Research Article

DOI: http://dx.doi.org/10.53043/2347-3894.acam11006

Bioinformatics-based Prediction of Character of Envelope Glycoprotein and Analysis of Epitopes of B- and T-cell of gp120

Meijie Liang[1]#, Wei Hong[1]# and Jiping Shao[1-3]*

[1]Hainan Medical University Haikou 571199 P.R. China

[2]Key Laboratory of Translation Medicine Tropical Diseases, Hainan Medical University Haikou 571199 P. R. China

[3]Hainan Medical University-The University of Hong Kong Joint Laboratory of Tropical Infectious Diseases, Hainan Medical University Haikou 571199 P. R. China

[#]Equal contribution

Asian Journal of Complementary and Alternative Medicine. Volume 11 Issue 01

Published on: 22/03/2023

***Author for Correspondence:** Jiping Shao, Hainan Medical University-The University of Hong Kong Joint Laboratory of Tropical Infectious Diseases, Hainan Medical University Haikou 571199 P. R. China

Cite this article as: Liang M, Hong W, Shao J. *Bioinformatics-based Prediction of Character of Envelope Glycoprotein and Analysis of Epitopes of B- and T-cell of gp120.* Asian Journal of Complementary and Alternative Medicine, Vol 11(1), 27-39:2023.

ABSTRACT

Objectives: To analyze the characteristics of gp120 gene sequence and protein structure of HIV-1 CRF01_AE and to predict B- and T-cell epitopes and to forecast binding sites between gp120 and mD1.22 by using bioinformatics in order to lay a foundation for the research and development of AIDS vaccines and peptide-based drugs.

Methods: The gene sequences and the amino acid sequences of gp120 were obtained from NCBI GeneBank database. Physicochemical property, membrane-panning region, PTMs sites of gp120 and secondary structure were analyzed by Expasy ProtParam, TMpred, NetPhos and GlycoMine and SOPMA, respectively. Epitopes of B-cell were predicted by VaxiJen v2.0, AllerTOP v2.0 and IEDB while epitopes of T-cell were predicted by IEDB. Protein-protein docking were forecast by Z-dock, pymol and so on.

Results: The calculation formula of gp120 is C1033H1633N299O319S9. The number of amino acids is 211, and the molecular weight is 23633.68. The theoretical isoelectric point (pI) is 7.76. It is estimated that 17 residues are phosphorylation sites, while gp120 has more than 5 glycosylation sites. The protein contains alpha helix (blue mark), beta turn (green mark), random coil (purple mark) and thin slice (red mark), which are 26.54%, 6.16%, 40.28% and 27.01% respectively.

We selected the top 10 possible epitopes of B cells, but only the top 5 epitopes of MHC-I. The most possible binding site between gp120 and mD1.22 was showed.

Conclusion: With the potential humoral immunity and cellular immunity, gp120 of HIV-1 is a promising target for HIV-1 treating. The interaction of gp120 with CD4 is the first step of HIV cycle which allows the entry of HIV into $CD4^+$ T-cells. This study has analyzed and predicted the binding sites between gp120 and mD1.22 and the character and B-cell and T-cell epitopes of gp120, could be a foundation for subsequent study of developing vaccine and polypeptide drugs.

Keywords: gp120; Bioinformatics; Characteristics of gp120; Prediction of gene epitopes; Binding site

INTRODUCTION

Nearly 17 years ago China launched its National HIV/AIDS Response Program, yet the epidemic still is not slowing. New cases and new deaths increase every year in 2005, 40 711 people living with human immunodeficiency virus (HIV; PLWH) were diagnosed and 5729 died, whereas in 2019, 148598 PLWH were diagnosed and 31522 died [1]. At present, AIDS is at low epidemic state but growth trend. The distribution of HIV-1 subtypes and recombinants changed over time in countries, regions, and globally [2]. Because of these reasons, the treatments had become more difficult. The development of anti-retroviral therapy (ART) is one of the medical milestones which could keep the person who got AIDS on track [3]. ART has significantly reduced AIDS-related morbidity and mortality. However, there are several drawbacks which including drug resistance, long-term drug therapy, toxicity and drug-drug interactions, poor bio-availability and lack of access to tissues and reservoirs in the current ART [4]. Therefore, to circumvent the problems mentioned above and effectively treat the HIV infection,

recent anti-AIDS therapeutic development efforts have been focusing on improving drug delivery and not just on discovering new chemical entities. Efforts have been made to design novel drug delivery systems for anti-HIV agents to reduce the dosing frequency, to enhance the bioavailability, to improve the Central Nervous System (CNS) penetration, to inhibit the CNS efflux and to deliver them to the target cells selectively with minimal side effects [4].

Nowadays, antibody and vaccine therapies are promising in the aera of virus. The conserved regions in HIV envelope glycoproteins gp41 and gp120 are involved in binding of the virus to the host cells [5]. Many targeting strategies have focused on binding to these protein sequences to achieve specificity, though the high rates of mutation exhibited by the retrovirus make it challenging to identify a suitable targeting sequence on the envelope [6]. The glycoprotein-120 (gp120) present in the outer envelope of HIV enables the entry of the virus particles into $CD4^+$ T-cells [7] resulting in damaging the human immune system. Therefore, the binding sites between gp120 and CD4 cells are particularly important in the treatment of HIV.

In a recent research, soluble expression, thermal stability and specificity of mD1.22 are significantly higher than mD1.2, and can bind and neutralize most of the isolates as a component of a bispecific multivalent protein [8]. So we forecast and analyzed the binding sites between gp120 and D1.22 by using bioinformatics in order to lay a foundation for the research and development of AIDS vaccines and peptide-based drugs. The identification of novel potent broadly neutralizing antibodies (bnAbs) against HIV-1 during the last several years gave new hopes to the old idea to use antibodies as anti-HIV-1 therapeutics. Attempts to use bnAbs alone or in combination or as components of chimeric antigen receptors (CARs), bispecific T cell engagers (BiTEs) and other bispecific proteins resulted in promising results both in vitro and in vivo [9-13]. And in advanced work [14], bispecific killer cells engagers (BiKEs), which can bind to natural killer (NK) cells through the activating receptor CD16A and guide them to cells expressing the HIV-1 envelope glycoprotein (Env), especially gp120, are a promising new weapon for elimination of infected cells and eradication of the virus. In presence of cells expressing HIV-1 envelope glycoproteins (Envs), these BiKEs induced cytokine production and killed Env-expressing cells. They also effectively mediated killing of chronically and acutely HIV-1 infected T cells by human peripheral blood mononuclear cells (PBMC).

Moreover, vaccine therapy, especially recombinant vaccine, is another strategy of treating viral infection. Last year, an outbreak of coronavirus disease, called COVID-19 started epidemic in late Dec. 2019 in Wuhan, the provincial capital of Hubei, China [15]. Numerous ideas of various vaccines had been put forward. One of those ideas is recombinant vaccine [16]. In this research, new vaccine candidate could be conceived by computerized analysis of viral sequences. The candidate vaccine could be selected eventually by retrieving the amino acid sequence of COVID-19, selecting of antigenic proteins, predicting and selecting of T-cell and B-cell epitopes, evaluating antigenicity and allergenicity, analyzing the physicochemical parameters, predicting secondary and tertiary structure and stimulating molecular dynamics [16]. The final designed proteins could direct following research.

