

Prediction of an Epitope-Based Vaccine Against Human Immunodeficiency Virus (HIV) an *in silico* Approach

Bashir Salim^{1*}, Nabila Mutassim Murad², Zahra Abdelmagid²

¹Faculty of Veterinary, University of Khartoum, Khartoum, Sudan

²School of Pharmacy, Ahfad University for Women, Omdurman, Sudan

***Corresponding author:** Bashir Salim, Department of Parasitology, Faculty of Veterinary Medicine, University of Khartoum, Sudan. Tel: +249912195616; E-Mail: bashirsalim@gmail.com

Citation: Salim B, Murad NM, Abdelmagid Z (2017) Prediction of an Epitope-Based Vaccine Against Human Immunodeficiency Virus (HIV) an *in silico* Approach. J Vaccines Immunol 2017: J110. DOI: 10.29011/2575-789X.000010

Received Date: 11 May, 2017; **Accepted Date:** 23 May, 2017; **Published Date:** 30 May, 2017

Abstract

Human Immunodeficiency Virus (HIV) is the causative agent of the Acquired Immunodeficiency Syndrome (AIDS). The disease is a major global public health threat that causes over 39 million deaths in 78 million cases ever since it was discovered. To date, there is no vaccine available for prevention, and most of the efforts were directed towards the conventional vaccines approaches. Thus, in the present study, an alternative newly emerging *in silico* approach for designing peptide-based vaccines has been sought. This immune informatics approach has several advantages regarding the safety, stability and the ease to manufacture. We designed a potential multi-epitope vaccine against HIV-1 purposely for the most endemic sub-Saharan African countries. Three essential structural genes that include envelope (*Env*), group specific antigen (*Gag*) and polymerase (*Pol*) protein sequences were retrieved from the databases, analyzed and verified through a total of 15 *in silico* tools. This resulted in nine antigenic and non-allergic potential epitopes capable to provoke both humoral and cell mediated immunity without induction of autoimmunity. These epitopes covered 76.57% of the world population with a high sequence conservancy varies from 80-97%. Furthermore, computational docking techniques were used to confirm the strong binding interactions of the epitopes with their specific HLA molecules. Albeit, the results are awaiting validations by *in vitro* and *in vivo* experiments, nonetheless, this study provides a useful insight for the developing of successful vaccines to prevent the devastating HIV infections.

Keywords: Allergenicity; Antigenicity; Docking; Epitope; HIV-1; HLA; Immuno informatics

Introduction

AIDS is number one leading cause of death in Africa and the fourth worldwide [1]. By the end of 2014, approximately 37 million people were living with HIV, of which, nearly 70% are in sub-Saharan Africa, and almost 1-2 million die of the disease every year [2]. This devastating disease infects approximately 3 million new cases annually, 50% of them occur in young adults (15-24 years of age). Moreover, AIDS has left approximately 14 million orphans worldwide [3]. The HIV is a double stranded ribonucleic acid (RNA) retrovirus that infects cells of the immune

system and mainly destructs the cluster of differentiation (CD4⁺) T lymphocytes. The use of Antiretroviral Therapy (ART) along with the behavioral changes have reduced the HIV/AIDS-related deaths globally and the number of new HIV infections respectively [2,4]. Ever since the HIV/AIDS has been discovered in 1981, scientists had strived to come up with a protective and safe vaccine using different vaccine technologies; either (i) protein subunit as in the first clinical trials; AIDSVAX® (VAX004 and VAX003) in 1998-2003 (ii) recombinant adenovirus vector like in HIV Vaccine Trials Network (HVTN) 503/ Phambili (iii) canary pox vector prime followed by a protein subunit boost as in RV144 Thai trial (vi) or DNA prime followed by recombinant adenovirus vector boost in HVTN 505. All trials showed no efficacy on their target popu-

lations, except RV144 that showed 32.1% efficacy in preventing HIV infection nonetheless it did not reduce the viral load [5].

In comparison to the conventional vaccines, that is very much laborious, time consuming and expensive, an alternative computational (*in silico*) approaches have recently been extensively studied for predicting epitopes with reduced cost and time [6]. Epitope Based Vaccines (EVs) is one of these approaches that has significantly progressed after the advancement and better understanding of the molecular basis of antigen recognition and their binding to the human HLA motifs through refined algorithms. These vaccines are easy to develop, chemically stable, highly specific, and free from any infectious or oncogenic potential hazard [7,8].

The rational design for this work has followed paradigm of the emerging publications Chinese researchers that have already taken a step forward towards this new multidisciplinary strategy. They designed an epitope-based peptide HIV vaccine for Chinese populations using immune informatics. Likewise, yet with different organisms such as in *Shigella* spp as recently described. Moreover, this approach has already proven its effectiveness in the case of multiple sclerosis [9] tuberculosis, Ebola, Respiratory Syncytial Virus, Corona virus [9,10,11]. The strategy of these studies was to make use of the newly emerging field that combine immunology and bioinformatics to simulate the interaction between antigens and host immune system. Therefore, we undertake this new approach, aiming to design an effective and safe HIV-1 vaccine that will protect humans against HIV/AIDs.

