Heliyon 10 (2024) e23183

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Investigating the role of hypothetical protein (AAB33144.1) in HIV-1 virus pathogenicity: A comparative study with FDA-Approved inhibitor compounds through *In silico* analysis and molecular docking

Md. Imran Hossain[a], Anika Tabassum Asha[a], Md. Arju Hossain[a], Shahin Mahmud[a,*], Kamal Chowdhury[b], Ramisa Binti Mohiuddin[c], Nazneen Nahar[a], Saborni Sarker[a], Suhaimi Napis[d], Md Sanower Hossain[e], A.K. M. Mohiuddin[a,**]

[a] *Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Tangail, 1902, Bangladesh*
[b] *Biology Department, Claflin University, 400 Magnolia St, Orangeburg, SC 29115, USA*
[c] *Department of Pharmacy, Mawlana Bhashani Science and Technology University, Tangail, 1902, Bangladesh*
[d] *Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor D.E., Malaysia*
[e] *Centre for Sustainability of Mineral and Resource Recovery Technology (Pusat SMaRRT), Universiti Malaysia Pahang Al-Sultan Abdullah, Kuantan 26300, Malaysia*

ARTICLE INFO

Keywords:
Protease
Hypothetical protein
HXB2 strain
AAB33144.1
Cupridoxins
Cytoplasmic membrane

ABSTRACT

Aim and objective: Due to the a lot of unexplored proteins in HIV-1, this research aimed to explore the functional roles of a hypothetical protein (AAB33144.1) that might play a key role in HIV-1 pathogenicity.

Methods: The homologous protein was identified along with building and validating the 3D structure by searching several bioinformatics tools.

Results: Retroviral aspartyl protease and retropepsin like functional domains and motifs, folding pattern (cupredoxins), and subcellular localization in cytoplasmic membrane were determined as biological activity. Besides, the functional annotation revealed that the chosen hypothetical protein possessed protease-like activity. To validate our generated protein 3D structure, molecular docking was performed with five compounds where nelfinavir showed (−8.2 kcal/mol) best binding affinity against HXB2 viral protease (PDB ID: 7SJX) and main protease (PDB ID: 4EYR) protein.

Conclusions: This study suggests that the annotated hypothetical protein related to protease action, which may be useful in viral genetics and drug discovery.

* Corresponding author.
** Corresponding author.
E-mail addresses: shahin018mbstu@gmail.com (S. Mahmud), akmmohiu@yahoo.com (A.K.M. Mohiuddin).

https://doi.org/10.1016/j.heliyon.2023.e23183
Received 18 November 2023; Accepted 28 November 2023
Available online 6 December 2023

1. Introduction

Human Immunodeficiency Virus (HIV) weakens the immune system of its hosts, which is in charge of fighting infections and diseases. HIV infection can result in Acquired Immunodeficiency Syndrome (AIDS), severely depleting the immune system [1]. A total of 39.0 million people were expected to have HIV at the end of 2022, per the Joint United Nations Programme on HIV/AIDS (UNAIDS). Among them, 630 000 people died from HIV-related causes and 1.3 million people acquired HIV in 2022. The global total of new HIV infections has fallen by 32 % since 2010 to an expected 1.5 million cases worldwide in 2021. Besides, the SDG objective 3.3 of eradicating the HIV epidemic by 2030 is supported by global HIV strategies from UNAIDS, the Global Fund, and WHO (https://www.who.int/news-room/fact-sheets/detail/hiv-aids) (updated on July 13, 2023).

The disease has spread more widely, causing devastation among additional individuals and driving up demand for vaccines and novel drugs. The US Food and Drug Administration (FDA) has approved 26 anti-HIV drugs since discovering HIV. Ten of these substances are inhibitors of the HIV protease. Patients with HIV are given the FDA-approved protease inhibitor drugs, but low viral resistance barriers can make the medication ineffective if taken carelessly [2,3]. Therefore, in-depth research on the HIV genome is necessary to identify new potential therapeutic and vaccine targets in HIV that may be given to patients safely. This study examined hypothetical proteins to identify novel HIV therapeutic and vaccination targets. Within the retroviral family, subfamily Orthoretrovirinae, the human immunodeficiency virus (HIV) is classified as belonging to the genus Lentivirus. HIV is divided into types 1 and 2 (HIV-1, HIV-2) based on genetic traits and variations in the viral antigens [4]. HIV-1's genetic diversity stems from its high mutation rate and ability to recombine with other subtypes. Because different subtypes of HIV may respond differently to treatment and have different immune responses, this genetic diversity has implications for developing HIV treatments and vaccines [5]. Approximately 70 envelope projections are found on each HIV-1 virion, making it an enveloped virus [6]. Dendritic cells (DCs) are the first to come into contact with HIV-1, and they play a key role in the virus's ability to infect CD4$^+$ T lymphocytes. The HIV-specific DC receptor DC-SIGN binds HIV-1 at its gp-120 region and facilitates the virus's transport to lymphoid tissue prior to cell infection [7].

HXB2 was the first HIV-1 strain to have its entire genome sequenced, allowing researchers better to understand the virus and its interactions with host cells [8]. HXB2 has six distinct mutations in the V3 loop region of the HIV-1 genome that have been found to boost in vivo replication, viral entry and pathogenesis. It has also been used to test the effectiveness of antiretroviral drugs in inhibiting HIV-1 replication [9,10]. The Gag-Pol precursor polyprotein is synthesized from protease protein encoded in the viral genome [11]. The conserved catalytic residues in protease protein allowed it to be identified as a member of the aspartic protease family Asp-Thr/Ser-Gly. Each of the two 99-residue subunits of the mature PR dimer, which is catalytically active, includes one copy of the catalytic triplet. Gag and Gag-Pol polyproteins have many cleavage sites recognized by protease, which hydrolyzes the peptide bond to liberate the separate structural proteins and enzymes. Infectious virus cannot be produced unless the cleavage sites are hydrolyzed in the proper sequential order [12,13].

HIV protease was immediately identified as a possible target for developing antiretroviral medicines due to its crucial function in viral replication [14]. The most successful AIDS treatment currently is a combination of HIV protease inhibitors, reverse transcriptase inhibitors, and integrase inhibitors, known as highly active antiretroviral therapy (HAART). The highest intrinsic antiviral activity is found in protease inhibitors compared to all HIV-1 medications. Nine antiviral protease inhibitors are currently authorized, but except for tipranavir, all are peptidomimetics. For patients failing first-line therapy with IN and RT inhibitors, darunavir, lopinavir, and atazanavir are advised in second-line regimens [15,16]. The HIV-1 protease is inhibited by protease inhibitors, which are substrate or transition state analogs. During the development of viruses, this enzyme breaks down viral polyproteins. Gp160 polyprotein is broken down by a cellular protease into the transmembrane and surface subunits complexes gp120/gp41 that mediate viral entry. The HIV-1 protease, on the other hand, cleaves the Gag and Gag-Pol polyproteins into many mature virion proteins, resulting in mature virions that can infect new cells [17]. Despite the significance of PIs in treating HIV-1, it is not known precisely which steps in the virus life cycle these medications prevent when used in a clinical setting. Research by Jiang, J. and Aiken, C. (2007) mentioned that the HIV-1 protease is necessary for viral infectivity [18]. Another research by Whitcomb, Jeannette M. et al. (2007) demonstrated that the stability of full-size unintegrated cDNA is impacted by protease inhibitors, not the beginning or progression of reverse transcription [19].

Atazanavir is one of several drugs that work as an aza peptide HIV 1 protease inhibitor, preventing the viral gag and gag-pol proteins from being cleaved in HIV-infected cells [20]. Besides, warfarin is metabolized by the cytochrome P450 (CYP) 3A4 enzyme, which is inhibited by ritonavir. Combined treatment of ritonavir with warfarin increases the risk of bleeding because ritonavir inhibits CYP3A4 [21]. Despite the significance of PIs in treating HIV-1, it is not known exactly which steps in the virus life cycle these medications prevent when used in a clinical setting.

In most genomes that have been fully sequenced, 50–60 % of the genes have been determined to have a function. Each organism's genome contains several hypothetical proteins, or genes with undefined functions [22]. Although HIV has a short genome, many of the proteins produced by HIV are still in the category of "hypothetical proteins" because the current understanding of their structures and biological functions is limited. An unanticipated but remarkably nonrandom connection between genomic coding regions and disease-associated HIV-1 insertions was found in thirty of 34 hypothetical and annotated genes [23]. Therefore, functional annotation of these HPs may make it possible to prioritize new therapeutic targets for the treatment of infectious disorders like those brought on by HIV. For several types of communicable diseases, such as chlamydia, tuberculosis, and shigellosis hypothetical proteins have been investigated as potential therapeutic targets [24]. The hypothetical proteins can be annotated *in-silico* at a low cost and quickly enough to investigate their function [25]. Multiple in-silico-based databases and software were utilized in this investigation to anticipate hypothetical protein (accession no. AAB33144.1) function, which could lead to identifying novel pharmacological targets for screening, drug development, and designing for treating HIV infections.

2. Material and methods

2.1. Sequence retrieval and similarity identification

This study searched for hypothetical proteins (HXB2 = viral protease) from the National Center for Biotechnology Information (NCBI) protein database to examine distant ancestors using BLAST (Basic Local Alignment Search Tool) programs. After applying filtering by setting organism: viruses; sequence of length: 90 to 100 amino acids and molecular weight: 9000 Da–10000 Da, we have found two hypothetical proteins (accession No. AAB33144.1 and AAB33143.1 with Gene ID: 913388 and 913387, respectively). The BLASTp program was used [26] to perform similarity searches of AAB33144.1 protein only against the non-redundant Swis-sProt/UniProt database, available at the NCBI protein Database (https://www.ncbi.nlm.nih.gov/) to identify functional proteins that potentially shared structural similarities with the uncharacterized protein. This initial step provided a starting point for speculating on the role of the putative protein we were interested in. Following retrieving the FASTA format sequence from the NCBI database, several prediction services were employed to characterize the hypothetical protein *in silico* (Table 1). The Protein Data Bank was found to not have the three-dimensional structures of these proteins. As a result, the present project involved creating 3D models of these HXB2 = viral protease proteins (accession No: AAB33144.1) and validating our protein structures. The flow diagram of our investigation is provided in Fig. 1.

2.2. Multiple sequence alignment and phylogeny analysis

In this study, we employed Clustal Omega tools for multiple sequence alignment because amino acid or nucleotide sequences could be aligned numerous times rapidly and precisely using this software (https://www.ebi.ac.uk/Tools/msa/clustalo/) [27]. The FASTA sequence of our query protein AAB33144.1 and homologous six hypothetical proteins including AAB33143.1, AAB33142.1, AAB33141.1, P20892.3, P12497.4, and P35963.3 were used to performing of multiple sequence alignment program. Multiple sequence alignments are essential for predicting protein structure and function, implying phylogenies, and performing other sequencing alignment programs. The state-of-the-art has been raised by newly developed systems, which are more accurate, can scale to hundreds of proteins, and can compare proteins with different domain designs [28]. For a phylogenetic tree analysis, we have used MEGA X (Molecular Evolutionary Genetics Analysis) [29] to generate a network file of the reported protein sequence and then display it through iTOL (Interactive Tree Of Life) (https://itol.embl.de/) web server [30]. A web-based programme called iTOL is used to view, manipulate, and annotate phylogenetic trees.

