

Research Article

Evolutionary Pattern of Asian HIV-1 Subtype B from 1990 to 2007: *In Silico* Analysis Based on Envelop Protein

Sobia Kanwal and Tariq Mahmood

Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan

Correspondence should be addressed to Tariq Mahmood, tmahmood.qau@gmail.com

Received 31 October 2011; Accepted 7 December 2011

Academic Editor: Esteban Martinez

Copyright © 2012 S. Kanwal and T. Mahmood. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

HIV-1 envelop gene is a major target for vaccine development. Envelop protein and its V3 loop is shown to be important determinant of HIV-1 pathogenicity. Herein, the evolutionary pattern of most prevalent HIV-1 subtype B in Asia is determined by analyzing envelop protein and V3 domain based on the 40 randomly selected sequences of HIV-1 from database (Los Alamos), divided into four groups since 1990–2007. Construction of envelop protein phylogeny by using MEGA 5 exhibit the active mutation pattern, increase in potential N-glycosylation sites which were predicted by using online software SignalP-NN. An online available tool Drawgram was used for multiple sequence alignment (MSA) of HIV-1 subtype B envelop region and V3 loop while the alignment was rechecked by using CLUSTAL W and further was analyzed for GPGX motif and conserved region in V3 loop. Variation at fourth position of the GPGX motif and 60% conservation was found in V3 loop. Hence, this diversifying pattern of envelop protein in the Asia formulates the HIV-1 strains more pathogenic during the period of 17 years. These findings might help in understanding significant structural and functional constrains of the mutant viral strains and ultimately in vaccine development.

1. Introduction

High genetic variability of HIV-1 (Human immunodeficiency virus I) is characterized and classified into three distinctive groups; group M (main), group O (outliner), and group N (non-M/non-O) [1–3]. Majority of the global epidemic of the infection is caused by group M which may be further subdivided into nine nonrecombinant subtypes (A-D, F-H, J, and K) and at least 34 intersubtypes recombinant forms, termed as circulating recombinant forms (CRF) [4] distributed on the basis of geographical locations. Subtype B is circulating in United States and Europe, in addition, two genetic forms of subtype B and CRF01-AE are dominant in South East Asia while subtype A is most prevalent in European countries [5].

Envelop gene plays an important role in entry and fusion of HIV-1 and is a major target for vaccine development studies. Envelop protein gp160 is encoded by envelop gene involved in host cell-HIV interaction. It is composed of two noncovalently bound subunits, the external glycoprotein gp120 and the transmembrane gp41. External glycoprotein gp120 plays an important role in entry of the virus into the

cell by binding to the specific receptors on the surface of the target cell. These receptors are CD4 and chemokine receptors known as HIV-1 coreceptors [6]. The chemokine receptors CXCR4 and CCR5 are the two major coreceptors used by HIV-1 isolates [7]. The binding of the chemokine receptor requires the presence of selected amino acids in gp120 (specifically within the V3 loop, and also in other regions), providing greater affinity to CCR5 or CXCR4, and therefore, the viral tropism [8–11]. The gp120 portion of envelop has been generally classified into five hypervariable regions (V1 to V5) with conserved regions interspersed [12]. Among these variable regions, the V3 is one of the most important determinants of viral tropism and coreceptor usage [13], as it contains major antigenic and neutralizing epitopes in HIV-1 which are well exposed upon CD4-binding [14]. This third variable region (V3 loop) of the HIV-1 gp120 also plays an important role in the mechanism controlling the fusion between envelop of HIV-1 and the cell membrane of the target cells [15]. Hence, it is considered as the fusion domain whose sole function would be to initiate the fusion process after an initial interaction between gp120 and CD4 on the surface of the CD4+ cells [7].