Materials and Methods

Sequence Retrieval

A number of 45 complete nucleotide sequences of the *Env* gene were retrieved from the National Centre for Biotechnology Information (NCBI) database(<http://www.ncbi.nlm.nih.gov/>), targeting the endemic countries of sub-Saharan Africa that include Cameron, South Africa, Kenya, Tanzania and Zambia. These nucleotides were subsequently translated into protein sequences via Expert Protein Analysis System (ExPASy) (<http://www.expasy.org/>) [12].

In addition to, retrieval of the globally available protein sequences of the other *Gag* and *Pol* genes of HIV-1 subgroup M were retrieved from Universal Protein Knowledgebase (<http://www.uniprot.org>) [13]. All these sequences were downloaded and stored in FASTA format for further analysis.

Sequence Alignment and Phylogeny

Multiple sequences alignment was performed using Biological sequence alignment Editor (BioEdit) v7.0.9.0 [14] and ClustalX2.0 [15], to look for conserved regions. Subsequently, a total of 45 *Env* protein sequences from different sub-Saharan African courtiers were phylogenetically analyzed using Molecular Evolu-

tionary Genetics Analysis (MEGA) v.6.06 software [16]. The dendrogram was generated using the unrooted, maximum likelihood method based on the JTT matrix-based model as the best model and 1000 bootstrap replicates.

Antigenicity of the Conserved Regions

Antigenicity is a term used to describe the recognition of the molecules by the antibodies and/or the immune system cells. To elicit the best antigenic conserved regions of *Env*, *Gag* and *Pol*, a VaxiJen v2.0 server [17] was used with the default prediction parameters and a threshold value of 0.4.

Epitopes Prediction

Prediction of CTL Epitopes: The potential CTL epitopes were predicted using an online server Net CTL v1.2 (<http://www.cbs.dtu.dk/services/NetCTL/>) [18]. A server that implements a prediction method integrating MHC-I binding, Transporters Associated with Antigen Presentation (TAP) and proteasomal cleavage. CTL epitopes restricted to 12 MHC class 1 supertypes (A1, A2, A3, A24, A26 B7, B8, B27, B39, B44, B58, and B62) were predicted from their antigenic conserved peptides. Their proteasomal cleavage prediction that depends on Artificial Neural Networks (ANN) and the TAP transport efficiency that depends on weight matrix. The threshold was set as 0.5 to give a specificity of 94% and sensitivity of 89%. During the analysis as the antigenicity might be lost VaxiJen v2.0 [17] again was used. On the basis of the resulted NetCTL prediction combinatorial scores (Comb) of > 0.9 and VaxiJen scores of ≥ 0.4 the best candidate epitopes were selected for further analysis.

MHC-I Binding Prediction: The FASTA formats of the selected antigenic NetCTL epitopes were analyzed by the Immune-Epitope Data Base - Analysis Resource (IEDB-AR) (<http://www.iedb.org/>). To calculate the half-maximal inhibitory concentration (IC50) values that predicts the peptide affinity to MHC-I, the recommended NetMHCpan [19] was applied in MHC-I processing predictions tool (Tenzer *et al.*, 2005) (<http://tools.iedb.org/processing/>) [20] with accuracy ranging from 81-97% as reported in [20,21]. In this processing the CTL epitopes were predicted from their reference protein sequences based on their ability to cleaved by proteasomal cleavage and be transported by TAP transporter. Here, all the MHC-I alleles were chosen with 9 amino acids as fixed length.

MHC-II Binding Prediction: In contrast to MHC-I, the antigenic conserve peptides were analyzed to predict their binding with MHC-II molecules, following the IEDB recommended method. This method utilizes the consensus method (Wang *et al.*, 2008 and 2010), that selects the best method of binding either the neural network-based alignment (NN-align) [22], Stabilization Matrix Alignment Method (SMM-align) [23], Combinatorial library or Sturniolo Method [24] for the prediction of the specific HLA-DP, DQ and DR. All the available alleles in MHC class II binding predicting tool (<http://tools.iedb.org/mhcii/>) were selected.

B-cell Epitopes Prediction: Linear B cell epitopes vary in length from 5-20 amino acids as has been reported in the B cell epitope Bcipep database (<http://www.imtech.res.in/raghava/bcipep/>). In order to predict these linear B cell epitopes, the conserve regions that have Vaxijen scores of ≥ 0.4 and are surface located according to transmembrane helix topology with Hidden Markov Model predictor TMHMM v0.2 server (<http://www.cbs.dtu.dk/services/TMHMM/>) [25], were submitted to Artificial neural network based B cell epitope prediction (ABCpred) database (<http://www.imtech.res.in/raghava/abcpred/>) [26].

Quality Assessment of the Epitopes

Allergenicity and Epitope Conservancy Assessment: AllerTOP v.2 [27] was used to predict the allergenicity of the epitopes, and the epitope conservancy analysis tool of IEDB (http://tools.iedb.org/tools/conservancy/iedb_input) [28] was utilized to test each potential epitope conservancy level by searching their identities in all protein sequences retrieved from the database.

Human Self-Peptides Cross-Reactivity Determination: To determine the presence of human peptide analogs identical to the predicted epitopes, NCBI BlastP program (<http://www.ncbi.nlm.nih.gov/BLAST/>) [29] was utilized.

