Indian Journal of Biochemistry & Biophysics
Vol. 62, November 2025, pp. 1198-1210
DOI: 10.56042/ijbb.v62i11.17129

Predictive analysis of structural interaction among HIV-1 proteins and Class I MHC molecule: A computational approach for therapeutic target

Ankita Bhaduri[1], Balaka Biswas[1], Soumyadeep Dey[1], Mousumi Saha[2] & Agniswar Sarkar[1]*

[1]Department of Biosciences, JIS University, Agarpara, Kolkata-700 109, West Bengal, India

[2]Department of Microbiology, Ballygunge Science College, University of Calcutta, Kolkata-700 019, West Bengal, India

Received 13 February 2025; *revised 03 July 2025*

Human immunodeficiency virus type 1 (HIV-1) remains a significant global health challenge because it can impair the host's immune system and establish long-lasting infections. HIV-1 evades immune detection by modulating major histocompatibility complex class I (MHC-I) molecules, which are essential for presenting viral antigens to cytotoxic T lymphocytes (CTLs). The specific molecular interactions involved in this process remain unclear, which poses a challenge for the advancement of targeted therapies. In this study, we employed a systematic computational approach to explore the structural relationships between selected HIV-1 accessory proteins (*nef, tat, rev,vpu*) and human MHC-I molecules. Analyses included protein sequences evaluation to identify conserved domains and structural motifs, as well as prediction of secondary structures, transmembrane topology, and 3D-modelling for prediction of potential interaction sites. Structural alignments and molecular docking simulations demonstrated substantial conformational compatibility between HIV-1 proteins and MHC-I molecules, particularly in regions critical for immune modulation. The outcomes of this research provide novel insights into the structural mechanisms that underpin HIV-1-mediated immune evasion. This study identifies protein interfaces and conserved motifs related to MHC-I interference, providing insight for designing antiviral strategies to improve immune recognition and control viral persistence.

Keywords: Bioinformatics, Immune response, Molecular modelling, Protein interaction, Viral evasion

Human immunodeficiency virus (HIV) belongs to the genus Lentivirus, within the family *Retroviridae* and the subfamily *Orthoretrovirinae*. Based on viral antigenic characteristics, HIV is classified into two types: HIV-1 and HIV-2. Phylogenetic analyses suggest that HIV was introduced into the human population between 1920 and 1940[1-3]. Various viral and host factors influence the outcome of HIV-1 infection and its progression in affected individuals. Key determinants of HIV-1 pathogenesis include cellular tropism and receptor-coreceptor interactions. Despite extensive research over the past three decades, the precise mechanisms underlying CD4+ T cell depletion while R5 and X4 viral strains persist in AIDS remain incompletely understood[4,5]. HIV infection typically remains asymptomatic in the early stages, with only minor alterations in the immune system. This asymptomatic phase lasts for approximately three months before seroconversion, during which HIV-specific antibodies become detectable in affected individuals. The progression and clinical manifestations of the disease vary significantly among individuals[6]. HIV infection consists of two main stages, with the initial primary or acute phase typically being asymptomatic, causing individuals to seem healthy. However, during this phase, the virus actively replicates within the lymphoid tissues and disseminates through the bloodstream to various organs. The advanced phase is characterized by increasing immunosuppression, making individuals vulnerable to opportunistic infections such as *Mycobacterium tuberculosis, Mycobacterium avium, Cytomegalovirus* (CMV), and *Candida* species. Persistent immune deterioration ultimately leads to the development of AIDS, which is defined by severe depletion of CD4+ T cells and a high HIV plasma viral load[7,8]. In the 1980s, patients with primary HIV infection exhibited early involvement of the central nervous system (CNS), and HIV was successfully isolated from cerebrospinal fluid (CSF)[9,10]. The primary target cells of HIV infection in the brain include perivascular macrophages, microglia, and astrocytes. Many antiretroviral drugs have limited penetration through the blood-brain barrier (BBB) during CNS infection, reducing their efficacy in eradicating HIV from this compartment. In the early

*Correspondence:
E-mail: ognish@gmail.com; agniswar.sarkar@jisuniversity.ac.in

stages of HIV infection, cytopathic effects (CPE) are not observed in infected astrocytes, allowing the virus to persist within the CNS and establish a viral reservoir. Consequently, the CNS serves as a significant obstacle to HIV eradication. HIV primarily targets CD4+ cells, particularly CD4+ T helper lymphocytes[11-14]. The virus utilizes the CD4 receptor along with co-receptors such as C-C chemokine receptor type 5 (CCR5) or C-X-C chemokine receptor type 4 (CXCR4) to infect host cells. This infection leads to host cell death, resulting in a progressive depletion of CD4+ T lymphocytes[15,16]. Since CD4+ T lymphocytes are critical regulators of the adaptive immune system, their depletion weakens immune function, ultimately leading to acquired immunodeficiency syndrome (AIDS). HIV is primarily transmitted through unprotected sexual intercourse with an infected individual[17]. Other transmission routes include mother-to-child transmission, the use of contaminated needles, and exposure to infected blood[18,19]. Populations at higher risk of HIV infection include men who have sex with men (MSM) and intravenous drug users. Both free-floating viral particles and HIV-infected CD4+ cells are present in body fluids such as semen, vaginal secretions, and blood, facilitating the transmission of infection to new host cells[20,21]. Approximately 38.4 million individuals worldwide are currently living with HIV infection. Since the onset of the HIV epidemic, an estimated 84.2 million people have been infected, and 40.1 million have succumbed to the disease [31-33]. Sub-Saharan Africa bears the highest burden of HIV, accounting for approximately 70% of the global cases, with numbers continuing to rise. HIV-1 is a spherical virus characterized by glycoprotein spikes (*gp120* and *gp41*) that protrude from the surface of its viral envelope[22,23]. The virus contains group-specific antigens (*gag*), which are essential for the formation of virions from infected cells, and pol proteins, which encode key enzymes required for viral replication, including reverse transcriptase, protease, and integrase[24,25]. Additionally, HIV-1 expresses regulatory proteins such as viral protein R (*vpR*), viral protein U (*vpU*), and virion infectivity factor (*vif*), which play crucial roles in nuclear import, viral replication, and the degradation of CD4 molecules, thereby enhancing viral pathogenesis[15,26]. HIV evades the host immune system by downregulating the expression of major histocompatibility complex (MHC) class I and II molecules, which are essential for antigen presentation and immune cell recognition. At the molecular level, HIV interferes with the transcription and translation of MHC class I and II genes, leading to a reduction in their surface expression on infected cells[27]. This process is primarily mediated by the HIV-1 accessory protein *"nef"*, which is critical for viral pathogenesis and represents a potential target for antiretroviral drug development. The protein *"nef"* interacts with the cytoplasmic tails of MHC class I and II molecules, directing them toward the endocytic degradation pathway. The mechanism by which *"nef"* down-regulates MHC involves its interaction with phosphofurin acidic cluster sorting protein 1 (PACS-1), which hijacks the ADP-ribosylation factor 6 (ARF6) endocytic pathway via phosphatidylinositol-3 kinase (PI3K), ultimately leading to MHC-I trafficking to the trans-Golgi network for degradation. Additionally, the HIV-1-*vpu* protein facilitates MHC-I degradation by interacting with the host protein β-TrCP, which recruits the E3 ubiquitin ligase complex to ubiquitinate and target MHC-I for proteasomal degradation. HIV-1-*tat* protein also contributes to immune evasion by repressing MHC-I promoter activity. As an extracellular protein, *"tat"* indirectly affects MHC-I by inhibiting dendritic cell-mediated phagocytosis of apoptotic cells. Furthermore, the HIV-1-*rev* protein modulates MHC-I expression through indirect mechanisms. The protein *"rev"* plays a crucial role in transporting specific HIV mRNAs necessary for viral protein synthesis. In asymptomatic HIV-1 infection, low *"rev"* activity leads to reduced expression of gag proteins, resulting in ineffective cytotoxic T lymphocyte (CTL)-mediated killing of infected cells that present gag epitopes[28,29].

