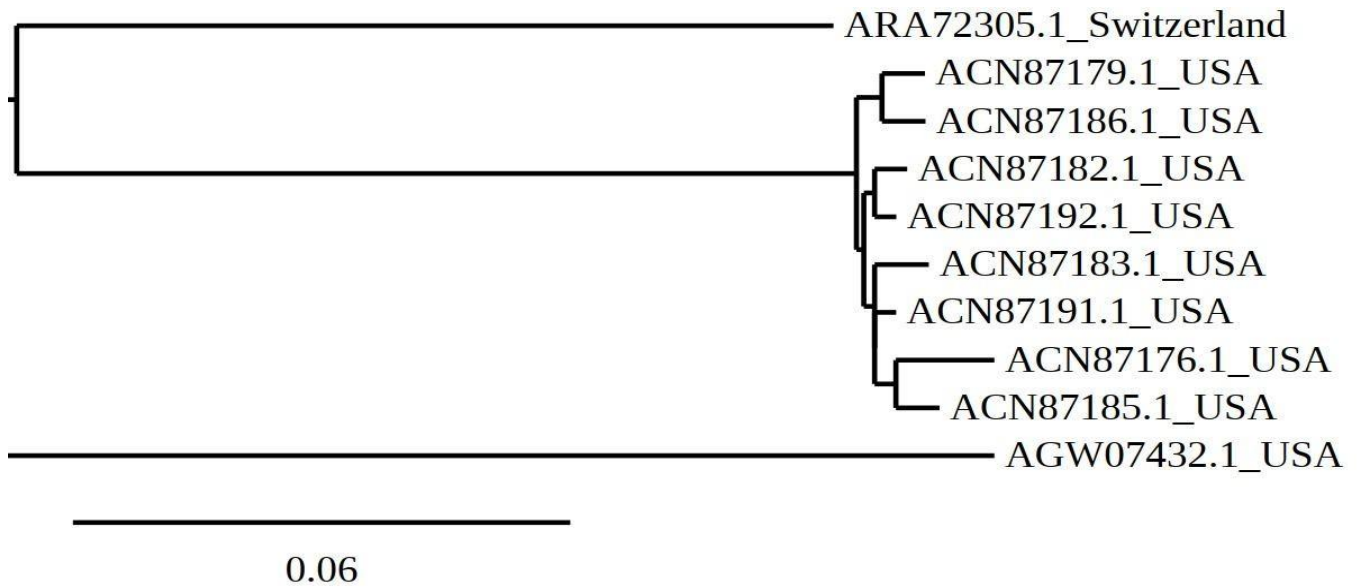




## Phylogenetic analysis:

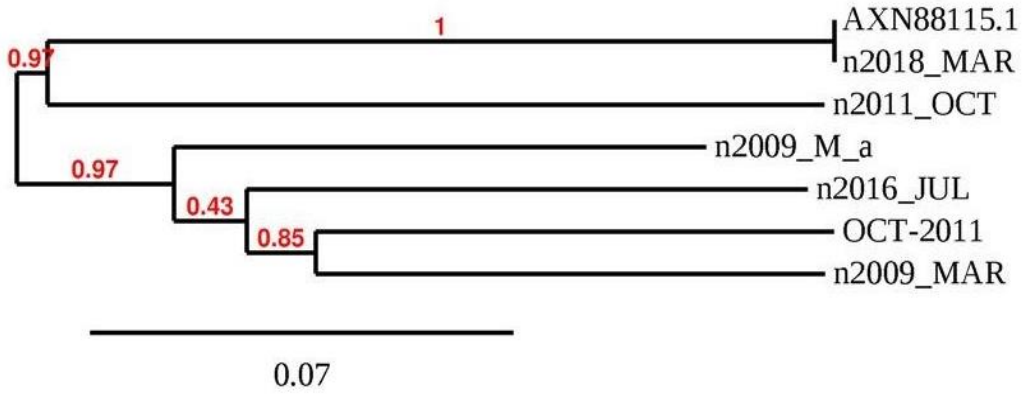
The retrieved sequences will be conducted in Phylogenetic to determine the common ancestor of each strain using different tools from (<http://www.phylogeny.fr>)