The Population Coverage Prediction

The human population coverage was predicted by the IEDB population coverage calculation tool (Bui et al., 2006) [30] in which each selected antigenic epitope was submitted to (http://tools.iedb.org/tools/population/iedb_input). The corresponding HLA alleles from MHC Class I and II resulted from the previous working steps, were submitted. The default parameters connected to the latest HLA genotypic allele frequency database (<http://www.allelefrequencies.net/>) [31] were set. This database contains frequencies of 3245 alleles (including both classes) for the world, 16 geographical areas, 21 ethnicities, 115 countries.

Molecular Docking and HLA-Epitopes Interaction

Computational Details: The molecular docking was carried out on a workstation 2.10 GHz Intel® core™ i7-4510U CPU, 4GB (RAM), Intel® HD Graphics, Generic PnP monitor, and operating under the windows 7 ultimate (x86_64bit). Structure construction, optimization, and visualization were carried out using the molecular modeling packages SYBYL-X suite of programs [32,33], and Molecular Operating Environment (MOE). Compounds were minimized at Tripos force field with Gasteiger-Hückel charges using Powell method.

Receptor Preparation for Docking: 3D-structures used in this study were retrieved from protein data bank (<http://www.rcsb.org>), all water molecules were removed, both polar and non-polar hydrogen atoms were added, side chains were repaired, chain termini were fixed, protonation types were set for the re-orientation of the hydrogen atoms which would be more favorable to hydrogen bonding. Atomic charges were assigned to the receptor using AMBER7 FF99 force field [32], and to the ligand using Gasteiger-Hückel method. The protein complex was minimized using AMBER7 FF99 force field. Finally, the 3D structure of the prepared protein was saved as MOL2 file [34].

org), all water molecules were removed, both polar and non-polar hydrogen atoms were added, side chains were repaired, chain termini were fixed, protonation types were set for the re-orientation of the hydrogen atoms which would be more favorable to hydrogen bonding. Atomic charges were assigned to the receptor using AMBER7 FF99 force field [32], and to the ligand using Gasteiger-Hückel method. The protein complex was minimized using AMBER7 FF99 force field. Finally, the 3D structure of the prepared protein was saved as MOL2 file [34].

Results and Discussion

Sequences Divergence and Phylogenetic Analysis

The *Env*, *Gag* and *Pol* aligned protein sequences revealed a number of 8, 10 and 14 conserved peptide regions respectively (Table 1).

Gene	Conserved regions	VaxiJen score ^a	Prediction of transmembrane helix ^b
ENV	LWVTVYYGVPVW	-0.032	
	HNWATHACVPT-DP	-0.322	
	NVSTVQCTHGI	0.5046	Inside
	AGSTMGAA-SITLTQARQLLS-GIVQQQ	0.8661	Outside
	LTVWGIKQLQ	1.3109	Inside
	FIMIVGGLIGLRI	1.2012	Outside
	RQGYSPLSFQT	1.1719	Outside
	DLRSLCLFSYH	0.9023	Outside
GAG	WEKIRLRPGG	1.4173	
	SQNYPIVQN	0.3632	
	SPRTLNAWVKV	-0.1749	
	EEKAFAFSPEVIPMF-SALSEGATPQDLN	0.6516	Outside
	GGHQAAAMQMLK	0.0504	
	PRGSDIAGTTS	1.1163	Inside
	GLNKIVRMYSP	-0.9406	
	QGPKEP-FRDYVDRF	-1.3698	
	TLLVQNANPDC	1.0613	
	EEMMTACQGVGGP	0.2338	

POL	FRKYTAFTIPS	0.5938	Inside
	EALLDTGADDTVL	0.0926	
	WPKPMIGGI-GFIKV	0.4334	Outside
	GCTLNFPIS	-0.4052	
	LKPGMDGP	0.2322	
	FAIKKKDS	1.5596	
	KWRKLVDFRELN-KRTQDFWEVQL-GIPHAGLK	0.6107-	Inside
		0.8898	
	VLDVGDAYFS	1.6272	Outside
	RYQYNVLPQG-WKGSPAIFQ	0.4147	Inside
	IYQYMDDLYVGS-DLE	0.7202	Inside
	TPDKKHQKEPP--	1.2037	
	VGKLNWASQIY	1.0397	Inside
	QWTYQIYQEP	0.1682	
	KNLKTGKYA	1.3042	

Table 1: The Conserved Antigenic Regions for *Env*, *Gag* and *Pol* Genes Of HIV, Antigenic Scores and the Location in The Virus Cell Membrane.

On the other hand, the dendrogram was generated from 45 *Env* protein sequences from five sub-Saharan African countries revealed clear phylogeography where sequences from each and every country were grouped separately from each other except one South African VIRT16255. Additionally, clade II also placed the neighboring countries Tanzania, Zambia and South Africa in one group. However, clade I that grouped Kenya and Cameroon in one group remained to be explained since these countries did not share any neighborhood as illustrated in (Figure 1).