2.3. Evaluation of physiochemical features

Chemical and physical properties like molecular weight (Mw), theoretical isoelectric (pI), number of negatively charged residues (Asp + Glu), number of positively charged residues (Arg + Lys), instability index, aliphatic index, grand average of hydropathicity (GRAVY), and so on were predicted by the Protparam (Protein Parameters) interface on the ExPASy (the Expert Protein Analysis System) website (https://web.expasy.org/protparam/) [31]. The structural and functional qualities of a protein are reflected in its physico-chemical properties. An accurate approach to computing physical and chemical properties, the ProtParam interface only employs a single sequence per analysis [32].

Table 1
Computational tools used for *in silico* analysis of hypothetical protein (accession no: AAB33144.1).

No	Server name	Reference	Purpose
1	BLASTp	[26]	Similarity search
2	Clustal omega	[27]	Multiple sequence alignment
3	MEGA X and iTOL	[29,30]	Phylogeny analysis
4	ProtParam	[31]	Physicochemical characterization
5	Virus-mPLoc	[34]	Subcellular localization
6	Genome Net	[36]	Motif discovery
7	Pfam	[37]	Family relationship identification
8	DeepCoil	[38]	Coiled-coil motif identification
9	PFP-FunDSeqE	[39]	Fold recognition
10	PSIPRED and SOPMA	[40,41]	Secondary structure prediction
11	HHpred	[42]	Tertiary structure prediction
	YASARA	[43]	Energy Minimization
12	PROCHECK	[44]	Structure verification
	Verify3D	[45]	Structure validation
	ERRAT	[46]	Structure validation
13	CASTp	[47]	Active Sites prediction
	RCSB PDB	[48]	3D structure information of protein
	PubChem	[53]	Collection of Ligand structure
14	PyRx and Biovia Discovery Studio Visualizer	[55,56]	Molecular docking and structure visualization

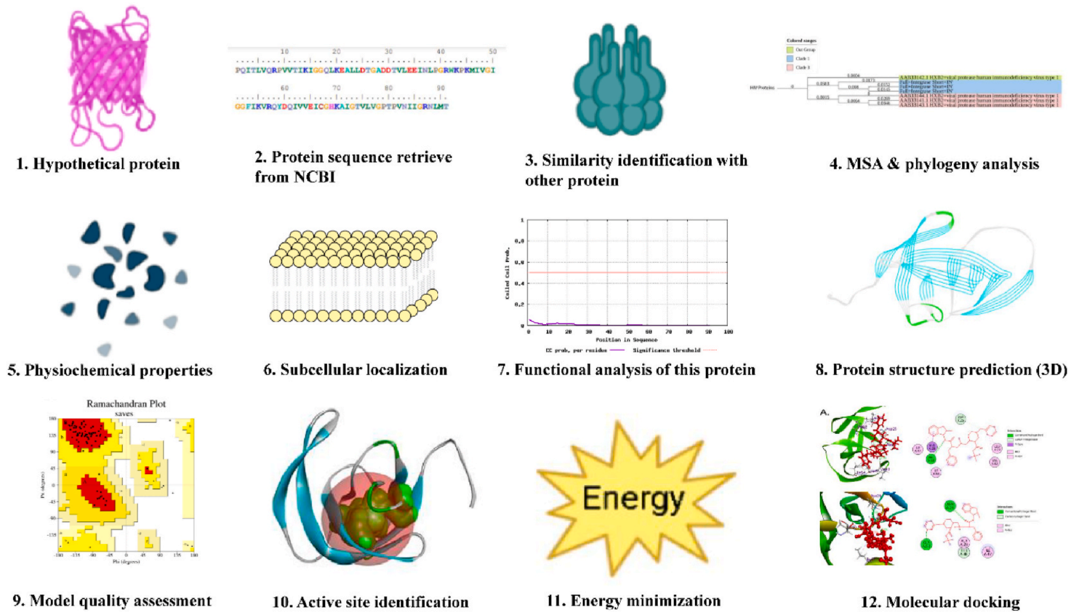
1. Hypothetical protein
PQITLVQRPVVTIKIGGQLKEALLDTGADDTVLEEINLPGRWKPKMIVGI
GGFIKVRQYDQIVVEICGHKAIGTVLVGPTPVNIIGRNLMT
2. Protein sequence retrieve from NCBI
3. Similarity identification with other protein
4. MSA & phylogeny analysis
5. Physiochemical properties
6. Subcellular localization
7. Functional analysis of this protein
8. Protein structure prediction (3D)
Ramachandran Plot
9. Model quality assessment
10. Active site identification
Energy
11. Energy minimization
A.
12. Molecular docking

Fig. 1. A schematic representation of the several steps involved in the functional and structural characterization of hypothetical protein HXB2 (accession no: AAB33144.1) of HIV-1.

2.4. Exploring subcellular localization

Predicting protein functions will benefit greatly from a clear understanding of how to predict subcellular localization from protein sequences. The proper transport of a protein to its ultimate location is essential to its function because proteins have evolved to operate best in a particular subcellular localization [33]. This study used the Virus-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/virus-multi/) web tool to predict subcellular localization. Due to their complexity, the Virus-mPLoc (Virus Multi-Location Predictor) predictor can recognize the multi-location virus proteins, which cannot be predicted using conventional methods. Additionally, it can provide more accurate predictions of viral protein location in a host cell [34].

2.5. The evaluation of operational biological properties

Protein domains are discrete molecular evolution units typically connected with certain features of the molecular and cellular function of the query protein sequence. Conducting a search at NCBI's conserved domain database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for conserved domain analysis [35]. Protein sequence motifs serve as the family-specific identifiers of proteins and are frequently employed in protein function prediction. Motif was used to look for protein motifs through Genome Net server (https://www.genome.jp/tools/motif/) [36]. Using the Universal Protein Resource Knowledgebase (UniProtKB) and the NCBI databases, millions of protein sequences have been annotated with Pfam (Protein family database) and Superfamily domain annotations to give the protein's evolutionary relationships [37]. The DeepCoil service [38] was used to identify the protein's coiled-coil conformation. In addition, the PFP-FunD SeqE server (Predicting protein Fold Pattern with Functional Domain and Sequential Evolution information) was utilized for protein folding pattern detection [39].

2.6. Structure prediction and homology modeling of hypothetical protein

The sequence of amino acids in FASTA format for this putative protein was gathered. Essentially, this was the protein's main structure. We utilized PSIPRED (PSI-blast-based secondary structure PREDiction) (http://bioinf.cs.ucl.ac.uk/psipred/) [40] is a well-known web service offering a broad range of protein prediction and annotation tools with a primary focus on protein structural annotations and to predict secondary structures, SOPMA (Self Optimized Prediction Method from Alignment) (https://npsa.lyon.inserm.fr) [41] was assessed.

Using a pairwise comparison profile of hidden Markov models (HMMs), the HHpred server (https://toolkit.tuebingen.mpg.de/tools/hhpred) at the Max Planck Institute for Developmental Biology in Tübingen was used to generate 3D structure which can be projected the building of homology modeling [42]. The 3D model was forecasted more precisely using the template with the highest score. As time passed, the YASARA (Yet Another Scientific Artificial Reality Application) energy reduction server further honed the 3D model [43].

2.7. Assessment of the quality of 3-dimensional models

Different evaluation criteria were used to assess the validity of the hypothetical protein model of HIV. The improved structure of the model was put through several tests in the last stage of homology modeling to check for internal consistency and reliability. The Psi/Phi Ramachandran plot from the PROCHECK study was inspected to assess the backbone conformation. Quality evaluation of the predicted three-dimensional structure was also performed using PROCHECK (https://servicesn.mbi.ucla.edu/PROCHECK/) [44], Verify3D (http://nihserver.mbi.ucla.edu/Verify 3D/) [45], and ERRAT (servicesn.mbi.ucla.edu/ERRAT/) [46].

2.8. Identifying active sites

For proteins to operate appropriately, their geometric and topological characteristics, such as surface pockets, inner cavities, and cross channels, are essential. The protein's active site was identified with the help of the Computed Atlas of Surface Topography of Protein (CASTp) (http://sts.bioe.uic.edu/), which is a web-based tool for identifying, outlining, and quantifying concave surface areas on protein 3D structures and mapping of functionally annotated residues [47].

2.9. Molecular docking protocol

2.9.1. Receptor and ligand preparation

The hypothetical viral protease protein (accession no: AAB33144.1) was selected for molecular docking as a receptor (PDB ID: 7SJX). Then the RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) server [48] was used to explore the three-dimensional (3D) structure of the main protease protein of HIV-1 (PDB ID: 4EYR) for comparing our binding affinity [49]. Crystallographic water molecules were removed from the 3D coordinate file after identifying and eliminating co-crystallized ligands from the structure of 4EYR.

We have chosen several anti-HIV FDA-approved drugs including Indinavir (CID: 5362440), Lopinavir (CID: 92727), Nelfinavir (CID: 64143), Saquinavir (CID: 441243), and Tipranavir (CID: 54682461) as a ligand for *in-silico* docking experiments which were collected from the literature review analysis [3]. The rationale for choosing this inhibitor over other molecules is to validate our generating 3D model of HIV and also compare it to the main protease of HIV. Because our generating 3D structure provided the same nature as the main protease and also reported FDA-approved drugs can potentially inhibit the main protease function. For instance, saquinavir binds strongly to the protease enzyme and competitively inhibits its activity necessary for virus maturation and proliferation [50]. Patients with HIV infection treated with nelfinavir had lower levels of Fas expression and Fas-mediated apoptosis as well as higher $CD4^{+}$ cell counts [51]. On the other hand, inhibiting the production of infectious virions allows the lopinavir coformulation to have an antiviral impact, which prevents further rounds of cellular infection [52].

PubChem (https://pubchem.ncbi.nlm.nih.gov) database [53] was used to retrieve the canonical smiles id of all the drugs, and then Online SMILES Translator and Structure File Generator web-based tool was used to convert the 3D SDF to PDB structure of the drugs [54]. The research was simplified even further by optimizing ligands and converting them to PDBQT format with the help of the graphical user interface version of PyRx [55].

