

Immunoinformatics Study of gp120 of Human Immunodeficiency Virus Type 1 Subtype CRF35_AD Isolated from Iranian Patients

Hossein Keyvani,¹ Nayeb Ali Ahmadi,^{2,*} Mohammad Mehdi Ranjbar,³ Saeid Ataei Kachooei,³ Khodayar Ghorban,⁴ and Maryam Dadmanesh⁵

¹Department of Virology, Iran University of Medical Sciences, Tehran, IR Iran

²Proteomics Research Center, Department of Medical Lab Technology, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran

³Department of Immunology, Razi Vaccine and Sera Research Institute, Karaj, IR Iran

⁴Department of Immunology, School of Medicine, AJA University of Medical Sciences, Tehran, IR Iran

⁵Department Infectious Diseases, School of Medicine, AJA University of Medical Sciences, Tehran, IR Iran

*Corresponding author: Nayeb Ali Ahmadi, Proteomics Research Center, Department of Medical Lab Technology, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran. Tel: +98-2122718528, Fax: +98-2122721150, E-mail: nayebalia@sbmu.ac.ir, nayebalia@yahoo.com

Received 2016 January 15; Revised 2016 July 19; Accepted 2016 August 02.

Abstract

Background: Human Immunodeficiency Virus (HIV)-1 infection is one of the most important infectious diseases in Iran, and common isolates belong to the new CRF35_AD subtypes. The Gp120 protein, which is located on the surface of the HIV envelope, plays a role in entrance into host cells and immune responses. As there isn't any clear analysis of the envelope protein of Iranian isolates regarding potential variations, structural and immunological properties, in the current study we attempted to research in this area.

Objectives: The present study was designed to demonstrate the immunoinformatics of gp120 of human immunodeficiency virus Type 1 subtype CRF35_AD, isolated from Iranian patients.

Methods: In this analytical perspective bioinformatics study, the steps were as follows; data collection and sequence classification (303 sequences), finding the mutational/conserved regions, evaluation of N-linked glycosylation sites, prediction of tertiary structures, model validation, and prediction of conformational and linear B-cell epitopes.

Results: High degrees of sequence variation in the CRF35_AD subtype and also more than 10 variation sites in gp120 protein segments were identified. The total N-linked glycosylation sites for selected complete *env* sequences in NX [ST] pattern and the NXS and NXT combination count revealed that most of the glycosylation sites were conserved. Tertiary structure was obtained by homology modeling, and the Ramachandran plot assessment showed model validity. Finally, mapped consensus discontinuous immunogenic regions (epitopes) were AA25-65, AA337-365 and AA443-505.

Conclusions: The obtained results provide a background for understanding CRF35_AD molecular characteristics, as well as design and development of effective HIV-1 vaccines and immunotherapeutic regimens against CRF35_AD subtype.

Keywords: Immunoinformatics, HIV-1, CRF35_AD Subtype, gp120, Iran

1. Background

Human immunodeficiency virus (HIV) is a member of the genus lentivirus and the family retroviridae, which causes acquired immunodeficiency syndrome (AIDS). This is a condition in which progressive failure of the immune system allows life-threatening opportunistic infections and cancers to thrive (1, 2). In Iran, the first case of HIV was detected in 1996, and the prevalence of HIV among the general population in Iran has remained low until now, but it has a high prevalence among injection drug users (IDUs). Measures adopted over the years resulted in slowed progression of the epidemic amongst IDUs, but the role of sexual transmission in the spread of HIV in Iran is alarmingly growing (3-5). Therefore, the HIV epidemic in Iran is

in the concentrated phase (6).

The primary envelope (*env*) gene product is a single polypeptide precursor named gp160, which is then cleaved into gp120 (~ 480 amino acids) and gp41 (~ 345 amino acids) in the endoplasmic reticulum by the cellular protease (7). The gp120 is an extracellular protein located on the surface of HIV envelope, while gp41 acts as an integral transmembrane protein. The entrance of HIV into cells is dependent on gp120, as it acts in attachment to specific cell surface receptors such as CD4 receptor and chemokine receptors (8, 9). Glycosylations of gp120 plays a vital role in mediating interactions of the virus with innate and adaptive components of the immune system (10).

Although there is little evidence on variations and

immunoinformatics characteristics of CRF35_AD subtype, there are noticeable data available for other subtypes. Envelope region of different HIV-1 isolates showed high variability (approximately 30%), and were mainly located in five hypervariable regions (HVRs), V1 to V5. These variations are related to the infidelity of reverse transcriptase enzyme that has no editing process for transcriptional errors (11).

Also, investigations on antibody mapping showed that the linear epitopes on the gp120 glycoprotein (located in the V2 and V3 regions, constitute the most highly exposed peptides on the HIV multimeric envelope glycoprotein complex (12). Antibodies against the V3 loop of gp120 neutralize HIV infection, however, these antibodies are mostly type-specific, and do not possess the potential of broad neutralization (12). Variation events in the V2 and V3 regions of the envelope glycoprotein increase the ability of the viral strains to infect different cell types (13). Besides, accumulation of variations results in changes in biological properties of viruses and these variations significantly affect pathogenesis of the disease (14, 15). The function of a protein is dependent on its tertiary (3D) structure, and *in vitro* methods for identification of protein's 3D structure by X-ray crystallography are time-consuming and very expensive in comparison with *in silico* prediction methods (16, 17). Moreover, clarification of 3D structure helps immunoinformatics and chemoinformatics, required for designing novel vaccine and inhibitory medicines (18).