Therefore, we consider that these procedures play important role in constructing antibodies and vaccine of treating HIV-1 envelope glycoprotein, especially gp120, as well. Here, we first analyze the formation and predict the epitopes, both T-cells and B-cells, of gp120. Primary classifications were divided into four groups, including M, N, O, and P, resulted from the HIV-1 extensive genetic variability. Group M accounts for the vast majority of global pandemic cases and includes 9 major subtypes, covering A1, A2, B, C, D, F1, F2, G, H, J, and K and circulating recombinant forms (CRFs) [17-21]. Initially within the same individual, multiple infections resulted in recombinant forms, while favoring recombination between two distinguishing genomic RNA upon coinfection of the same cell and production of heterozygous virions [22].

The envelope glycoproteins (Envs) from human immunodeficiency virus type 1 (HIV-1) mediate viral entry. The binding of the HIV-1 gp120 glycoprotein to CD4 triggers conformational changes in gp120 that allow high-affinity binding to its coreceptors. In contrast to all other Envs from the same phylogenetic group, M, which possess a serine (S) at position 375, those from CRF01_AE strains possess a histidine (H) at this location. This residue is part of the Phe43 cavity, where residue 43 of CD4 (a phenylalanine) engages with gp120 [18].

Originally, CRFs once peculiarly affected Southern Asia. CRF01 was first found in central Africa and went through varieties. On the occasion, in late 1980s, Asian CRF01 began epidemic from Thailand [19,20]. Over the past decade or so, CRF01_AE, a subtype of CRF01, was observed with an increasingly swift growth rate. Of all national infections, accounting for 42.5% in China in 2007 [21] while 95% of those in Thailand in 2004 [22]. Genome of CRF01_AE composed of the gag-pol region, most accessory genes deriving clade A while the env gene deriving E [23]. A considerable viral fitness may make a difference resulted from altering or

exchanging of genes fragments which encoding crucial proteins, like envelope (Env) [24-25]. Furthermore, immune system evasion ability and virulence change can provide virus evolutionary advantages as well [26]. In this study, the sequence of CRF01_AE has been retrieved and other character been analyzed. Afterward, secondary and tertiary structure also be predicted. Moreover, tertiary structure has been refined and validated.

MATERIALS AND METHODS

Retrieval of protein sequences

At the very beginning, the amino acid sequences of gp120 were retrieved from the National Centre for Biotechnology Information (NCBI) at www.ncbi.nlm.nih.gov. The sequences were performed in FASTA format and for succeeding analysis.

1. **Analysis of structure-founded of gp120**

The primary structure of protein is the marshalling sequence of amino acid residue. All of the residues connect via covalent linkage. The secondary structure is the way of polypeptide chain folding and coiling. The secondary structure can be classified to type-H(α-helix), type-E(β-sheet)and type-C(coil). Expasy PortParam was used to predict the constitution of amino acid, the atomic number, molecular formula, relative molecule mass, isoelectric point and stability of gp120 [27].

1. **Analysis of physicochemical property**

The composition, quantities of atoms, molecular formular, relative molecular mass, amount of positive and negative charges, isoelectric point and stability of amino acid of gp120 were analyzed by using Expasy ProtParam tool at http://www.expasy.org/tools/protscale.html [28].

1. **Analysis of membrane-panning region**

Generally speaking, transmembrane proteins do not express in prokaryotic system. If existing transmembrane domain in the sequence, eukaryotic system is needed for protein expression. In this research, the character of transmembrane domain of gp120 were forecast by using TMpred at https://embnet.vital-it.ch/software/TMPRED_form.html [29].

1.1 **The PTMs site of gp120**

Posttranslational modifications(PTMs) are the basis of cell regulation, which engages in cell signaling transduction and plays an important role in genetic expression, the effect of activation and inactivation of enzyme, the process in protein folding and keeping protein stabilizing as well as in the interaction between two proteins. Here, we predicted phosphorylation site and glycosylation site. The phosphorylation site of gp120 was forecasted by using NetPhos at http://www.cbs.dtu.dk/services/NetPhos/ [30] and glycosylation site was predicted at GlycoMine at http://glycomine.erc.monash.edu/Lab/GlycoMine [31].

1.2 **Analysis of secondary structure**

The secondary structure is the way of polypeptide chain folding and coiling. The secondary structure can be classified to type-H(α-helix), type-E(β-sheet)and type-C(coil). The secondary structure were analyzed by using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) [32].

1. **Antigenicity evaluation and Prediction of epitope of B-cell**

Antigenicity of gp120 (CRF01_AE) was predicted via using the VaxiJen v2.0 server. The VaxiJen classified antigens based on auto cross-covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties which is a novel alignment-independent method and overcome the limitations of alignment-dependent sequence methods [33]. The prediction of vaccine candidate allergenicity is essential. The allergenicity of the designed protein was computed by AllerTOP (http://www.ddg-pharmfac.net/AllerTOP/). AllerTOP method predicts recombinant protein allergenicity on auto cross-covariance ACC that describe residue hydrophobicity, size, abundance, helix and b-strand forming propensities [34]. AllerTOP v.2 has the highest accuracy (88.7%) compared to several servers for allergen prediction [34].

Immunogenicity is the ability to activate specific immune response of a substance or a molecule [31]. Stronger immunogenicity occurs on effective T-cell epitope [35]. Therefore, immunogenicity was screened by using IEDB at http://tools.iedb.org/immunogenity/. As a consequence, the immunogenicity of epitope accessed as positive were deemed to be potential immunogen [36].

The specificity of antigens depends on special sequence and space structure of amino acid residue located on molecule surface, which is called epitope or antigenic determinant. The epitope could be classified into continuous epitope and noncontinuous one, related to the kinds of amino acid, marshalling sequence, special genes, space structure, the nature of charge and hydrophilia of protein. These epitopes could be recognized by lymphocyte to trigger immune response, after which antigen-antibody reaction occurs. This is the basis of specificity of antigen. B-cell epitope forecasting is the key progress of design of antibody and vaccine.

Selecting the best B-cell epitope from xx protein by using online tool BepiPred-2.0(http://www.cbs.dtu.dk/services/

BepiPred/) [37]. BepiPred-2.0 is viewed as the most reliable tool to predict the epitope of B-cell. Simultaneously using multiple forecast tools could enhance the precise of prediction. LBtope assay(http://crdd.osdd.net/raghava//lbtope/) [38], is another reliable B-cell epitope prediction tool. ABCpred is based on Artificial neural network model, tool of forecasting B-cell epitope, could seek both Bcipep and Swiss-Prot database, accuracy rate could be at 66%.

2. **Prediction of epitope of T-cell**

T-cell epitope prediction aims to identify the shortest peptides within an antigen that are able to stimulate either CD4 or CD8 T-cells [39].

T-cell epitope is a short-peptide that submitted to T-cell receptor after APC dealt. T-cell epitope is significant for research to explore the mechanism and progress of immune cell action, produce subunit polypeptide vaccine and gene. Cytotoxic T cell (TCL) recognize the epitope combined with class 1 MHC molecule leads to killing the target cell. It's one of important immune defense reactions against tumor, transplant and kinds of infection. In this research, predicting T-cell epitope in any special protein could be achieved by using IEDB tool at http://tools.iedb.org/immunogenity/.