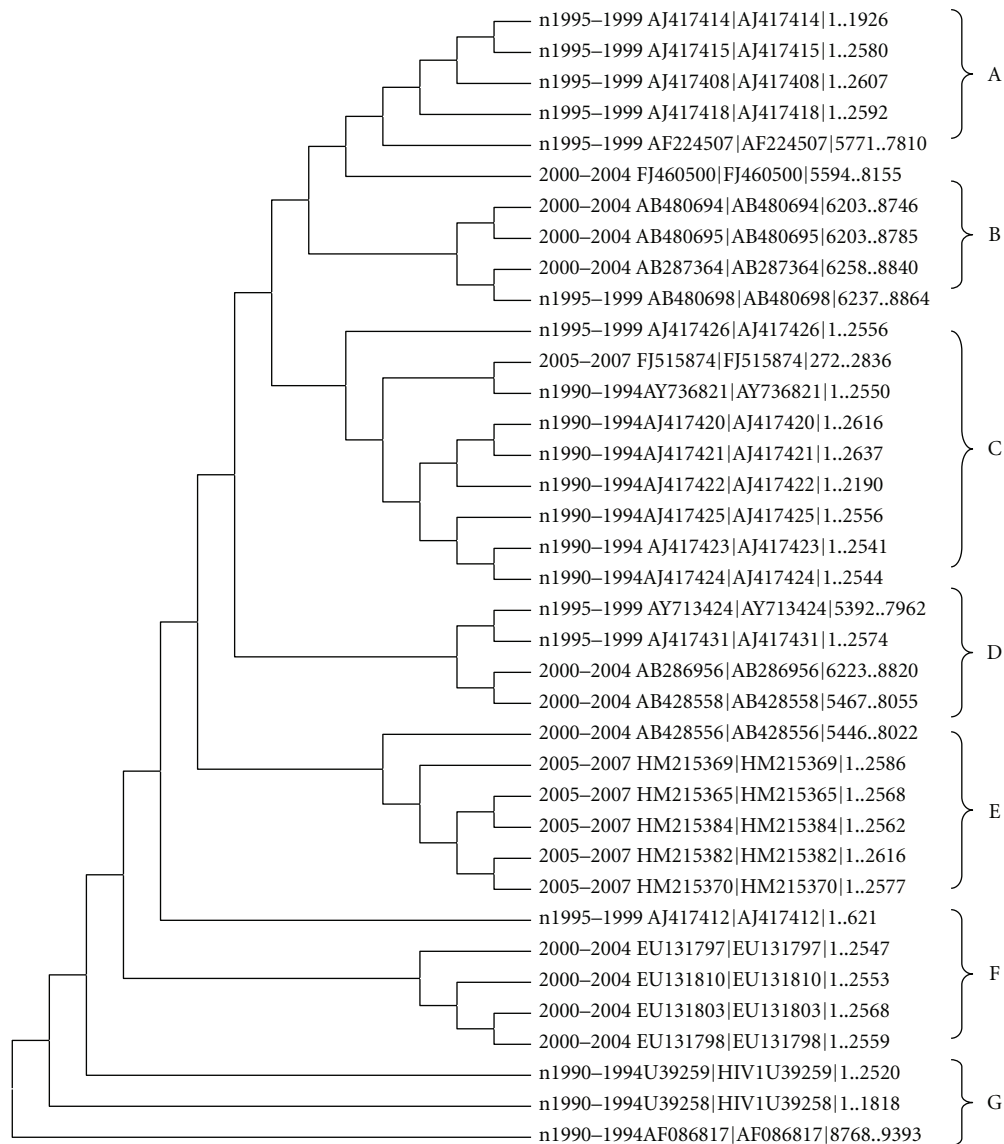


FIGURE 1: Neighbour-joining phylogenetic tree (constructed using MEGA 5) based on amino acid sequences of envelop protein from HIV-1 subtype B in Asia.