Since there are not much data available on variability and immunoinformatics characteristics of Iranian isolates of HIV-1 (CRF35_AD subtype), the present study can be valuable in providing background data required for development of preventive vaccines and therapeutic regimens.

2. Objectives

The present study concentrated on finding mutational/conserved regions, generation of 3D structure of the gp120 protein by homology modeling, then validation of this model, and finally analysis of *in silico* immunological properties of gp120 protein in Iranian isolates, CRF35_AD subtype.

3. Methods

3.1. Data Collection and Sequence Classification

In this analytical perspective bioinformatics study, the complete coding region of putative *env* protein sequences and partial gp120 protein sequences of 303 CRF35_AD isolates, which were assumed to be Iranian HIV-1 subtypes, were retrieved from the National center for biotechnology information (NCBI) database (<http://ncbi.nlm.nih.gov/>), and datasets were built.

3.2. Application of Hybrid Approach for Finding the Mutational/Conservative Regions

The retrieved sequences were aligned using the ClustalW algorithm (19), and analyzed and trimmed in the Bioedit 7.7.9 software. Subsequently, short sequences and gaps were excluded from datasets. The Shannon entropy analysis (20), the Simpson index (21) and the Wu-Kabat variability coefficient (22, 23) approaches were exploited for evaluation of variability, and finding mutational/conservative regions, among sequences. The Shannon entropy analysis (20) is the most sensitive method for evaluation of variations. The Simpson index is another variation analysis method for calculation of diversities. The Wu-Kabat variability coefficient is a well-established descriptor of susceptibility of an amino acid position to evolutionary replacements (23).

3.3. Prediction of N-Linked Glycosylation Sites

For predicting N-linked glycosylation sites, which is one of the Post-Translational Modifications (PTMs), we took advantage of the GlycoSite tool at Los Alamos HIV database (24). This tool highlights and tallies predicted Nx [ST] patterns, where x can be any amino acid. In glycosylation processes, an oligosaccharide chain binds to asparagine (N) occurring in the tri-peptide sequence N-X-S or N-X-T, where X can be any amino acid except Pro (24). Accession numbers of the used sequences were BAM37429, BAM37384, BAM37456 and BAM37411, named A, B, C and D, respectively (6). Of course, we randomly selected these four sequences.

3.4. Prediction of Tertiary Structure (Three-Dimensional (3D) Structure)

The BLASTP (protein basic local alignment search tool) and PSIBLAST in protein data bank (PDB) were applied for selection of the most homologous sequences to target sequences as template, and for constructing 3D model of gp120 protein of CRF 35_AD (accession number BAM37429) by the homology modeling method. The mentioned parameters for selecting template were resolution below 3 angstrom related to its x-ray crystallography, R value below 0.3, similarity of template with targeted sequence above 70% and also low E value and root-mean-square deviation (RMSD). For gaining target-template alignment, ClustalW, using a gap penalty of 10 and a gap extension penalty of 0.05, was employed. Target-template alignment and loop modeling were used for subsequent model generation by the MODELLER software version 9v7. Then the 3D model energy minimized in the Swiss-Pdb Viewer software v4.1.

3.5. Homology Models Validation

Evaluation of model validity quality is an essential step in homology modeling. A geometric evaluation of the modeled 3D structure was performed using the Ramachandran plot (25).

3.6. Prediction of Conformational and Linear B-Cell Epitope Based on Physico-Chemical Profiles

We took advantage of DiscoTope 2.0, ElliPro and CBTOPE servers for conformational predictions, also called discontinuous, B cell epitopes from protein 3D structures. Kolaskar and Tongaonkar prediction method is a physico-chemical approach for B-cell epitope prediction based on a semi-empirical method, developed on physico-chemical properties of amino acid residues. This approach has enough efficiency to detect antigenic peptides with about 75% accuracy (26).

4. Results

4.1. Variation Plots for Finding Mutational and Conservational Sites

In the present study, according to the server suggestions, positions with $H > 2.0$ in Shannon entropy plot, $D > 0.6$ in Simpson plot and variability > 15 in Wu-Kabat are considered variable, whereas those below this threshold were supposed to be semi-conserved or conserved regions.

The entropy of the *env* of the Iranian HIV-1 isolates showed different regions in protein sequences (showed by arrows in Figure 1, stretch between amino acid 80 - 97, 130 - 200, 310, 332 - 353, 375 - 378, 410 - 425, 474 - 488, 640 - 645 and 662 - 674) with high entropy (or hypervariable regions, HVRs), and in gp120 separately (Figure 2, amino acid positions 48, 64, 77, 93, 96, 99, 103, 111, 117 and 123). Wu-Kabat showed low sensitivity and four HVR regions in *env* and five in gp120. Results of Shannon and Simpson variation plots were very similar to each other.

The vast majority of positions in gp41 segment (Figure 1), exhibited low to moderate entropy (≤ 2.0) (except two highly variable regions), indicating lower probability of mutations occurring over time in comparison with gp120 (13 highly variable regions) of CRF35_AD subtype. As it is clearly demonstrated in Figure 3, in case of CRF 35 AD, based on all sequences registered in NCBI Los Alamos until now, in Iranian isolates of HIV-1 that belong to different subtypes, there are 10 or more sites of variability, most of which are located in COOH termini and loop of gp120 protein. These highly variable regions are demonstrated in 3D structure of *env* protein of CRF35_AD subtype that was constructed by homology modeling in this study (Figure 3).