3. **Obtaining and preparing the structure of immune receptors**

Swiss-model (https://swissmodel.expasy.org) is used for homology modelling [44]. The amino acid sequence of gp120 was input in FASTA format, the templates would be provided subsequently. While the templates search was completed, templates were ranked according to expected quality of the resulting models, as estimated by Global Model Quality Estimate (GMQE) [45].

After selecting the templates, gp120 was prepared by autodock software where hydrogen atoms and charge were added to the structure, while deleting water from the molecule, to make the it ready for the next step [46].

4. **Protein-Protein docking**

In a recent research, mD1.22 included broad and potent neutralizing activity, high speci-ficity, stability, solubility, and affinity for gp120 and a small molecular size.Therefore,the binding sites between gp120 and mD1.22 lay a foundation for the research and development of AIDS vaccines and peptide-based drugs [8]. After preparation of the gp120, the Z-dock webserver [47] with default parameters was used to predict the best positions and orientations where gp120 can have the highest number of favorable interactions with mD1.22 [8].

The ZDOCK Server provides a fast and effective means to produce models of protein–protein complexes and symmetric multimers, via a user-friendly web interface. In addition to generating and viewing structures of docking models through the server's tools and interface, users have the option of submitting ZDOCK Server output files directly to several available docking refinement and post-processing tools (linked from the server page). In the future, possible developments include clustering and postprocessing functionality directly on the server itself, in addition to other improvements based on user requests.[48]

Results

1. **Sequence of gp120 (CRF01_AE)**

In this study, the sequence of gp120 envelope glycoprotein of CRF01_AE subtype of HIV-1 was obtained from NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank). The sequence in FASTA format was showed in the following table (Table 1). The gene ID (FM865531.1, total 636 bp) and

Table 1: FASTA format of gp120 from NCBI.

Protein	Accession number	FASTA
gp120	FM865531.1	>FM865531.1 HIV-1 CRF01_AE partial env gene for gp160, strain 07CN.HN085, genomic RNA TGTTAAATGGCAGTCTAGCAGAAGAGGAAATAATAATCAGATCTGAGAATCTCACAAA-CAATGCCAAGACCATAATAGTGCACCTTAATAAATCTGTAGTAATCGATTGTACCA-GACCCTCCAACAATACAAGAACAAGTATAACTATAGGACCGGGACAAGTATTCTATAGAA-CAGGAGAAATAATAGGAGACATAAGAAAAGCATATTGCGAAATTAATAGAACAAATTG-GTTTGAAACTTTAAACCAGGTAGCTAAAAAATTAAAAGAGCACTTTCCTAATAAGA-CAATAAGCTTTCAACCACACTCAGGAGGAGATCTAGAAATTACAACGCATCACTTTAATTG-TAAAGGGGAATTTTTCTATTGCAATACAAGCCACCTGTTTACGTACAAAGAAAATGAAAC-CAGAGGGGAAAGTACTAGCACTATCATACTTCCATGCAAGATAAAGCAAATTATAAACCT-GTGGCAGGGAGTAGGACAAGCAATGTGTGCTCCTCCCATCAGTGGAAGAATTAATTGTG-TATCAAATATTACAGGAATACTATTAACAAGAGATGGGGGTGATAATGGTAATGATACTGAC-GAGATCTTCAGACCTGGAGGAGGAAGTATGAGGGACAATTGGAGAA

accession number (CAR96232.1, total 211 amino acids) were given on NCBI.

2. Structure foundation of gp120

2.1 Analysis of physiochemical property of gp120

Analyze HIV-1 CRF01 with Expasy PortParam_ Open reading frame (ORF) of AE strain gp120. The formula of gp120 is C1033H1633N299O319S9, and the total number is 3293. The number of amino acids is 211, and the molecular weight is 23633.68. Among them, the number of isoleucine (Ile) is 25, accounting for 11.4% of the total protein, and other amino acids account for less than 10%. The ratio of the total number of negatively charged and positively charged residues is 22:23. The theoretical isoelectric point (pI) is 7.76. The instability index is 28.38, indicating that the protein is stable. Its aliphatic index is 79. The total average value of hydrophilicity is -0.457, indicating that gp120 is a hydrophobic protein (Figure 1).

2.2 Analysis of transmembrane domain of gp120

TMpred is used to predict the transmembrane region and orientation of gp120. The calculation results were shown in Table 2 and Figure 2. In TMpred, scores only more than 500 points are considered meaningful. We saw that the scores for the results from the table are 456 and 354, respectively. So these results showed that the possible transmembrane helical region of gp120 was not specific, so gp120 was not a transmembrane protein.

2.3 Analysis of ptms sites of gp120

NetPhos and glycosamine are used to predict phosphorylation sites and glycosylation sites, which are one of ptms respectively. The calculation results were shown in Figure 3, Table 3 and Table 4. As shown in the figure, potential phosphorylation sites existed at the sites of markers, where protein phosphorylation mediates the activity of substates. At the same time, N - or o-glycosylation can convert proteins into glycoproteins, which play an important role in the protection and stability of cells. Here, we gave some high score results of glycosylation sites.

Table 2: The helices of gp120.

Orientation	Start position	End position	Score	Strongly prefered
Inside to outside helices	153	171	456	no
Outside to inside helices	156	184	364	no