2.9.2. Molecular docking and binding interaction visualization

In order to locate lead compounds with specified biological functions, small-molecule libraries are docked to macromolecules using the open-source software PyRx (Python Prescription Toolkit for Radiological Analysis) via Autodock vina wizard interface [55]. At the beginning of docking, the ligands were thought to be flexible, while the protein was thought to be rigid. The Auto Grid engine found in PyRx was used to generate the configuration file for the grid's specifications. All findings with a relative root-mean-square deviation (RMSD) of less than 1.0 kcal/mol were grouped as promising candidates for the preferable binding. For this reason, the most negative binding energy was attributed to the ligand with the highest binding propensity. During docking performance, grid box was generated by setting exhaustiveness = 8; center_x:y:z: = −10.7703: 0.6456: 21.8096; and dimension x:y:z = 38.2173907089: 32.3865090561: 36.173718605. In the end, we used the Biovia Discovery Studio visualizer version 21 [56] to analyze the hydrophobic and hydrogen bond interactions.

Table 2
Six homologous proteins were explored from non-redundant UniProt KB and SwissProt database based on high scoring against our AAB33144.1 query sequence.

Accession no	Protein	Organism	Identity	Query Cover	Maximum score (bit score)	E-value
AAB33144.1	HXB2 = viral protease	HIV-1 (Ro 31–8959 resistant isolate Ro34)	100 %	100.00 %	180	9×10^{-57}
AAB33142.1	HXB2 = viral protease	HIV-1 (Ro 31–8959 resistant isolate Ro32)	100 %	100.00 %	180	1×10^{-56}
AAB33143.1	HXB2 = viral protease	HIV-1 (Ro 31–8959 resistant isolate Ro33)	98.90 %	100.00 %	178	4×10^{-56}
AAB33141.1	HXB2 = viral protease	HIV-1 (Ro 31–8959 resistant isolate Ro31)	96.70 %	100.00 %	176	4×10^{-55}
P20892.3	Gag-Pol polyprotein	HIV - 1 (OYI ISOLATE)	92.31 %	100.00 %	175	3×10^{-51}
P12497.4	Gag-Pol polyprotein	HIV- 1 (NEW YORK-5 ISOLATE)	92.31 %	173	174	4×10^{-51}
P35963.3	Gag-Pol polyprotein	HIV-1 (YU-2 isolate)	92.31 %	173	173	9×10^{-51}

3. Result and discussion

3.1. Multiple sequence alignment and phylogenetic tree construction

Similarities with other viral protease proteins were found by BLASTp searches against the nonredundant and SwissProt databases (Table 2). The FASTA sequences of the hypothetical protein (AAB33144.1) and the related annotated proteins were compared using multiple sequence alignment. A single template protein known as the crystal structure of an HIV-1 protease was found to mimic the sequences of the viral HXB2 = viral protease closely.

Phylogenetic analysis was also conducted to verify the homology evaluation between the proteins at the complex and subunit levels. In our study, the phylogenetic tree indicated that our query protein (Accession No. AAB33144.1) was closely related to AAB33141.1 and AAB33143.1 with a branch distance of 0.0289 and 0.0046, respectively (Fig. 2). By dint of phylogeny analysis, the rapid evolution of HIV makes it possible to examine the networks of HIV-1 transmission, and forensic investigations [57].

3.2. Physiochemical features of the hypothetical protein

The protein contains 91 amino acids; the ones found in the most incredible quantity are (in descending order) Ala (3), Arg (4), Asn (3), Asp (4), Cys (1), Gln (5), Glu (4), Gly (11), His (1), Ile (12), Leu (8), Lys (6), Met (2), Phe (1), Pro (6), Ser (0), Thr (7), Trp (1), Try (1), Val (11), and Pyl (0). The theoretical pI of the protein was 9.02, and its determined molecular weight was 9855.71 Da, so it is negatively charged. Understanding a protein's role in examining its molecular evolution requires a comparative assessment of its physicochemical features. The physicochemical and structural characteristics of proteins also control their co-receptor binding affinities that affect viral tropism and are concentrated in certain areas [58]. In Fig. 3 and Table 3, we presented the data that fully describe the physiochemical characteristics, including the frequency and prevalence of amino acids of the reported hypothetical protein. Besides, in Fig. 3, Aliphatic amino acid (I) showed higher frequency and tRNA synthetase class (Y) showed lower frequency in terms of amino acid compositions.

3.3. Subcellular localization

Query protein AAB33144.1 HXB2 = viral protease [human immunodeficiency virus type 1 (HIV 1), Ro 31–8959 resistant isolate Ro34, Peptide Mutant, 91 aa] and its predicted locations are host cell membrane and host cytoplasm. HIV hides itself from the host immune system by delivering its genome into the cytoplasm of the host cell through a convoluted series of activities. The HIV protein envelope (Env) interacts with the main cellular receptor CD4 to begin cell infection and this subsequent interaction causes the viral and host cell membranes to fuse [59].

3.4. The implication of the hypothetical protein in biological function

This hypothetical protein sequence was discovered to contain one distinct domain using the conserved domain (CD) search tool & three non-specific domains. Retroviral aspartyl protease (accession no: cd05482) at 11 to 91 amino acid was a pepsin/retropepsin-like aspartic protease family protein-specific domain. Aspartic protease known as HIV protease (HIV-PR), breaks down the viral gag polyprotein into smaller proteins as the HIV matures. One catalytic aspartate from each chain makes up the HIV-PR active site, shielded by two b-hairpin structures known as flaps. The flap curl tips of the HIV protease control the mechanism of ligand binding in the active sites [60]. On the other hand, non-specific domains, including Retropepsin of human endogenous retroviral components belonging to the RTVL H family (accession no: cd06095) at 11 to 91 amino acid, retroviral aspartyl protease (accession no: pfam00077) at 7 to91 amino acid, and aspartyl protease (accession no: pfam13650) at 11 to57 amino acid (Table 4). *M. N. L. Nalam, A. Peeters, T. H. M. Jonckers, I. Dierynck, and C. A. Schiffer (2007)* conducted a study and concluded that retropepsins are encoded by the retroviral genome and form a component of the polyprotein precursor that the viral protease cleaves during maturation. However, the HIV-1 protease, a key drug target for treating HIV/AIDS, is one of the most well-studied retropepsins [61]. Retropepsin is a protease enzyme encoded due to the RTVL H human endogenous retroviral gene**.** These retrovirus-like elements are left over from ancient retroviruses that integrated into the human genome and were passed down to subsequent generations [62]. While L. *Bénit, P. Dessen, and T. Heidmann (2001)*

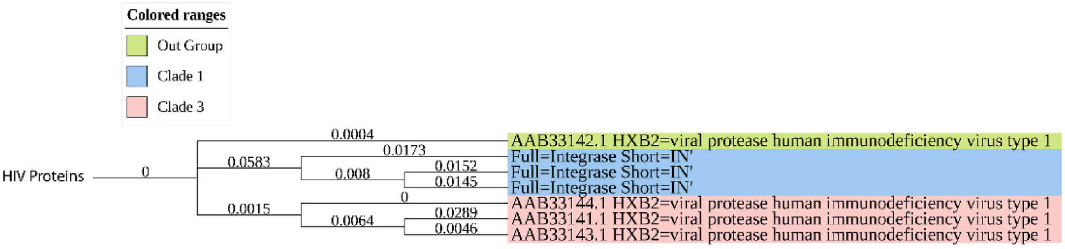
Colored ranges
🟩 Out Group
🟦 Clade 1
🟥 Clade 3
HIV Proteins
0
0.0583
0.0015
0.0004
0.0173
0.008
0.0152
0.0145
0
0.0064
0.0289
0.0046
AAB33142.1 HXB2=viral protease human immunodeficiency virus type 1
Full=Integrase Short=IN'
Full=Integrase Short=IN'
Full=Integrase Short=IN'
AAB33144.1 HXB2=viral protease human immunodeficiency virus type 1
AAB33141.1 HXB2=viral protease human immunodeficiency virus type 1
AAB33143.1 HXB2=viral protease human immunodeficiency virus type 1

Fig. 2. Phylogenetic structures based on the genetic relatedness of hypothetical protein proteases. The phylogenetic tree provided closely related two proteins, AAB33141.1 and AAB33143.1, with a branch distance of 0.0289 and 0.0046, respectively, in our query sequence.

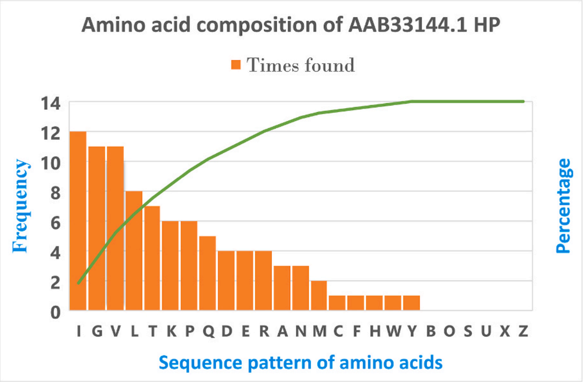

Fig. 3. A Pareto plot represents the prevalence of amino acid composition of AAB33144.1 hypothetical proteins in descending order of frequency with a cumulative line on a secondary axis as a percentage of the total. Aliphatic group (G, A, V, L, I); Aromatic (F, W, Y); Sulphur (C and M); Basic (K, R, H); Acidic (B, D, E, N, Q, Z); Aliphatic Hydroxyl (S and T) and tRNA synthetase class (Z, E, Q, R, C, M, V, I, L, Y, W). Here, letter I indicates the highest frequency, and Y indicates the lowest frequency of amino acids.

Table 3
Physiochemical properties analysis of hypothetical protein (accession no: AAB33144.1) and closely related protein (accession no: AAB33143.1).

Property	Value of AAB33144.1 protein
Number of amino acids	91 aa
Molecular weight	9855.71 K Da
Theoretical pI	9.02 (basic)
Total number of negatively charged residues (Asp + Glu)	08
Total number of positively charged residues (Arg + Lys)	10
Ext. coefficient	6990 M − 1 cm-1
Instability index (II)	35.11 (stable protein)
Aliphatic index	124.07
Grand average of hydropathicity (GRAVY)	0.257 (hydrophobic)

Table 4
List of specific & non-specific domains and superfamily of the hypothetical protein (AAB33144.1).

Name	Accession	Description	Interval (amino acid)	E-value
Domain				
RVP	Pfam00077	Retroviral aspartyl protease: single domain aspartyl proteases from retroviruses	7 to 91	8.84×10^{-30}
HIV retropepsin like	Cd5482	Retropepsins of RTVL_H family of human endogenous retrovirus-like elements.	11 to 91	1.04×10^{-25}
HIV retropepsin like	Cd05482	The Aspartyl protease family consists of predicted aspartic proteases	11 to 91	0.000655
Asp protease 2	Cd05482	The Aspartyl protease family consists of predicted aspartic proteases.	11 to 57	0.00836
Superfamily rowhead				
Pepsin retropepsin like superfamily	C11403	Pepsin-like aspartate proteases are found in ceels and retroviruses. Retroviral pepsin and pepsin-like enzymes are half as long as their cellular equivalents. The actual alignment was detected with superfamily member pfam00077	7 to 91	$\mathbf{8.84 \times 10^{-30}}$

demonstrated that endogenous retroviruses are commonly thought to be “junk” DNA with no known function, research indicates that some endogenous retroviral sequences may play essential roles in regulating gene expression and immune function [63].