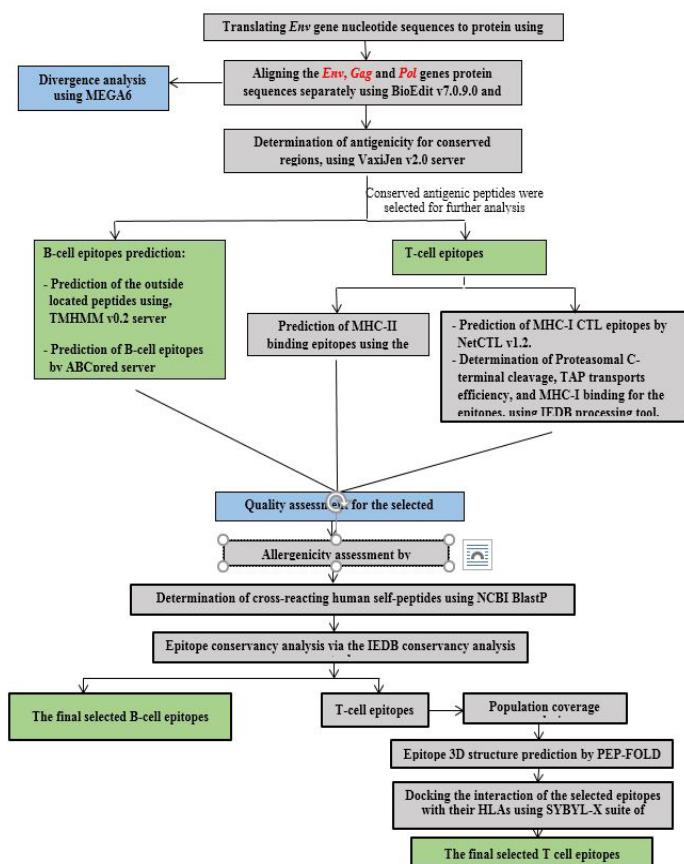


Figure 1: Schematic Representation of the Workflow.

Antigenicity Determination

All the obtained conserved regions were subjected to the antigenicity assessment to ensure their ability to provoke human immunity. This resulted in 4, 6, and 10 antigenic conserved peptides for *Gag*, *Env* and *Pol* genes respectively as indicated in (Table 1).

Epitopes Prediction

Prediction of CTL Epitopes: The above mentioned conserved antigenic peptides generated 5, 10 and 14 predicted CTL epitopes for the *Gag*, *Env* and *Pol* genes respectively. Based on the comb score, we have selected the top 14 epitopes and further subjected them to a second round of antigenicity assessment that might be lost during the step of epitope prediction. This assessment revealed that all except one Pol epitope remain antigenic (Table 2), while the rest of the epitopes we choose to neglect, were interacting with 1 or 2 already involved alleles.

Gene	Conserve regions	NetCTL		Epitope Vaxijen score
		Epitopes	Comb score ^a	
ENV	NVSTVQCTHGI	STVQCTHGI	0.9404	0.7002
		TMGAASITL	1.103	1.2925
	AGSTMGAASITLTVQAR-QLLSGIVQQQ	GAASITLTV	0.9672	1.1486
		LTVQARQLL	0.5361	0.6422
	LTVWGIKQLQ	LTVWGIKQL	0.6392	1.1169
		MIVGGLIGL	1.0993	0.6959
	FIMIVGGLIGLRI	IVGGLIGLR	0.915	1.8458
	RQGYSPLSFQT	RQGYSPLSF	1.3837	1.3895
		GYSPLSFQT	0.6519	1.56
	DLRSLCLFSYH	LRSLCLFSY	1.2737	0.937
GAG	EEKAFSPEVIPMFSALSE-GATPQDLN	FSPEVIPMF	1.4734	0.7154
		IPMFSALSE	0.9036	0.563
		SEGATPQDL	1.3975	0.4863
		AFSPEVIPM	0.5858	1.2855
	PRGSDIAGTTS	GSDIAGTTS	0.6061	0.7489
POL	FRKYTAFTIPS	FRKYTAFTI	1.0157	1.0715
		RKYTAFTIP	0.5148	0.8899
	WPKPMIGGIGGFIKV	KMIGGIGGF	1.499	0.4211
		KPKMIGGIG	0.5057	0.6622
	KWRKLVDRELN-KRTQDFWEVQLGIPH-PAGLK	RKLVDFREL	0.711	1.0169
		TQDFWEVQL	1.7491	0.5995
		LGIPHPAGL	0.5124	0.215
		KRTQDFWEV	0.6643	1.2846
		WEVQLGIPH	0.7864	1.0969
	VLDVGDAYFS	VLDVGDAYF	0.8585	1.6303
	RYQYNVLPQGWKG-SPAIFQ	RYQYNVLPQ	0.8997	0.8655
	IYQYMDDLYVGSDLE	YMDLYVGS	0.8364	1.0388
	VGKLNWASQIY	KLNWASQIY	1.4382	0.8895
		GKLNWASQI	0.6035	1.3742

• The bolded sequences are the selected CTL epitopes.

a. Combinatorial score: The higher the score, the more qualified the epitope.

Table 2: The Conserved Antigenic Peptides and the Produced Epitopes Using Netctl V2.1 Server.

MHC-I binding prediction: The standard IC₅₀ threshold used in the previous studies was 500 nM. However, most of the selected epitopes in this study were having a binding affinity of < 250 nM to increase the level of confidence. Collectively, 7 epitopes were found to interact with 13 different HLA alleles, in which (MIVGGLIGL and KLNWASQIY) were found to be interacting with 4 different alleles, (KMIGGIGGF) were interacting with 3 different alleles (Table 3).