Overall, entropy analysis revealed numerous highly variable regions and very low conserved regions, especially in gp120 segment.

4.2. N-Linked Glycosylation Sites

The total potential in N-linked glycosylation sites (PNGSs) in NX [ST] pattern for A, B, C and D isolates was 30, 29, 33, and 28, respectively. Also, the NXS and NXT combination count for AC; BAM37429, BAM37384, BAM37456 and BAM37411 were 11 and 19, 12 and 17, 7 and 26, and 13 and 15, respectively (see details in Figure 4). According to the results, most of the glycosylation sites are conserved, although there are differences in number of glycosylations between isolates that will effect potency of virus in infecting Iranian hosts. Glycosylation is more frequent in gp12 segment, in amino acids ~ 130 to 400.

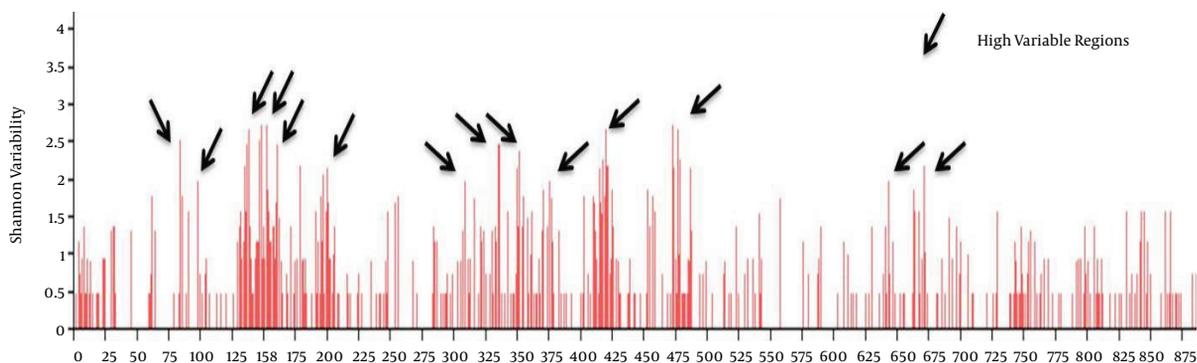
4.3. 3D Structure and Model Validation

By homology modeling followed with energy minimization, 3D structures of gp120 protein of CRF35_AD subtype were predicted based on 4TVP template. The gp120 3D model with a total number of 522 residues was validated using the Ramachandran plot. Assessment of the plot showed that 84.9% of residues (443 amino acids) were in the most favored regions, 8.8% (46 amino acids) in allowed regions, and 6.3% (33 amino acids) in the outlier region. The overall percentage of residues in favored and allowed regions was 93.7. The results obtained for quality of predicted model are acceptable because most of the assessed residues are located in the favored and allowed regions of the Ramachandran plot.

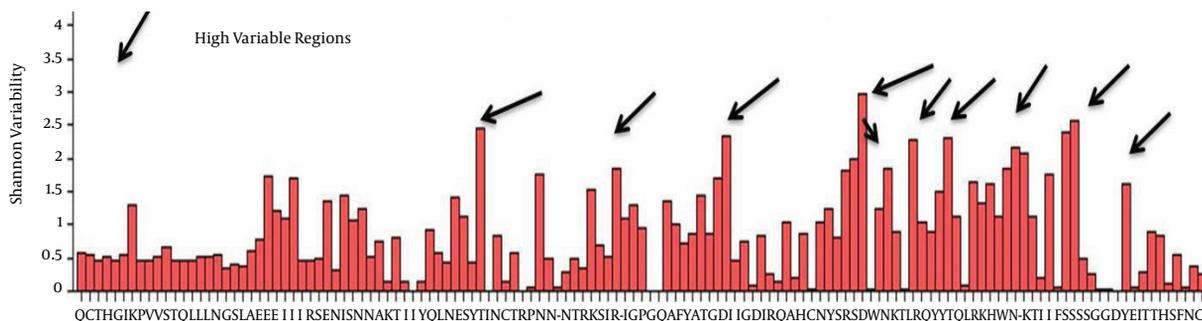
4.4. Prediction of Conformational and Linear B-Cell Epitopes Based on Physico-Chemical Properties

Identified conformational and linear B-cell epitope out of 524 total residues for gp120, subtype CRF35_AD, by hybrid approach, is presented in Table 1 represents in 3D structure (Figure 5). Also, results of DicoTope and ElliPro servers, and mapping consensus immunogenic epitopes were visualized and represented in 3D modeled structures (Figure 5).

Immunogenic epitopes are represented in Ribbon (DicoTope) and line and spheres (ElliPro) by software default in upper Figure 5. As there are arranged epitopes in Table 1, some predicted epitopic regions by DiscoTope and ElliPro software, and also in linear epitope prediction results, are overlapping (or consensus). In lower figures by two representations (ribbon & ribbon and surface representation), consensus immune region based on the results of conformational (DiscoTope and ElliPro) and Linear (Kolaskar and Tongaonkar) is depicted.