Further, gp120 is also extensively glycosylated protein, containing 24–26 potential glycosylation sites (24 for HIV-INL4-3 strain, 26 for HIV-ISF2) [16]. Glycosylation of envelop gene is crucial in all phases of HIV biology that track from proper envelop folding, processing of virus transmission, and the immune evasive mechanisms [17–19]. Therefore, the systematic analysis of the glycosylation profiles of envelop could provide precious insights in designing well envelop (env) immunogens. A decreased number of N-linked glycosylation sites in gp120, especially within and around the V3 region, have been demonstrated during evolution from R5 to the X4 phenotype [20, 21].

Several amino acids residues are extremely conserved in V3 loop due to their functional importance among HIV-1

variants. However, the vast majority of the HIV-1 sequences, and their putative common SIVcpz ancestors, have the GPGX signature pattern (motif) at the central portion of V3 loop of the envelop gene [22]. This conserved motif in a variable region exhibits purifying selection pressure. Some lineages, however, have been found worldwide with alternate signature patterns, such as the GWGR subtype B variant in which the proline residue is substituted by a tryptophan [23, 24]. This variant was first recorded in Japan in early 1990s [25] but a serologic study suggests its presence in Brazil since 1983 [17].

The present study is the first attempt to track the evolutionary pattern of HIV-1 in Asia over the period of 17 years. The biological activity and pathogenicity of the most

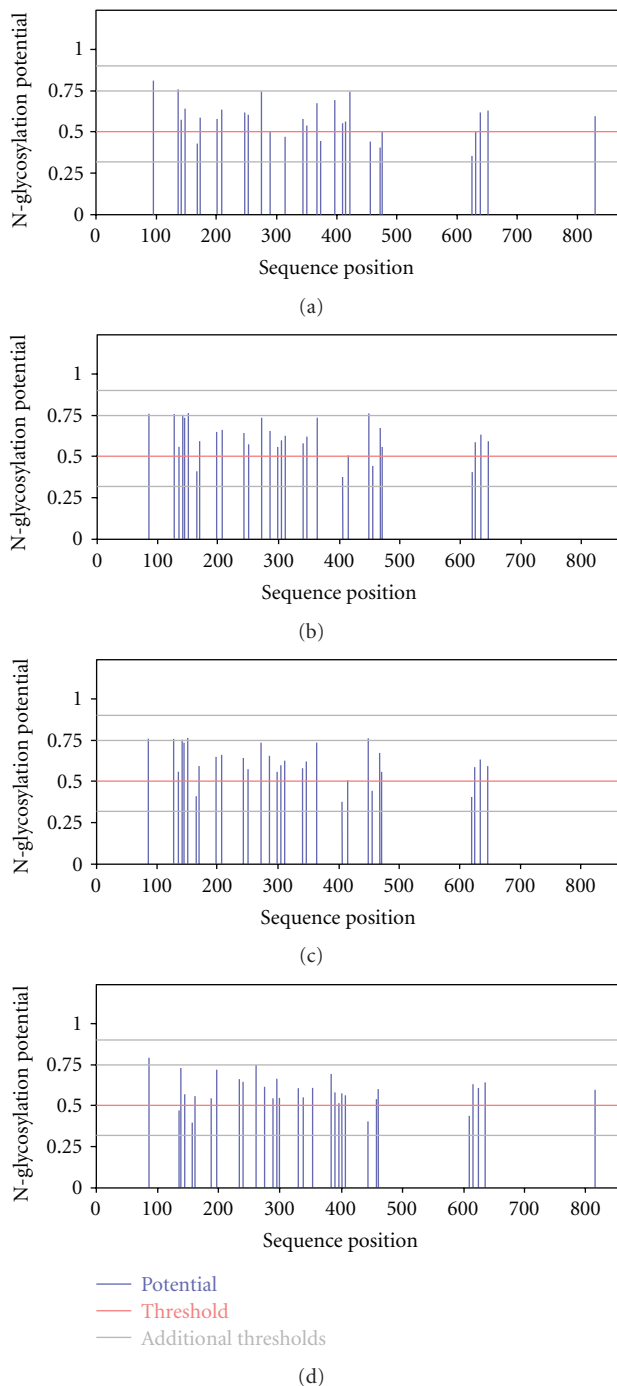


FIGURE 2: Potential N-glycosylation sites predicted from SignalP-NN. (a) Group I (1990–1994). (b) Group II (1995–1999). (c) Group III (2000–2004). (d) Group IV (2005–2007).

prevalent HIV-1 subtype B in Asia was determined based on the available sequences of HIV-1 in database (Los Alamos) during the time frame of 1990–2007.