Gene	MHC-I Epitopes	Total Score a	MHC IC ₅₀ b [nM]	Alleles
ENV	STVQCTHGI	-0.93	213.43	HLA-A*02:06
		0.01	24.61	HLA-A*68:02
	TMGAASITL	-0.5	333.68	HLA-A*02:06
		-0.44	287.11	HLA-A*02:03
		-0.13	142.33	HLA-A*02:01
	GAASITLTV	-1.27	291.48	HLA-A*02:03
		-1.13	207.1	HLA-A*02:06
		-1.08	185	HLA-A*68:02
	MIVGGLIGL	0.42	72.33	HLA-A*68:02
		0.67	40.06	HLA-A*02:01
		0.68	39.18	HLA-A*02:03
		1.17	12.77	HLA-A*02:06
	IVGGLIGLR	-0.89	197.86	HLA-A*31:01
		-0.64	113.74	HLA-A*68:01
	RQGYSPLSF	0.64	93.97	HLA-A*32:01
		1	40.31	HLA-B*15:01
GAG	FSPEVIPMF	0.07	359	HLA-A*02:06
	IPMFSALSE	-2.33	329.6	HLA-B*35:01
POL	KMIGGIGGF	-0.11	183.54	HLA-A*02:06
		0.68	29.98	HLA-A*32:01
		0.8	23	HLA-B*15:01
	TQDFWEVQL	-0.07	155.84	HLA-A*02:06

Abbreviations: MHC IC₅₀: The Half Inhibitory Concentration of the Major Histocompatibility Complex; HLA: Human Leukocyte Antigen

• The bolded sequences are the selected epitopes.

a. The total score combines the proteasomal cleavage, TAP and MHC I binding.

b. The lower the value the higher the binding affinity.

Table 3: The MHC-I Processing Epitopes of the Different Genes that have IC₅₀<500 nM.

MHC-II binding prediction: The 15mers conserved antigenic regions mention above (3.2) revealed 1, 9 and 13 predicted MHC-II binding epitopes for *Pol*, *Gag* and *Env* respectively (see Appendix 1, 2 and 3). The best candidate epitopes having a binding affinity of < 250 nM and percentile rank of ≤ 20 % were subjected to antigenicity assessment. This reduces the number of antigenic epitopes to 1, 3 and 4 that were considered the most potential antigenic MHC-II binding epitopes and were further tested for allergenicity assessment.

B-Cell epitopes prediction: The conserved antigenic outside-located peptides (Table 1) predicted 23 potential B-cell epitopes by a default threshold of 0.51 as shown in (Table 4).

Gene	Conserved Peptides	B cell epitopes	ABCpred score a	VaxiJen score b
ENV	AGSTMGAASITLT-VQARQLLSGIVQQQ	MGAASITLTV LTVQARQLLS TLTVQARQLLSG AGSTMGAASITL AGST- MGAASITLTV TLTVQARQLLSGIV AASITLTVQARQLLSG STMGAASITLTVQAR- QLL AGSTMGAASITLT- VQARQLL	0.65 0.60 0.65 0.55 0.76 0.75 0.71 0.71 0.60	1.2124 0.4829 0.9316 0.8003 0.9030 0.7711 1.0835 1.1175 0.9187
	RQGYSPLSFQT	RQGYSPLSFQ	0.59	1.5463
	DLRSLCLFSYH	DLRSLCLFSY	0.58	0.9207
GAG	EEKAESPEVIPMF-SALSEGATPQDNLN	PEVIPMFSAL EVIPMFSALSEG EKAFSPEVIPMF IPMFALSEGATPQ KAESPEVIPMFSAL PEVIPMFSALSEGATP VIPMFSALSEGAT- PQDNLN AFSPEVIPMFSALSE- GATPQ	0.68 0.69 0.65 0.82 0.68 0.86 0.64 0.87	0.5563 0.4009 0.5106 0.3827 0.7288 0.4076 0.5954 0.5696
	TLLVQNANPDC	LLVQNANPDC	0.67	1.0281
POL	WKPKMIGGIGGFIKV	WKPKMIGGIG WKPKMIGGIGGF KPKMIGGIGGF	1 1 2	1.3298 0.7847 0.0626
<p>a. The higher the ABCpred score, the higher prospecton to be a B cell epitope. b. The epitopes having Vaxijen score ≥ 0.4 score was only considered for further analysis.</p>				

Table 4: Predicted Linear B-Cell Epitopes from ABCpred Server.

Quality Assessment for the Epitopes

Allergenicity and Epitopes Conservancy Assessment: Despite the fact that most researchers used AllerHunter server (Muh. et al., 2009), that has 96%, 78.2% and 87.1% as specificity, sensitivity and accuracy respectively in assessing allergens as well as non-allergens. In this study, we preferred to use AllerTOP v.2 for its highest accuracy 88.7% among those servers. Surprisingly, only 1, 5 and 14 epitopes out of the 7, 8 and 23 epitopes of MHC-I, MHC-II and B-cell respectively were found to be non-allergic (Table 5 and 6).