In this study, we aimed to clarify the interactions between HIV-1 accessory proteins and MHC-I molecules to establish a mechanistic relationship. Understanding these interactions may offer valuable insights for developing novel therapeutic strategies against HIV-1 infection.

Materials and Methods

Sequence retrieval and analysis of biological roles

HIV-encoded accessory proteins (*nef, tat, rev*, and *vpu*) and sequences from human class I MHC molecules were retrieved from the National Center for Biotechnology Information (NCBI) protein database. These sequences were obtained in FASTA format. To identify regions that are conserved both structurally and functionally, the retrieved sequences were

subjected to analysis via the NCBI Conserved Domains Database (NCBI-CDD). This comparative analysis provided insights into the biological functions and stability of the proteins involved.

Protein secondary structure prediction and analysis

The protein sequences were analyzed using the Phyre2 Protein Fold Recognition Server, a computational tool for predicting and evaluating protein structures[30]. Phyre2 provided insights into the predicted 3D structure, domain architecture, and functional annotations of the proteins. Additionally, it identified regions of disorder and related superfamilies. To determine evolutionary relationships and conserved regions, Multiple Sequence Alignment (MSA) was performed using Clustal Omega[31]. Position-specific iterated BLAST (PSI-BLAST) was employed to identify homologous sequences for comparison, facilitating the detection of structural similarities and differences. The closest homologous sequence identified by PSI-BLAST was used as a control to enhance structural comparisons. The physicochemical properties of the proteins, including sequence length, molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, and grand average of hydropathicity (GRAVY), were analyzed using ExPASy (Expert Protein Analysis System)[32] and represented in Table 1. These parameters provided insights into the stability, solubility, and overall biochemical characteristics of the proteins. Additionally, a phylogenetic tree was constructed using Clustal Omega to further elucidate evolutionary relationships among the analyzed proteins. To predict transmembrane helices, TMHMM 2.0, which utilizes a hidden Markov model, was applied[33]. Identifying these integral membrane segments is crucial, as they are often less accessible to antibodies, thereby playing a key role in epitope mapping. TM analysis of HIV and MHC are elaborately explained in (Table 2). FASTA sequences of MHC class I molecules were retrieved from the NCBI database and analyzed using Gene Threader to determine pH values and the amino acid composition of these proteins. The highest and lowest abundance of specific amino acids was assessed, providing valuable insights into the structural stability and functional properties of MHC class I proteins.

Structural domains prediction and protein properties analysis

The protein sequences were submitted to the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB) for an in-depth structural analysis. The RCSB-PDB provides detailed three-dimensional (3D) structural information, including the presence of water molecules, ligands, and bonds, along with experimental data such as X-ray diffraction and crystallizationscores[34]. The PDB files were analyzed using PDBsum, which generates structural summaries with functional annotations. PDBsum offers additional

Table 1 — Physicochemical analysis of HIV-1 accessory proteins, including amino acid count, molecular weight, theoretical isoelectric point (pI), and other key properties. The *"nef"* protein is stable and slightly hydrophobic, whereas *"tat"* and *"rev"* are unstable with hydrophilic characteristics. The *"vpu"* protein is highly stable and exhibits strong hydrophobicity

Protein Name	Amino Acids	Molecular Weight (KDa)	Theoretical pI	Instability Index	Status	Aliphatic Index	GRAVY	Ext. Coefficient ($M^{-1}cm^{-1}$)
nef	216	24.57	5.72	39.34	Stable	69.07	-0.598	50670 (Cystine); 50420 (Reduced)
tat	101	11.53	9.05	71.60	Unstable	34.85	-1.318	15845 (Cystine); 15470 (Reduced)
rev	116	12.76	9.41	86.42	Unstable	74.91	-0.766	8605 (Cystine); 8480 (Reduced)
vpu	82	92.42	4.69	28.56	Stable	143.78	0.454	12490

Table 2 — Template modelling (TM) analysis of HIV-1 proteins with homologous sequences. The *"rev"* protein has the highest RMSD (Root Mean Square Deviation) value, indicating poor geometric alignment, whereas *"vpu"* has the lowest RMSD value, suggesting the best geometric alignment among the analyzed proteins

Protein Name	Homology PDB Code	Chain 1 Name	Chain 2 Name	Chain 1 Length (residues)	Chain 2 Length (residues)	Aligned Length	RMSD	TM-score (Chain 1)	TM-score (Chain 2)
nef	7syd	A759095	B759095	105	614	67	4.93	0.32953	0.08674
tat	7syd	A663313	B663313	86	614	57	4.31	0.33162	0.07660
rev	7syd	A759018	B759018	439	614	219	7.70	0.27945	0.22022
vpu	7syd	A973033	B973033	81	614	52	3.78	0.38097	0.07350

contextual information, such as the number of water molecules, the presence of pores, tunnels, and ligands, which are essential for identifying potential epitope regions[35]. The Ramachandran plots derived from PDBsum were meticulously analysed to assess the distributions of phi (ϕ) and psi (ψ) dihedral angles within the protein structure. This analysis is essential for evaluating protein stability and identifying any disallowed conformations that could result in unfavourable steric interactions. Furthermore, it offers valuable insights into bond lengths and angles, thereby enhancing the structural validation of the protein[36]. The PDB files were further analyzed using TM-align for comparative studies of homologous protein structures identified through PSI-BLAST. Identical structures extracted from PSI-BLAST were utilized in the comparative analysis. Discovery Studio 2024 Client was employed for a more detailed examination of the protein structures obtained from TM-align. TM-align was also used to predict the binding affinity between the HIV-1 proteins and human MHC class I molecules[37]. TM-score and RMSD (Root Mean Square Deviation) values were assessed to evaluate structural similarity and predict the potential effectiveness of epitopes. To confirm the findings and ensure the accuracy of the structural and epitope predictions, PyMOL was used for re-analysis of the structural results. Finally, Chimera was employed to visualize the interactions between HIV-1 proteins and MHC class I molecules.