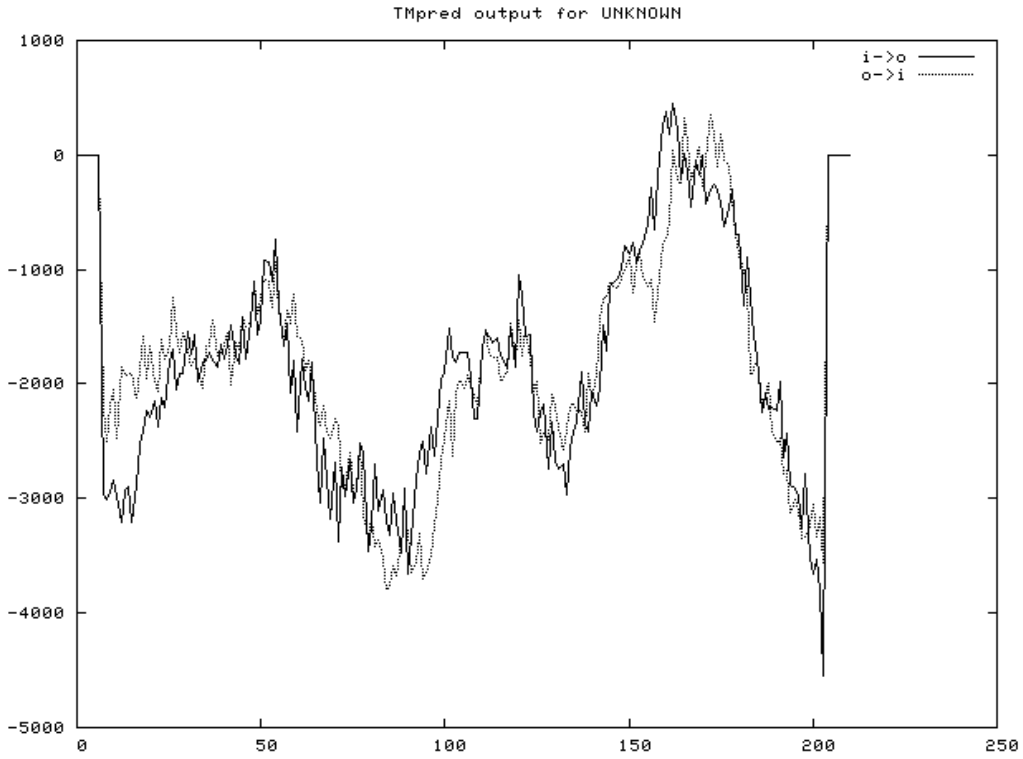

Figure 1: The analysis of gp120 transmembrane.

```
LNGSLAEEEIIIRSENLTNNAKTIIVHLNKSVVIDCTRPSNNTRTSITIG   #    50
PGQVFYRTGEIIGDIRKAYCEINRTNWFETLNQVAKKLKEHFPNKTISFQ   #   100
PHSGGDLEITTHHFNCKGEFFYCNTSHLFTYKENETRGESTSTIILPCKI   #   150
KQIINLWQGVGQAMCAPPISGRINCVSNITGILLTRDGGDNGNDTDEIFR   #   200
PGGGSMRDNWR                                          #   250
...S.............T.....................S.....S....   #    50
..................Y.....................................   #   100
..S.................Y..T....T.....T...S.S........   #   150
.....................................................T.....   #   200
....S......
```

Figure 2: Phosphorylation potential of gp120.

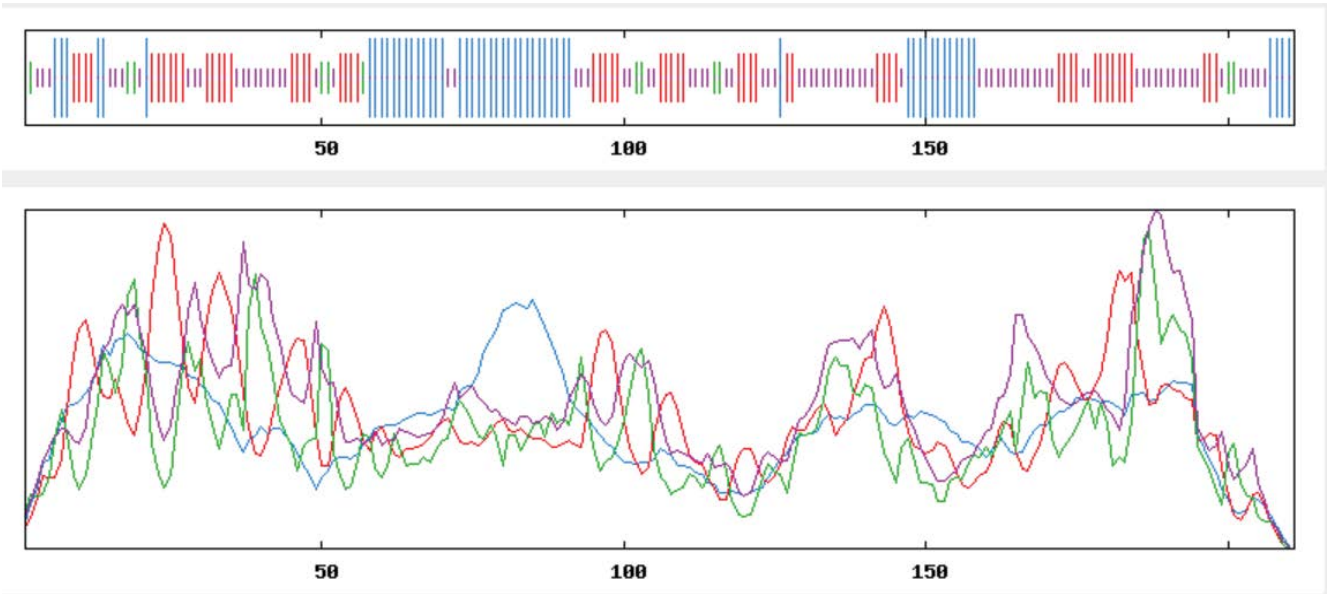

Figure 3: Secondary structure of gp120.

Table 3: N-glycosylation of gp120.

Rank	Position	Site	Motif	Score
1	19	N	IRSENLTNNAKTIIV	0.486
2	29	N	KTIIVHLNKSVVIDC	0.479
3	41	N	IDCTRPSNNTRTSIT	0.456
4	82	N	TNWFETLNQVAKKLK	0.425
5	94	N	KLKEHFPNKTISFQP	0.419
6	20	N	RSENLTNNAKTIIVH	0.403
7	178	N	GRINCVSNITGILLT	0.394
8	16	N	EIIIRSENLTNNAKT	0.378
9	124	N	KGEFFYCNTSHLFTY	0.372

Table 4: O-glycosylation of gp120.

Rank	Position	Site	Motif	Score
1	140	S	ENETRGESTSTIILP	0.507
2	142	S	ETRGESTSTIILPCK	0.506
3	136	T	FTYKENETRGESTST	0.477
4	125	T	GEFFYCNTSHLFTYK	0.450
5	37	T	KSVVIDCTRPSNNTR	0.444
6	103	S	TISFQPHSGGDLEIT	0.423
7	143	T	TRGESTSTIILPCKI	0.423
8	126	S	EFFYCNTSHLFTYKE	0.419
9	141	T	NETRGESTSTIILPC	0.397

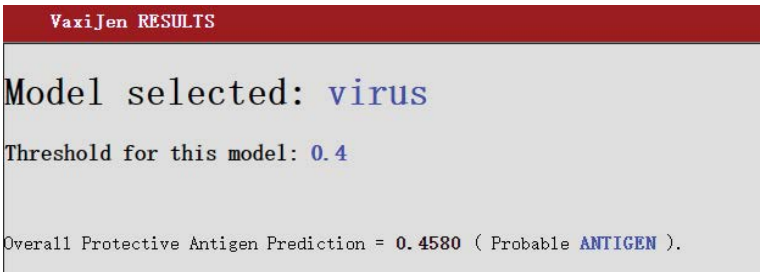

Figure 4: Antigen prediction of gp120.

Table 5: Possible epitopes of B-cell.

Rank	Sequence	Start position	Score
1	TDEIFRPGGGSMRDNW	195	0.94
2	IGDIRKAYCEINRTNW	62	0.92
3	SVVIDCTRPSNNTRTS	31	0.92
4	TSTIILPCKIKQIINL	141	0.92
5	TGILLTRDGGDNGNDT	180	0.91
6	YCEINRTNWFETLNQV	69	0.86
7	YRTGEIIGDIRKAYCE	56	0.82
8	GQAMCAPPISGRINCV	161	0.81
9	TISFQPHSGGDLEITT	96	0.80
10	EIIIRSENLTNNAKTI	9	0.80

2.2 **Analysis of secondary structure of gp120**

Online service SOMPA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) was used to predict secondary structure of gp120 glycoprotein. The result was showed (Figure 3).The protein contains alpha helix (marked with blue), beta turn (marked with green), random coil (marked with purple) and thin slice (marked with red), which are 26.54%, 6.16%, 40.28% and 27.01% respectively.

3. Evaluation of antigenicity of gp120 and Prediction of B-cell epitopes of gp120

The antigenicity of gp120 was predicted using VaxiJen v2.0, and the results were shown in Figure 4. The results showed that gp120 might be an antigen.