Figure 1. Demonstration of Variability Plots by Shannon Method for Complete Envelope Protein Sequences (CRF35_AD)

Values were computed from the alignments of *env* sequences. This plot shows variations along protein sequences. The x-axis represents protein sequence positions, and the y-axis depicts the Shannon computed results.

Figure 2. Variability in Shannon plot for Partial gp120 Protein Sequences

Values were computed from the alignments of *env* sequences. This plots shows variations along protein sequences of Iranian isolates of HIV-1. The x-axis shows the protein sequence positions, and the y-axis represents the Shannon computed results.

4.5. Mapping Consensus Discontinuous Immunogenic Regions (Epitopes)

Based on conformational and linear results of servers, transmembrane topology and N-Linked glycosylation sites comparisons, finally three more consensus immunogenic regions by several predictors in Iranian gp120 protein were selected.

These regions were as follow;

AA25-65 (NH₃-ICSAVENLWVTVYYGVPVW RDAETTLFCASDAKAYETEAH-COOH), AA337-365 (NH₃-WNNTLGQVAAQLRKHWNT IIFSNPSSGGD-COOH) and AA443-505 (NH₃-TGLLLTRDGGGGNQ SQNETFRPGGDMRD NWRSELYKYKVVKIEPLGVA PTKAKRRVVQREKRAV-COOH). Also they were represented in Figure 5.

5. Discussion

There is no longer doubt that HIV-1 infection is the most important infectious disease in developing countries in-

cluding Iran. There are a lot of concerns about the HIV/AIDS situation and route of transmission of the infection in Iran. The main concern is the potential of transmission between drug users through contaminated injection. Recently, the transmission pattern has shifted towards the sexual transmission, showing an increase of doubling, according to Iranian CDC (27, 28).

To the best of our knowledge, there is no other published immunoinformatics and variation analysis on gp120 protein of CRF35_AD subtype, the dominant Iranian isolate analyzed in the present study. Therefore, this study aimed to address the variations in Iranian isolates.

The envelope protein of the CRF35_AD subtype, the main Iranian HIV-1 isolate, contained 17 HVRs and gp120, separately 10 regions in all reported sequences, which reflects the relatively high degree of sequence variations among the studied isolates. Since gp120 segment has a pivotal role in the ability of HIV-1 to enter CD4⁺ T cells, it is quite sensible for the immune system neutralizing anti-

Table 1. Predicted Conformational and Linear B-cell Epitopes for Iranian Isolate of HIV-1 CRF 35_AD gp120 Protein

Server Name	Protein Name	Start-End Positions
The Disco Tope server	Gp120	110-112, 114-122, 137-139, 156-163, 249-258, 321, 324-332, 338, 366-379, 396, 424-436, 466-472, 519-523
CBTOPE	Gp120	Linear: 14-15, 25, 27, 32-37, 39-40, 45-56, 61, 76, 83-85, 89-90, 95, 123, 130, 151, 172-174, 180, 186-189, 205-225, 230, 233, 272-279, 292, 298, 312-319, 330-334, 353, 357-364, 376-382, 385-391, 397-412, 429-432, 436-437, 439, 441, 444-454, 479, 483, 484-485, 503-505, 518-524 1-21, 195-197, 459-498, 501-502, 505, 508-50
ElliPro	Gp120	Conformational^a: (1) A:I1, A:C2, A:S3, A:A4, A:V5, A:E6, A:N7, A:L8, A:W9, A:V10, A:T11, A:V12, A:Y13, A:Y14, A:G15, A:V16, A:P17, A:V18, A:W19, A:R20, A:D21, A:P195, A:A196, A:G197, A:F198, A:K459, A:I460, A:E461, A:P462, A:L463, A:G464, A:V465, A:A466, A:P467, A:T468, A:K469, A:A470, A:K471, A:R472, A:R473, A:V474, A:V475, A:Q476, A:R477, A:E478, A:K479, A:R480, A:A481, A:V482, A:G483, A:L484, A:G485, A:A486, A:L487, A:I489, A:G490, A:G493, A:A494, A:A495, A:G496, A:S497, A:T498, A:A501, A:A502, A:T505, A:V508, A:Q509 (2) A:K206, A:E207, A:A241, A:K242, A:E243, A:E244, A:I245, A:S249, A:E250, A:N251, A:I252, A:S253, A:N254, A:N255, A:A256, A:K257, A:V307, A:S308, A:R309, A:S310, A:E311, A:N313, A:N314, A:L316, A:G317, A:Q318, A:A320, A:A321, A:Q322, A:L323, A:R324, A:K325, A:H326, A:W327, A:N328, A:K329, A:T330, A:I331, A:I332, A:F333, A:S334, A:N335, A:S360, A:G361, A:F363, A:N364, A:S365, A:T366, A:W367, A:N368, A:T369, A:N370, A:G371, A:S372, A:E373, A:G374, A:S375, A:T376, A:D377, A:T378, A:S379, A:G380, A:N381, A:I382, A:R424, A:D425, A:G426, A:G427, A:G428, A:G429, A:N430, A:Q431, A:S432, A:Q433, A:N434, A:E435, A:T436 (3) A:E36, A:T37, A:E38, A:K91, A:P92, A:C93, A:V94, A:K95, A:L96, A:T97, A:P98, A:L99, A:C100, A:V101, A:T102, A:L103, A:N104, A:C105, A:T106, A:N107, A:A108, A:N109, A:I110, A:T111, A:M112, A:T113, A:S114, A:I115, A:T116, A:N117, A:T118, A:T119, A:E120, A:D121, A:M122, A:G124, A:I126, A:K127, A:N128, A:C129, A:S130, A:F131, A:N132, A:T133, A:T134, A:T135, A:E136, A:L137, A:D139, A:K140, A:R141, A:K142, A:K143, A:V144, A:Y145, A:S146, A:L147, A:F148, A:L151, A:D152, A:V153, A:V154, A:K155, A:I156, A:D157, A:D158, A:N159, A:N160, A:S161, A:N162, A:N163, A:S164, A:D165, A:R167, A:L168, A:I169, A:N170, A:C171, A:N172, A:T173, A:S174, A:Q178, A:C180, A:P181, A:K182, A:R273, A:P274, A:G275, A:N276, A:N277, A:T278, A:R279, A:R280, A:S281, A:I282, A:H283, A:I284, A:G285, A:P286, A:G287, A:Q288, A:A289, A:F290, A:Y291, A:A292, A:A293, A:T294, A:N295, A:I296, A:I297, A:G298, A:D299, A:I300, A:R301, A:Q302, A:P405, A:P406, A:I407, A:R408, A:G409, A:E410, A:I411 (4) A:S30, A:D31, A:A32, A:K33, A:A34, A:Y35, A:V49, A:P50, A:T51, A:D52, A:P53, A:N54, A:P55, A:Q56 (5) A:R511, A:Q512, A:L513, A:L514, A:S515, A:Q520, A:Q521, A:N522, A:N523, A:L524 (6) A:E57, A:I58, A:P59, A:L60, A:E61, A:N62, A:V63, A:T64, A:E65, A:E66, A:N204, A:E205, A:G212, A:P213, A:C214, A:K215, A:N216
Linear Epitope Prediction by Kolaskar and Tongaonkar Antigenicity		
IEDB	Gp120	4-19, 25-32, 41-52, 80-88, 90-105, 142-156, 166-174, 176-204, 214-239, 258-266, 268-274, 281-294, 302-308, 317-324, 353-360, 383-391, 400-406, 418-423, 452-468, 472-477, 481-495

