

# In-silico Structural Modeling of Human Immunodeficiency Virus Proteins

Amir Elalouf 

Department of Management, Bar-Ilan University, Ramat Gan, Israel.

Biomedical Engineering and  
Computational Biology  
Volume 14: 1–14  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11795972231154402



**ABSTRACT:** Human immunodeficiency virus (HIV) is an infectious virus that depletes the CD4<sup>+</sup> T lymphocytes of the immune system and causes a chronic life-treating disease—acquired immunodeficiency syndrome (AIDS). The HIV genome encodes different structural and accessory proteins involved in viral entry and life cycle. Determining the 3D structure of HIV proteins is essential for new target position finding, structure-based drug designing, and future planning for computational and laboratory experimentations. Hence, the study aims to predict the 3D structures of all the HIV structural and accessory proteins using computational homology modeling to understand better the structural basis of HIV proteins interacting with host cells and viral replication. The sequences of HIV capsid, matrix, nucleocapsid, p6, reverse transcriptase, invertase, protease, gp120, gp41, virus protein r, viral infectivity factor, virus protein unique, RNA splicing regulator, transactivator protein, negative regulating factor, and virus protein x proteins were retrieved from UniProt. The primary and secondary structures of HIV proteins were predicted by ExPasy ProtParam and SOPMA web servers. For the homology modeling, the MODELLER predicted the 3D structures of HIV proteins using templates. Then, the modeled structures were validated by the Ramachandran plot, local and global quality estimation scores, QMEAN scores, and Z-scores. Most of the amino acid residues of HIV proteins were present in the most favored and generously allowed regions in the Ramachandran plots. The local and global quality scores and Z-scores of the HIV proteins confirmed the good quality of modeled structures. The 3D modeled structures of HIV proteins might help further investigate the possible treatment.

**KEYWORDS:** 3D-structure, human immunodeficiency virus, homology modeling, protein structures validation, MODELLER

**RECEIVED:** October 1, 2022. **ACCEPTED:** January 16, 2023.

**TYPE:** Original Research

**FUNDING:** The author received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Amir Elalouf, Department of Management, Bar-Ilan University, Max and Anna, Ramat Gan 5290002, Israel. Email: amir.elalouf@biu.ac.il

## Introduction

Human immunodeficiency virus (HIV) is an infectious virus that depletes the CD4<sup>+</sup> T lymphocytes of the host immune system, which leads to a chronic life-treating disease—acquired immunodeficiency syndrome (AIDS).<sup>1</sup> According to UNAIDS, in 2021, approximately 38 million people were living with HIV, and 650 000 people died from AIDS-related illnesses.<sup>2</sup> CD4<sup>+</sup> receptors of helper T cells such as macrophages, eosinophils, and neutrophils facilitate the HIV binding and entry to the cells.<sup>3</sup> The genome of HIV has 9 genes that encode 15 viral proteins. The polyprotein of HIV includes Gag proteins (capsid, matrix, nucleocapsid, and p6), Pol proteins (reverse transcriptase, invertase, protease), Envelope glycoproteins (Gp120, Gp41), and accessory proteins (virus protein r, viral infectivity factor, virus protein unique, RNA splicing regulator, transactivator protein, negative regulating factor, and virus protein x proteins) cleaved by cellular proteases.<sup>4</sup> When the HIV-1 virus enters a cell, the envelope glycoproteins facilitate the fusion of the virus and cellular membranes.<sup>5</sup>

In HIV, capsid protein (p24) forms the conical capsid, matrix protein (p17) forms the inner membrane layer, nucleocapsid (p7) intricates in the formation of RNA complex, and p6 involves in the release of virus particles.<sup>6</sup> Protease (p10) involves the proteolytic cleavage of precursor protein resulting in structural proteins and viral enzymes. Reverse transcriptase (p51) transcribes the RNA of HIV into DNA. Integrase (p32)