Presentation Type	Gene	Non-allergic Epitopes	Epitope conservancy	Alleles
MHC-I	ENV	MIVGGLIGL	93.33%	HLA-A*68:02 HLA-A*02:01 HLA-A*02:03 HLA-A*02:06
MHC-II	ENV	GAASITLTQARQLL	93.33%	HLA-DRB4*01:01 HLA-DRB1*07:01 HLA-DRB1*01:01 HLA-DQA1*01:02/DQB1*06:02 HLA-DQA1*05:01/DQB1*03:01
		STMGAASITLTQAR	93.33%	HLA-DQA1*01:02/DQB1*06:02 HLA-DQA1*05:01/DQB1*03:01
	GAG	WKPKMIGGIG	97.92%	HLA-DRB1*04:01 HLA-DRB1*04:05 HLA-DRB1*15:01 HLA-DRB1*09:01 HLA-DRB1*01:01
		IPMFSALSEGATPQD	81.82%	HLA-DRB1*09:01 HLA-DQA1*05:01/DQB1*03:01 HLA-DRB1*01:01
	POL	WKPKMIGGIGGFIKV	97.92%	HLA-DQA1*05:01/DQB1*03:01

Table 5: Establishing the Epitopes that Pass the Allergenicity Assessment, the Conservancy, and the Corresponding Alleles.

Presentation Type	Gene	Non-allergic Epitopes	Epitope conservancy
	ENV	TLTVQARQLLSGIV AASITLTQARQLSG AGSTMGAASITLTQARQLL DLRSLCLFSY	93.33% 93.33% 93.33% 95.56%
	GAG	PEVIPMFSALSEGATP LLVQNANPDC	84.85% 93.94%
	POL	WKPKMIGGIG	97.92%

Table 6: The Non-Allergic B-Cell Epitopes and the Conservancy.

whereas the allergic epitopes are shown in (Appendix 4). On the other hand, as shown in the same tables, most of the Env gene epitopes that bind to MHC-I, MHC-II and B-cell gave 93.33% sequence similarity to the reference protein. Whereas, the Gag gene epitopes vary in their conservancy from 81-84%, and the Pol gene epitopes revealed 97.92% sequences similarity.

Human Self-Peptides Cross-Reactivity Analysis: Only one single B-cell epitope has been detected to have putative conserved domain identical to human peptide among all the non-allergic epitopes. Thus, it was eliminated from the epitopes pool to avoid triggering an autoimmune response.

Population Coverage

The global population coverage of the six non-allergic MHC-I & II candidate epitopes was 76.57%. Concerning the continents, the overall population coverage is presented in (Figure 2). In our target African sub-Saharan countries, the covering was ranging from 63.48% and 59.49% in Cape Verde and São Tomé and Príncipe respectively, passing through Sudan with 53.57%, while the least recorded percentage was 15.73% in Equatorial Guinea. Interestingly, we detected the highest coverage in Northern Ireland (89.01%) followed by Austria (88.98%) and England (88.39 %).

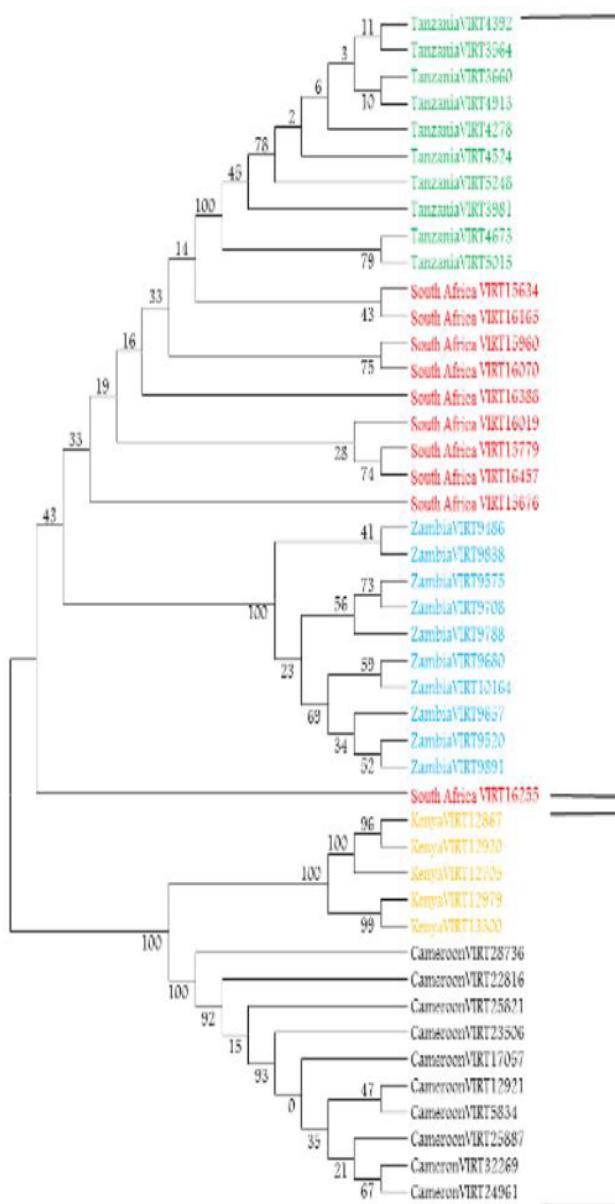


Figure 2: A Dendrogram Demonstrating 45 *Env* Protein Sequences Divergence Of HIV-1 Using Unrooted and Bootstrapped Maximum Likelihood Method. This Dendrogram Shows Two Phylogenetically Distinct Clades, Clade (I) Grouped Kenya (Yellow) And Cameroon (Black), Clade (II) Grouped Tanzania (Green), Zambia (Blue) and South Africa (Red).