^aEpitopes arranged based on their scores.

bodies to bind to the sites located in variable regions of this protein. Consequently, the rate of mutations in the HVRs of gp120 are high, to provide the virus the chance of escaping from antiviral antibodies (29). Increasing gp120 variability helps viral replication, therefore causes viral fitness in individuals infected by diverse HIV-1 variants (30).

Yao and colleagues analyzed five HVR regions from HIV-1 subtypes A, B, C, D, G, and H. The results revealed high gp120 variation, and the majority of HIV-1 gp120s had 496-515 amino acids (length polymorphism) and 21-30 PNGSs. Also, among five HVR regions, the V3 had the lowest rate of polymorphism and heterogeneity and more PNGSs, while V1 and V4 regions showed high rates of polymorphism, heterogeneity and more PNGSs (31). This assumption of structure is suggestive of the need of strategies for reducing the polymorphism, heterogeneity, and PNGSs in the four HVRs to be considered in HIV-1 vaccine development strategies

for effectively stimulating immune response (31).

Also, in evaluation of sequence diversity in HVRs of the external glycoprotein of HIV-1, extensive variation was found between sequences from different individuals, and various amino acid substitutions in the HVRs, which changed the number and positions of potential PNGSs (32).

Understanding adaptive evolution of HIV in populations is essential in monitoring the spread of infection and developing effective vaccines (33). Rapid mutations of HIV-1 allow the virus to evade the adaptive host defense responses. Also, each HIV-1 subtype has variability in its env and other proteins, which could have different immune responses not cross-reactive with other subtypes (34).

Another cause of increasing fitness of HIV to its individual (after mutation) is variability potential in N-linked glycosylation sites (PNGSs) (35). The total PNGSs in NX [ST] pattern was between 28 and 33, and for four isolates in com-

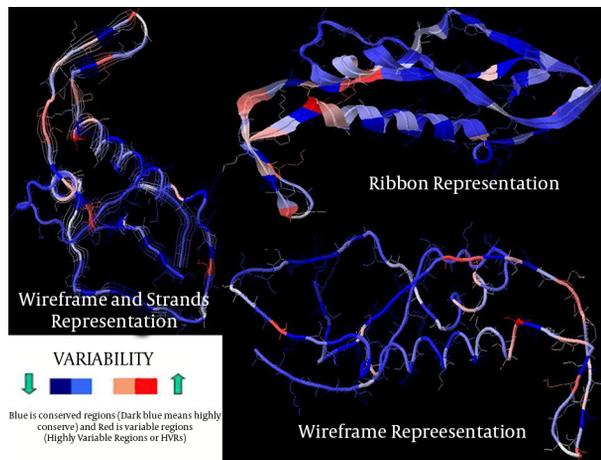


Figure 3. Hypervariable Regions (HVR) Represented by Different Representation Patterns Along gp120 Protein of Iranian Isolates of HIV-1

combination of NXS and NXT, total PNGSs was 7 to 13 and 15 to 26, respectively, which revealed high glycosylation, conservation and potency of virus in infecting the host.