figuer16: *Phylogenetic analysis of p10*



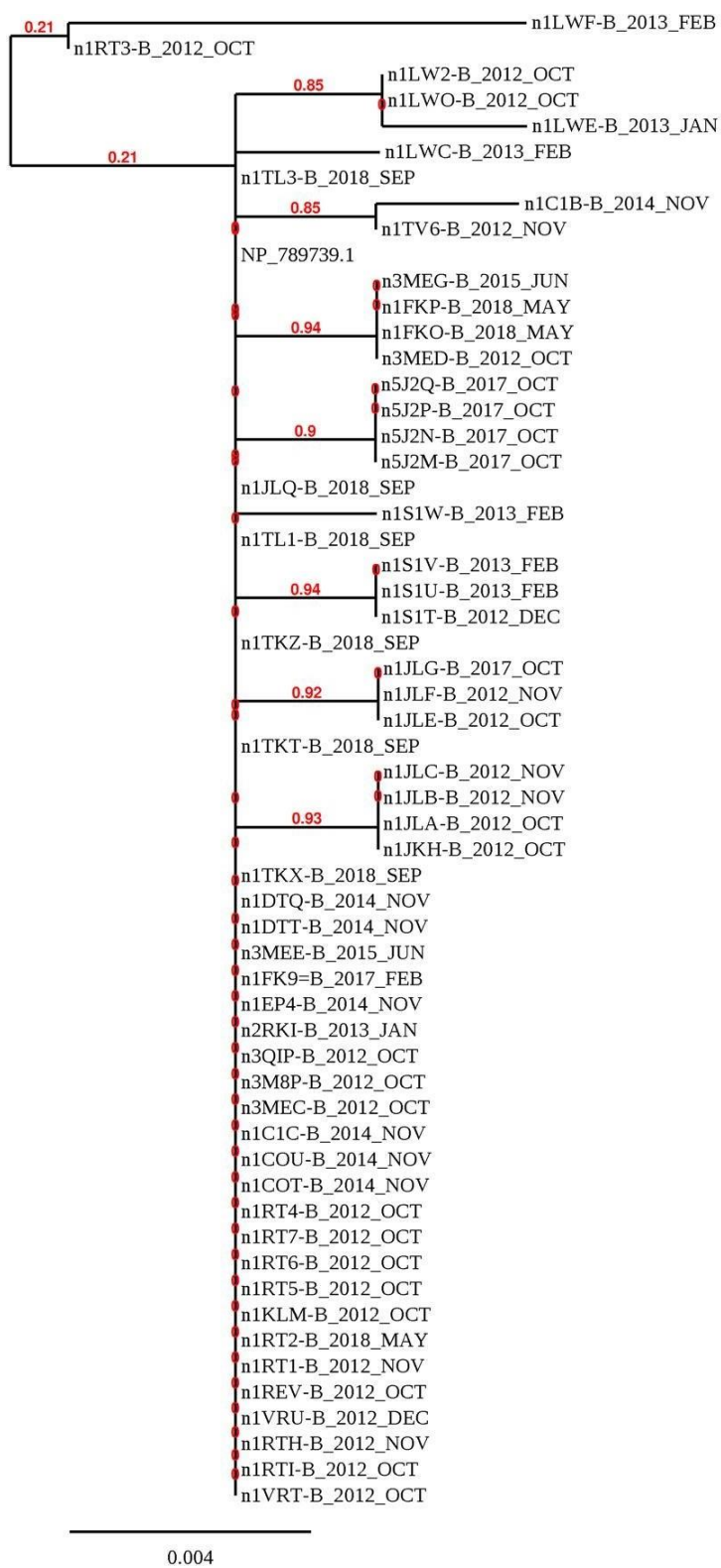


Figuer17: Phylogenetic analysis for retrieved sequences p21





Figuer 18: Phylogenetic analysis for retrieved sequence p51





## Discussion:

In this study we aimed to predict possible peptide vaccine for HIV, regarding epitopes that would elicit an antibody immune response "B-cell epitopes <sup>708</sup>QGYSP<sup>712</sup> , <sup>708</sup>QGYSP<sup>712,73</sup>CVPTDPNPQ<sup>81</sup>, <sup>346</sup>FKNL<sup>349</sup> , is proposed as they got results above thresholds in Bepipred linear epitope prediction, Emini surface accessibility and Kolaskar and Tongaonkar antigenicity prediction methods and have shown 100% conservancy among the (HIV P10,P21, P51) retrieved sequences.