Molecular docking and HLA-epitopes interaction analysis

According to the binding energy illustrated below in (Table 7).

Peptide	Protein (PDB ID)	Docking Energy Kcal/mole
MHC-I (MIVG-GLIGL)	HLA*02.01 (4U6Y)	-10.2
MHC-II (PEVIPMF-SALSEGAT)	HLA*1.01.01 (1AQD)	-7.1
MHC-II (GAA-SITLTVQARQLL)	HLA*1.01.01 (1AQD)	-5.9

Table 7: The Docking Results of the Three Selected Epitopes Performed By SYBELL-1.

II

The lower the energy, the higher the binding affinity of the ligand to the receptor. Since the hydrogen bonds are considered strong bonds, the more the hydrogen interactions found between the ligand and the receptor, the higher the binding affinity. Although the single MHC-I epitope (MIVGGLIGL) that bind to the HLA*02.01 was having only two hydrogen bonds with lengths of 1.93 and 2.93 Å shown in (Figure 3 and 4),

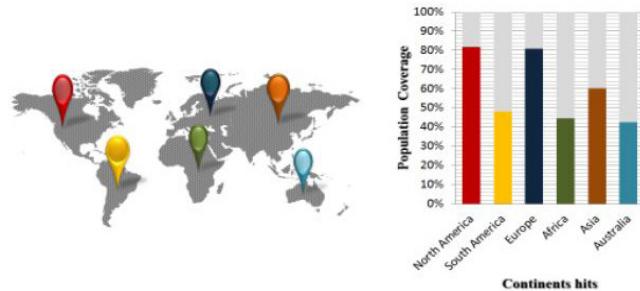


Figure 3: The IEDB worldwide epitopes population coverage.

I

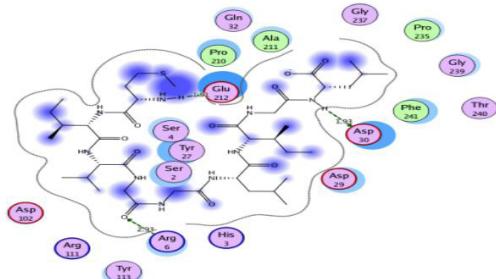


Figure 4: The Docking Of MHC-I Epitope (MIVGGLIGL) Shown as Chemical Structure Ligand With HLA-A*02.01 (PDB ID: 4u6y) Amino Acid Residues. The Hydrogen Bonds Are Marked By (H) And the Ionic Bonds By (I).

however, it has the lowest binding energy of -10 Kcal/mole out of the three selected epitopes. In a comparison made between the other two MHC-II epitopes that bind to the same HLA-DB1:01:01 shown in (Figure 5 and 6).

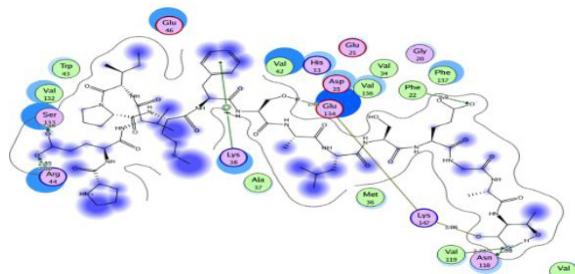


Figure 5: The Docking Of MHC-II Epitope (PEVIPMFSALSEGAT), The Chemical Structure Ligand, With HLA-DB*1.01.01 (PDB ID: 1aqd) Amino Acid Residue. The Hydrogen Bonds Are Marked By (H) and the Ionic Bonds By (I).

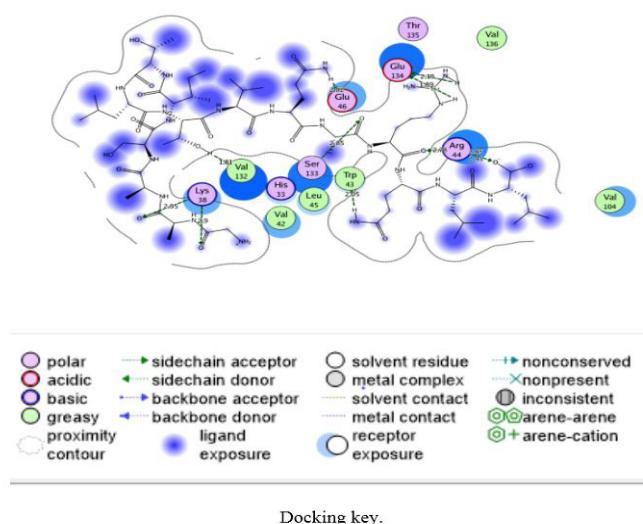


Figure 6: The Docking Of MHC-II Epitope (GAASITLTVQARQLL), The Chemical Structure Ligand, With HLA-DB*1.01.01 (PDB ID: 1aqd) Amino Acid Residues. The Hydrogen Bonds Are Marked By (H) and the Ionic Bonds By (I).

