International Journal of Biotechnology and Biomedical Sciences
Volume 1, Number 1; July- December, 2015 pp. 6-9
© Krishi Sanskriti Publications
http://www.krishisanskriti.org

Computational Prediction of Gag Epitopes Specific for HLA-DRB1*11 Alleles

Jemmy Christy H[1*] and Alex Anand D[2]

[1]Department of Bioinformatics, Sathyabama University
[2]Department of Biomedical Engineering, Sathyabama University
E-mail: [1]jemmyjacob@gmail.com, [2]danielalexanand@gmail.com

Abstract—*Immunoinformatics approach based studies of peptide-MHC complexes and characterizations of the peptide-MHC interaction would be ideal solution for clade specific vaccine construction for HIV. CD8+ T cell induction and maintenance as well as B cell response are based on CD4+ T cells role,thus epitopes restricted to these cells play a tremendous role HIV disease control. Various compuatational based algorithms like Stabilized Matrix Method and Neural Network were adopted to identify Gag epitope based consensus vaccine candidates restricted to HLA- DRB1*11 and were designated as P1-P6. De nova based designed three dimensional epitope structures and were assessed for their binding efficacy towards HLA-DRB1*11 allele using Cluspro docking method.Based on binding energy potential of docked complexes,molecular interactions like conventional hydrogen bonding, non classical carbon hydrogen bonding ,salt bridges epitopes were ranked.Current studies would provide the insights for designing Gag based vaccine constructs restricted to HLA-DRB1*11 allele for HIV therapy.*

1. INTRODUCTION

Acquiredimmunodeficiencysyndrome (AIDS) was first reported in the United States in 1981 and has since become a major worldwide epidemic. AIDS is caused by the human immunodeficiency virus, or HIV. Human immunodeficiency virus (HIV), a retrovirus that belongs to the Lentiviridae family[1].75 million people were HIV positive and nearly 36 million deaths happens worldwide due to HIV infection[2].Among the Indian population nearly 20.89 lakh people living with HIV/AIDS according to National AIDS control organization[3].Combination of antiretroviral drugs included in Highly active anti retro viral therapy (HAART) and used as therapeutic option to treat HIV individuals, but their side effects and resistance development seems to be a major obstacle[4].However HAART therapy and leads to a declination in morbidity and mortality of HIV infected individuals, but cannot eradicate the virus [4].Thus HIV vaccine based research has been a key area in science and numerous resources have been directed for HIV vaccines development[5-10].Architecture of the HIV-1 genome is complex and comprised of three functional groups of genes like structural genes (Gag, Env ,Pol) ,regulatory genes (Tat,Rev) and accessory genes (Vpu,Vpr,Vif,Nef)[11] Designing of vaccine for globally HIV infected people is not possible, due to genetic diversity of HIV and clade /subtype differences seen among the affected population [12, 13].There were three distinct groups: M (Major), O (Outliers), and N (non-M and non-O) of HIV-1 circulating in a global level but group M is the most predominant around the world ,and M group has nine subtypes: A–D, F–H, J, and K [14].Clade HIV-1C is responsible for Global HIV burden [15],Thus our studies restricted to analysis to HIV-1 subtype C-Gag specific CD4+ epitopes. Vaccine design need to consider two major key factors like hyper mutation capacity of HIV and HLA polymorphism, thus it is necessary to analyze the conserved fragments in the Gag sequences among Indian populations well as their affinity towards HLA alleles [16,17]. Resistance or slow progression to HIV/AIDS observed among the HIV patients with HLA-A02, HLA-A11, HLAB27, HLA-B*2705, HLA-B51, HLA-B*5701 of polymorphic genes nearly 500 allelic variants were listed. Heterodimeric HLA class II DRB1 allele restricted CD4+ T cell responses in HIV disease outcome among the HIV controllers is thrust area of HIV research [29-30].Our studies based on DRB1*11 a commonly distributed alleles among south Indian population and assessment of the binding affinity towards Gag based CD4+ epitopes [31]. This systematic studies would be helpful in selection of potent T cell epitopes restricted to DRB1*11.

2. METHODOLOGY

2.1 Analysis of Gag amino acid sequence conservancy

Gag protein sequences were retrived from HIV Sequence database [32] and conserved fragment of Gag sequences among Indian sample were retrieved from the literature survey of our earlier work [33]. Based on the conservancy score for each amino acid position of Gag sequence conserved fragment region considered for Epitope prediction.