There were more epitopes that was addressed as potentially promising epitopes as they bound the highest number of alleles in MHC I .For p10, <sup>47</sup>EANTTLFCA<sup>55</sup>, <sup>53</sup>FCASDAKAY<sup>61</sup>, <sup>55</sup>,ASDAKAYET<sup>63</sup> ,from p21 <sup>38</sup>YYGVPVWKE<sup>46</sup>, <sup>10</sup> PQEVFLVNV<sup>18</sup> and <sup>29</sup>AAGSTMGAA<sup>37</sup>, and <sup>63</sup>EWEFVNTPP<sup>71</sup> epitope was found to bind 51 MHCI Alleles ,<sup>70</sup>PPLVKLWYQ<sup>78</sup> epitope was found to bind 50 MHC I alleles and ,<sup>79</sup>LEKEPIVGA<sup>87</sup> epitope was found to bind 51 MHCI alleles.

All epitopes were found to have a high binding affinity and low binding energy to MHC class I molecule in the structural level, but The<sup>63</sup> EWEFVNTPP<sup>71</sup>,

<sup>70</sup>PPLVKLWYQ<sup>78</sup> ,, <sup>79</sup>LEKEPIVGA<sup>87</sup> epitopes were visualized because they bind the largest number of alleles, its importance as a component of the designed peptide vaccine. For the MHC II binding peptides, <sup>119</sup>IISLWDQSL<sup>127</sup>, <sup>108</sup>CVKLTPLCV<sup>116</sup>for p10 , <sup>38</sup>YYGVPVWKE<sup>46</sup>, <sup>16</sup>FNMMWKNNMV<sup>30</sup>, <sup>17</sup>LLQYWSOEL<sup>31</sup>,from p21, 174 conserved epitopes were predicted to interact with different human MHC II alleles. We postulated the best two; the 9-mer peptide (core)<sup>7</sup>WKGSPAIFQ<sup>21</sup> that showed affinity for 27 MHC- II molecules, <sup>11</sup>FLWMGYEL<sup>25</sup> that showed affinity for 43 MHC- II molecules ,and <sup>58</sup>WEFVNTPPL<sup>72</sup> that showed affinity for 37 MHC- II molecules .

population coverage all epitopes from the reverse transcriptase proteins which selected as epitopes with high affinity to interact with MHC class I and MHC class II, subjected to IEDB population coverage set against the whole world population. In p10 six epitopes of MHC I&MHC II was high affinity give 99.68% whole world , In p21 six epitopes of MHC I&MHC II was high





affinity give 99.62% whole world ,and six epitopes of MHC I&MHC II was high affinity give 99.76% whole world in p51.

**QGYSP** from 704 to 708 was found to have greatest score among all other peptides ,and one peptides had got results above thresholds in Bepipred linear epitope prediction, Emini surface accessibility and Kolaskar and Tongaonkar antigenicity prediction methods and have shown 100% conservancy among the (HIV P10,P21) retrieved sequences.

The peptides found in the present study may prove more immunogenic as compared to the earlier reported peptides. Predicted peptides might show the physicochemical instability, to overcome this limitation, several structural as well as physical modification strategies are available to enhance the poor physicochemical stability of peptides. these strategies are including peptidomimetic approach, prodrug approach, analogue formations, hydrophobic ion pairing, conjugation with fatty acids and use of substitute methods of drug administration.

Researchers have been working to gather data linked to HIV to understand its biology, transmission and patho-physiology in order to eliminate the disease completely. In the near future, we anticipate that predicted epitopes have therapeutic potential with an outstanding scope.

In recent studies, novel antigenic epitopes of some essential and vital proteins revealed that can victoriously elicit response of immune system therefore becoming great peptide vaccines targets and protecting host from virus attack. So, the current research was conducted to predict antigenic epitopes of reverse transcriptaseP51 (RT) HIV. We carry out sequence, structure, and conservation analysis as well as homology modeling of reverse transcriptase P51 (RT) of HIV.

Research discloses T and B cell epitopes which are antigenic and conserved among the HIV isolates of different countries. These epitopes are capable to induce a particular immunologic response. Interestingly few of the antigenic epitopes proposed and investigated in this work might present a preliminary set of peptides for future vaccine development against HIV for control and prevention of this devastating epidemic .