Three motifs, including retroviral aspartyl protease (PF00077), aspartyl protease (PF13650), and gag-polyprotein putative aspartyl

Table 5
List of motifs of the hypothetical protein (accession no: AAB33144.1).

Pfam	Position (Independent E-value)	Description
RVP	7 to91 ($\mathbf{9.1 \times 10^{-24}}$)	Retroviral aspartyl protease (PF00077)
Asp_protease_2	11 to56 (0.0058)	Aspartyl protease (PF13650)
gag-asp_protease	11 to34 (0.018)	gag-polyprotein putative aspartyl protease (PF13975)

protease (PF13975), were also identified through Pfam via motif server (Table 5). The enzyme retroviral aspartyl protease is essential for the human immunodeficiency virus and other retroviruses' ability to reproduce (HIV). Several drugs, including Saquinavir, ritonavir, and lopinavir are protease inhibitors that target the retroviral aspartyl protease and prevent viral polyprotein cleavage, thereby inhibiting virus replication [64]. Aspartyl protease 2, also known as the amyloid precursor protein, is cleaved by the enzyme beta-secretase 1 (BACE1) (APP) and is a key enzyme in the production of beta-amyloid [65]. Gag-asp protease is an enzyme that is essential for retroviral particle maturation. The structure and function of gag-asp protease in HIV-1, the causative agent of AIDS, have been extensively studied. *J. Fanfrlik, A. K. Bronowska, J. Rezac, O. Přenosil, J. Konvalinka, and P. Hobza (2010)* mentioned that the aspartyl protease recognizes and cleaves specific amino acid sequences in the Gag polyprotein to produce mature viral proteins [66].

Superfamily search revealed a pepsin_retropepsin_like superfamily (accession no: cl11403) with e-value = 8.84e-30 (Table 5). Pepsin-like aspartate proteases found in cells and retroviruses belong to the same family, but the latter is twice as lengthy. Two domains with comparable topological characteristics are found in eukaryotic pepsin-like proteases. Although structurally linked along a 2-fold axis, the N and C-terminal domains share little in sequence homology outside of the active site region. This indicates an ancient duplication occurrence leading to the evolution of the enzymes [67]. Retroviral and eukaryotic proteases share atypical active site pattern (Asp-Thr/Ser-Gly-Ser), as do eukaryotic pepsin-like proteases' N and C-termini. Pepsin-like aspartate proteases found in retroviruses, retrotransposons, and retroelements make up the retropepsin-like family [68].

The PFP-FunDSeqE method for recognizing fold patterns in proteins found that the protein sequence contained a "cupredoxins" fold. Azurin exhibits structural similarities to the variable regions of the copper-containing, electron-transporting proteins known as cupredoxins. To prevent infectious agents from entering host cells and consequently inhibit parasitemia or viral proliferation, A study by *A. Chaudhari* et al. *(2006)* revealed that in the case of parasites and viruses like *P. falciparum* and HIV-1, azurin binds to a wide range of surface or envelope proteins [69]. In the coiled-coil graph in Fig. 4, the X-axis indicates the positions in the protein of amino acids number, and the Y-axis represents the probability score of the coiled-coil. We have found the number of predicted coiled-coil domains with threshold 1(P), 10(V), 50(I), and 90(M) amino acid positions of protein sequences. Because of their structural simplicity, the coils served as attractive scaffolds for developing functional biomaterials. The N-terminal heptad repeats region (NHR) of the gp41 subunit forms a core parallel trimeric coiled-coil structure that is surrounded by three antiparallel C-terminal heptad repeat (CHR) helices to create a very stable six-helix-bundle (6HB). The viral and cell membranes are brought together during this energetically advantageous folding phase, which makes it easier for them to fuse and, as a result, allows the viral material to enter the cell [70]. The coiled-coil structure of NHR has been targeted by the development of several peptides and small molecules, although relatively few of these have advanced to the clinical stage [71].

3.5. Protein structure prediction

The percentages of protein alpha helices (7.69 %), beta turns (7.69 %), extended strands (45.05 %), and random coils (39.56 %) are listed below based on data analyzed by the SOPMA secondary structure prediction server (Table S1 and Fig. 5). The HHpred server was used to make the 3D layout prediction. To create functional models, 30 % sequence homology is typically needed. In general, homology modeling is an important tool in the field of structural biology, allowing researchers to learn more about the connections between protein structure and function, create unknown proteins 3D structure, and improve up the process of drug development [72]. The server's prediction of the protein's 3D structure perfectly matched the template with the best score = 52.25, probability = 98.7 %, E-value = 0.0000017, resolution = 8.20 Å, Ramachandran outliers = 0.4 % and sidechain outliers = 7.6 %. Template 7SJX is a Human

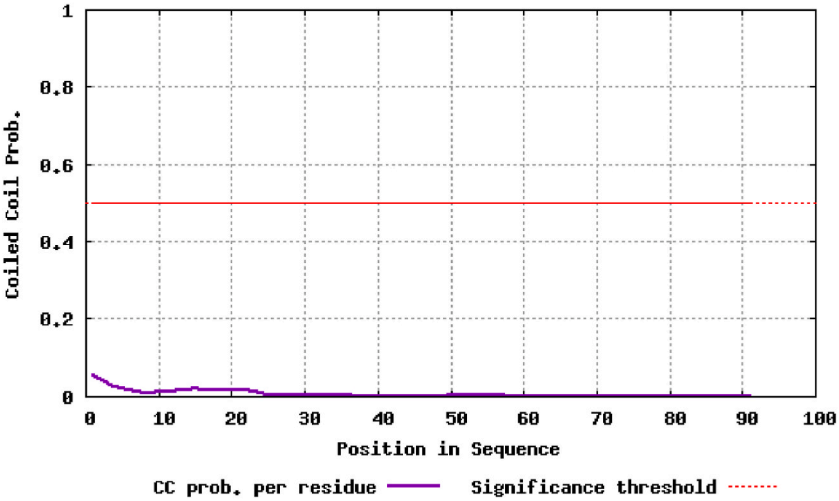

Fig. 4. Functional analysis (coiled-coil) interaction of hypothetical protein (AAB33144.1). The X-axis represents the position in sequence and the Y-axis represents the coiled-coil probability. This Figure provided a 0.08 probability score with the number of predicted coiled-coil domains at 1(P), 10 (V), 50(I), and 90(M) amino acid positions, which didn't cross the threshold level of 0.5.

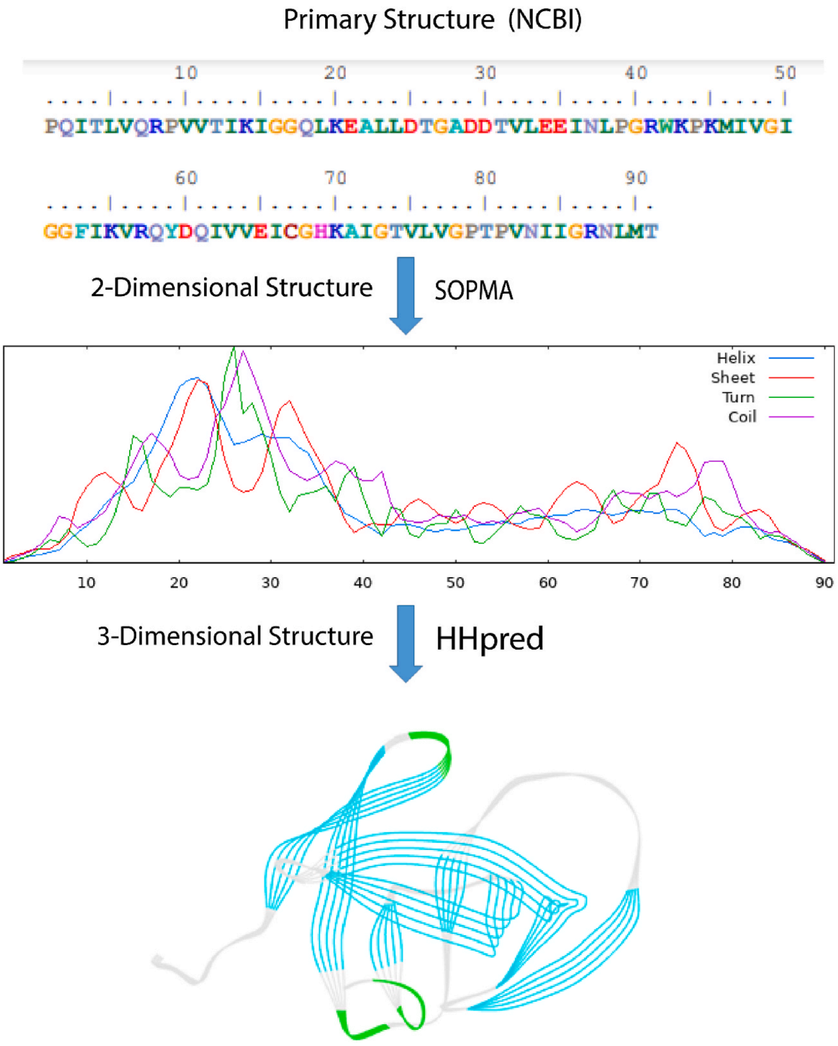

Fig. 5. Primary, secondary & three sdimensional structure prediction of protease-like nature the hypothetical protein (AAB33144.1). The primary structure provided the FASTA sequence of 91 amino acids; the Secondary structure provided the frequency of protein alpha helices (7.69 %), beta turns (7.69 %), extended strands (45.05 %), and random coils (39.56 %).

immunodeficiency virus type 1 BH10 polyprotein silence structure that was selected as a template to generate a 3D model of the protein. There are two chains (A and B) in the 1053-residue cryo-EM structure, with chain A as a blueprint for the model construction. The cryo-electron microscopy structure of the HIV-1 Pol polyprotein [73] provides insight into virion maturation.

3.6. 3D model validation

After getting the sequence from a database, the SAVES v6.0 online server was used to validate the model following the PROCHECK, ERRAT and Veryfy3D servers. PROCHECK is used to develop the Ramachandran plot of the phi/psi distribution in the model and check for non-GLY residues in the disallowed regions [74]. Ramachandran plot indicates a valid model has a minimum coverage of 98.65 % residues in the most favorable areas and only 1.4 % residues in the additional allowed region (Table S2 and Fig. 6). The total number of residues was 90; among them, the number of non-glycine and non-proline residues was 72. Once a 3D model of the target sequence was created, it was also checked using the structure validation servers Verifiy3D and ERRAT. The ERRAT server predicted a total quality factor of 56.94 for the model, and the Verify3D graph indicates that on the 3D-1D scale >0.1, fewer than 80 % of amino acids have achieved and key parameters related to overall quality factors were discovered to range from 98.6 to 99 % [75]. Using energy minimization, the best models were refined, and it was discovered that the models (Energy Minimization score = 8444 kcal/mol) were identical to the validated models. The total outcomes showed that the projected 3-dimensional structure of HXB2 = viral protease proteins is acceptable.