2.2 Gag epitope prediction and assessment of Population Coverage

Based on low percentile rank of Immune Epitope Database (IEDB) predicted Gag epitopes that are restricted to HLA-

DRB1*11 allele were selected [34].IEDB prediction server includes various methods like Consensus

Method, combinatorial library, NN-align [35] netMHCII-2.2 [36], SMM-align (netMHCII-1.1) [37], Sturniolo[38], and NetMHCIIpan [39]. Resulting output includes units of IC_{50}nM for combinatorial library and SMM_align.Thus lower IC_{50}nM values indicate higher affinity. In General peptides with IC_{50}values <50 nM are considered high affinity, <500 nM intermediate affinity and <5000 nM low affinity. Raw score values of Sturniolo output indicates higher score in turn implies higher affinity. NetMHCIIpan method is used when Consensus and other methods such as SMM_align, NN_align,COMBLIB and/or Sturniolo are not available for a particular allele.However, if only one or two of these methods are available, NetMHC II pan can be used as second or third method. Low percentile ranked epitopes were screened and assessed for population coverage among Indian population. Population coverage of the conserved gag epitopes with the corresponding HLA-DRB1*11 allele were analyzed based on population coverage analysis tool of IEDB [40] depending on allele frequencies.net database, a huge population dataset on the web [41].

2.3 3D Modelling of Gag Epitopes structure

I-Tasser was used to model the three dimensional structure of DRB1*11 allele restricted Gag Epitopes [42]. Template search based on locally implemented meta server LOMETS was explored in I Tasser server,and TM-align server allows fragment assembly simulation ,thus final function predictions are concluded from the consensus hits among the top structural matches along with function scores calculated based on the confidence score of I–TASSER structural models [42]. Evaluation of structural similarity between target and template models carried out based on TM-Score and sequence identity in the structurally aligned regions.

2.4 HLA-DRB1*07 allele and Gag Epitope affinity analysis

Cluspro server was used to assess the Promiscuous Gag Epitopes binding affinity for HLADRB1*11 Allele [43], Cluspro is a fully integrated Docking server implemented with PIPER and FFT (Fast Fourier Trans-form) based rigid docking program. Complete protocol includes two stages , generation of low energy docked complexes based on pairwise interaction potential as first stage and clustering of docked complexes and low energy clusters assessment using SDU (Semi-Definite programming based Underestimation) program which predicts clusters stability using medium range optimization algorithm as second stage and finally stable clusters are further refined using Monte-Carlo simulation[43]. Balanced, Electrostatic favored, Hydro phobic favored and Vdw+Elec modules results were retrived and top ranked models were selected and visualized using DS Visulalizer 4.0 [44] to assess interaction and visualization of HLA and epitope interaction.

3. RESULTS AND DISCUSSION

3.1 Retrieval and analysis of Gag amino acid sequence and conservancy

Indian patients Gag protein sequences were retrieved from HIV sequence database and consensus fragment were Retrieved based on literature our earlier work [33].

3.2 HLA-DRB1*11 allele population coverage assessment and Gag epitopes prediction

IEDB server predicted HLA-DRB1*11 allele restricted Gag epitopes were retrieved, low percentile ranked epitopes were selected based on their IC_{50} value and named as promiscuous epitopes in Gag protein and were listed in **Table.1**

Table 1: CD_4 + promiscuous epitope and prediction scores.

Epitopes	Percentile Rank	Net mhc IIscore	Smm align score	Sturniolo score
IYKRWIILGLNKIVR	0.53	8.4	6	33
YKRWIILGLNKIVRM	0.53	7.9	6	31
KRWIILGLNKIVRMY	0.53	9.1	7	41
GKKQYRLKHLVWASR	0.77	10.1	7	47
NEEAAEWDRLHPLPA	0.77	10.4	7	52

3.3 De nova modeling of Gag Epitopes structure

C score of I-TASSER predictions used to assess the model quality of Gag Epitope models,and C-score which is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is generally in the range of [-5,2], if a C-score of higher value signifies a model with a high confidence and vice-versa [45]. Further TM-score and RMSD Values reported in Column 3 & 4 in the [**Table 2**] are the estimated values of based on their correlation with C-score [46]. I-TASSER generates full length model of proteins by excising continuous fragments from threading alignments and then reassembling them using replica-exchanged Monte Carlo simulations. Low temperature replicas (decoys) generated during the simulation are clustered by SPICKER and top five cluster centroids are selected for generating full atomic models. The cluster density is defined as the number of structure decoys at a unit of space in the SPICKER cluster. A higher cluster density means the structure occurs more often in the simulation trajectory and therefore signifies a better quality model. The values in the second last columns of the above mentioned table represents the number of structural decoys that are used in generating each model. The last column represents the density of cluster [46].