Results and Discussion

Analysis of physicochemical properties

The physicochemical properties of MHC-I were investigated by retrieving sequences in FASTA format from the NCBI database and analyzing them using Gene Threader. The analysis indicated that the pH values ranged from highly acidic for KAI2541543.1 to slightly basic for QOY57179.1, as summarized in Table 3. Methionine content varied between 2.08 and 2.59, predominantly stabilizing at 2.11, which suggests its conserved role in protein structural integrity. In contrast, leucine levels ranged from 9.09 to 12.93, with a predominant value of 9.94, highlighting its hydrophobic properties that are crucial for maintaining protein stability. These amino acid compositions significantly influence protein stability, binding interactions, and enzymatic activity[38]. The NCBI-CDD was employed for domain structure analysis, revealing that the HIV-1-*nef* protein (Table 4) interacts with host cellular proteins involved in signal transduction and cell activation. This interaction occurs through binding to a subset of

Table 3 — The physicochemical properties of MHC-I molecules reveal significant details about their stability and structural characteristics. The pH values of these molecules indicate a highly acidic to slightly basic environment. This variance in pH could suggest different functional roles or stability under various conditions. Additionally, the amino acid composition shows that methionine and leucine maintain relatively constant levels across most sequences, suggesting structural stability. However, accession number >QOY57179.1 presents notable variation, indicating potential deviations in structure or function that merit further investigation

Sl. No.	Accession no. (MHC class 1)	pH	Methionine	Leucine
1.	>WES13122.1	6.71	2.11	9.94
2.	>KAI2541543.1	7.00	2.11	9.94
3.	>QOY57257.1	6.20	2.08	9.09
4.	>QOY57179.1	0.61	2.59	12.93
5.	>QOY57178.1	6.00	2.11	9.94
6.	>QOY57177.1	6.00	2.11	9.94
7.	>QOY57176.1	6.00	2.11	9.94
8.	>QOY57175.1	6.00	2.11	9.94
9.	>QOY57174.1	6.00	2.11	9.94
10.	>QOY57173.1	6.00	2.11	9.94

Table 4 — Domain structure analysis and Open Reading Frame (ORF) predictions for HIV-1 accessory proteins (*nef, tat, rev*, and *vpu*). The study, conducted using the NCBI Conserved Domain Database (NCBI-CDD) and ORF Finder, includes evaluations of E-values, domain intervals, and the largest and smallest ORFs identified

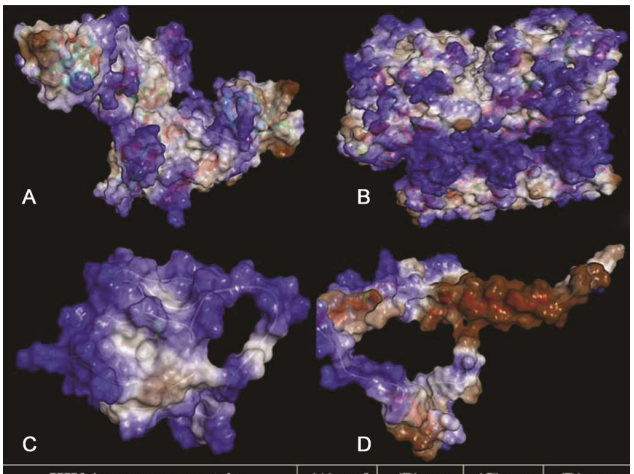
Protein Name	Super family	Domain	Accession number	E-value	Interval	Largest ORF length (aa)	Smallest ORF length (aa)
nef	F-protein	+F-protein	Pfam00469	1.12e-137	2-215	ORF1-215	ORF3-34
tat	*tat*	+*tat*	Pfam00539	8.87e-20	2-63	ORF1-100	ORF3-49
rev	*rev*	+*rev*	Pfam00424	3.05e-22	1-90	ORF1-115	ORF4-35
vpu	*vpu*	+*vpu*	Pfam00558	3.61e-17	23-79	ORF1-81	ORF2-73

the SRC kinase family, which in turn enhances viral pathogenicity. Graphical analysis indicates that the *"nef"* aligns with the F protein superfamily, exhibiting a conserved region essential for its function. Specifically, the Pfam0469 domain spans residues 2–215 of the *"nef"* sequence, with a highly significant E-value of 1.12e-127, highlighting its functional importance[39].

The HIV-1-*tat* protein interacts with both tat-RNA and host cellular proteins. Graphical analysis indicates that *"tat"* is a member of the tat superfamily, featuring a conserved region within the Pfam domain that spans residues 2 to 63 of the *tat*-sequence, with an E-value of 8.87e-20. Similarly, the HIV-1 rev protein plays a crucial role in facilitating the nuclear export of unspliced or partially spliced viral pre-mRNAs. Recent studies suggest that the *"rev"* protein also interacts with host cellular proteins. Graphical analysis confirms that *"rev"* is classified within the rev superfamily and contains a conserved region within the Pfam domain, covering residues 1 to 90 of the *rev* sequence, with an E-value of 3.05e-22[40,41].

The HIV-1-*vpu* protein features an N-terminal transmembrane domain and engages with host cellular proteins. Graphical analysis categorizes *"vpu"* within the *vpu*-superfamily and reveals a conserved region in the Pfam domain, encompassing residues 23-79 of the *vpu* sequence, with an E-value of 3.16e-17. The NCBI-ORF was employed to identify potential open reading frames within selectable DNA, protein, or RNA sequences. The analysis indicated that the HIV-1-*nef* protein ORF1 encodes the longest protein, comprising 215 amino acids, while the other ORFs yield proteins ranging from 39 to 208 amino acids. In a similar vein, the HIV-1-*tat* protein ORF1 encodes a 100-amino-acid protein, with additional ORFs encoding proteins between 49 and 64 amino acids[42,43]. For the HIV-1 *"rev"* protein, ORF1 encodes a sequence of 115 amino acids, with other ORFs varying from 35 to 81 amino acids. Furthermore, the HIV-1-*vpu* protein ORF1 translates into an 81 amino acid protein, with an alternative ORF encoding a 73 amino acid protein. Comprehensive data regarding domain structure analysis and ORF identification for these four proteins are compiled in (Table 4). The physicochemical properties of HIV-1 proteins were evaluated using the Expasy-Swiss Bioinformatics Resource Portal, with various characteristics summarized in Table 1, and their GRAVY (Grand Average of Hydropathy) values illustrated in (Fig. 1).

The *"nef"* protein consists of 216 amino acids, with a molecular weight of 24,566.72 Da and a theoretical isoelectric point (pI) of 5.72. The instability index of 39.34 indicates its stability. The aliphatic index of *"nef"* is 69.07, and its GRAVY value (Fig. 1) is -0.598, suggesting slight hydrophobicity. The extinction coefficients are 50,670 $M^{-1}cm^{-1}$ for cystine-containing forms and 50,420 $M^{-1}cm^{-1}$ in the reduced state. In comparison, *"tat"* is a smaller protein, comprising 101 amino acids with a molecular weight of 11,538.02 Da. It has a basic pI of 9.05 and a relatively high instability index of 71.60, classifying it as unstable. The low aliphatic index (34.85) and a GRAVY value of -1.318 (Fig. 1) indicate significant hydrophilicity. The extinction coefficients for Tat are 15,845 $M^{-1}cm^{-1}$ (cystine-containing) and 15,470 $M^{-1}cm^{-1}$ (reduced). The *"rev"* protein consists of 116 amino acids, with an instability index of 86.42 and a high aliphatic index of 74.91, making it highly unstable. It has a molecular weight of 12,759.34 Da and a basic pI of 9.41. The GRAVY value of -0.766 (Fig. 1) suggests slight hydrophilicity. Its extinction coefficients are 8,605 $M^{-1}cm^{-1}$ (cystine-containing) and 8,480 $M^{-1}cm^{-1}$ (reduced). Finally, *"vpu"* is the smallest out of the



HIV-1 accessory proteins	(A) *nef*	(B) *tat*	(C) *rev*	(D) *vpu*
GRAVY Value	-0.598	-1.318	-0.766	0.454

Fig. 1 — Presents an analysis of the Grand Average of Hydropathy (GRAVY) values for HIV-1 accessory proteins. Hydrophobicity measurements were conducted using Discovery Studio for the proteins *"nef"*, *"tat'*, *"rev"*, and *"vpu"*. (A) The GRAVY value for *"nef"* is -0.598, indicating slight hydrophobicity and a moderate affinity for lipid bilayers; (B) *"tat"*; (C) *"rev"* display GRAVY values of -1.318 and -0.766, respectively, consistent with hydrophilic profiles and lower likelihood of membrane association. In contrast; and (D) *"vpu"* has a GRAVY value of 0.454, signifying a higher degree of hydrophobicity and a stronger propensity for interaction with lipid membranes

four proteins, comprising 82 amino acids with a molecular weight of 9,241.87 Da, and an instability index of 28.56, classifying it as stable [63,64]. With a GRAVY value of 0.454 (Fig. 1),*"vpu"* is highly hydrophobic. Notably, it has the highest aliphatic index (143.78) among the four proteins and the extinction coefficient for *"vpu"* is 12,490 $M^{-1}cm^{-1}$.