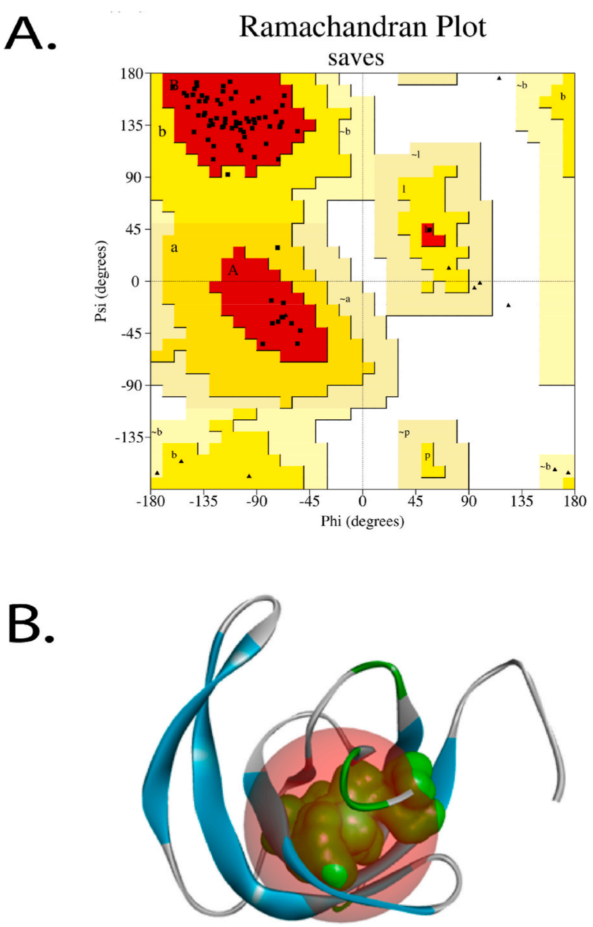

Fig. 6. The PROCHECK via SAVES online server confirms the accuracy and validation of the modeled 3-D structure shown in Ramachandran plot (A) Here, the Ramachandran plot indicates a valid model had a minimum coverage of 98.65 % residues in the most favorable areas and only 1.4 % residues in the additional allowed region. The putative protein's active region, indicated by a subscript (B), is generated through Biovia Discovery Studio Visualizer. Here, the crimson sphere represents the putative protein's active site (AAB33144.1).

3.7. *Active site prediction*

CASTp online server has been demonstrated to be a useful tool for a wide range of studies, including the investigation of signaling receptors, the discovery of cancer therapeutics, the understanding of drug action mechanisms, and the study of immune disorder diseases [47]. The anticipated active site of the hypothetical protein (AAB13344.1), which consists of 17 amino acids, was found to have a surface area of 214.675 A2 and a volume of 205.981 A3 (Fig. 6). The residues Asp 25, Gly 27, ALA 28, Asp 29, Asp 30, Val 32, Ile 47, Val 48, Gly 49, Glu 50, Phe 53, Ile 54, Thr 80, Pro 81, Val 82, and Ile 84 were found to be involved in the active site confirmed by CASTp online server. *Schimer, Jiri* et al. *(2012) conducted* a computational structure based on novel protease inhibitors of HIV by using benzodiazepine analogue and concluded that benzodiazepine can bind to the protease active site of Ala28/Ala28 amino acid position [76].

3.8. *Binding affinity analysis*

Additionally, we have used a computational strategy for repurposing existing drugs in the research to validate the hypothetical protein (AAB13344) model of HIV type-1. We have collected five drug compound which is FDA approved and potential inhibitors of HIV protease. To validate our protein with specific drug action, PyRx, a virtual tool, evaluated the five compounds based on their binding affinity with HXB2=HIV type-1 protease (PDB ID = 4EYR) and a hypothetical protein (accession no: AAB33144.1 and PDB ID: 7SJX).

We have chosen the main viral protease (PDB ID: 4EYR) the because it is already inhibited by several anti-HIV FDA-approved drugs [3], and our identified hypothetical protein (accession no: AAB33144.1) is similar to main viral protease. To effectively treat HIV infection and AIDS, protease inhibitors (PIs) are known to be utilized extensively. Then, we validated the 3D structure of the hypothetical protein by comparing the binding energy of five FDA drugs against the main viral protease. Molecular docking is a

Table 6

List of the ligands, their binding affinity and amino acids that participate in the interaction.

Name of potential target (PDB ID)	Ligands (PubChem ID)	Binding affinity (Kcal/mol)	Amino Acid Involved Interaction	
			Hydrogen Bonds Interaction	Hydrophobic Bonds Interaction
Hypothetical protein				
HXB2 (AAB33144.1)	Indinavir (5362440)	−7.9	VAL 48 and ASP 25	LEU 23, ALA 28, ILE 47, ILE54 and VAL 82
HXB2 (AAB33144.1)	Lopinavir (92727)	−7.8	ASP 25, ASP 29, and ASP 30, GLY 49, GLY 27 and ALA28	ARG 8, LEU 23, VAL 48, ILE 50, ILE 54, PRO 81, and VAL 82
HXB2 (AAB33144.1)	Nelfinavir (64143)	−8.2	ASP 25, GLY 52, THR 80 and PRO 81	LEU 23, ALA 28, ILE 47, ILE 54, and VAL 82
HXB2 (AAB33144.1)	Saquinavir (441243)	−7.6	ASP 25	LEU 23, ILE 50 and ILE 54
HXB2 (AAB33144.1)	Tipranavir (54682461)	−7.6	ASP 25	LEU 23, ALA 28, ILE 50, ILE 54 and PRO 81
Main Protease protein of HIV				
Protease (4EYR)	Indinavir (5362440)	−7.4	ASN 25, GLY 52 and THR 80	ALA 28, and ILE 47
Protease (4EYR)	Lopinavir (92727)	−7.7	ASN 25, and GLY 27	ALA 28, VAL 32, PRO 81, and VAL 84
Protease (4EYR)	Nelfinavir (64143)	−7.3	ASN 25, THR 80	ALA 28, ASP 30,VAL 32, PRO 81, and VAL 84
Protease (4EYR)	Saquinavir (441243)	−8.1	ARG 8, ASN 25	ALA 28, ASP 30, ILE 47, PRO 81 and VAL 84
Protease (4EYR)	Tipranavir (54682461)	−7.2	ASN 25, THR 80	ALA 28, VAL 32, PRO 81 and VAL 84

computational technique used to find small molecules that attach specifically to a given protein (receptor). The binding energy of a molecule indicates how tightly and strongly it will attach to its target protein. As a potential drug, a compound with a lower binding energy is favored [77].

Docking results of HIV-1 protease (AAB33144.1) with reported five compounds, Indinavir (CID: 5362440), Lopinavir (CID: 92727), Nelfinavir (CID: 64143), Saquinavir (CID: 441243), and Tipranavir (CID: 54682461) were presented in Table 6. Among them, the best-docked score for nelfinavir is (−8.2 kcal/mol) and indinavir (−7.9 kcal/mol). Besides, VAL 48, ASP 25, ILE 47, ASP 29, GLY 49, GLY 27, ALA 28, ILE 50, GLY 52, THR 80, PRO 81, VAL 82, ILE 54, VAL 32 residues were located in the binding area of HIV type-1 protease (HXB2) hypothetical protein (AAB33144.1).

At the active region of a hypothetical protein, indinavir showed a binding affinity of −7.9 kcal/mol for the amino acid residues VAL 48, ASP 25, and ILE 47, etc. On the other hand, for the original protease protein (PDB ID: 4EYR), indinavir showed (−7.4 kcal/mol) binding affinity with ASN 25, GLY 52 and THR 80, ALA 28, and ILE 47 amino acids that make up the binding pocket and take a role in the interaction (Fig. 7A). A synthetic peptidomimetic competitive inhibitor of the HIV aspartyl protease, indinavir prevents the viral particles from completing the process of maturation into infectious virions by cleaving the gag and pol gene products into their functional components [78].

Lopinavir showed binding affinity (−7.8 kcal/mol) with ASP 29, GLY 49, GLY 27, ALA 28, and ILE 50 amino acid residues, which interacted at the binding pocket of our hypothetical protein. Lopinavir also expressed (−7.7 kcal/mol) binding affinity with ASN 25, GLY 27, ALA 28, VAL 32, PRO 81, and VAL 84 amino acid residues, which participate in the interaction at the binding pocket of major protease protein (Fig. 7B). Both HIV-1 and HIV-2 proteases are highly inhibited by lopinavir and indinavir. It forms a stable complex with the protease enzyme, blocking its ability to cleave viral polyproteins into functional proteins by binding to its active site. The resulting viral particles are too immature to infect new cells and are harmless [79,80].

For hypothetical protein, nelfinavir showed (−8.2 kcal/mol) exhibited GLY 52, THR 80, PRO 81, and VAL 82 interacting amino acid residues in a protein's active domain. Nelfinavir, on the other hand, exhibited binding affinity with ASN 25, THR 80, ALA 28, ASP 30, VAL 32, PRO 81, and VAL 84 residues in the major protein Protease (PDB ID: 4EYR) (−7.3 kcal/mol) in Fig. 7C. The HIV-1 protease enzyme, which breaks viral polyproteins into their functional subunits, is inhibited by nelfinavir's binding to it. This prevents the development of infectious viral particles and stops HIV-1 replication by stopping the virus' maturation [81].

Saquinavir showed a binding affinity with ASP25, LEU23, ILE50, and ILE54 (−7.6 kcal/mol) for a putative protein. Additionally, Saquinavir showed binding affinity (−8.1 kcal/mol) with the ARG 8, ASN 25, ALA 28, ASP 30, ILE 47, PRO 81, and VAL 84 amino acid residues of the main protein protease (4EYR) in Fig. 7D. The kinetics of Saquinavir bound to HIV-1 protease and discovered that it works as a competitive inhibitor of the enzyme by tightly binding to the active site, obstructing substrate access, and eventually halting the production of contagious viral particles [82].

Tipranavir showed (−7.6 kcal/mol) binding affinity with ASP 25, LEU 23, ALA 28, ILE 50, ILE 54, and PRO 81 amino acid residues of our hypothetical protein. In addition, tipranavir also exhibited (−7.2 kcal/mol) binding affinity with ASN 25, THR 80, ALA 28, VAL 32, PRO 81, and VAL 84 residues in a protein's binding pocket that are involved in the relationship (Fig. 7E). Tipranavir works by attaching to the HIV-1 protease enzyme's active region and preventing it from doing its job. In particular, tipranavir binds to the protease enzyme to create a complex that stops it from dissecting the viral polyprotein into its constituent parts, preventing the spread of new infectious virus particles [83].

4. Conclusions

Due to our poor understanding of their structures and biological activities, several HIV virus-produced proteins still fall under the category of "hypothetical proteins." This study determined the structure and biological function of the hypothetical protein (accession no: AAB13344.1) of HIV type 1 (HIV-1) by employing the series of *in-silico* approach and finally validated 3D structure of the protein through PyRx-based molecular docking.