2. Material and Methods

Evolutionary pattern of HIV-1 subtype B was tracked by analyzing the envelop protein and its V3 domain.

2.1. Dataset. The sequences used in the study were divided into four groups, that is, each group constitutes the sequences from four years:

- Group I: 1990–1994,
- Group II: 1995–1999,
- Group III: 2000–2004,
- Group IV: 2005–2007.

In total, 10 sequences of HIV-1 subtype B from each group were obtained from Los Alamos HIV database (<http://hiv-web.lanl.gov/>). Two type of analysis were performed based on envelop protein and V3 domain of the downloaded sequences.

2.2. Phylogenetic Analysis Based on Envelop Protein. Forty sequences of envelop protein were randomly selected and aligned using CLUSTALW (parameters reading DNA Pairwise Parameters; Gap opening penalty: 10, Gap extension penalty: 0.1, Multiple Parameters; Gap opening penalty: 10, Gap extension penalty: 0.2, protein weight matrix: BLOSUM, Residue specific penalties: ON, Hydrophilic penalties: ON, Gap separation distance: 4, End gap separation: OFF, negative matrix OFF, Delay divergent cutoff: 30%). To obtain phylogenetic tree alignment was analyzed in MEGA 4 tool (parameters reading; Tree inference Method: neighbor joining, Include sites: Gaps/missing data: complete deletion, substitution model: Amino p-distance, Substitutions to include: All).

2.3. Prediction of N-Glycosylation Sites in Envelop Protein. N-linked glycosylation sites of HIV-1 envelop protein were predicted by using online software SignalP-NN.

2.4. Alignment of V3 Domain. ClustalW was used for the alignment of V3 motifs to check the level of conservation of this region. The conserved region was calculated manually.

2.5. Analysis of GPGX Motif. HIV-1 envelop protein sequences were aligned by using ClustalW of online software Drawgram to find out conserved GPGX motif.

2.6. Determination of Conserved Region. Used Sequences were aligned through multiple sequence alignment (MSA) of ClustalW from online software Drawgram.

3. Results and Discussion

3.1. Phylogenetic Profile of Envelop Protein. Phylogenetic analysis of the envelop protein sequences in Asian strains of subtype B was performed (Figure 1). The phylogenetic tree representing the sequence relationships by amino acid substitutions identified distinct genetic subgroups exhibiting that viruses within these lineages were genetically closely related.

Isolates from 1990s make the root of the tree and also exhibits the early evolution of the isolates. Strains from groups year 2000 to 2004 and from 2005 to 2007 clustered together and strains from 1990s act as their root. However,