Computer-assisted structural prediction and analysis

Structural modelling and transmembrane topology prediction of HIV-1 proteins

The structural modelling of HIV-1 proteins was conducted using template-based approaches in Phyre2, revealing distinct fold types with varying degrees of confidence and coverage, as illustrated in (Fig. 2A). The HIV-1-*nef* protein was modelled using the template c4en2B, which represents the HIV-1-*nef* in complex with the MHC-I cytoplasmic domain and the µ1-adaptin subunit of the AP-1 adaptor. The model has 100% confidence for 69% of the sequence, covering 149 residues with high precision. Furthermore, 95% of the sequence (205 residues) can be modelled with greater than 90% confidence using multiple templates. Similar findings were reported by Singh *et al*. (2009)[44]. The HIV-1-tat protein was modelled using the template cltivA, specifically within the A chain, with 100% confidence over 85% of the sequence (86 residues). However, 59% of the sequence is predicted to be disordered. These results are consistent with findings by Das *et al*. (2011)[40]. The HIV-1-*rev* protein was modelled using the template cbsyA, covering residues 1–69 with 100% confidence, though only 49% of the sequence (57 residues) could be modelled. Additionally, 60% of the sequence is predicted to be disordered, indicating that a substantial portion of the protein lacks a defined structure. Similar observations were reported by Heaphy *et al.* (1990)[45]. The HIV-1-*vpu* protein was modelled using the template c2n28A, based on solid-state NMR analysis. This model achieved 100% confidence over 99% of the sequence (81 residues), suggesting that the structure is suitable for in-depth analysis. A study by Sharpe *et al*. (2006) also highlighted the structural dynamics of the HIV-1-*vpu*transmembrane domain using solid-state NMR with magic-angle spinning[46].

Structural Analysis Using RCSB

The 3D structures and detailed properties of the HIV-1 proteins were obtained from the RCSB Protein

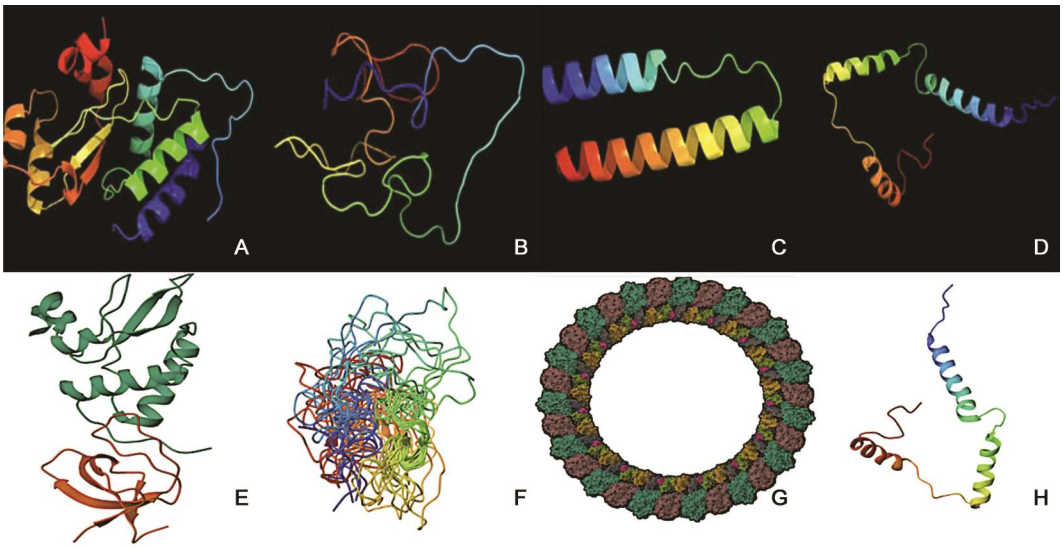


Fig. 2 — Structural modelling of HIV-1 proteins was performed using template-based methods in Phyre2. Each protein displayed distinct fold types, with varying levels of confidence and sequence coverage. (A) The HIV-1 *"nef"* protein was modelled using the template c4en2B, with 100% confidence over 69% of its sequence, including residue 205; (B) The *"tat"* protein used the template cltivA (A chain), achieving 100% confidence for 85% of the sequence, though 59% of its 86 residues were predicted to be disordered; (C) The *"rev"* protein was modelled with the template cbsyA (residues 1–69), with approximately 60% expected to be disordered; (D) The *"vpu"* protein structure was generated using the template c2n28A, based on solid-state NMR analysis. (E-H) Experimental structural data were obtained from the RCSB database, where E. the HIV-1 SF2 *"nef"* in complex with the Fyn SH3 R96L variant was determined by X-ray diffraction; (F) The *"tat"* protein structure highlights its role in transcription activation; (G) The *"rev"* protein, in complex with tubulin, was visualised by single-particle electron microscopy at a resolution of 3.53 Å; and H. The *"vpu"* protein structure, resolved via Solid-state NMR provides further insights into its structural and functional properties

Data Bank (PDB), as depicted in (Fig. 2B). The analysis of the HIV-1-SF2-*nef* protein, in complex with the Fyn SH3 R96L domain, was determined using X-ray diffraction, yielding a resolution of 3.32 Å. The structure exhibited R-values of 0.300 (R-Free), 0.286 (R-Work), and 0.288 (R-Observed), suggesting a moderate-resolution structure. The HIV-1-*tat* protein, essential for transcription activation, was constructed using NMR spectroscopy and expressed in *Escherichia coli*. The HIV-1 Rev protein was analyzed in complex with tubulin using single-particle electron microscopy at a resolution of 3.53 Å. The HIV-1-*vpu* protein was characterized using solid-state NMR, providing insights into its structural and functional properties.

Transmembrane topology prediction

The topology of transmembrane helices in the *nef, tat, rev,* and *vpu* proteins was predicted using the TMHMM2.0 server. The WEBSEQUENCE analysis of *"nef"* (216 amino acids) revealed no predicted transmembrane helices, with a transmembrane helix probability of 0.00082. The N-terminus inside the membrane probability was 0.09818, confirming that *"nef"* is likely a cytosolic or extracellular protein. The analysis of *"tat"* (101 amino acids) also showed no predicted transmembrane helices, with a very low transmembrane probability of 0.00017. The N-terminus inside the membrane probability was 0.08880, indicating that *"tat"* is an extracellular protein. The *"rev"* protein (116 amino acids) showed no transmembrane helices. However, the probability of the N-terminus being membrane-associated was 0.21021, suggesting a potential but minimal interaction with membranes. The *"vpu"* protein (82 amino acids) exhibited transmembrane helices, with 23.05494 amino acids located within these helices, indicating a significant membrane-associated region. The N-terminus inside the membrane probability was 0.03057, suggesting that the N-terminus is likely extracellular. All WEBSEQUENCE analyses and their detailed results are presented in (Fig. 3A).