The sub-cellular location of the host cell membrane, the presence of protease enzymes with specific aspartyl protease domains, the presence of cupredoxin folding pattern, and the protease activity nature of the identified protein, among other characteristics highlight the importance of this protein in HIV pathogenesis. On the other hand, drug resistance in HIV protease is a subtly altered balance of recognition events between the relative affinity of the HIV protease to bind inhibitors. These properties of the hypothesized protein also will improve understanding of HIV-1 pathogenesis and its drug design. To hypothetical protein (AAB13344.1) 3D structure validation, we have worked on five effective FDA-approved inhibitor compounds (tipranavir, indinavir, Saquinavir, nelfinavir, and lopinavir) against HIV-1 main proteases and a hypothetical protein (AAB13344.1). Among all the reported protease inhibitors, nelfinavir showed (−8.2 kcal/mol) the best binding interaction score compared to the main protease protein. The characteristics of the structure and function of hypothetical proteins discussed in this study have the potential to offer significant insights into disease mechanisms and facilitate drug discovery efforts against HIV-1 in the near future. Therefore, extensive *in vitro* research is required to test and verify the possibilities and discover the role of proteins in health sciences.

Data availability statement

All data included in the study will be available for everyone as per journal policy.

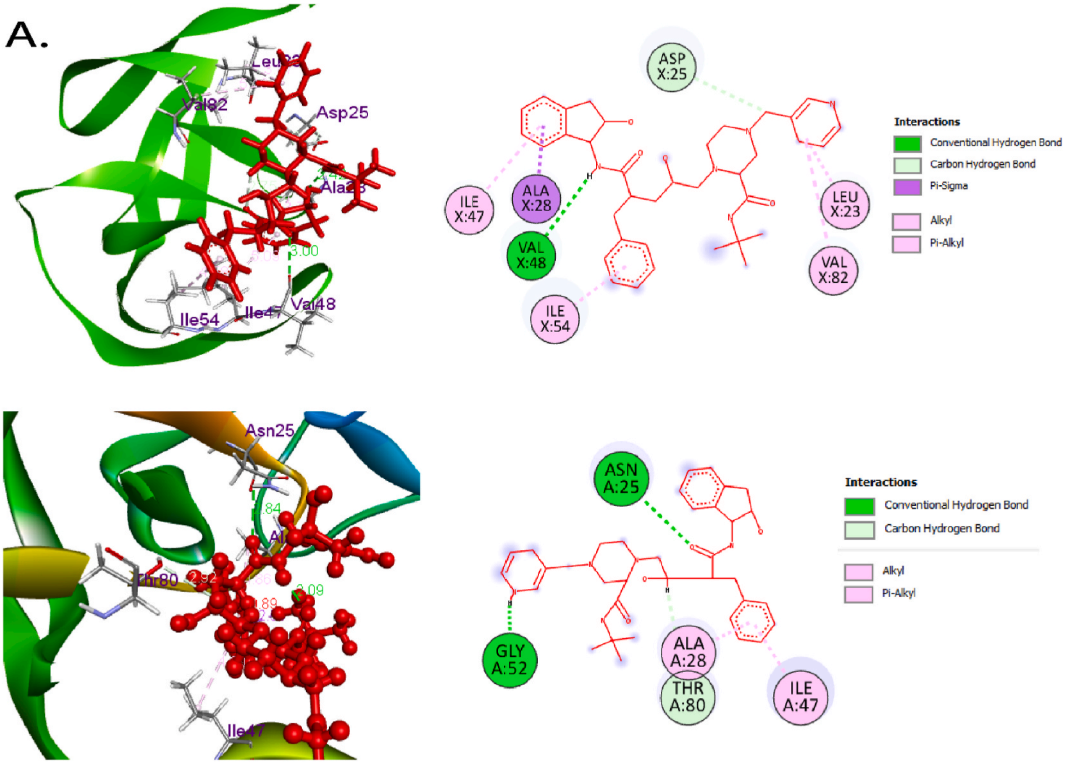

Fig. 7A. Binding affinity analysis of Indinavir drug compound against hypothetical and protease proteins.

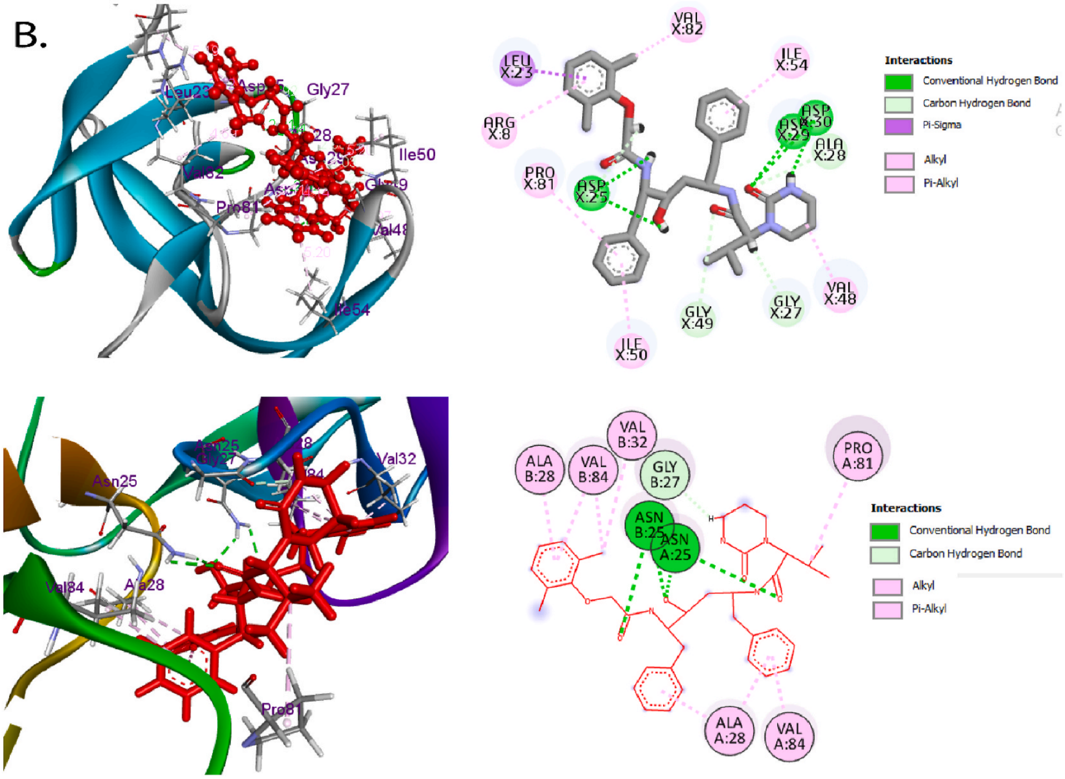

Fig. 7B. Binding affinity analysis of Lopinavir drug compound against hypothetical and protease proteins.

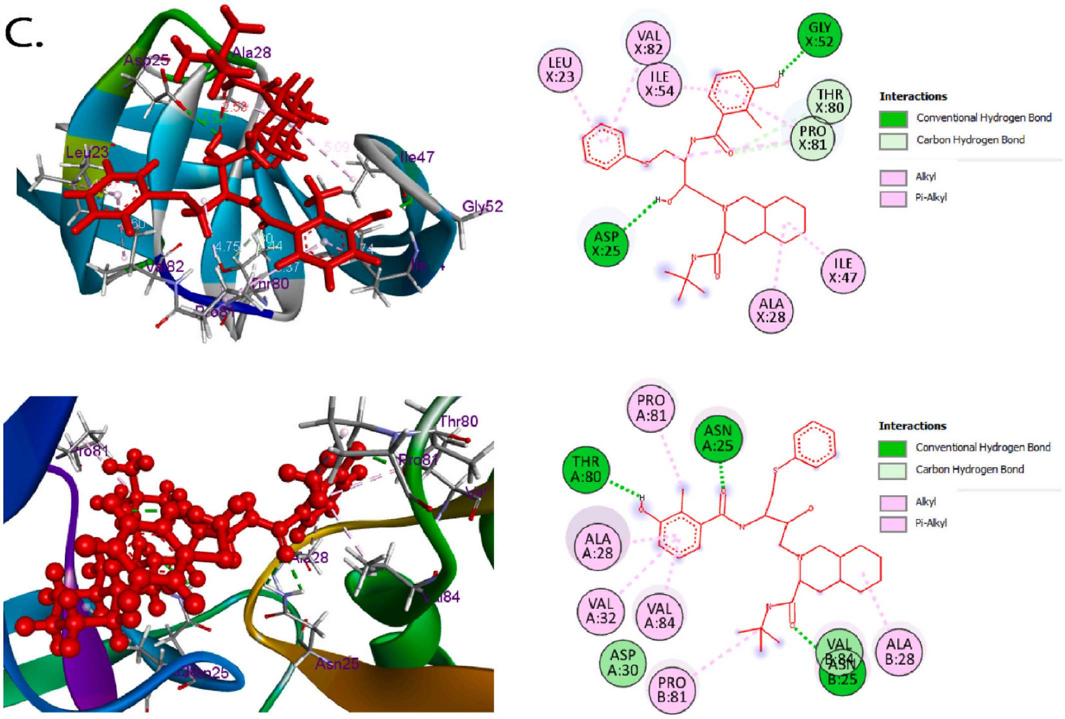

Fig. 7C. Binding affinity analysis of Nelfinavir drug compound against hypothetical and protease proteins.

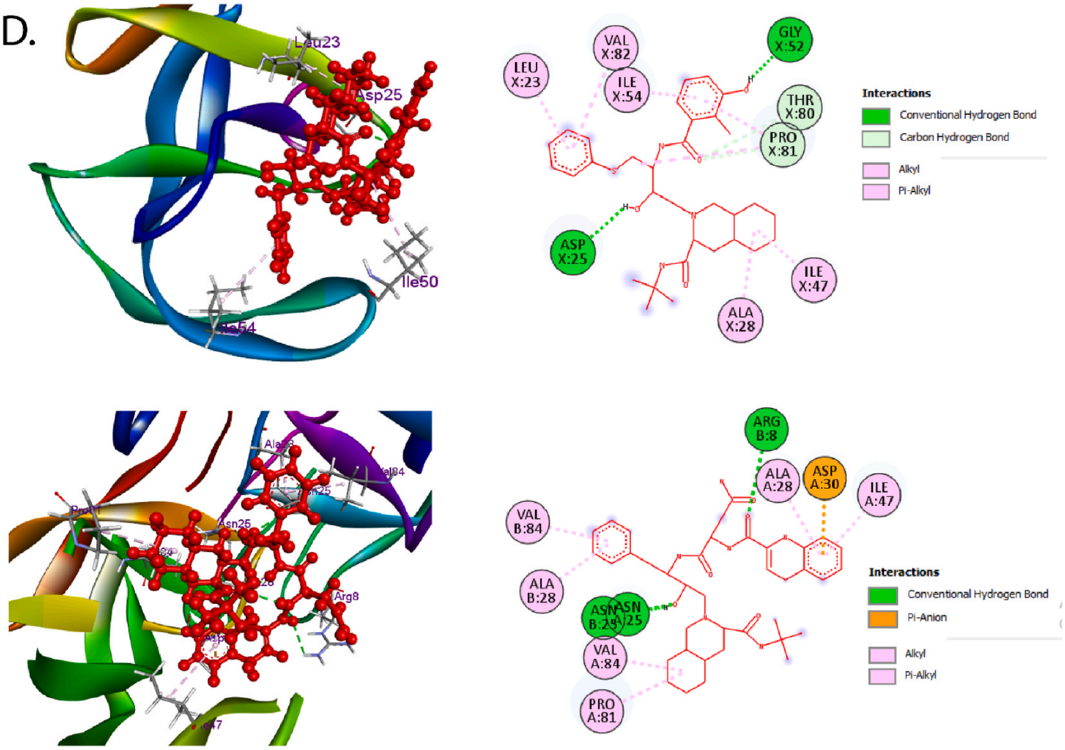

Fig. 7D. Binding affinity analysis of Saquinavir drug compound against hypothetical and protease proteins.