The complete protein sequences in CRF35_AD of HIV-1 isolated from Iranian patients had 28 to 33 sites of glycosylation. The PNGSs prepares binding of long-chain carbohydrates to high variability regions of gp120, without blocking vital sites for binding to cellular receptors, thus it seems that the number of PNGSs in *env* might affect fitness of the virus by increasing or decreasing sensitivity to neutralizing antibodies via a protective coat of oligosaccharides that inhibits antibody binding to gp120 (28, 29). These large carbohydrate chains bound to gp120 might hide antibody-binding sites (29). If the PNGS number decreases significantly, the virus is in easy access of neutralizing antibodies. Also, higher glycan density promotes viral ability to evade antibodies, and thus promotes higher viral fitness (35, 36). Most studies have focused on the V3 loop (29, 35, 36), which is an important yet not exclusive determinant of viral tropism and cell entry.

By utilizing the hybrid approach, evaluation of trans-membrane topology and N-Linked glycosylation sites, the comparison of three consensus immunogenic regions for Iranian gp120 protein (CRF35_AD) was achieved. The regions included; AA₂₅₋₆₅ (NH₃-ICSAVENLWVTVYYGVPVWRDAETTLFCASDAKAYETEAH-COOH), AA₃₃₇₋₃₆₅ (NH₃-WNNTLGQVAAQLRKHWNKTIIFSNSGGD-COOH) and AA₄₄₃₋₅₀₅ (NH₃TGLLLRDGGGGNQS QNETFRPGGGDMRDNRSELYKYKVVKIEPLGV APTKAKRRVVQREKRAV-COOH).

Finding of HIV-1 immunogens, capable of inducing high-titer antibodies that neutralize a broad spectrum of

HIV-1, has a major priority for HIV-1 vaccine development (37, 38). As previous studies have clearly indicated for other subtypes, Gp120 has five conserved (C1 ± C5) and HVRs (V1 ± V5) (39). Evaluation of antigenic conservation of epitopes of isolates from clades A, B, D, F, G and H showed that only the V3 and C5 structures are shared and well exposed to the immune system (39). These regions are also reported in CRF35_AD and obtained consensus epitopes are suitable for using in chimeric and polytopic vaccines and additionally production of immunotherapeutic antibodies (elicit antibodies capable of neutralizing virus infectivity).

Clarification of the structure-function correlation in viral proteins is important for identification of potential vaccine targets to be exploited as preventive and therapeutic regimes (38). The predictions of features, such as trans-membrane domains, glycosylation sites, secondary and tertiary structure, are crucial for analyzing the structure-function relationship of proteins encoded in viral genomes (38).

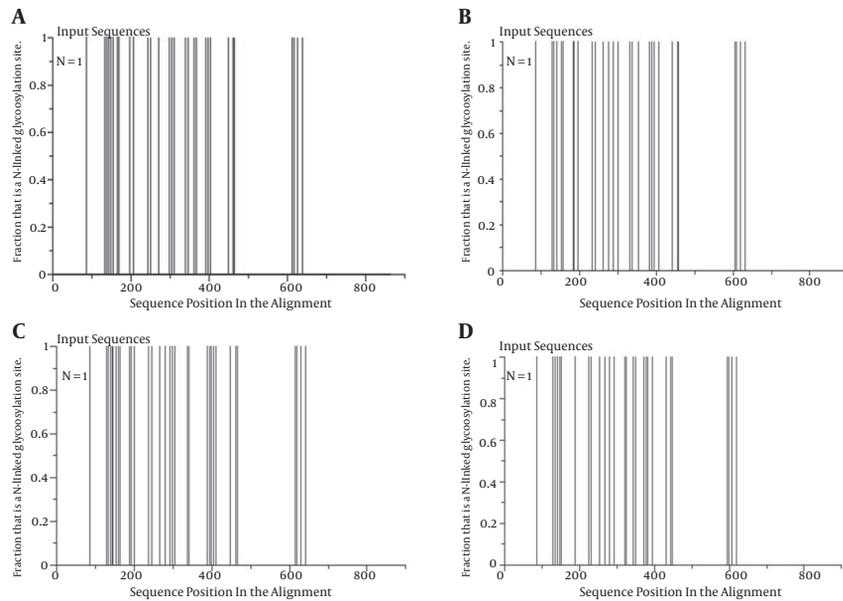
Ideally, results of the current study are useful for understanding CRF35_AD molecular characteristics (such as HVRs), levels and sites of glycosylation and its structure, as well as development of specific and effective envelope-based vaccines by identified consensus immunogenic candidate regions. The data of the present study can be analyzed along with data obtained from more isolates from Iran and also from other countries. This can be assumed as a disadvantage of the present study and should be followed as a broader study in the future.