The predicted B cell epitopes are sorted according to the scores obtained by the trained recurrent neural network. The higher the score of peptide segment, the higher the possibility of epitope. All peptides shown here were higher than the selected threshold (0.5). Here we selected the top 10 sequences with the highest score, which may be candidate sequences for b cell epitopes (Table 5).

4. **Prediction of T-cell of gp120**

IEDB was used to predict epitopes of MHC-I. The results are displayed on Table 6. The results showed different HLA alleles. The higher the score of sequences, the higher the possibility of epitope. We chose the sequence with the top 5 scores, which may be the epitope of MHC-I.

5. **Obtained and prepared the structure of immune receptors**

One of the prediction of templates of gp120, 6ieq, is the coordinates and structure factors for the ConM SOSIP.v7 structure in complex with Fab PGT124 and Fab 35O22 [49]. We obtained the structure of gp120 from gp160 which was provided from Swiss-model. The conformation of gp120 will be displayed in figure. As a mutant of mD1.2, the difference between mD1.22 and mD1.2 is the change of 55 residue, A in mD1.2 and V in mD1.22 [8] (Figure 5).

6. **Protein-Protein docking**

In order to determine whether these two proteins adhere to the interaction, we use Z-Dock for docking research. The final docking output file uses the default Swiss-PdbViewer to analyze hydrogen bonds. FireDock provides an additional analysis tool to quantify long-range and short-range electrostatic interactions, van der Waals binding and p-p superposition energy.

Here, we showed the most possible binding site between gp120 and mD1.22. The interaction occurs on the residue. Here, we showed that different residues can interact with mD1.22 (Figure 6).

B: mD1.22 react with LYS-252, ASN-262, GLU-293, AGR-444 and THR-446 of gp120; a. outline of complex-2; b. binding sties of gp120 which could interact with mD1.22; c. the length of hydrogen bond between ARG-444 and mD1.22 is 3.1; d. the length of hydrogen bond between GLU-293 and mD1.22 is 3.0; e. the length of hydrogen bonds between THR-446 and mD1.22 are 2.1 and 3.2; f. the length of hydrogen bonds between LYS-252 and mD1.22 is 2.7; g. the length of hydrogen bonds between ASN-262 is 2.8

Our results showed that possible binding sites between gp120 and mD1.22 were LYS-252, ASN-262, GLU-293, AGR-444 and THR-446. These residues all bound with mD1.22 in a hydrogen-bonded manner. The longer the hydrogen bond length between them and mD1.22, the more stable the binding.

Table 6: Possible epitopes of MHC-I.

Allele	Start	End	Length	Peptide	Score
HLA-B*57:01	149	157	9	KIKQIINLW	0.99
HLA-B*40:01	138	146	9	GESTSTIIL	0.99
HLA-A*03:01	87	95	9	KLKEHFPNK	0.98
HLA-1*24:02	121	129	9	FYCNTSHLF	0.97
HLA-A*68:01	79	87	9	ETLNQVAKK	0.96

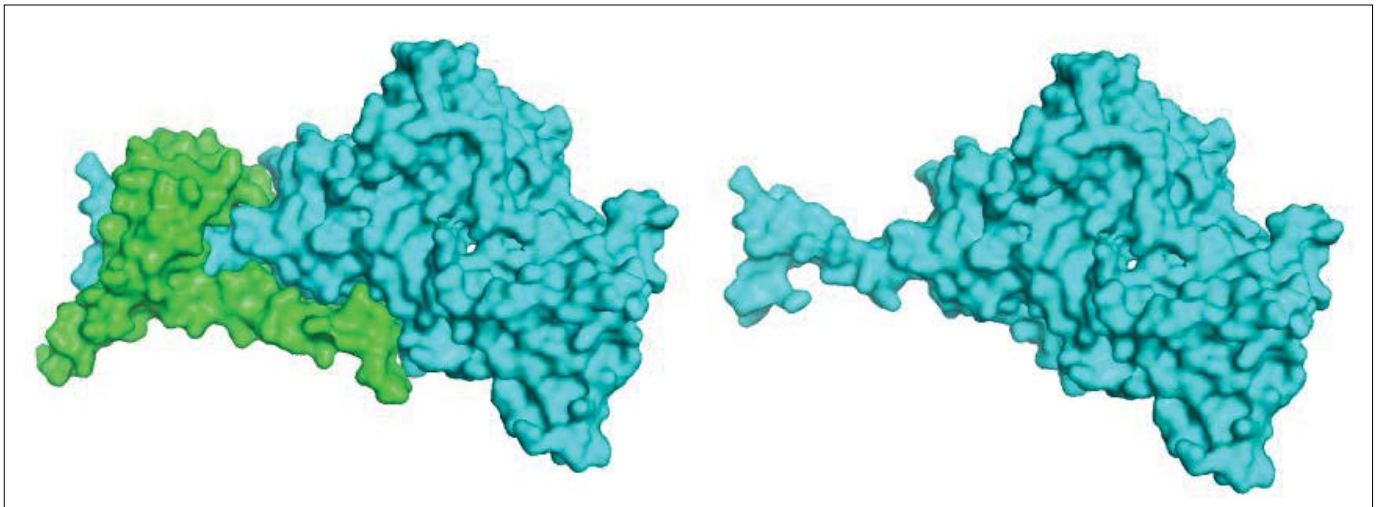

Figure 5: Model of complex of gp120-mD1.22 and gp120. (Cyan for gp120 and green for mD1.22) A: model of complex of gp120 and mD1.22. B: model of gp120.