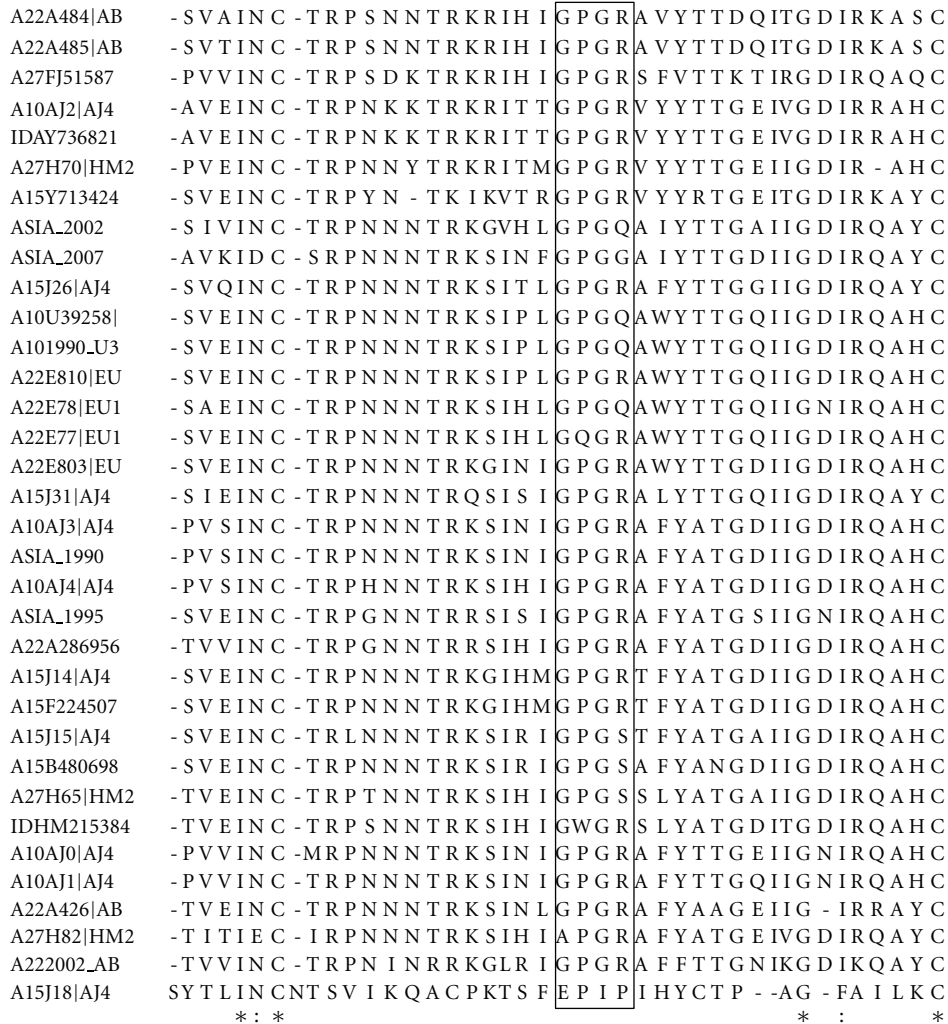


FIGURE 3: Multiple sequence alignment showing the variation a GPGX motif.

strains from 1990 to 1994 make separate cluster. The clustering did not correlate closely with chronological origin of the sequences. It was observed from the clustering pattern that strains from 1990–1995 remained resistant to selection pressure and act as the root of the tree, while strains from other groups exhibit the high mutation rate demonstrated by their branching pattern. The differences in the evolutionary characteristics found could be attributed to a different selection pattern of the strains by the nature.

Consequently, high mutation rate and divergence have been shown by almost all groups while some viral lineages have shown extensive branching pattern which may be due to negative selection of the strains by the nature (Figure 1, Clade B, C, D, and E). An increase in envelope sequence heterogeneity has also been observed in an infected population over the period of time [8, 9]. This increase in sequence heterogeneity may help in viral escape from host immune response. Other clades of the tree exhibit the strong selection and remain less divergent over the period of time (Figure 1, Clade A and F). The viral robustness is a key factor affecting the viral virulence which measures the capacity of a virus to

cause disease [22]. Genetic analysis of envelope gene, particularly the third variable domain (V3), showed that the inter-host (intrasubtype) variation increased over the course of the AIDS epidemic [26].

3.2. Number of Predicted N-linked Glycosylation Sites. Increase in N-linked glycosylation sites was observed through the period of 17 years from 1990 to 2007. This might serve to protect viruses from neutralizing antibodies. Glycosylation by the host cell can vigorously maneuver the folding, stability, and biological function of virus-encoded proteins [27, 28]. The predicted glycosylation site calculated by using the software SignalP-NN in the period of 1990–1994 is 20, from 1995–1999 is 25, from 2000–2004 is 26, and from 2005–2007 is 30 (Figure 2).