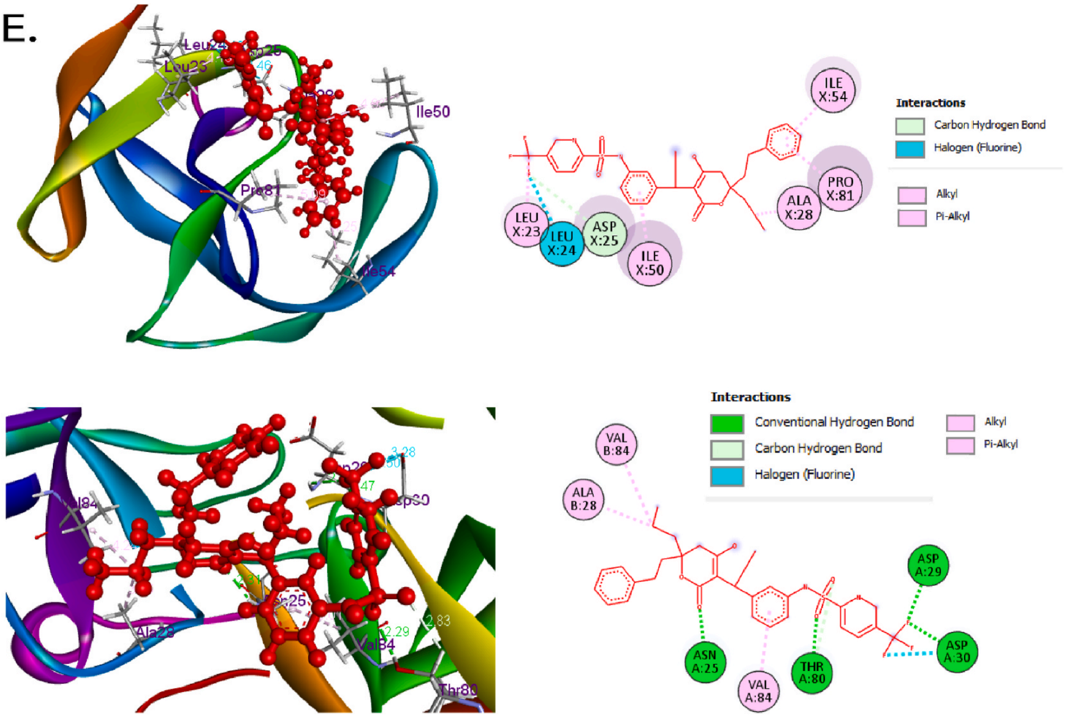

Fig. 7E. Binding affinity analysis of Tipranavir drug compound against hypothetical and protease proteins.

Funding information

None/No funding source.

CRediT authorship contribution statement

Md Imran Hossain: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft. **Anika Tabassum Asha:** Conceptualization, Data curation, Methodology, Writing – original draft. **Md Arju Hossain:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft. **Shahin Mahmud:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Kamal Chowdhury:** Formal analysis, Methodology, Validation, Writing – review & editing. **Ramisa Binti Mohiuddin:** Formal analysis, Methodology, Validation, Writing – original draft. **Nazneen Nahar:** Formal analysis, Methodology. **Saborni Sarker:** Formal analysis, Methodology. **Suhaimi Napis:** Formal analysis, Investigation, Methodology, Writing – review & editing. **Md Sanower Hossain:** Data curation, Formal analysis, Methodology, Writing – review & editing. **A.K.M. Mohiuddin:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

All the authors declared no conflict of interest.

Acknowledgement

None.

Abbreviations

HIV	Human Immunodeficiency Virus
AIDS	Acquired Immunodeficiency Syndrome
FDA	Food & Drug Administration
RTVL:	Reverse Transcriptase Visible Light
ART	Antiretroviral Treatment
NCBI	National Center for Biotechnology Information

FASTA	First Alignment
Blast$_p$:	Basic Local Alignment Search Tool
ITOL:	Interactive Tree of Life
GRAVY	Grand Average of hydropathicity
ExPASy	Expert Protein Analysis Tools
PI	Isoelectric Point
Virus mPloc	Virus Multi-Location Predictor
CDD	Conserved Domain Database
PSIPRED	PSI-blast based secondary structure PREDiction
RMSD	Root Mean Square Deviation

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23183.

References

[1] F. Tiruneh, L. Chewaka, D. Abdissa, Statistical Joint Modeling for Predicting the Association of CD4 Measurement and Time to Death of People Living with HIV Who Enrolled in ART, Southwest Ethiopia, HIV/AIDS-Research Palliat, Care, 2021, pp. 73–79.

[2] S.B. Lloyd, S.J. Kent, W.R. Winnall, The high cost of fidelity, AIDS Res. Hum. Retrovir. 30 (1) (2014) 8–16.

[3] Z. Lv, Y. Chu, Y. Wang, HIV Protease Inhibitors: a Review of Molecular Selectivity and Toxicity, HIV/AIDS-Research Palliat. care, 2015, pp. 95–104.

[4] G.A.C. Blood, Human immunodeficiency virus (HIV), Transfus. Med. Hemotherapy 43 (3) (2016) 203.

[5] A.J. Smith, Q. Li, S.W. Wietgrefe, T.W. Schacker, C.S. Reilly, A.T. Haase, Host genes associated with HIV-1 replication in lymphatic tissue, J. Immunol. 185 (9) (2010) 5417–5424.

[6] C.T. Lemke, et al., Distinct effects of two HIV-1 capsid assembly inhibitor families that bind the same site within the N-terminal domain of the viral CA protein, J. Virol. 86 (12) (2012) 6643–6655.

[7] T.B.H. Geijtenbeek, Y. Van Kooyk, DC-SIGN: a novel HIV receptor on DCs that mediates HIV-1 transmission, Dendritic. Cells Virus Infect. (2003) 31–54.

[8] M. Alizon, S. Wain-Hobson, L. Montagnier, P. Sonigo, Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients, Cell 46 (1) (1986) 63–74.

[9] F. Clavel, M. Guyader, D. Guétard, M. Sallé, L. Montagnier, M. Alizon, Molecular cloning and polymorphism of the human immune deficiency virus type 2, Nature 324 (6098) (1986) 691–695.

[10] L. Su, et al., Identification of HIV-1 determinants for replicationin vivo, Virology 227 (1) (1997) 45–52.

[11] J. Konvalinka, H.-G. Kräusslich, B. Müller, Retroviral proteases and their roles in virion maturation, Virology 479 (2015) 403–417.

[12] S.C. Pettit, J.N. Lindquist, A.H. Kaplan, R. Swanstrom, Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates, Retrovirology 2 (2005) 1–6.

[13] L. Deshmukh, V. Tugarinov, J.M. Louis, G.M. Clore, Binding kinetics and substrate selectivity in HIV-1 protease– Gag interactions probed at atomic resolution by chemical exchange NMR, Proc. Natl. Acad. Sci. USA 114 (46) (2017) E9855–E9862.

[14] I.T. Weber, Y.-F. Wang, R.W. Harrison, HIV protease: historical perspective and current research, Viruses 13 (5) (2021) 839.

[15] B.L. Jilek, et al., A quantitative basis for antiretroviral therapy for HIV-1 infection, Nat. Med. 18 (3) (2012) 446–451.

[16] D.S. Clutter, M.R. Jordan, S. Bertagnolio, R.W. Shafer, HIV-1 drug resistance and resistance testing, Infect. Genet. Evol. 46 (2016) 292–307.

[17] S.A. Rabi, et al., Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance, J. Clin. Invest. 123 (9) (2013) 3848–3860.

[18] J. Jiang, C. Aiken, Maturation-dependent human immunodeficiency virus type 1 particle fusion requires a carboxyl-terminal region of the gp41 cytoplasmic tail, J. Virol. 81 (18) (2007) 9999–10008.

[19] J.M. Whitcomb, et al., Development and characterization of a novel single-cycle recombinant-virus assay to determine human immunodeficiency virus type 1 coreceptor tropism, Antimicrob. Agents Chemother. 51 (2) (2007) 566–575.

[20] D.R. Goldsmith, C.M. Perry, "Atazanavir," Drugs 63 (2003) 1679–1693.

[21] A.K. Ghosh, H.L. Osswald, G. Prato, Recent progress in the development of HIV-1 protease inhibitors for the treatment of HIV/AIDS, J. Med. Chem. 59 (11) (2016) 5172–5208.

[22] S. Sivashankari, P. Shanmughavel, Functional annotation of hypothetical proteins–A review, Bioinformation 1 (8) (2006) 335.

[23] K.D. Mack, et al., HIV insertions within and proximal to host cell genes are a common finding in tissues containing high levels of HIV DNA and macrophage-associated p24 antigen expression, JAIDS J. Acquir. Immune Defic. Syndr. 33 (3) (2003) 308–320.

[24] H.A. Walters, et al., Hypothetical proteins play a role in stage conversion, virulence, and the stress response in the Entamoeba species, Exp. Parasitol. 243 (2022), 108410.

[25] B. Ghebremedhin, Human adenovirus: viral pathogen with increasing importance, Eur. J. Microbiol. Immunol. 4 (1) (2014) 26–33.

[26] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L. Madden, NCBI BLAST: a better web interface, Nucleic Acids Res. 36 (suppl_2) (2008). W5–W9.

[27] F. Sievers, D.G. Higgins, Clustal omega, Curr. Protoc. Bioinforma. 48 (1) (2014) 3–13.

[28] R.C. Edgar, S. Batzoglou, Multiple sequence alignment, Curr. Opin. Struct. Biol. 16 (3) (2006) 368–373, https://doi.org/10.1016/j.sbi.2006.04.004.

[29] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, Mega X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (6) (2018) 1547.