Acknowledgments

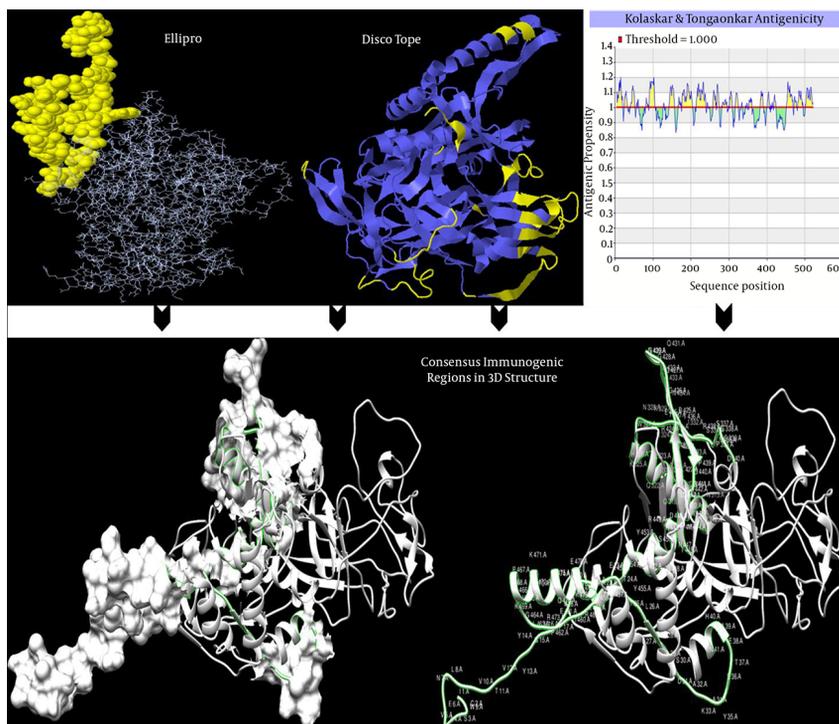
The authors thank all their colleagues, who cooperated in this investigation. Also, the authors specially thank Jahanbakhsh, since he, for first time, introduced the full genome sequences of Iranian HIV subtypes for future researches.

Footnote

Authors' Contribution: All authors had equal contributions in this study (regarding: research, analysis, bioinformatics, interpretation and writing of the article).

Figure 4. Locations of N-linked Glycosylation Sites in Four Envelope Proteins of Iranian Drug Resistant Isolates

X-axis, position of N- glycosylation in envelope protein; Y-axis, fraction that is an N-linked glycosylation site; A, BAM37429; B, BAM37384; C, BAM37456; D, BAM37411.

Figure 5. 3D Representation of Immunogenic Regions for Iranian HIV-1 CRF35_AD Isolate gp120 Protein That Was Predicted by Homology Modelling and Visualization Using the SwissPdb Viewer Software v4

Upper photos show predicted dominant epitopes by DiscoTope, ElliPro and Kolaskar and Tongaonkar antigenicity methods. Also, below the figure, consensus immunogenic region among servers are represented.