Structural validation using Ramachandran plots

Figure 3B presents Ramachandran plots, which graphically depict φ-ψ torsional angle distributions, highlighting allowable and disallowed conformations in protein residues. In these plots,red regions represent the most favoured conformations, indicating structural stability. Brown and yellow regions correspond to less-favoured but still allowed conformations. White regions denote disallowed conformations due to steric hindrance. Proteins with PDB codes 7d7s and 7u0f exhibited dense clusters within favoured regions, signifying well-refined structures. However, 1tiv displayed several outliers, such as HIS13 and VAL26, suggesting potential structural anomalies. The 2n28 model had minimal outliers, indicating a high-quality structure.

Validation of protein structure

Multiple sequence alignment (MSA) and phylogenetic analysis

The multiple sequence alignment (MSA) of the four HIV-1 proteins, shown in (Fig. 4A), highlights both conserved and variable regions. The sequence NP_057855.1 exhibits a unique structural role, as it aligns less extensively in certain regions due to the presence of a transmembrane domain. This sequence is also hydrophobic, containing characteristic stretches such as "LIVVWVSIVI" and "EYRKIL". Residues marked with an asterisk (*) indicate fully conserved positions, which are crucial for maintaining structural and functional integrity. The sequence AAB03750.1 exhibits extensive alignment in highly conserved regions but shows less overlap with functional motifs compared to other sequences. The phylogenetic tree in (Fig. 4B) reveals evolutionary relationships among the sequences. AAF80535.1 and AAF80536.1 share a common node with the shortest branch distance, indicating a close evolutionary relationship. In contrast, AAB03750.1 and NP_057855.1 diverge earlier, forming distinct clusters with longer branch lengths.

Structural comparison using TM-align

The TM-align analysis assessed structural similarity among HIV-1 proteins relative to control proteins, as summarized in Table 2. The results indicate varying degrees of alignment: HIV-1-*rev* has the highest number of aligned residues (219) but also the highest RMSD value (7.70 Å), indicating poor geometric alignment. However, it has the highest sequence identity (0.073) among the analyzed proteins. HIV-1-*vpu* demonstrates the best geometric fit, with the lowest RMSD value (3.78 Å) and the highest TM-score for Chain 1 (0.38097). However, it has the shortest aligned length (52 residues) and the lowest TM-score for Chain 2 (0.07350). HIV-1-*tat* exhibits a moderately aligned length (57 residues) and an RMSD value of 4.31 Å. HIV-1 *nef* shows weak alignment, with the lowest sequence identity (0.030) and low TM-scores. These findings suggest that

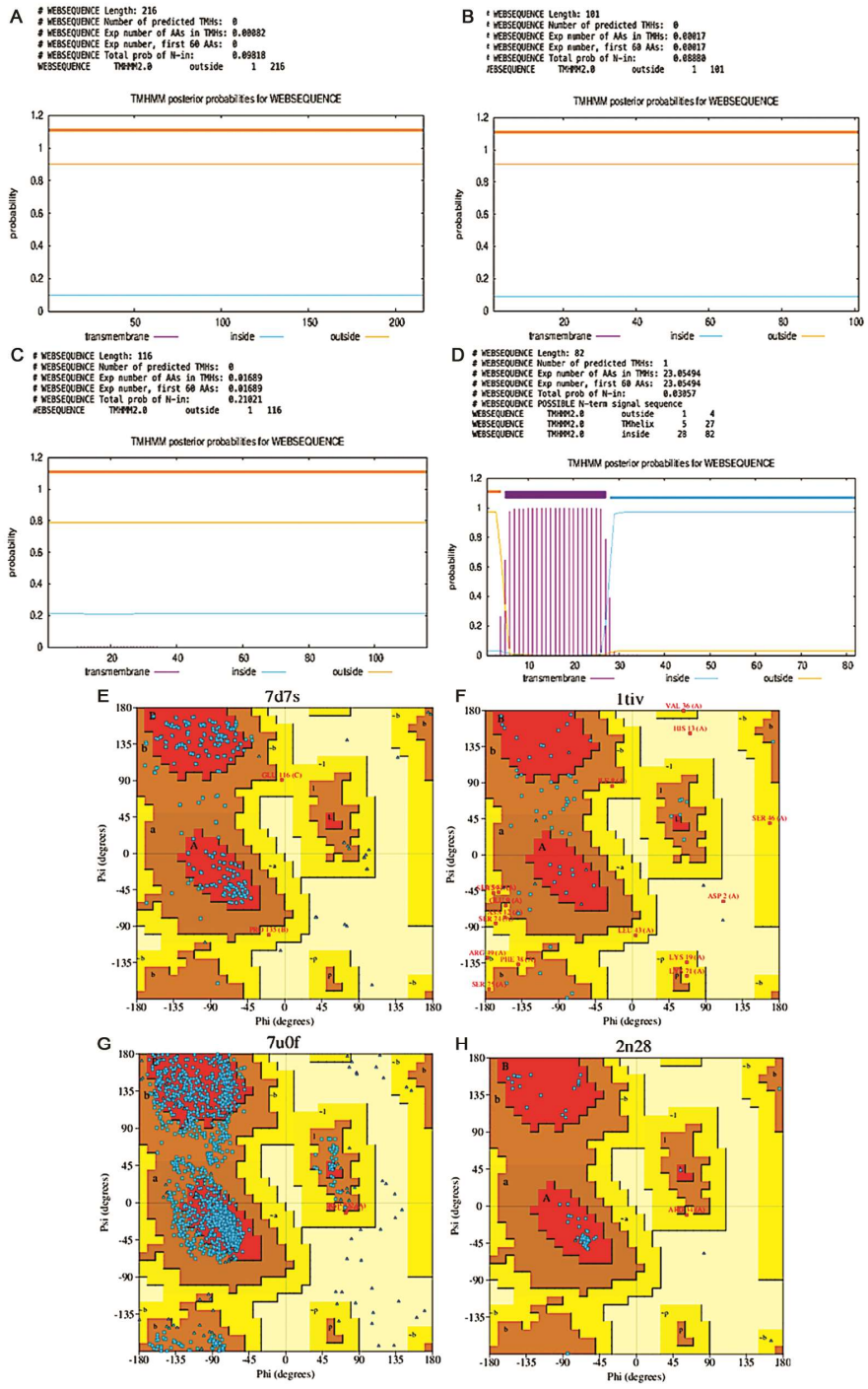


Fig. 3 — Topology Prediction and Structural Validation of HIV-1 Accessory Proteins. (A-D) TMHMM2.0 topology predictions show that*"nef"* (216 aa), *"tat"* (101 aa), and *"rev"* (116 aa) lack predicted transmembrane helices, while *"vpu"* (82 aa) contains a single transmembrane helix; and (E-H) Ramachandran plot analysis *via* the PDB SUM database demonstrates that the structures with PDB codes 7d7s and 7u0f have a high proportion of residues in favourable conformational regions, indicating well-refined models. In contrast, the model for PDB code 1tiv exhibits several outliers (*e.g*., HIS13, Val26), suggesting reduced stereochemical quality, whereas the 2n28 structure displays minimal outliers, consistent with a higher-quality fold. TMHMM2.0 identifies transmembrane regions based on hydrophobicity and statistical likelihood, supporting the classification of *"nef"*, *"tat"*, and *'rev*" as non-membranous and *"vpu"* as membrane-associated, in line with their biological roles. Ramachandran plot clustering in favourable regions further validates the structural quality and reliability of the protein models for subsequent functional analyses