The (PEVIPMFSALSEGAT) epitope was found to be better in the binding by having a lower energy -7.1 Kcal/mole (Table 5), with two hydrogen bonds with the lengths of 1.74 and 1.98 Å, three electrostatic ionic bonds ranging from 2.85 - 3.06 Å in their lengths, two ionic-dipole bonds with lengths of 2.78 and 2.9 Å, a single charge-transfer interaction between lysine 38 amino acid residue of the receptor with the ligand benzene ring and a hydrophobic Van der Waals (VdW) interaction with Valine 42, Glycine 134, Valine 136 and Phenylalanine 22. Whereas (GAA-SITLTQARQLL) epitope was found to be having the highest binding energy of -5.9 Kcal/mole, five hydrogen bonds vary in their length from 1.81-2.58 Å, three ionic bonds ranging in length from 2.02-2.93 Å, three ionic-dipole interactions vary in length between 2.78-2.95 Å and VdW interaction with Valine 132, Valine 42 and leucine 45.

Conclusion

In spite the fact that epitopes prediction by means of conventional experimental techniques such as recombinant vaccines, sub-unit protein and DNA vaccines, to stimulate both humoral and T cell immunity are feasible, but they are considered as time-consuming, overpriced, and not applicable in the large-scale screening. Alternatively, advances in computer modeling known as *in silico* studies, provide a systematic and high throughput candidate peptides that can hypothetically interact with HLA alleles via high population coverage, therefore, minimizing the high volume of laboratory work with higher probabilities of success. A successful HIV vaccine along with the availability of ART is required to have a free world of HIV by 2030 as presumed by WHO and UNAIDS, 2015 [2,4]. Hence, in this study we managed to identify potential epitopes that have the ability to provoke the immune system against HIV-1. We have predicted a cohort of six epitopes from the conserved regions of HIV-1 *Env*, *Gag* and *Pol* genes that interact with 13 different HLA alleles from MHC-I and II providing CTL responses in human, beside additional seven epitopes that are able to elicit B-cell responses. All the predicted epitopes have a conservancy of > 81% and global population coverage of 76.57%. However, *in vivo* and *in vitro* studies are remained to determine their effectiveness as a proposed multi-epitope vaccine.

References

1. Levinson W (2014) Review of Medical Microbiology and Immunology. In:13. McGraw-Hill Medical, New York 2014: 825-826.
2. UNAIDS (2015) How AIDS Changed Everything: 88-98.
3. Abbas AK, Lichtman AH, Pillai S (2015) Cellular and Molecular Immunology. In: 8. Elsevier Saunders Incorporation Philadelphia 2015: 450-463.
4. WHO (2015) Data and Statistics.
5. Cohen YZ, Dolin R (2013) Novel HIV vaccine strategies: overview and perspective. Therapeutic Advances in Vaccines 1: 99-112.
6. Mahela K, Ramana J (2016) Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: a cooperative genomics and immunoinformatics approach. Molecular Bio-Systems 12: 890-901.
7. Holland J and Domingo E (1998) Origin and evolution of viruses. Virus Genes 1: 13-21.
8. Sette A, Newman M, Livingston B, McKinney D, Sidney J, et al. (2002) Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. *Tissue Antigens* 59: 443-451.
9. Bourdette DN, Edmonds E, Smith C, Bowen JD, Guttmann CR, et al. (2005) A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Multiple Sclerosis* 5: 552-561.
10. Khan MA, Hossain MU, Islam MM, Rakib-Uz-Zaman SM, Morshed MN (2015) Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immune informatics study. *Scandinavian Journal of Immunology* 1: 25-34.

11. Oany AR, Al Emran A, Jyoti TP (2014) Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an *in silico* approach. *Drug Design, Development and Therapy* 8: 1139-1149.
12. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, et al. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* 40: W597-W603.
13. The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Research* 43: D204-D212.
14. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
15. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.
16. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729.
17. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 8: 4.
18. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, et al. (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8: 424-436.
19. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, et al. (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One* 2: 1-10.
20. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, et al. (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences* 62: 1025-1037.
21. Schellens IM, Kesmir C, Miedema F, van Baarle D, Borghans JA (2008) An unanticipated lack of consensus cytotoxic T lymphocyte epitopes in HIV-1 databases: the contribution of prediction programs. *AIDS* 22: 33-37.
22. 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.
23. Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8: 238-250.
24. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, et al. (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nature Biotechnology* 6: 555-561.
25. Sonnhammer ELL, von Heijne G, Krogh, A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In Proceeding of Sixth International Conference on Intelligent Systems for Molecular Biology 6: 175-182.
26. Saha S, Raghava GP (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 1: 40-48.
27. Dimitrov I, Bangov I, Flower DR, Doytchinova I (2014) AllerTOP v.2-a server for *in silico* prediction of allergens. *Journal of Molecular Modeling* 20: 2278-2284.
28. Bui HH, Sidney J, Li W, Fusseder N, Sette A (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 1: 361-367.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.
30. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 7: 153-158.
31. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, et al. (2015) Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acid Research* 28: 784-788.
32. Jain AN (2003) Surflex fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry* 46: 499-511.
33. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera-a visualization system for exploratory research and analysis, *Journal of Computational Chemistry* 25: 1605-1612.
34. Shen Y, Maupetit J, Derreumaux P, Tuffery P (2014) Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *Journal of Chemical Theory and Computation* 10: 4745-4758.