Table 2: 3D structure of Gag epitopes and prediction scores

EPITOPE	C SCORE	TM SCORE	RMSD
IYKRWIILGLNKIVR	1.31	0.52	2.1 A0
YKRWIILGLNKIVRM	1.34	0.55	2.1 A0
KRWIILGLNKIVRMY	1.43	0.54	2.4 A0

GKKQYRLKHLVWASR	0.28	0.75	0.5 A0
NEEAAEWDRLHPLPA	0.53	0.65	1.3 A0

3.4 HLA-DRB1*07 allele Gag Epitope Interaction analysis

Protein Data Bank was used to retrieve the HLA-DRB1*11 allele structure [47]. PIPER docking program implemented in Cluspro server was used to assess the HLA and Gag Epitopes interaction based on their binding energy scores generated from an energy function.Scoring of binding energy potential is based on sum of potential terms of shape complementaity, electrostatics, desolvation contributions, and Decoys as reference states (DARS) [48].

Low binding energy scored docked complexes were selected for HLA and epitope interaction assessment and those epitopes considered as promiscuous epitope candidate for vaccine construction. For our analysis clusters lowest binding energy values of balanced, electrostatic favored clusters, hydrophobic-favored and VdW+Elec clusters were included then binding energy scores of promiscuous epitopes P1-P6 binding efficiency were analyzed. DS visualizer 4.0 based observation of interaction between HLA and epitope interaction concluding that there were 3 subcategories of hydrogen bonding like Conventional Hydrogen Bond ,Carbon Hydrogen Bond and Salt Bridge interactions implies the stability of interaction and would aid CD_4+ regulated immune response against HIV infection [49].Hydrogen bonds donor and acceptors atoms of HLA and Gag epitopes and their bonding distance were listed in the **[Table- 3]**,carbon Hydrogen Bond interactions were considered as weaker since the donor is a polarized carbon atom and these interactions were determined using the same geomentric criteria used for classical hydrogen bonds with the exception of the default distance criterion being 3.8 Å.[50]. Promiscous epitopes were ranked based on the binding affinity pattern among the 6 predicted epitopes, we concluding that P1, P2 and P4 ,P6 could be considered as the potential epitopes, since their binding pattern and residue interactions were within the binding groove of DRB1*11 allele thus could confer stable molecular interaction.

Table 3: Gag epitopes binding affinity for DRB1*11 allele

GAG Epitopes	Binding Energy (kcal/mol)	Interacting Atoms	Bond Distance (A0)
IYKRWIILGLNKIVR	-716	CYS11: H – LYS3: O	2.72051
		CYS11: SG – ARG4: O	2.05172
		LEU28: H – ILE13: O	1.50211
YKRWIILGLNKIVRM	-817.2	ARG5:HH12-LYS2: O	2.10325
		ARG5:HH22-LYS2: O	2.09819
		ARG123:HH1-ILE12: O	1.79247
KRWIILGLNKIVRMY	-966.2	CYS44:SG-ARG:OG	2.67014
		CYS44: SG – LEU6: O	1.74318
		GLY49:H-ARG2:OG	1.56721
GKKQYRLKHLVWASR	-739.5	CYS44:H-TYR7:OH	2.6704
		LYS8:HH11-:GLN178:OE1	1.2973
		LYS8:HH12-GLY180: O	1.3422
NEEAAEWDRLHPLPA	-896.4	ALA5:HG – VAL114:O	2.13447
		ARG9:HE21-VAL114: O	2.44648
		ARG9:HE21-SER117:OG	2.64351

4. CONCLUSION

Our studies based on various computational methods assessed both epitope prediction and their affinity towards DRB1*11 using docking studies ,thus provides the structural insight of Gag epitopes namely P1- IYKRWIILGLNKIVR,P2-YKRWIILGLNKIVRM,P3- KRW IILGLNKIVRMY, P4-GKKQYRLKHLVWASR,P5-NEEAA EWDRLHPLPA. So Gag based epitopes would be the ideal considered as vaccine construct for HIV infection since their stable molecular interaction namely conventional hydrogen bonding and salt bridges with the HLA ,could aid both humoral and cell-mediated immunity. Anchor residues of ag epitope conferred greater affinity with the binding pocket residues of HLA-DRB1*11 which is a more highly distributed allele population in India. To evaluate their efficiency as vaccine candidate to construct an ideal HIV vaccine for Indian population further Invitro and in vivo studies are needed.