[30] I. Letunic, P. Bork, Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation, Nucleic Acids Res. 49 (W1) (2021), https://doi.org/10.1093/nar/gkab301.

[31] E. Gasteiger, et al., " Protein Identif. Anal. Tools ExPASy Server, [Google Sch, 2005. "Humana Press.

[32] V.K. Garg, et al., MFPPI–multi FASTA ProtParam interface, Bioinformation 12 (2) (2016) 74.

[33] C. Yu, Y. Chen, C. Lu, J. Hwang, Prediction of protein subcellular localization, Proteins Struct. Funct. Bioinforma. 64 (3) (2006) 643–651.

[34] H.-B. Shen, K.-C. Chou, Virus-mPLoc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites, J. Biomol. Struct. Dyn. 28 (2) (2010) 175–186.

[35] A. Marchler-Bauer, et al., CDD: NCBI's conserved domain database, Nucleic Acids Res. 43 (D1) (2015). D222–D226.

[36] T.L. Bailey, N. Williams, C. Misleh, W.W. Li, MEME: discovering and analyzing DNA and protein sequence motifs, Nucleic Acids Res. 34 (suppl_2) (2006) W369–W373.

[37] J. Mistry, et al., Pfam: the protein families database in 2021, Nucleic Acids Res. 49 (2021), https://doi.org/10.1093/nar/gkaa913. D1.

[38] J. Ludwiczak, A. Winski, K. Szczepaniak, V. Alva, S. Dunin-Horkawicz, DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences, Bioinformatics 35 (16) (Aug. 2019) 2790–2795, https://doi.org/10.1093/bioinformatics/bty1062.

[39] H.-B. Shen, K.-C. Chou, Predicting protein fold pattern with functional domain and sequential evolution information, J. Theor. Biol. 256 (3) (Feb. 2009) 441–446, https://doi.org/10.1016/j.jtbi.2008.10.007.

[40] D.W.A. Buchan, D.T. Jones, The PSIPRED protein analysis workbench: 20 years on, Nucleic Acids Res. 47 (W1) (2019). W402–W407.

[41] C. Geourjon, G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Bioinformatics 11 (6) (1995) 681–684.

[42] J. Söding, A. Biegert, A.N. Lupas, The HHpred interactive server for protein homology detection and structure prediction, Nucleic Acids Res. 33 (suppl_2) (2005) W244–W248.

[43] H. Land, M.S. Humble, YASARA: a tool to obtain structural guidance in biocatalytic investigations, Protein Eng. methods Protoc. (2018) 43–67.

[44] R.A. Laskowski, M.W. MacArthur, J.M. Thornton, PROCHECK: Validation of Protein-Structure Coordinates, 2006.

[45] M. von Grotthuss, J. Pas, L. Wyrwicz, K. Ginalski, L. Rychlewski, Application of 3D-jury, GRDB, and Verify3D in fold recognition, Proteins Struct. Funct. Bioinforma. 53 (S6) (2003) 418–423.

[46] M. Lengths, M. Angles, Limitations of structure evaluation tools errat, Quick Guidel. Comput. Drug Des. 16 (2018) 75.

[47] W. Tian, C. Chen, X. Lei, J. Zhao, J. Liang, CASTp 3.0: computed atlas of surface topography of proteins, Nucleic Acids Res. 46 (W1) (Jul. 2018) W363–W367, https://doi.org/10.1093/nar/gky473.

[48] P.W. Rose, et al., The RCSB Protein Data Bank: redesigned web site and web services, Nucleic Acids Res. 39 (suppl_1) (2010) D392–D401.

[49] Z. Liu, et al., Insights into the mechanism of drug resistance: X-ray structure analysis of multi-drug resistant HIV-1 protease ritonavir complex, Biochem. Biophys. Res. Commun. 431 (2) (Feb. 2013) 232–238, https://doi.org/10.1016/j.bbrc.2012.12.127.

[50] M. Pereira, N. Vale, Saquinavir: from HIV to COVID-19 and cancer treatment, Biomolecules 12 (7) (2022) 944.

[51] C.M. Perry, J.E. Frampton, P.L. McCormack, M.A.A. Siddiqui, R.S. Cvetković, Nelfinavir: a review of its use in the management of HIV infection, Drugs 65 (2005) 2209–2244.

[52] A. Chandwani, J. Shuter, Lopinavir/ritonavir in the treatment of HIV-1 infection: a review, Ther. Clin. Risk Manag. 4 (5) (2008) 1023–1033.

[53] S. Kim, et al., PubChem substance and compound databases, Nucleic Acids Res. 44 (D1) (2016) D1202–D1213.

[54] F. Oellien, M.C. Nicklaus, Online SMILES translator and structure file generator, Natl. Cancer Inst. 29 (2004) 97–101.

[55] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx, Methods Mol. Biol. (2015), https://doi.org/10.1007/978-1-4939-2269-7_19.

[56] B.L. Jejurikar, S.H. Rohane, Drug Designing in Discovery Studio, 2021.

[57] A.B. Abecasis, M. Pingarilho, A.-M. Vandamme, Phylogenetic analysis as a forensic tool in HIV transmission investigations, Aids 32 (5) (2018) 543–554.

[58] K. Bozek, T. Lengauer, S. Sierra, R. Kaiser, F.S. Domingues, Analysis of physicochemical and structural properties determining HIV-1 coreceptor usage, PLoS Comput. Biol. 9 (3) (2013), e1002977.

[59] C.B. Wilen, J.C. Tilton, R.W. Doms, HIV: cell binding and entry, Cold Spring Harb. Perspect. Med. 2 (8) (2012).

[60] M. Mahanti, S. Bhakat, U.J. Nilsson, P. Söderhjelm, Flap dynamics in aspartic proteases: a computational perspective, Chem. Biol. Drug Des. 88 (2) (2016) 159–177. Wiley Online Library.

[61] M.N.L. Nalam, A. Peeters, T.H.M. Jonckers, I. Dierynck, C.A. Schiffer, Crystal structure of lysine sulfonamide inhibitor reveals the displacement of the conserved flap water molecule in human immunodeficiency virus type 1 protease, J. Virol. 81 (17) (2007) 9512–9518.

[62] V. Blikstad, F. Benachenhou, G.O. Sperber, J. Blomberg, Endogenous retroviruses: evolution of human endogenous retroviral sequences: a conceptual account, Cell. Mol. Life Sci. 65 (2008) 3348–3365.

[63] L. Bénit, P. Dessen, T. Heidmann, Identification, phylogeny, and evolution of retroviral elements based on their envelope genes, J. Virol. 75 (23) (2001) 11709–11719.

[64] A. Muhlrad, D. Pavlov, Y.M. Peyser, E. Reisler, Inorganic phosphate regulates the binding of cofilin to actin filaments, FEBS J. 273 (7) (2006) 1488–1496.

[65] P. Gehlot, S. Kumar, V.K. Vyas, B.S. Choudhary, M. Sharma, R. Malik, Guanidine-based β amyloid precursor protein cleavage enzyme 1 (BACE-1) inhibitors for the Alzheimer's disease (AD): a Review, Bioorg. Med. Chem. (2022), 117047.

[66] J. Fanfrlik, A.K. Bronowska, J. Rezac, O. Přenosil, J. Konvalinka, P. Hobza, A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands, J. Phys. Chem. B 114 (39) (2010) 12666–12678.

[67] L.L. Palese, Conformations of the HIV-1 protease: a crystal structure data set analysis, Biochim. Biophys. Acta, Proteins Proteomics 1865 (11) (2017) 1416–1422.

[68] M. Nijhuis, N.M. van Maarseveen, J. Verheyen, C.A.B. Boucher, Novel mechanisms of HIV protease inhibitor resistance, Curr. Opin. HIV AIDS 3 (6) (2008) 627–632.

[69] A. Chaudhari, et al., Azurin, Plasmodium falciparum malaria and HIV/AIDS: inhibition of parasitic and viral growth by Azurin, Cell Cycle 5 (15) (2006) 1642–1648.

[70] M. Cano-Muñoz, S. Cesaro, B. Morel, J. Lucas, C. Moog, F. Conejero-Lara, Extremely thermostabilizing core mutations in coiled-coil mimetic proteins of HIV-1 gp41 produce diverse effects on target binding but do not affect their inhibitory activity, Biomolecules 11 (4) (2021) 566.

[71] D. Yu, et al., Molecular mechanism of HIV-1 resistance to sifuvirtide, a clinical trial–approved membrane fusion inhibitor, J. Biol. Chem. 293 (33) (2018) 12703–12718.

[72] C.N. Cavasotto, S.S. Phatak, Homology modeling in drug discovery: current trends and applications, Drug Discov. Today 14 (13–14) (2009) 676–683.

[73] J.J.E.K. Harrison, et al., Cryo-EM structure of the HIV-1 Pol polyprotein provides insights into virion maturation, Sci. Adv. 8 (27) (Jul. 2022) eabn9874, https://doi.org/10.1126/sciadv.abn9874.

[74] R.A. Laskowski, N. Furnham, J.M. Thornton, The Ramachandran plot and protein structure validation, in: Biomolecular Forms and Functions: a Celebration of 50 Years of the Ramachandran Map, World Scientific, 2013, pp. 62–75.

[75] N.M. Atre, H.H. Pilley, S.R. Nagmote, A. Khan, V. Changole, G.S. Deshpande, Comparative protein structure analysis of HXB2= viral protease from HIV-1 genome, J. Comput. Intell. Bioinforma. 4 (1) (2011) 101–109.

[76] J. Schimer, et al., Structure-aided design of novel inhibitors of HIV protease based on a benzodiazepine scaffold, J. Med. Chem. 55 (22) (2012) 10130–10135.

[77] J. Fan, A. Fu, L. Zhang, Progress in molecular docking, Quant. Biol. 7 (2) (2019) 83–89.

[78] T.R. Cressey, N. Plipat, F. Fregonese, K. Chokephaibulkit, Indinavir/ritonavir remains an important component of HAART for the treatment of HIV/AIDS, particularly in resource-limited settings, Expert Opin. Drug Metab. Toxicol. 3 (3) (2007) 347–361.

[79] L.P. Greg, P.F. David, "Rituximab," Drugs 63 (8) (2003) 803–843.

[80] R. Sittl, Transdermal Buprenorphine: a Viewpoint by Reinhard Sittl, Springer, 2003.

[81] K. Buriánková, et al., Molecular basis of intrinsic macrolide resistance in the Mycobacterium tuberculosis complex, Antimicrob. Agents Chemother. 48 (1) (2004) 143–150.

[82] C.-H. Shen, Y.-C. Chang, J. Agniswamy, R.W. Harrison, I.T. Weber, Conformational variation of an extreme drug resistant mutant of HIV protease, J. Mol. Graph. Model. 62 (2015) 87–96.

[83] L. Menéndez-Arias, Molecular basis of human immunodeficiency virus type 1 drug resistance: overview and recent developments, Antivir. Res. 98 (1) (2013) 93–120.