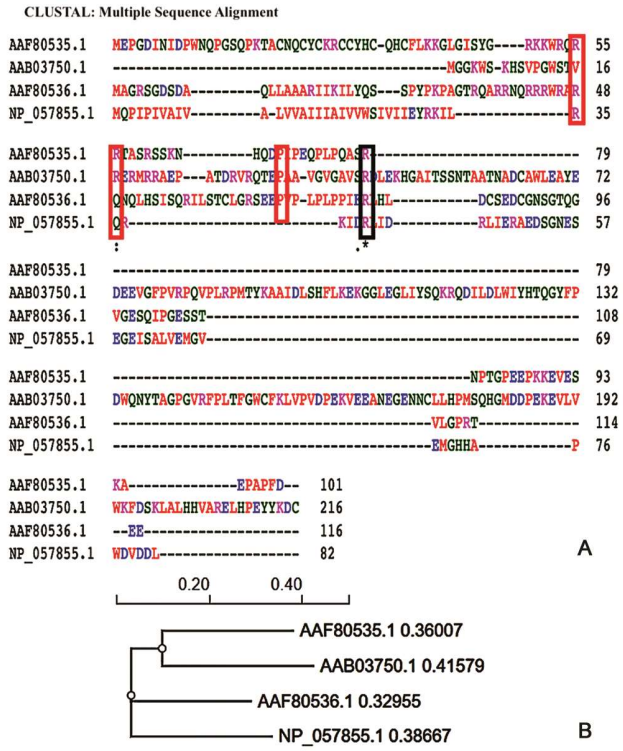


Fig. 4 — Multiple Sequence Alignment (MSA) and Phylogenetic Analysis of HIV-1 Proteins. (A) MSA of HIV-1-*nef*proteins (AAF80535.1), *"tat"* (AAB03750.1), *"rev"* (AAF80536.1), and *"vpu"* (NP_057855.1) was performed using CLUSTAL ω (1.2.4). The alignment highlights conserved regions (*) and semi-conserved residues (:) with gaps (-) between sequences. Acidic residues (Asp, Glu) are marked in red, while basic residues (Lys, Arg) are marked in blue. Green-highlighted residues, such as Ser and Thr, are potential hydrogen bond donors. The most conserved region is outlined with a black box, while semi-conserved regions are marked in red boxes. Based on the alignment, *"nef"* and *"tat"* share several conserved residues and exhibit fewer gaps, indicating close evolutionary relationships, which is supported by phylogenetic analysis. *"vpu"*, being shorter and more divergent, has fewer conserved residues and more gaps, with the presence of transmembrane regions; and (B) The phylogenetic tree, constructed based on sequence homology, illustrates the evolutionary relationships between *nef, tat, rev*, and *vpu. "nef"* and "*tat"* form a sister group, while "*rev"* and "*vpu"* share a common ancestor

"vpu" is structurally compact, whereas *"rev"* has extensive alignment but significant structural deviations.

Comparative structural analysis with MHC Class I proteins

The HIV-1 proteins were aligned with MHC Class I molecules for a comparative structural analysis, as shown in (Table 5). The analysis revealed several key findings. Regarding similarities, the RMSD values ranged from 2.5 to 6.7 Å, with Chain 2 demonstrating higher TM-scores in most cases, indicating moderate structural similarity. On the other hand, significant differences were observed in the number of residues present in the HIV-1 proteins, which varied from 81 to 439. This variation led to disparities in the aligned residues, ranging from 27 to 133. The analysis also highlighted variability in RMSD and TM-scores. Some protein pairs showed poor alignment with high RMSD values, while others aligned well with low RMSD values. TM-scores for Chain 1 fluctuated, reflecting greater structural diversity, whereas Chain 2 exhibited more consistency. Furthermore, molecular interactions between HIV-1 proteins and MHC Class I molecules were analyzed using Chimaera, as depicted in (Fig. 5). Figure 5A shows the interaction between Nef (HIV-1) and MHC-I (PDB: 1je6), Figure 5B depicts the interaction between Tat (HIV-1) and MHC-I (PDB: 1b3j), Figure 5C illustrates the interaction between HIV-1-*rev* and MHC-I (PDB: 8tlz), and Figure 5D presents the interaction between HIV-1-*vpu* and MHC-I (PDB: 8tlz).

Structural visualization using PyMOL highlighted key secondary structural elements

The green regions in the molecular structures represent the protein backbone, while alpha-helices are depicted as coiled regions. Beta-sheets appear as flat, extended regions, and loops connect these

Table 5 — TM analysis of HIV-1 proteins with MHC Class I molecules. RMSD values vary among proteins but generally remain high, suggesting poor geometric alignment. The TM-score of chain 1 indicates structural diversity among the protein complexes

MHC-1 PDB code	HIV-1 PDB code	Chain 1 name	Chain 2 name	Length of chain 1	Length of Chain 2	Aligned length	RMSD value	TM score (chain 1)	TM score (chain 2)
1b3j	7D7S	A235675	B235675	264	105	72	5.02	0.18094	0.33467
1b3j	1tiv	A806120	B806120	264	86	50	4.04	0.13910	0.29820
1b3j	7u0f	A887608	B887608	264	439	132	6.02	0.29667	0.20347
1b3j	2n28	A282003	B282003	264	81	39	3.15	0.12281	0.31867
1hyr	7D7S	A750766	B750766	128	105	67	5.54	0.26130	0.29343

(*Contd.*)

Table 5 — TM analysis of HIV-1 proteins with MHC Class I molecules. RMSD values vary among proteins but generally remain high, suggesting poor geometric alignment. The TM-score of chain 1 indicates structural diversity among the protein complexes (*Contd.*)