The N-linked glycosylation (NLG) of viral envelope proteins, through the formation of a "glycan shield," is one of the major mechanisms for inhibiting or minimizing virus neutralizing antibody response [29] which promotes viral persistence and immune evasion. Further, the N-linked glycosylation pattern has also been shown to have a varied

TABLE 1: Amino acids frequency at the fourth position of the GPGX tetramer of V3 loop of HIV-1 subtype B.

Amino acid	Percentage of subtype B isolates in Asia
Arginine (R)	73%
Glutamin (Q)	11%
Serine (S)	8%
Others	8%

effect on the stimulation of neutralizing antibodies in HIV-1 [30, 31]. Consequently, alterations in glycosylation events influencing co-receptor utilization will also allow the virus to escape the controlling responses of the CC chemokines *in vivo*, which are believed to play an effective role in inhibiting viral replication and slowing disease progression [32].

3.3. *Analysis of V3 Domain.* Multiple sequence alignment (MSA) describes the variation at fourth position of the tetrapeptide GPGX (Figure 3). Arginine remains predominant at the fourth position while presence of serine, glutamine, and other amino acids at this position is highly nonsignificant (Table 1).

In HIV-1, V3 is involved in coreceptor usage [13]. This region also contains major antigenic and neutralizing epitopes in HIV-1 which are well exposed upon CD4-binding [14–17]. Residues in the V3 crown, including the GPG motif, are important for soluble gp120-CD4 complex binding to cell surface receptor CCR5 [33]. The tip motif represents a target for neutralizing antibodies [34]. Therefore, sequence variation in this motif may have an impact on virus infectivity and disease progression.

3.4. *Consensus of V3 Loop.* The conserved amino acids were calculated after aligning the V3 loop sequences. Out of 35 amino acids of the V3 loop, 21 remain the same during the studied period. Hence, the 60% of amino acids of the V3 loop remain conserved (Figure 4). The conserved residues is marked (*) Figure 4.

According to previous research, some of the conserved or relatively conserved residues are normally essential for the biological functions of the HIV-1 [34]. The deletion or substitution of some residues in this loop was found to affect gp120 coreceptor interactions and may further influence the HIV-1 target entry into the host.

4. Conclusion

The present study is the first report that demonstrated the evolutionary pattern of HIV-1 during the period of 17 years from 1990 to 2007. The pathogenicity and infectivity of the HIV-1 subtype B is increased that showed the rapid evolutionary and modulatory mechanism. The analyzed data also provides useful information on the surveillance of HIV-1 infection in Asia and highlights the evolutionary pattern of virus populations which may be under varied selection pressures during different periods of time. However, further

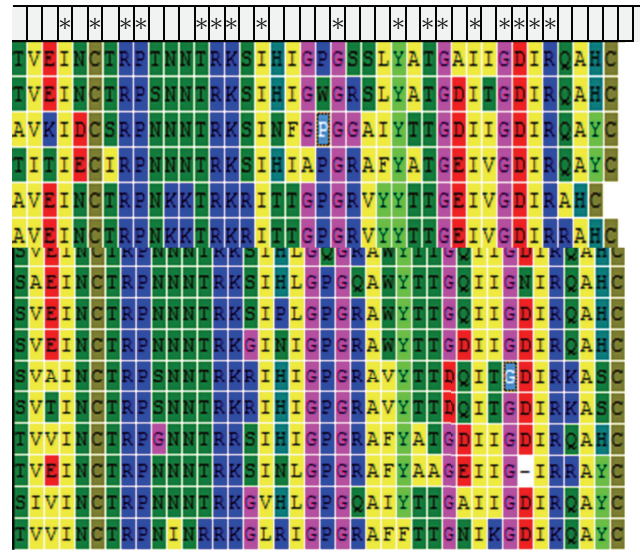


FIGURE 4: Aligned sequences exhibiting the consensus region of V3 loop (indicated by *).

studies are required to get more insights of all these mechanisms and for developing effective therapies and vaccines against HIV-1 infection.