REFERENCES

[1] Moore,M.D.,W. S Hu,W.S., *AIDS Reviews*., 11,2, 2009,pp. 91–102.

[2] United Nations Programme on HIV/AIDS. “Global Report fact sheet: the global AIDS epidemic, 2013”. http://www.unaids.org/en/resources/campaigns/globalreport2013/factsheet/

[3] NACO annual report 2013-14 , http://naco.gov.in/NACO/Quick_ Links/Publication/Annual_Report/NACO_Annual_Report/Annual_Report_2013-14/

[4] N Kumarasamy,N.,Patel,A.,Pujari, S., *Indian J Med Res*, 134,2011,pp.787–800.

[5] Cohen,J.,” Shots in the dark: the wayward search for an AIDS vaccine”, New York: W.W. Norton & Company; 2001,pp.228–369.

[6] Miedema,F., *AIDS*, 22,14,2008,pp.1699–703.

[7] Grady ,C., “The Search for an AIDS vaccine: ethical issues in the development and testing of a preventive HIV vaccine.” Bloomington and Indianapolis (IN): Indiana University Press; 1995.

[8] Spearman.P., *Curr HIV Res*,1,1,2001,pp101–20.

[9] Munier,CM.,Anderson,CR.,Kelleher, AD ., *Drugs*,71,4,2011,pp.387–414.

[10] HIV Vaccines and Microbicides Resource Tracking Working Group. Investing to end the AIDS epidemic: a new era for HIV prevention research & development,http://www.hivresourcetracking.org/sites/default/files/July%202012%20Investing%20to%20End%20the%20AIDS%20Epidemic%20A%20New%20Era%20for%20HIV%20Prevention%20Research%20&%20Development.pdf,2013.

[11] Costin, J.M., *Virology Journal*, vol. 4, article 100, 2007.

[12] Barouch,DH., *Nature*, 455,2008,pp:613–619.

[13] Walker,BD.,Ahmed,R.,Plotkin,S., *Nat Med*, 17.2011,pp.1194–1195.

[14] Thomson,M,M.,Pérez-Álvarez,L.,Nájera,R., *Lancet Infect Dis.*; 2,8, 2002 Aug,pp. 461-71.

[15] Hemelaar,J.,Gouws,E.,Ghys,PD.,Osmanov,S., *AIDS*, 25,2011,pp.679–689.

[16] Brusic,V.,August,J,T., *Pharmacogenomics*, 5, 6,2004,pp.597–600.

[17] Ovsyannikova,I,G.,Jacobson,R,M.,Poland,G,A., *Pharmacogenomics*,5,4,2004,pp.417–427.

[18] MacDonald,K,S.,Fowke.,K,R.,Kimani,K,R., Dunand,V,A.,Nagelkerke,N,J.,Ball,T,B.,Oyugi,J.,Njagi,E., Gaur,LK.,Brunham.,R,C.,Wade,J.,Luscher,M,A., Krausa,P.,Rowland-Jones,S,, Ngugi,E.,Bwayo,JJ., Plummer,F,A., *J Infect Dis*,,181,2000,pp.1581–1589.

[19] Singh,P.,Kaur,G.,Sharma,G.,Mehra,N,K.,” “ *Vaccine*,,26,2008,pp.2966–2980.

[20] Selvaraj,P.,Swaminathan.,S,Alagarasu,K.,Raghavan., Narendran, S., Narayanan, P,J.,.*Acquir Immune Defic Syndr*, 43,2006, pp.497–499.

[21] Kaslow,RA.,Dorak,T.,Tang,JJ., *J Infect Dis*, 191,2005,pp.S68–S77.

[22] McNeil,AJ.,Yap,PL.,Gore,SM.,Brettle,RP.,McColl,M.,Wyld,R., Davidson,S.,Weightman,R.,Richardson,A,M., Robertson,J,R., *QJM*, 89,1996,pp.177–18.

[23] Neumann-Haefelin,C.,*Dtsch Med Wochenschr*, 136,2011,pp.320–324.