protein integrates the proviral DNA into the host genome. Tat (p14) activates the viral gene transcription. Rev (p19) regulates the export of non-spliced and partially spliced viral mRNA.<sup>7</sup> Nef (p27) influences HIV replication, enhancement of infectivity of viral particles, and downregulation of CD4<sup>+</sup> on target cells. Vif (p23) is critical for infectious virus production.<sup>7</sup> Vpr (p15) interacts with p6, facilitates virus infectivity, and affects the cell cycle. Vpu (p16) controls CD4<sup>+</sup> degradation and modulates intracellular trafficking. Vpx (p15) is involved in the early steps of virus replication of HIV-2.<sup>8</sup> Gp120 (surface glycoprotein) facilitates the attachment of the virus to the target cell. *In vitro* studies confirmed that the recombinant Gp120 interacts directly with CD4<sup>+</sup>/co-receptors and activates the cellular pathways involving actin regulation, proliferation, cell adhesion, and increased chemokines and cytokines.<sup>9</sup> Gp41 (transmembrane protein) intricates the anchorage of gp120, the fusion of viral and cell membranes. Gp120 interferes with CD4<sup>+</sup> costimulatory functions and induces apoptosis in cells.<sup>10</sup> HIV-1 enters the target host cells through gp120 and gp41 trimeric complexes of viral envelope glycoproteins.<sup>11</sup> When gp120 binds to the CD4<sup>+</sup> glycoprotein on target cells, it undergoes conformational changes that allow it to bind to either the CCR5 or CXCR4 chemokine receptors.<sup>12</sup> Gp41 plays a significant role in driving the fusion process by forming a fusogenic 6-helix bundle structure.<sup>13</sup> It has an ectodomain, a membrane-spanning segment, and a cytoplasmic tail.<sup>14</sup> After receptor binding, the



transmembrane envelope glycoprotein gp41 exposes its ectodomain, causing additional conformational changes that result in the fusion of the viral membrane and host cell membrane.<sup>12</sup>

The structural basis of HIV envelope protein interacting with host cells must be understood to identify new drugs that prevent HIV entry.<sup>15</sup> The prediction of reliable 3D protein structures has been successfully achieved by computational methods, allowing scientists to understand the protein's behavior and function, its interactions with its ligands, and the effects of specific insertions, mutations, and deletions on its conformation and function.<sup>16,17</sup> In previous studies, the simulated structures were collected by NMR spectroscopy and X-ray to reveal common structural motifs in V3 loops of HIV Gp120 protein.<sup>18</sup> By comparative modeling and simulated annealing, the amino acid sequence of the V3 loop and its conformation in 3D folds and local geometry has been determined.<sup>18</sup>

The structural characteristics of proteins play a crucial role in elucidating the molecular mechanisms underlying biological processes.<sup>19</sup> Crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy are generally used to develop quality protein structures, but these processes are time-consuming, expensive, and difficult to perform for membrane proteins.<sup>20</sup> An alternative technique to develop protein structures is *in-silico* 3D structure prediction using homology modeling.<sup>20</sup> Homology modeling predicts the 3D structures of the target proteins by using the similarity between the target and template sequences. This computational technique is efficiently helpful in predicting membrane proteins such as HIV gp120 and gp41 that are difficult to crystallize. Further, homology modeling provides a clear understanding to study the receptor-ligand interaction. Single or multiple template modeling offers profound insight into unsolved structures but clinically significant drug-targeting proteins. As a result of homology modeling studies, various problems associated with crystallizing proteins that play essential roles in major disease pathways are often overcome and open the door to *in-silico* models providing more structural insight when experimental structures are lacking.<sup>21,22</sup> Different studies have reported 3D-modeled structures of different proteins.<sup>23-26</sup>

Nevertheless, to our knowledge, no study has previously studied the 3D-modeled structures of all the HIV structural and accessory proteins. Yet, homology modeling is done by different online and offline servers and software. For instance, MODELLER is well known for its comparative tertiary structure prediction capabilities.<sup>27</sup> PROCHECK, Quantitative Model Energy Analysis (QMEAN), and Protein Structure Analysis (ProSA) servers are standard web servers for validating the modeled protein 3D structures.<sup>28,29</sup> This study aims to use a computational method to predict the 3D structures of all HIV structural and accessory proteins to gain a better understanding of how HIV proteins interact with host cells and replicate. Further, different online tools validated the 3D modeled structures of HIV proteins to build other plots and graphs.

## Methods

### *Proteins sequences*

The RNA-based HIV-1 genome comprises all the information of 16 proteins required for the replication and structural assembly of the new virions. For example, the *gag* gene of the HIV genome transcribes 4 structural proteins, that is, capsid, matrix, nucleocapsid, and p6. In comparison, the *pol* gene transcribes proteins such as reverse transcriptase, invertase, and protease necessary for the replication of HIV. Further, surface glycoproteins gp120 and gp41 are transcribed from the *env* gene. All the structural and accessory proteins for HIV were listed, and their sequences were downloaded from UniProt (<https://www.uniprot.org/>) in FASTA format.