References

- [1] F. Simon, P. Mauciere, I. Loussert-Ajaka et al., “Identification of a new human immunodeficiency virus type 1 distinct from group M and group O,” *Nature Medicine*, vol. 4, no. 9, pp. 1032–1037, 1998.
- [2] A. Ayoub, S. Souquière, B. Njinku et al., “HIV-1 group N among HIV-1-seropositive individuals in Cameroon,” *AIDS*, vol. 14, no. 16, pp. 2623–2625, 2000.
- [3] Y. Takebe, R. Uenishi, and X. Li, “Global molecular epidemiology of HIV: understanding the genesis of AIDS pandemic,” *Advances in Pharmacology*, vol. 56, pp. 1–25, 2008.
- [4] M. M. Thomson and R. Najera, “Molecular epidemiology of HIV-1 variants in the global aids pandemic: an update,” *AIDS Reviews*, vol. 7, no. 4, pp. 210–224, 2005.
- [5] A. M. Geretti, “HIV-1 subtypes: epidemiology and significance for HIV management,” *Current Opinion in Infectious Diseases*, vol. 19, no. 1, pp. 1–7, 2006.
- [6] H. Deng, R. Liu, W. Ellmeier et al., “Identification of a major co-receptor for primary isolates of HIV-1,” *Nature*, vol. 381, no. 6584, pp. 661–666, 1996.
- [7] T. Dragic, V. Litwin, G. P. Allaway et al., “HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5,” *Nature*, vol. 381, no. 6584, pp. 667–673, 1996.
- [8] S. S. Hwang, T. J. Boyle, H. K. Lyerly, and B. R. Cullen, “Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1,” *Science*, vol. 253, no. 5015, pp. 71–74, 1991.
- [9] T. L. Hoffman and R. W. Doms, “HIV-1 envelope determinants for cell tropism and chemokine receptor use,” *Molecular Membrane Biology*, vol. 16, no. 1, pp. 57–65, 1999.
- [10] C. C. Huang, S. N. Lam, P. Acharya et al., “Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4,” *Science*, vol. 317, no. 5846, pp. 1930–1934, 2007.