MHC-1 PDB code	HIV-1 PDB code	Chain 1 name	Chain 2 name	Length of chain 1	Length of Chain 2	Aligned length	RMSD value	TM score (chain 1)	TM score (chain 2)
1hyr	1tiv	A875145	B875145	128	86	48	4.38	0.21857	0.27610
1hyr	7u0f	A513624	B513624	128	439	86	5.05	0.36872	0.14649
1hyr	2n28	A645489	B645849	128	81	33	4.43	0.15648	0.21309
7f16	7D7S	A699461	B699461	123	105	58	4.50	0.27831	0.30856
7f16	1tiv	A886194	B886194	123	86	48	4.83	0.21567	0.27080
7f16	7u0f	A163492	B163492	123	439	85	5.56	0.33142	0.13644
7f16	2n28	A496417	B496417	123	81	41	5.22	0.16263	0.20088
7f17	7D7S	A702652	B702652	124	105	58	4.48	0.27729	0.30907
7f17	1tiv	A736846	B736846	124	86	48	4.47	0.22272	0.27730
7f17	7u0f	A842674	B842674	124	439	80	5.19	0.35037	0.13583
7f17	2n28	A799427	B799427	124	81	40	5.25	0.15760	0.19762
7f18	7D7S	A83198	B83198	123	105	67	5.49	0.26781	0.29339
7f18	1tiv	A266640	B266640	123	86	47	4.73	0.21850	0.27315
7f18	7u0f	A324359	B324359	123	439	83	5.41	0.34736	0.13728
7f18	2n28	A407283	B407283	123	81	32	4.51	0.15757	0.21096
7fyf	7D7S	A715714	B715714	124	105	66	5.41	0.26572	0.29254
7fyf	1tiv	A760670	B760670	124	86	44	4.23	0.21282	0.26660
7fyf	7u0f	A457897	B457897	124	439	84	5.31	0.35725	0.35725
7fyf	2n28	A130796	B130796	124	81	27	3.71	0.14543	0.19255
7f19	7D7S	A100639	B100639	124	105	67	5.51	0.26667	0.29362
7f19	1tiv	A146201	B146201	124	86	44	4.24	0.21304	0.26733
7f19	7u0f	A272537	B272437	124	439	78	5.13	0.33390	0.13166
7f19	2n28	A241790	B241790	124	81	39	4.95	0.16671	0.21291
1je6	7D7S	A649161	B649161	275	105	64	4.99	0.15801	0.30690
1je6	1tiv	A650389	B650389	275	86	61	4.88	0.14784	0.30306
1je6	7u0f	A161977	B161977	275	439	133	6.74	0.26758	0.19107
1je6	2n28	A303914	B303914	275	81	38	2.68	0.12040	0.33658
8tlz	7D7S	A621551	B621551	164	105	67	5.16	0.23593	0.31261
8tlz	1tiv	A977319	B977319	164	86	54	4.52	0.20079	0.29488
8tlz	7u0f	A773940	B773940	164	439	124	5.47	0.40436	0.19880
8tlz	2n28	A281086	B281086	164	81	39	2.90	0.18471	0.31404
8tm0	7D7S	A189687	B189687	253	105	61	4.87	0.16309	0.30145
8tm0	1tiv	A940468	B940468	253	86	46	4.71	0.12538	0.26752
8tm0	7u0f	A613449	B613449	253	439	129	6.14	0.29296	0.19607
8tm0	2n28	A100257	B100257	253	81	52	4.82	0.13782	0.28785

alpha-helices and beta-sheets. Ligands are shown as multicoloured stick-like structures, interacting with the proteins. In the binding affinity analysis, the distances between HIV-1 proteins and MHC Class I proteins were measured. For the *nef*-MHC-I (1je6) pair, the ligand-protein distances were 2.6, 3.2, and 3.5 Å. For *tat*-MHC-I (1b3j), the distances were 2.1, 3.2, and 3.4 Å. For *rev*-MHC-I (8tlz), the distances were 2.6, 2.7, 3.1, and 3.1 Å. Lastly, for *vpu*-MHC-I (8tlz), the distances were 3.3, 3.0, and 3.4 Å. Smaller binding distances suggest stronger interactions and higher binding affinity. According to this analysis, the interaction strength between these HIV-1 proteins and MHC Class I proteins is notably high, particularly in comparison to other structural assessments presented in (Table 4).

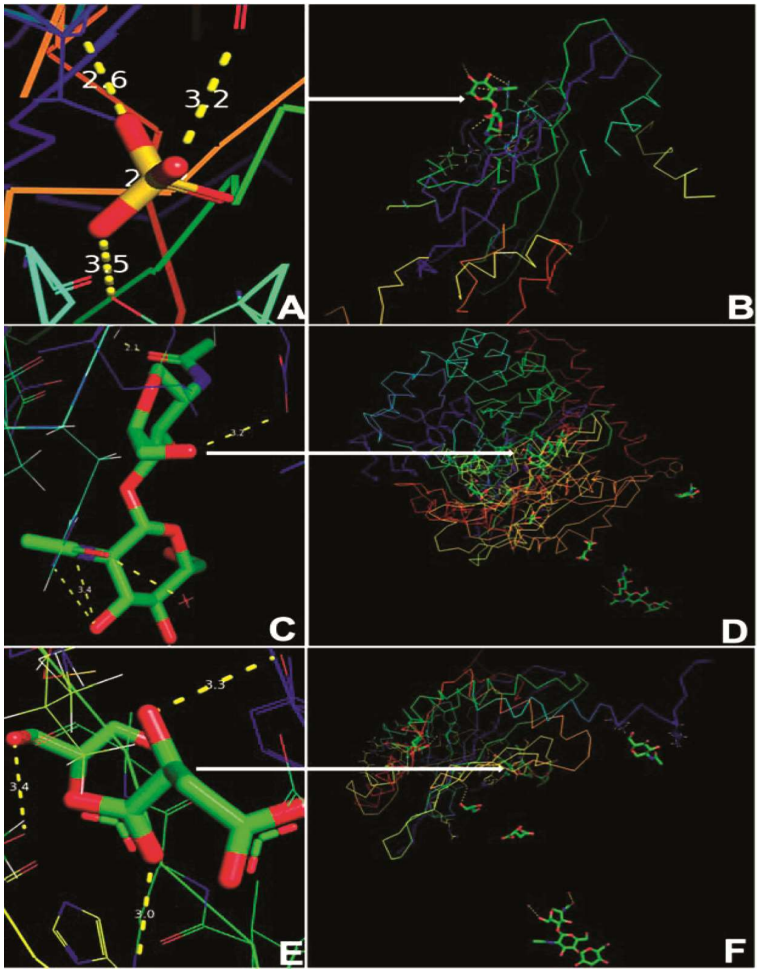


Fig. 5 — Protein-Protein Interactions Between HIV-1 Accessory Proteins and MHC Class I Molecules Visualized Using PyMOL. (A-B) Structural representations illustrate the interaction interface between HIV-1 *"nef"* and MHC class I (PDB ID: 1JE6). Hydrogen bonds between interacting residues are highlighted, with distances ranging from 2.6 Å to 3.5 Å; (C-D) Interaction between HIV-1 *"tat"* and MHC class I (PDB ID: 1B3J) is depicted, demonstrating hydrogen bond distances spanning 2.1 Å to 3.4 Å; and (E-F) The structural interface between HIV-1 *"rev"* and MHC class I (PDB ID: 8TLZ) is shown, with observed hydrogen bond distances ranging from 2.6 Å to 3.1 Å. All panels display key interaction regions, emphasizing the contributions of hydrogen bonding and van der Waals forces to the stable association of HIV-1 accessory proteins with MHC class I molecules

Conclusion

This research highlights the complex interactions between HIV-1 and MHC-I molecules, demonstrating the importance of computational methods in enhancing our understanding of these interactions. By employing multiple sequence alignment, phylogenetic analysis, TM-align, and molecular docking, this study provides essential insights into the structural stability and characteristics of HIV-1 proteins. Key structural parameters such as RMSD, TM-scores, and binding distances reveal how variations in protein conservation and structural deviations may influence HIV-1 pathogenesis. While computational approaches show significant potential for predicting protein interactions, further experimental studies are necessary to validate these predictions and improve therapeutic strategies. The findings of this study could play a crucial role in advancing modern treatments for HIV-1 by refining the datasets used for modelling protein interactions and enhancing therapeutic predictions.

References

1 German Advisory Committee Blood (Arbeitskreis Blut), Subgroup 'Assessment of Pathogens Transmissible by Blood', Human immunodeficiency virus (HIV). *Transfus Med Hemother*, 43 (2016) 203.

2 Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF & Shaw GM, Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature*, 397 (1999) 436.