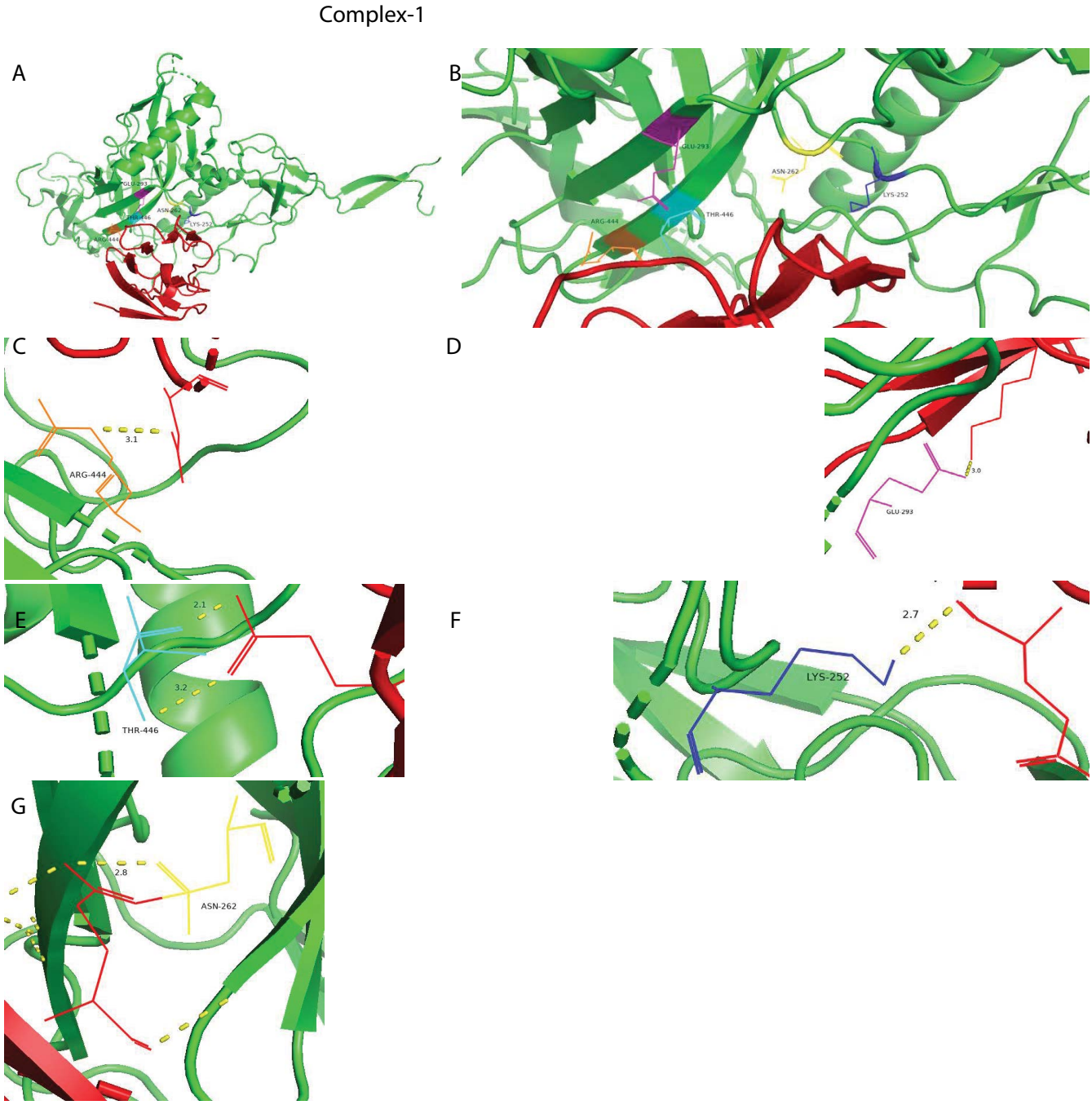

Figure 6: Possible binding site between gp120 and mD1.22 with different compounds and length of hydrogen bond. (green for gp120 and red for mD1.22).

DISCUSSION

In this study, we obtained the gp120 sequence on NCBI. Then, we analyzed the physical and chemical properties of gp120, including molecular formula, molecular weight, theoretical isoelectric point, instability index, etc. Then, we analyzed the transmembrane domain of gp120. Unfortunately, there is no transmembrane region on gp120, which means that gp120 cannot cross the membrane. Then, the phosphorylation sites and glycosylation sites were predicted by NetPhos and glycosamine, respectively. There were 14 phosphorylation sites and several glycosylation sites. However, the n-glycosylation sites did not meet the expectations, while the two o-glycosylation sites could pass through. After that,

we also analyzed the secondary structure of gp120, which contains helix, rotation, random helix and slice, accounting for 26.54%, 6.16%, 40.28% and 27.01% respectively, that is, the main secondary structure is coil. In addition, we predicted the antigenicity of gp120, and the results showed that the protein was highly possible as an antigen.

Finally, we also predicted and analyzed the characteristics of gp120 epitope, and gave the sequence of the top 10 points, which may be the candidate epitope of B cells. With the practical purpose that, forecasting of B-cell epitope is for the sake of promoting B-cell epitope identification to take place of the antigen, for producing antibody or being used in structure-function studies [42]. What's more, the aiming of forecasting t-cell epitope is to identify the shortest peptides within an antigen which are capable to excite parts of T-cells, either CD4 or CD8 [39].

The pathogen is not a whole, but is recognized by b cells or t cells, and its molecular component is antigen. Specific receptors exist on the surface of b cells or t cells, which can distinguish these antigens. Through these receptors, not only b cells and t cells need to be activated, but also the innate immune system needs to be activated. As the second activation signal, it is necessary for antigen recognition.

When recognizing linear b cell epitopes, antibodies can recognize denatured antigens. Loss of recognition for conformational B-cell epitopes would occur after antigen denaturing. Accounting for about a percentage of 90, B-cell epitopes are conformational. As a matter of fact, however, linear B-cell epitopes simply existed in minorities of native antigens [43].

Gp120, engaging in the process of human immunodeficiency virus intruding target cell, is a vitally important glycoprotein on surface of HIV-1. Bioinformatics is based on mathematical and computer methods, which could predict function and some other useful information of a protein, with large quantities of information and strong purpose. In this study, we used a combination methods of Expasy ProtParam tool, SOPMA, TMpred, ABCpred, IEDB and so on to predict the character of surface of gp120. Although gp120 is not a transmembrane protein, it is still a hot spot in succeeding research for the vaccine or antibody of HIV-1. Because of the various results here, and consisting the error, we need to verify the results with in vitro and in vivo experiments [40,41].

Biophysical characteristics of HIV Env, such as affinity for CD4 may be important predictors of its in vivo efficacy and may serve as important surrogate markers for screening Env structures as potential vaccine candidates. And the high affinity binding site for human immunodeficiency virus (HIV) envelope glycoprotein gp120 resides within the amino-terminal domain (D1) of CD4 [50]. Moreover, sCD4 was competent to replace membrane-bound CD4 to trigger infection mediated by several HIV-1 envelopes [51]. In a recent research,the engineered cavity-altered single-domain sCD4 (mD1.22) has a unique combination of excellent properties, including broad and potent neutralizing activity, high specificity, stability, solubility, and affinity for the HIV-1 envelope glycoprotein gp120, and small molecular size [8]. So we forecast the binding sites between gp120 and mD1.22 by using bioinformatics. CD4 makes numerous contacts with conserved gp120 residues and is connected with the gp120 inner domain and the coreceptor binding site via a water-filled solvent channel [52,53] Finally, we found possible residues and the strength of hydrogen bonds between residues and mD1.22.

Possible residues, including LYS-252, ASN-262, GLU-293, AGR-444 and THR-446, could come into reaction with mD1.22, we had displayed the strength of hydrogen bond here.

GP120 of HIV-1 has potential humoral and cellular immunity, and is a promising target for HIV-1 treatment. This study analyzed and predicted the characteristics of gp120, b cell and t cell epitopes, which could lay a foundation for the subsequent development of vaccines and peptide drugs.

REFERENCES

1. Wu Z, McGoogan JM, Detels R. The Enigma of the Human Immunodeficiency Virus (HIV) Epidemic in China. Clin Infect Dis. 2021 Mar 1;72(5):876-881. doi: 10.1093/cid/ciaa835. PMID: 32569373; PMCID: PMC7935386.
2. Hemelaar J, Elangovan R, Yun J, Dickson-Tetteh L, Fleminger I, Kirtley S, Williams B, Gouws-Williams E, Ghys PD; WHO–UNAIDS Network for HIV Isolation Characterisation. Global and regional molecular epidemiology of HIV-1, 1990-2015: a systematic review, global survey, and trend analysis. Lancet Infect Dis. 2019 Feb;19(2):143-155. doi: 10.1016/S1473-3099(18)30647-9. Epub 2018 Nov 30. Erratum in: Lancet Infect Dis. 2020 Mar;20(3):e27. PMID: 30509777.
3. Saag MS, Benson CA, Gandhi RT, Hoy JF, Landovitz RJ, Mugavero MJ, Sax PE, Smith DM, Thompson MA, Buchbinder SP, Del Rio C, Eron JJ Jr, Fätkenheuer G, Günthard HF, Molina JM, Jacobsen DM, Volberding PA. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2018 Recommendations of the International Antiviral Society-USA Panel. JAMA. 2018 Jul 24;320(4):379-396. doi: 10.1001/jama.2018.8431. PMID: 30043070; PMCID: PMC6415748.
4. Ramana LN, Anand AR, Sethuraman S, Krishnan UM. Targeting strategies for delivery of anti-HIV drugs. J Control Release. 2014 Oct 28;192:271-83. doi: 10.1016/j.jconrel.2014.08.003. Epub 2014 Aug 10. PMID: 25119469; PMCID: PMC7114626.

5. Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC. Atomic structure of the ectodomain from HIV-1 gp41. Nature. 1997 May 22;387(6631):426-30. doi: 10.1038/387426a0. PMID: 9163431.

6. Agrawal-Gamse C, Lee FH, Haggarty B, Jordan AP, Yi Y, Lee B, Collman RG, Hoxie JA, Doms RW, Laakso MM. Adaptive mutations in a human immunodeficiency virus type 1 envelope protein with a truncated V3 loop restore function by improving interactions with CD4. J Virol. 2009 Nov;83(21):11005-15. doi: 10.1128/JVI.01238-09. Epub 2009 Aug 19. PMID: 19692476; PMCID: PMC2772790.

7. D.G. Myszka, R.W. Sweet, P. Hensley, M. Brigham-Burke, P.D. Kwong, W.A.Hendrickson, R. Wyatt, J. Sodroski, M.L. Doyle, Energetics of the HIV gp120–CD4 binding reaction, Proc. Natl. Acad. Sci. 97 (2000) 9026–9031.

8. Chen W, Feng Y, Prabakaran P, Ying T, Wang Y, Sun J, Macedo CD, Zhu Z, He Y, Polonis VR, Dimitrov DS. Exceptionally potent and broadly cross-reactive, bispecific multivalent HIV-1 inhibitors based on single human CD4 and antibody domains. J Virol. 2014 Jan;88(2):1125-39. doi: 10.1128/JVI.02566-13. Epub 2013 Nov 6. PMID: 24198429; PMCID: PMC3911630.

9. Pegu A, Asokan M, Wu L, Wang K, Hataye J, Casazza JP, Guo X, Shi W, Georgiev I, Zhou T, Chen X, O'Dell S, Todd JP, Kwong PD, Rao SS, Yang ZY, Koup RA, Mascola JR, Nabel GJ. Activation and lysis of human CD4 cells latently infected with HIV-1. Nat Commun. 2015 Oct 20;6:8447. doi: 10.1038/ncomms9447. PMID: 26485194; PMCID: PMC4633990.

10. Sloan DD, Lam CY, Irrinki A, Liu L, Tsai A, Pace CS, Kaur J, Murry JP, Balakrishnan M, Moore PA, Johnson S, Nordstrom JL, Cihlar T, Koenig S. Targeting HIV Reservoir in Infected CD4 T Cells by Dual-Affinity Re-targeting Molecules (DARTs) that Bind HIV Envelope and Recruit Cytotoxic T Cells. PLoS Pathog. 2015 Nov 5;11(11):e1005233. doi: 10.1371/journal.ppat.1005233. PMID: 26539983; PMCID: PMC4634948.

11. Zhen A, Kamata M, Rezek V, Rick J, Levin B, Kasparian S, Chen IS, Yang OO, Zack JA, Kitchen SG. HIV-specific Immunity Derived From Chimeric Antigen Receptor-engineered Stem Cells. Mol Ther. 2015 Aug;23(8):1358-1367. doi: 10.1038/mt.2015.102. Epub 2015 Jun 8. PMID: 26050990; PMCID: PMC4817874.

12. Leibman RS, Riley JL. Engineering T Cells to Functionally Cure HIV-1 Infection. Mol Ther. 2015 Jul;23(7):1149-1159. doi: 10.1038/mt.2015.70. Epub 2015 Apr 21. PMID: 25896251; PMCID: PMC4817793.

13. Liu L, Patel B, Ghanem MH, Bundoc V, Zheng Z, Morgan RA, Rosenberg SA, Dey B, Berger EA. Novel CD4-Based Bispecific Chimeric Antigen Receptor Designed for Enhanced Anti-HIV Potency and Absence of HIV Entry Receptor Activity. J Virol. 2015 Jul;89(13):6685-94. doi: 10.1128/JVI.00474-15. Epub 2015 Apr 15. PMID: 25878112; PMCID: PMC4468509.

14. Li W, Wu Y, Kong D, Yang H, Wang Y, Shao J, Feng Y, Chen W, Ma L, Ying T, Dimitrov DS. One-domain CD4 Fused to Human Anti-CD16 Antibody Domain Mediates Effective Killing of HIV-1-Infected Cells. Sci Rep. 2017 Aug 22;7(1):9130. doi: 10.1038/s41598-017-07966-3. PMID: 28831040; PMCID: PMC5567353.

15. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020 Feb 15;395(10223):470-473. doi: 10.1016/S0140-6736(20)30185-9. Epub 2020 Jan 24. Erratum in: Lancet. 2020 Jan 29;: PMID: 31986257; PMCID: PMC7135038.

16. Enayatkhani M, Hasaniazad M, Faezi S, Guklani H, Davoodian P, Ahmadi N, Einakian MA, Karmostaji A, Ahmadi K. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. J Biomol Struct Dyn. 2020 May 2:1-16. doi: 10.1080/07391102.2020.1756411. Epub ahead of print. PMID: 32295479; PMCID: PMC7196925.

17. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373:123–126. https://doi.org/10.1038/373123a0

18. Zoubchenok D, Veillette M, Prévost J, Sanders-Buell E, Wagh K, Korber B, Chenine AL, Finzi A. 2017. Histidine 375 modulates CD4 binding in HIV-1 CRF01_AE envelope glycoproteins. J Virol 91:e02151-16. https://doi.org/10.1128/JVI.02151-16.

19. Zhang M, Foley B, Schultz AK, Macke JP, Bulla I, Stanke M, Morgenstern B, Korber B, Leitner T. 2010. The role of recombination in the emergence of a complex and dynamic HIV epidemic. Retrovirology 7:25. https://doi.org/10.1186/1742-4690-7-25.

20. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. 2000. Timing the ancestor of the HIV-1 pandemic strains. Science 288:1789 –1796. https://doi.org/10.1126/science.288.5472.1789.

21. Hemelaar J, Gouws E, Ghys PD, Osmanov S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS

22. W13–W23. https://doi.org/10.1097/01.aids.0000247564.73009.bc.20. Carr JK, Salminen MO, Koch C, Gotte D, Artenstein AW, Hegerich PA, St Louis D, Burke DS, McCutchan FE. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J Virol 70:5935–5943.

23. Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefeuvre P, Martin DP, Robertson DL, Negroni M. 2009. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. PLoS Pathog 5:e1000418. https://doi.org/10.1371/journal .ppat.1000418.

24. Archer J, Pinney JW, Fan J, Simon-Loriere E, Arts EJ, Negroni M, Robertson DL. 2008. Identifying the important HIV-1 recombination breakpoints. PLoS Comput Biol 4:e1000178. https://doi.org/10.1371/journal.pcbi.1000178.