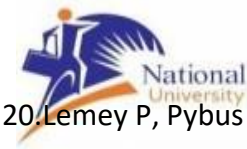


## References

1. Luciw PA. Human immunodeficiency viruses and their replication. In: Fields BN, editor. *Virology*. 3rd ed. Philadelphia: Lippincott-Raven; 1996. pp. 1881–1952. [[Google Scholar](#)]
2. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*. 1999;397:436–441. [[PubMed](#)] [[Google Scholar](#)]
3. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*. 2011;1:a006841. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
4. Faria NR, Rambant A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, P  pin J, Posada D, Peeters M, Pybus OG, Lemey P. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science*. 2014;346:56–61. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
5. Levy JA. *HIV and the Pathogenesis of AIDS*. 3rd ed. Washington: ASM Press; 2007. [[Google Scholar](#)]
6. Sauter D, Unterweger D, Vogl M, Usmani SM, Heigele A, Kluge SF, Hermkes E, Moll M, Barker E, Peeters M, Learn GH, Bibollet-Ruche F, Fritz JV, Fackler OT, Hahn BH, Kirchhoff F. Human tetherin exerts strong selection pressure on the HIV-1 group N Vpu protein. *PLoS Pathog*. 2012;8:e1003093. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
7. Vincenzi E, Poli G. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *Tissue Antigens*. 2013;81:61–71. [[PubMed](#)] [[Google Scholar](#)]
8. Kuiken C, Leitner T, Hahn B, Mullins J, Wolinsky S, Foley B, Apetrei C, Mizrahi I, Rambaut A, Korber B. HIV Sequence Compendium 2012. Los Alamos, NM: Los Alamos National Lab; [www.hiv.lanl.gov](http://www.hiv.lanl.gov). [[Google Scholar](#)]
9. Keele BF, Van Heuverswyn F, Li Y, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*. 2006 published online May 25, 2006. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
10. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull*. 2001;58:19–42. [[PubMed](#)] [[Google Scholar](#)]



11. Thomson MM, Najera R. Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev.* 2005;7:210–24. [[PubMed](#)] [[Google Scholar](#)]
12. Blackard JT, Cohen DE, Mayer KH. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. *Clin Infect Dis.* 2002;34:1108–14. [[PubMed](#)] [[Google Scholar](#)]
13. Steain MC, Wang B, Dwyer DE, Saksena NK. HIV-1 co-infection, superinfection and recombination. *Sex Health.* 2004;1:239–50. [[PubMed](#)] [[Google Scholar](#)]
14. Yerly S, Jost S, Monnat M, et al. HIV-1 co/super-infection in intravenous drug users. *AIDS.* 2004;18:1413–21. [[PubMed](#)] [[Google Scholar](#)]
15. Chohan B, Lavreys L, Rainwater SM, Overbaugh J. Evidence for frequent reinfection with human immunodeficiency virus type 1 of a different subtype. *J Virol.* 2005;79:10701–08. [[PMC free article](#)][[PubMed](#)] [[Google Scholar](#)]
16. Smith DM, Richman DD, Little SJ. HIV superinfection. *J Infect Dis.* 2005;192:438–
44. [[PubMed](#)] [[Google Scholar](#)]
17. McCutchan FE, Hoelscher M, Tovanabutra S, et al. In-depth analysis of a heterosexually acquired human immunodeficiency virus type 1 superinfection: evolution, temporal fluctuation, and intercompartment dynamics from the seronegative window period through 30 months postinfection. *J Virol.* 2005;79:11693–704. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]
18. Gottlieb GS, Nickle DC, Jensen MA, et al. Dual HIV-1 infection associated with rapid disease progression. *Lancet.* 2004;363:619–22. [[PubMed](#)][[Google Scholar](#)]
19. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis.* 2011;11:45–56. [[PubMed](#)] [[Google Scholar](#)]



20. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci U S A*. 2003;100:6588–6592. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

21. Mourez T, Simon F, Plantier JC. Non-M variants of human immunodeficiency virus type 1. *Clin Microbiol Rev*. 2013;26:448–461. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

22. Locatelli S, Peeters M. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. *AIDS*. 2012;26:659–673. [[PubMed](#)] [[Google Scholar](#)]

23. Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairoj AS, Brown TM, Salemi M, Vandamme AM, Kalish ML. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol*. 2003;77:6359–6366. [[PMC free article](#)] [[PubMed](#)] [[Google Scholar](#)]

24. UNAIDS Global report: UNAIDS report on the global AIDS epidemic 2013 [www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS\\_Global\\_Report\\_2013\\_en.pdf](http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf)

25. Inciardi JA, Williams ML. Editor’s introduction: the global epidemiology of HIV and AIDS. *AIDS Care*. 2005;17 (suppl 1):S1–8. [[PubMed](#)] [[Google Scholar](#)]

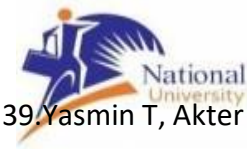
26. Sudan National AIDS program (SNAP), Situation Analysis, Behavioural and Epidemiological Surveys and Response Analysis, Sudan, FMOH SNAP, 2002.