3 Sharp PM & Hahn BH, Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*, 1 (2011) a006841.

4 Naif HM, Pathogenesis of HIV infection. *Infect Dis Rep*, 5 (2013) e6.

5 Fauci AS, 25 years of HIV. *Nature*, 453 (2008) 289.

6 Cunningham AL, Li S, Juarez J, Lynch G, Alali M & Naif H, The level of HIV infection of macrophages is determined by interaction of viral and host cell genotypes. *J Leukoc Biol*, 68 (2000) 311.

7 Moir S, Chun TW & Fauci AS, Pathogenic mechanisms of HIV disease. *Annu Rev Pathol*, 6 (2011) 223.

8 El-Atrouni W, Berbari E & Temesgen Z, HIV-associated opportunistic infections: Bacterial infections. *J Med Liban*, 54 (2006) 80.

9 Ho DD, Sarngadharan MG, Resnick L, Dimarzoveronese F, Rota TR & Hirsch MS, Primary human T-lymphotropic virus type III infection. *Ann Intern Med*, 103 (1985) 880.

10 Chiodi F, Albert J, Olausson E, Norkrans G, Hagberg L & Sönnerborg A, Isolation frequency of human immunodeficiency virus from cerebrospinal fluid and blood of patients with varying severity of HIV infection. *AIDS Res Hum Retroviruses*, 4 (1988) 351.

11 Lassmann H, Schmied M, Vass K & Hickey WF, Bone marrow-derived elements and resident microglia in brain inflammation. *Glia*, 7 (1993) 19.

12 Hickey WF, Basic principles of immunological surveillance of the normal central nervous system. *Glia*, 36 (2001) 118.

13 Eisfeld C, Reichelt D, Evers S & Husstedt I, CSF penetration by antiretroviral drugs. *CNS Drugs*, 27 (2013) 31.

14 Li GH, Henderson L & Nath A, Astrocytes as an HIV reservoir: Mechanism of HIV infection. *Curr HIV Res*, 14 (2016) 373.

15 Joseph SB, Arrildt KT, Sturdevant CB & Swanstrom R, HIV-1 target cells in the CNS. *J Neurovirol*, 21 (2015) 276.

16 Fenwick C, Joo V, Jacquier P, Noto A, Banga R & Perreau M, T-cell exhaustion in HIV infection. *Immunol Rev*, 292 (2019) 149.

17 Ng'eno BN, Kellogg TA, Kim AA, Mwangi A, Mwangi M & Wamicwe J, Modes of HIV transmission among adolescents and young adults aged 10–24 years in Kenya. *Int J STD AIDS*, 29 (2018) 800.

18 Belov A, Yang H, Forshee RA, Whitaker BI, Eder AF & Chancey C, Modeling the risk of HIV transfusion transmission. *J Acquir Immune Defic Syndr*, 92 (2023) 173.

19 Ball LJ, Puka K, Speechley M, Wong R, Hallam B & Wiener JC, Sharing of injection drug preparation equipment is associated with HIV infection: A cross-sectional study. *J Acquir Immune Defic Syndr*, 81 (2019) e99.

20 Jewanraj J, Ngcapu S & Liebenberg LJP, Semen: A modulator of female genital tract inflammation and a vector for HIV-1 transmission. *Am J Reprod Immunol*, 86 (2021) e13478.

21 Cavarelli M & Le Grand R, The importance of semen leukocytes in HIV-1 transmission and the development of prevention strategies. *Hum VaccinImmunother*, 16 (2020) 2018.

22 Gupta S, Granich R & Williams BG, Update on treatment as prevention of HIV illness, death, and transmission: Sub-Saharan Africa HIV financing and progress towards the 95-95-95 target. *Curr Opin HIV AIDS*, 17 (2022) 368.

23 Sakuragi J, Morphogenesis of the infectious HIV-1 virion. *Front Microbiol*, 2 (2011) 242.

24 McGettigan JP, Naper K, Orenstein J, Koser M, McKenna PM & Schnell MJ, Functional human immunodeficiency virus type 1 (HIV-1) Gag-Pol or HIV-1 Gag-Pol and Env expressed from a single rhabdovirus-based vaccine vector genome. *J Virol*, 77 (2003) 10889.

25 Hill M, Tachedjian G & Mak J, The packaging and maturation of the HIV-1 Pol proteins. *Curr HIV Res*, 3 (2005) 73.

26 Staudt RP, Alvarado JJ, Emert-Sedlak LA, Shi H, Shu ST & Wales TE, Structure, function, and inhibitor targeting of HIV-1 Nef-effect or kinase complexes. *J Biol Chem*, 295 (2020) 15158.

27 Goulder PJ & Watkins DI, Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nat Rev Immunol*, 8 (2008) 619.

28 Wonderlich ER, Leonard JA & Collins KL, HIV immune evasion: Disruption of antigen presentation by the HIV Nef protein. *Adv Virus Res*, 80 (2011) 103.

29 Khan N & Geiger JD, Role of viral protein U (Vpu) in HIV-1 infection and pathogenesis. *Viruses*, 13 (2021) 1466.

30 Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJ, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 10 (2015) 845.

31 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K & Li W, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7 (2011) 539.

32 Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD & Bairoch A, ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31 (2003) 3784.

33 Eddy SR, What is a hidden Markov model? *Nat Biotechnol*, 22 (2004) 1315.

34 Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D & Dutta S, The RCSB Protein Data Bank: New resources for research and education. *Nucleic Acids Res*, 41 (2013) D475.

35 Laskowski RA, Jabłońska J, Pravda L, Vařeková RS & Thornton JM, PDBsum: Structural summaries of PDB entries. *Protein Sci*, 27 (2018) 129.

36 Wiltgen M, Algorithms for structure comparison and analysis: Homology modelling of proteins. In: Encyclopedia of Bioinformatics and Computational Biology. *Elsevier*, (2019) 38.

37 Pawar SS & Rohane SH, Review on Discovery Studio: An important tool for molecular docking. *Asian J Res Chem*, 14 (2021) 86.

38 Zhang Y & Skolnick J, TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, 33 (2005) 2302.

39 Geyer M, Fackler OT & Peterlin BM, Structure-function relationships in HIV-1 Nef. *EMBO Rep*, 2 (2001) 580.

40 Das AT, Harwig A & Berkhout B, The HIV-1 Tat protein has a versatile role in activating viral transcription. *J Virol*, 85 (2011) 9506.

41 Lebedev A, Kim K, Ozhmegova E, Antonova A, Kazennova E & Tumanov A, Rev protein diversity in HIV-1 group M clades. *Viruses*, 16 (2024) 759.

42 González ME, Vpu protein: The viroporin encoded by HIV-1. *Viruses*, 7 (2015) 4352.

43 Ahmad N & Venkatesan S, Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. *Science*, 241 (1988) 1481.

44 Singh RK, Lau D, Noviello CM, Ghosh P & Guatelli JC, An MHC-I cytoplasmic domain/HIV-1 Nef fusion protein binds directly to the mu subunit of the AP-1 endosomal coat complex. *PLoS One*, 4 (2009) e8364.

45 Heaphy S, Dingwall C, Ernberg I, Gait MJ, Green SM & Karn J, HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located within the Rev response element region. *Cell*, 60 (1990) 685.

46 Sharpe S, Yau WM & Tycko R, Structure and dynamics of the HIV-1 Vpu transmembrane domain revealed by solid-state NMR with magic-angle spinning. *Biochemistry*, 45 (2006) 918.