References

- Weiss RA. How does HIV cause AIDS?. *Science*. 1993;**260**(5112):1273–9. [PubMed: 8493571].
- Douek DC, Roederer M, Koup RA. Emerging concepts in the immunopathogenesis of AIDS. *Annu Rev Med*. 2009;**60**:471–84. doi: 10.1146/annurev.med.60.041807.123549. [PubMed: 18947296].
- Modjarrad K, Mohraz M, Madani N. AIDS epidemic: Iran needs global support to fight HIV. *Nature*. 2013;**494**(7437):314. doi: 10.1038/494314c. [PubMed: 23426315].
- Haghdoust AA, Mostafavi E, Mirzazadeh A, Navadeh S, Feizzadeh A, Fahimfar N, et al. Modelling of HIV/AIDS in Iran up to 2014. *J AIDS HIV Res*. 2011;**3**(12):231–9. doi: 10.5897/jahr11.030.
- Rahimi-Movaghar A, Amin-Esmaili M, Haghdoust AA, Sadeghirad B, Mohraz M. HIV prevalence amongst injecting drug users in Iran: a systematic review of studies conducted during the decade 1998–2007. *Int J Drug Policy*. 2012;**23**(4):271–8. doi: 10.1016/j.drugpo.2011.09.002. [PubMed: 22000694].
- Jahanbakhsh F, Ibe S, Hattori J, Monavari SH, Matsuda M, Maejima M, et al. Molecular epidemiology of HIV type 1 infection in Iran: genomic evidence of CRF35_AD predominance and CRF01_AE infection among individuals associated with injection drug use. *AIDS Res Hum Retroviruses*. 2013;**29**(1):198–203. doi: 10.1089/AID.2012.0186. [PubMed: 22916738].
- Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature*. 1992;**360**(6402):358–61. doi: 10.1038/360358a0. [PubMed: 1360148].
- de Witte L, Bobardt M, Chatterji U, Degeest G, David G, Geijtenbeek TB, et al. Syndecan-3 is a dendritic cell-specific attachment receptor for HIV-1. *Proc Natl Acad Sci U S A*. 2007;**104**(49):19464–9. doi: 10.1073/pnas.0703747104. [PubMed: 18040049].
- Doms RW, Moore JP. HIV-1 membrane fusion: targets of opportunity. *J Cell Biol*. 2000;**151**(2):9–14. [PubMed: 11038194].
- Madsen J, Kliem A, Tornøe I, Skjoldt K, Koch C, Holmskov U. Localization of lung surfactant protein D on mucosal surfaces in human tissues. *J Immunol*. 2000;**164**(11):5866–70. [PubMed: 10820266].
- Chirmule N, Pahwa S. Envelope glycoproteins of human immunodeficiency virus type 1: profound influences on immune functions. *Microbiol Rev*. 1996;**60**(2):386–406. [PubMed: 8801439].
- Nara PL, Garrity RR, Goudsmit J. Neutralization of HIV-1: a paradox of humoral proportions. *FASEB J*. 1991;**5**(10):2437–55. [PubMed: 1712328].
- Stamatatos L, Cheng-Mayer C. Evidence that the structural conformation of envelope gp120 affects human immunodeficiency virus type 1 infectivity, host range, and syncytium-forming ability. *J Virol*. 1993;**67**(9):5635–9. [PubMed: 8350416].
- Levy JA. Pathogenesis of human immunodeficiency virus infection. *Microbiol Rev*. 1993;**57**(1):183–289. [PubMed: 8464405].
- Groenink M, Fouchier RA, Broersen S, Baker CH, Koot M, van't Wout AB, et al. Relation of phenotype evolution of HIV-1 to envelope V2 configuration. *Science*. 1993;**260**(5113):1513–6. [PubMed: 8502996].
- Paital B, Kumar S, Farmer R, Tripathy NK, Chainy GB. In silico prediction and characterization of 3D structure and binding properties of catalase from the commercially important crab, *Scylla serrata*. *Interdiscip Sci*. 2011;**3**(2):110–20. doi: 10.1007/s12539-011-0071-z. [PubMed: 21541840].
- Tramontano A, Leplae R, Morea V. Analysis and assessment of comparative modeling predictions in CASP4. *Proteins*. 2001;**Suppl 5**:22–38. [PubMed: 11835479].
- Jeffs SA, Shotton C, Balfe P, McKeating JA. Truncated gp120 envelope glycoprotein of human immunodeficiency virus 1 elicits a broadly reactive neutralizing immune response. *J Gen Virol*. 2002;**83**(Pt 11):2723–32. doi: 10.1099/0022-1317-83-11-2723. [PubMed: 12388808].
- Thompson JD, Higgins DG, Gilson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;**22**:4673–80.
- Shannon CE. A mathematical theory of communication. *Bell Sys Tech J*. 1948;**27**:623–56.
- Simpson EH. Measurement of Diversity. *Nature*. 1949;**163**:688.
- Kabat EA, Wu TT, Bilofsky H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J Biol Chem*. 1977;**252**(19):6609–16. [PubMed: 408353].
- Garcia Boronat M, Diez Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res*. 2008;**36**(2):35–41.
- Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, Kuiken C, et al. Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology*. 2004;**14**(12):1229–46.
- Lovell SC, Davis IW, Arendall W3, de Bakker PI, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins*. 2003;**50**(3):437–50. doi: 10.1002/prot.10286. [PubMed: 12557186].
- Saha S, Raghava GPS, editors. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. International Conference on Artificial Immune Systems. 2004; Italy. Springer; pp. 197–204.
- Center for Disease Control. Ministry of Health and Medical Education of the I.R. Iran. HIV/AIDS in Iran (Cumulative Statistics). [In persian]. Tehran: Office of the Deputy for Public Health; 2003.
- American-Iranian.Org. . American Iranian Council: Drug Use and HIV/AIDS in Iran Available from: <http://www.american-iranian.org/beta/publications.php?PressRelease=1&PressReleaseID=56>.
- Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, et al. The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature*. 1998;**393**(6686):705–11. doi: 10.1038/31514. [PubMed: 9641684].
- Novitsky V, Lagakos S, Herzig M, Bonney C, Kebaabetswe L, Rossenkhan R, et al. Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. *Virology*. 2009;**383**(1):47–59. doi: 10.1016/j.virol.2008.09.017. [PubMed: 18973914].
- Yao N, Zhang C, Liu Q, Liu J, Zhang C. Polymorphism characteristics of HIV-1 gp120 and 5 hypervariable regions. *Turk J Med Sci*. 2015;**45**(1):47–54. [PubMed: 25790529].
- Simmonds P, Balfe P, Ludlam CA, Bishop JO, Brown AJ. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. *J Virol*. 1990;**64**(12):5840–50. [PubMed: 2243378].
- Leitner T. In: Genetic subtypes of HIV-1. Myers G, Korber B, Foley B, editors. New Mexico: Los Alamos National Laboratory; 1997. Human retroviruses and AIDS 1996: A compilation and analysis of nucleic acid and amino acid sequences.
- Leth-Larsen R, Floridon C, Nielsen O, Holmskov U. Surfactant protein D in the female genital tract. *Mol Hum Reprod*. 2004;**10**(3):149–54. doi: 10.1093/molehr/gah022. [PubMed: 14981140].
- Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, Zhu H, et al. Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses. *Virology*. 2008;**374**(2):229–33. doi: 10.1016/j.virol.2008.01.029. [PubMed: 18314154].
- Frost SD, Wrin T, Smith DM, Kosakovsky Pond SL, Liu Y, Paxinos E, et al. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci U S A*. 2005;**102**(51):18514–9. doi: 10.1073/pnas.0504658102. [PubMed: 16399909].
- Zhou T, Xu L, Dey B, Hessel AJ, Van Ryk D, Xiang SH, et al. Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature*. 2007;**445**(7129):732–7. doi: 10.1038/nature05580. [PubMed: 17301785].

38. Yan Q. Bioinformatics databases and tools in virology research: an overview. *In Silico Biol.* 2008;**8**(2):71-85. [PubMed: [18928197](#)].
39. Azizi A, Anderson DE, Torres JV, Ogrel A, Ghorbani M, Soare C, et al. Induction of broad cross-subtype-specific HIV-1 immune responses by a novel multivalent HIV-1 peptide vaccine in cynomolgus macaques. *J Immunol.* 2008;**180**(4):2174-86. [PubMed: [18250424](#)].