- [11] A. A. Nabatov, G. Pollakis, T. Linnemann, A. Kliphuis, M. I. Chalaby, and W. A. Paxton, "Inpatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies," *Journal of Virology*, vol. 78, no. 1, pp. 524–530, 2004.
- [12] A. A. Nabatov, P. Georgios, L. Thomas, K. Aletta, M. I. M. Chalaby, and A. P. William, "Inpatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 Regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal," *Journal of Virology*, vol. 78, no. 1, pp. 524–530, 2004.
- [13] R. F. Speck, K. Wehrly, E. J. Platt et al., "Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop," *Journal of Virology*, vol. 71, no. 9, pp. 7136–7139, 1997.
- [14] C. D. Rizzuto, R. Wyatt, R. N. Hernández-Ramos et al., "A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding," *Science*, vol. 280, no. 5371, pp. 1949–1953, 1998.
- [15] L. M. Walker, S. K. Phogat, P. Y. Chan-Hui et al., "Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target," *Science*, vol. 326, no. 5950, pp. 285–289, 2009.
- [16] E. O. Freed, D. J. Myers, and R. Risser, "Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain," *Journal of Virology*, vol. 65, no. 1, pp. 190–194, 1991.
- [17] G. Pollakis, S. Kang, A. Kliphuis, M. I. M. Chalaby, J. Goudsmit, and W. A. Paxton, "N-Linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization," *Journal of Biological Chemistry*, vol. 276, no. 16, pp. 13433–13441, 2001.
- [18] X. Zhu, C. Borchers, R. J. Bienstock, and K. B. Tomer, "Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells," *Biochemistry*, vol. 39, no. 37, pp. 11194–11204, 2000.
- [19] A. Land and I. Braakman, "Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum," *Biochimie*, vol. 83, no. 8, pp. 783–790, 2001.
- [20] C. A. Derdeyn, J. M. Decker, F. Bibollet-Ruche et al., "Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission," *Science*, vol. 303, no. 5666, pp. 2019–2022, 2004.
- [21] T. B. Geijtenbeek and Y. van Kooyk, "DC-SIGN: a novel HIV receptor on DCs that mediates HIV-1 transmission," *Current Topics in Microbiology and Immunology*, vol. 276, pp. 31–54, 2003.
- [22] B. Chohan, D. Lang, M. Sagar et al., "Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels," *Journal of Virology*, vol. 79, no. 10, pp. 6528–6531, 2005.
- [23] Y. Liu, M. E. Curlin, K. Diem et al., "Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses," *Virology*, vol. 374, no. 2, pp. 229–233, 2008.
- [24] P. D. Kwong, M. L. Doyle, D. J. Casper et al., "HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites," *Nature*, vol. 420, no. 6916, pp. 678–682, 2002.
- [25] X. Wei, J. M. Decker, S. Wang et al., "Antibody neutralization and escape by HIV-1," *Nature*, vol. 422, no. 6929, pp. 307–312, 2003.
- [26] C. L. Kuiken, G. Zwart, E. Baan, R. A. Coutinho, J. A. R. van den Hoek, and J. Goudsmit, "Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 19, pp. 9061–9065, 1993.
- [27] E. Johnson, J. M. Sauvron, and R. C. Desrosiers, "Conserved, N-linked carbohydrates of human immunodeficiency virus type 1 gp41 are largely dispensable for viral replication," *Journal of Virology*, vol. 75, no. 23, pp. 11426–11436, 2001.
- [28] P. Kolchinsky, E. Kiprilov, P. Bartley, R. Rubinstein, and J. Sodroski, "Loss of a single N-linked glycan allows CD4-independent human immunodeficiency virus type 1 infection by altering the position of the gp120 V1/V2 variable loops," *Journal of Virology*, vol. 75, no. 7, pp. 3435–3443, 2001.
- [29] K. Schonning, B. Jansson, S. Olofsson, J. O. Nielsen, and J. E. Hansen, "Resistance to V3-directed neutralization caused by an N-linked oligosaccharide depends on the quaternary structure of the HIV-1 envelope oligomer," *Virology*, vol. 218, no. 1, pp. 134–140, 1996.
- [30] B. Dash, A. McIntosh, W. Barrett, and R. Daniels, "Deletion of a single N-linked glycosylation site from the transmembrane envelope protein of human immunodeficiency virus type 1 stops cleavage and transport of gp160 preventing env-mediated fusion," *Journal of General Virology*, vol. 75, no. 6, pp. 1389–1397, 1994.
- [31] E. Fenouillet, J. C. Gluckman, and I. M. Jones, "Functions of HIV envelope glycans," *Trends in Biochemical Sciences*, vol. 19, no. 2, pp. 65–70, 1994.
- [32] A. Goffard, N. Callens, B. Bartosch et al., "Role of N-linked glycans in the functions of hepatitis C virus envelope glycoproteins," *Journal of Virology*, vol. 79, no. 13, pp. 8400–8409, 2005.
- [33] Q. Hu, J. O. Trent, G. D. Tomaras et al., "Identification of ENV determinants in V3 that influence the molecular anatomy of CCR5 utilization," *Journal of Molecular Biology*, vol. 302, no. 2, pp. 359–375, 2000.
- [34] S. S. Hwang, T. J. Boyle, H. K. Lyerly, and B. R. Cullen, "Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1," *Science*, vol. 253, no. 5015, pp. 71–74, 1991.