[24] Pereyra F., Jia ,X., McLaren, PJ.,Telenti, A., de Bakker, P,I., Walker, B,D., Ripke,S., Brumme, C,J., Pulit S,L., Carrington M et al ., “International HIV Controllers Study *“*, *Science*, ,330,2010,pp.1551–1557.

[25] SA Migueles; MS Sabbaghian;WL Shupert;MP Bettinotti;FM Marincola;L Martino;CW Hallahan;SM Selig;D Schwartz;J Sullivan;M Connors M.*Proc Natl Acad Sci U S A*,2000, 97:2709–2714.

[26] H Tomiyama;T Sakaguchi;K Miwa;S Oka;A Iwamoto;Y Kaneko;M Takiguchi M. *Hum Immunol* ,1999,60:177–186.

[27] Y Zhang;Y Peng; H Yan;K Xu;M Saito;H Wu;X Chen;S Ranasinghe;N Kuse;T Powell;Y Zhao; W Li;X Zhang;X Feng;N Li;A Leligdowicz;X Xu;M John;M Takiguchi;A McMichael;S Rowland-Jones;T Dong. *J Immunol*, 2011, 187:684–691.

[28] S Ranasinghe; S Cutler; I Davis;R Lu; DZ Soghoian;Y Qi; J Sidney;G Kranias;MD Flanders ,M Lindqvist;B Kuhl;G, Alter;SG Deeks;BD Walker;X Gao;A Sette;M Carrington;H Streeck . Nat Med, 2013, 19(7):930-3.

[29] Christy and Anand: *BMC Infectious Diseases*,2014,14 (Suppl 3):O4.

[30] C Kuiken;B Korber and R W Shafer. *AIDS Reviews*,2003, 5(1)52–61.

[31] K Balakrishnan; C Rathika; R Kamaraj. *Int J Hum Genet*. 2012, 12(1): 45-55.

[32] C Kuiken;B Korber and R W Shafer.*AIDS Reviews*,2003, 5(1)52–61.

[33] Jemmychristy,H.,Alex anand,D.,.*Int J Pharm Bio Sci*,5,4,2014,pp.1212-1222.

[34] R Vita;L Zarebski;J.A Greenbaum et al., *Nucleic Acids Research*,2010, 38, D854– D862.

[35] P Wang;J Sidney J;Y Kim et al.,*BMC Bioinformatics*,2010, 11, 568.

[36] M Nielsen;O Lund, *BMC Bioinformatics*.2009, 10:296.

[37] Nielsen M, Lundegaard C, Lund O. *BMC Bioinformatics.* 2007,8:238.

[38] Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. *Nat Biotechnol.* 1999, 17(6):555-561,

[39] Karosiene,E.,Rasmussen,M.,Blicher,T.,Lund,O., Buus,S.,Nielsen,M.,*mmunogenetics*,65,10,2005,pp.711.

[40] Bui,H,H.,Sidney,J.,Dinh,K., Southwood,S., Newman,M,J.,Sette,A., *BMC Bioinformatics*.2006, Mar 17; 7:153.

[41] Gonzalez-Galarza,F,F,Christmas,S.,Middleton,D., Jones,A,R.,.*Nucleic Acids Research*, 39, 1,2011,pp. D913–D919.

[42] Zhang,Y., *BMCBioinformatics*,2008,9:40.

[43] Kozakov,D, Clodfelter,K,H.,; Vajda,S., Camacho,C,J., *Biophys J*, 89,2005,pp.867–875.

[44] Accelrys Software Inc., Discovery Studio Modeling Environment, Release 4.0, San Diego: Accelrys Software Inc., 2013.

[45] Ambrish Roy., Alper Kucukural., Yang Zhang.,, *Nature Protocols*, 5, 2010,pp.725-738.

[46] Ambrish Roy., Jianyi Yang., Yang Zhang., *Nucleic Acids Research*,40,2012,pp.W471-W477.

[47] Zhang.,Skolnick.,,*Proteins*, 57,2004,pp.702-710

[48] Kozakov.,D, Hall,D,R.,Beglov,D.,Branke,,R., Comeau,S,R.,Shen,Y.,Li;J,K.,Zheng,Vakili.,P, Paschalidis I,C et al.,*Proteins*,78,2010,pp.3124-3130.

[49] Bissantz,C.,Kuhn,B.,Stahl,M., *J. Med. Chem*, , 53, 2010,pp.5061-5084.

[50] Pierce,A,C, Sandretto,K,L.,Bemis,G,W., *PROTEIN: Structure, Function, and Genetics*, 49,2002.pp.567-576.