27. Sudan National AIDS program (SNAP), “Sudan federal ministry of health, Global AIDS Response Progress Reporting 2010–2011,”

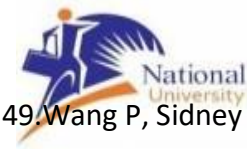
28. Sudan, “Assessing Risks to Stability, Richard Downie, Brian Kennedy, Center for Strategic and International S 13, 2011



29. Jump up to: a b c d e f g h i j k l m n o p WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children (PDF). Geneva: World Health Organization. 2007. pp. 6–16. ISBN 978-92-4-159562-9. Archived (PDF) from the original on October 31, 2013.
30. Diseases and disorders. Tarrytown, NY: Marshall Cavendish. 2008. p. 25. ISBN 978-0-7614-7771-6. Jump up to: a b c d e f g h i j k l m n o Mandell, Bennett, and Dolan (2010).
31. Jump up to: a b c d e f g h i j k l m n o p q r Vogel M, Schwarze-Zander C, Wasmuth JC, Spengler U, Sauerbruch T, Rockstroh JK (July 2010).
32. Gwyther, Lize. "Primary HIV/AIDS Care. A Practical Guide for Primary Health Care Personnel in the Clinical and Supportive Care of People with HIV/AIDS." (2008): 270.
33. Hu, Duoyi. "Stabilization of GP120 Outer Domain for HIV-1-1 Vaccine Immunogen Design." (2016).
34. Elliott T, Casey A, Lambert PA, Sandoe J. Lecture Notes: Medical Microbiology and Infection. John Wiley & Sons; 2012 May 29.
35. Blankson J. Control of HIV-1 replication in elite suppressors. *Discovery medicine*. 2010 Mar 24;9(46):261-6.
36. Walker BD. Elite control of HIV Infection: implications for vaccines and treatment. *Topics in HIV medicine: a publication of the International AIDS Society, USA*. 2007;15(4):134-6.
37. Holmes CB, Losina E, Walensky RP, Yazdanpanah Y, Freedberg KA. Review of human immunodeficiency virus type 1-related opportunistic infections in sub-Saharan Africa. *Clinical infectious diseases*. 2003 Mar 1;36(5):652-62.
38. Hall T, Carlsbad Ca (2011) BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences* 2: 60-61.



39. Yasmin T, Akter S, Debnath M, Ebihara A, Nakagawa T, et al. (2016) In silico proposition to predict cluster of B- and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane associated proteins of *C jejuni*. *In Silico Pharmacology* 4: 1-10.
40. Larsen JE, Lund O, Nielsen M (2006) Improved method for predicting linear B-cell epitopes. Jens Erik Pontoppidan Larsen, Ole Lund\* and Morten Nielsen. *BioMed Central* 2: 1-7.
41. Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of Hepatitis A Virus-Neutralizing Antibody by a Virus-Specific Synthetic Peptide. *J Virol* 55: 836-839.
42. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS* 276: 172-174.
43. Patronov A, Doytchinova I (2013) T-cell epitope vaccine design by immunoinformatics. Royal Society Publishing 1-13.
44. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, et al. (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science* 12: 1007-1017.
45. Tallmadge RL, Campbell JA, Miller DC, Antczak DF (2010) Analysis of MHC class I genes across horse MHC haplotypes. *Immunogenetics* 62: 159-172.
46. Szalai G, Antczak DF, Gerber H, Lazary S (1994) Molecular cloning and characterization of horse DQB cDNA. *Immunogenetics* 40: 458.
47. Brinkmeyer-Langford CL, Murphy WJ, Childers CP, Skow LC (2010) A conserved segmental duplication within ELA. *Anim Genet* 41: 186-195.
48. Nielsen M, Lund O (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10: 296.



49. Wang P, Sidney J, Kim Y, Sette A, Lund O, et al. (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics 11:568.

50. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC bioinformatics. 2006 Dec;7(1):153

51. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nature protocols. 2012 Aug;7(8):1511.

52. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612