25. Finzi A, Xiang SH, Pacheco B, Wang L, Haight J, Kassa A, Danek B, Pancera M, Kwong PD, Sodroski J. 2010. Topological layers in the HIV-1 gp120 inner domain regulate gp41 interaction and CD4-triggered conformational transitions. Mol Cell 37:656 – 667. https://doi.org/10.1016/j.molcel.2010.02.012.

26. Burke DS. 1997. Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis 3:253–259. https://doi.org/10.3201/eid0303.970301.

27. MR, W., et al., Protein identification and analysis tools in the ExPASy server. Methods in molecular biology (Clifton, N.J.), 1999. 112: p. 531-52

28. Huercha, et al., Sequence identification and expression profile of seven Dermacentor marginatus glutathione S-transferase genes. Experimental & applied acarology, 2020. 82(2): p. 295-308

29. TQ, L., et al., [The highly expressed secreted phosphoprotein 1 gene in prostate cancer metastasis: a microarray-based bioinformatic analysis]. Zhonghua nan ke xue = National journal of andrology, 2014. 20(11): p. 984-90.

30. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol. 1999 Dec 17;294(5):1351-62. doi: 10.1006/jmbi.1999.3310. PMID: 10600390.

31. Li F, Li C, Wang M, Webb GI, Zhang Y, Whisstock JC, Song J. GlycoMine: a machine learning-based approach for predicting N-, C- and O-linked glycosylation in the human proteome. Bioinformatics. 2015 May 1;31(9):1411-9. doi: 10.1093/bioinformatics/btu852. Epub 2015 Jan 6. PMID: 25568279.

32. A, S., P. A and P. M, In silico analysis and homology modeling of strictosidine synthase involved in alkaloid biosynthesis in catharanthus roseus. Journal, genetic engineering & biotechnology, 2020. 18(1): p. 44.

33. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Jan 5;8:4. doi: 10.1186/1471-2105-8-4. PMID: 17207271; PMCID: PMC1780059..

34. Dimitrov I, Flower DR, Doytchinova I. AllerTOP--a server for in silico prediction of allergens. BMC Bioinformatics. 2013;14 Suppl 6(Suppl 6):S4. doi: 10.1186/1471-2105-14-S6-S4. Epub 2013 Apr 17. PMID: 23735058; PMCID: PMC3633022

35. UK, A., T. M and R. MM, Immunoinformatics Approach for Epitope-Based Peptide Vaccine Design and Active Site Prediction against Polyprotein of Emerging Oropouche Virus. Journal of immunology research, 2018. 2018: p. 6718083.

36. L, P., B. M and P. LB, An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. Journal of immunology research, 2016. 2016: p. 6760830.

37. MC, J., et al., BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic acids research, 2017. 45: p. W24-W29.

38. H, S., A. HR and R. GP, Improved method for linear B-cell epitope prediction using antigen's primary sequence. PloS one, 2013. 8(5): p. e62216.

39. Ahmed RK, Maeurer MJ. T-cell epitope mapping. Methods Mol Biol. 2009;524:427-38. doi: 10.1007/978-1-59745-450-6_31. PMID: 19377963.

40. Michel LV, Kalmeta B, McCreary M, Snyder J, Craig P, Pichichero ME. Vaccine candidate P6 of nontypable Haemophilus influenzae is not a transmembrane protein based on protein structural analysis. Vaccine. 2011 Feb 11;29(8):1624-7. doi: 10.1016/j.vaccine.2010.12.082. Epub 2011 Jan 5. PMID: 21215345; PMCID: PMC3070289.

41. Michel LV, Snyder J, Schmidt R, Milillo J, Grimaldi K, Kalmeta B, Khan MN, Sharma S, Wright LK, Pichichero ME. Dual orientation of the outer membrane lipoprotein P6 of nontypeable haemophilus influenzae. J Bacteriol. 2013 Jul;195(14):3252-9. doi: 10.1128/JB.00185-13. Epub 2013 May 17. PMID: 23687267; PMCID: PMC3697637.

42. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and Methods for T- and B-Cell Epitope Prediction. J Immunol Res. 2017;2017:2680160. doi: 10.1155/2017/2680160. Epub 2017 Dec 28. PMID: 29445754; PMCID: PMC5763123.

43. Van Regenmortel MH. What is a B-cell epitope? Methods Mol Biol. 2009;524:3-20. doi: 10.1007/978-1-59745-450-6_1. PMID: 19377933.

44. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018 Jul 2;46(W1):W296-W303. doi: 10.1093/nar/gky427. PMID: 29788355; PMCID: PMC6030848.

45. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W252-8. doi: 10.1093/nar/gku340. Epub 2014 Apr 29. PMID: 24782522; PMCID: PMC4086089.

46. Laederach A, Dowd MK, Coutinho PM, Reilly PJ. Automated docking of maltose, 2-deoxymaltose, and maltotetraose into the soybean beta-amylase active site. Proteins. 1999 Nov 1;37(2):166-75. doi: 10.1002/(sici)1097-0134(19991101)37:2<166::aid-prot3>3.0.co;2-e. PMID: 10584063.

47. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins. 2003 Jul 1;52(1):80-7. doi: 10.1002/prot.10389. PMID: 12784371.

48. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics. 2014 Jun 15;30(12):1771-3. doi: 10.1093/bioinformatics/btu097. Epub 2014 Feb 14. PMID: 24532726; PMCID: PMC4058926.

49. Sliepen K, Han BW, Bontjer I, Mooij P, Garces F, Behrens AJ, Rantalainen K, Kumar S, Sarkar A, Brouwer PJM, Hua Y, Tolazzi M, Schermer E, Torres JL, Ozorowski G, van der Woude P, de la Peña AT, van Breemen MJ, Camacho-Sánchez JM, Burger JA, Medina-Ramírez M, González N, Alcami J, LaBranche C, Scarlatti G, van Gils MJ, Crispin M, Montefiori DC, Ward AB, Koopman G, Moore JP, Shattock RJ, Bogers WM, Wilson IA, Sanders RW. Structure and immunogenicity of a stabilized HIV-1 envelope trimer based on a group-M consensus sequence. Nat Commun. 2019 May 29;10(1):2355. doi: 10.1038/s41467-019-10262-5. PMID: 31142746; PMCID: PMC6541627.

50. Truneh A;Buck D;Cassatt D R;Juszczak R;Kassis S;Ryu S E;Healey D;Sweet R;Sattentau Q A region in domain 1 of CD4 distinct from the primary gp120 binding site is involved in HIV infection and virus-mediated fusion. [J] Journal of Biological Chemistry,1991

51. Esser U;Speck R F;Deen K C;Atchison R E;Sweet R;Goldsmith M A Molecular function of the CD4 D1 domain in coreceptor-mediated entry by HIV type 1. [J] AIDS Research and Human Retroviruses,2000

52. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393:648–659.

53. Madani N, Perdigoto AL, Srinivasan K, Cox JM, Chruma JJ, LaLonde J, Head M, Smith AB, III, Sodroski JG. 2004. Localized changes in the gp120 envelope glycoprotein confer resistance to human immunodeficiency virus entry inhibitors BMS-806 and #155. J Virol 78:3742–3752.