### *Primary structure prediction*

The primary structures of all the HIV structural and accessory proteins were predicted in the ExPasy ProtParam web server (<https://web.expasy.org/protparam/>).<sup>30</sup> In the primary structure analysis, different physicochemical characteristics such as aliphatic index (AI), base pairs length, extinction coefficient (Ec), grand average of hydropathy (GRAVY), instability index (II), molecular weight, theoretical isoelectric point (pI), and the total number of positive (+R) and negative (-R) residues were calculated.

### *Secondary structure prediction*

The secondary structures of the HIV structural and accessory proteins were predicted in an online server SOPMA ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_sopma.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)). All the parameters of SOPMA were set by default, such as the numbers of conformational states to 4 (helix, sheet, turn, and coil), similarity threshold to 8, and window width to 17.<sup>31</sup>

### *Homology modeling*

HHpred online server (<https://toolkit.tuebingen.mpg.de/tools/hhpred>) was used for the homology modeling of all the targeted HIV proteins. HHpred used MODELLER after PIR alignment of the targeted sequence with the template(s) having maximum sequence similarity and identity.<sup>27,32</sup> The MODELLER predicted the structure of the query sequence of the protein(s) with homologous protein(s) by using the far more conserved protein structure approach. Further, advanced structure modeling and loop refinement were performed to establish a 3D model of the protein structures. All the modeled structures were visualized in Discovery Studio Visualizer 2021.

### *3D structure validation of modeled proteins*

All the modeled structures of HIV proteins were validated using methods such as the Ramachandran plot, local and

global quality estimation scores, QMEAN scores, and Z-scores. PROCHECK online server (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) was used to plot the Ramachandran plot for the conformational, stereochemical, and structural quality validation of the 3D modeled HIV structural and accessory proteins.<sup>33</sup> In the Ramachandran plot, about 90% of the amino acid residues in the most favored regions confirm the good quality of the 3D modeled structures. Similarly, for the estimation of both global (ie, for the entire structure) and local (ie, per residue) errors, the QMEAN server (<https://swissmodel.expasy.org/qmean/>) was used to calculate the composite scoring function of 3D modeled structures of HIV proteins.<sup>34</sup> The overall quality of an entire model is calculated by a global score.<sup>35</sup> Based on the alignment and template used in constructing the model, the Global Model Quality Estimation can be expressed as a number between 0 and 1, which reflects the expected model accuracy. The denominator of this number is the coverage of the target sequence. Reliability increases with higher numbers.<sup>36–38</sup> For the local quality estimation of each residue, the model structure on the x-axis compares the expected similarity to the native structures on the y-axis. Residues that show a local quality score above 0.6 are considered good.<sup>20</sup> The QMEAN score above  $-4.0$  confirms the good quality of the predicted 3D structures. Further, the ProSA tool (<https://prosa.services.came.sbg.ac.at/prosa.php>) evaluated the overall quality of each modeled structure of the proteins with a Z-score. The Z-score compared the quality of the 3D modeled proteins with all the experimental NMR and X-ray crystallographic protein structures.<sup>28</sup> The QMEAN Z-score estimates the degree of nativeness, such as inter-atomic packing, back-bone geometry, and unexpected solvent accessibility of the modeled structure with the quality of experimental structures for native proteins of similar size.<sup>28,29</sup>

## Results

### *Protein sequences*

The sequences of all the HIV structural (capsid, matrix, nucleocapsid, p6, gp120, gp41, reverse transcriptase, integrase, and protease) and accessory (virus protein r, viral infectivity factor, virus protein unique, RNA splicing regulator, transactivator protein, negative regulating factor, and virus protein x) proteins were retrieved from the UniProt database. Table 1 represents all the genes, proteins, their UniProtKB ID, and amino acid sequences of all the structural and accessory proteins of HIV.

### *Primary structures of proteins*

The ProtParam tool of the ExPasy server predicted different physicochemical characteristics of HIV structural and accessory proteins. Table 2 represents the results of different physicochemical parameters of HIV structural and accessory proteins.

### *Secondary structure prediction*

Table 3 represents different values of the predicted parameters necessary for the secondary structures of HIV structural and accessory proteins using the SOPMA online server. All the secondary structures of HIV proteins showed alpha helix regions (except transactivator protein), extended strands, beta turns, and random coil. However, no HIV proteins confirmed the presence of 310 helices, pi helix, beta bridge, bend region, ambiguous states, and other states regions in their secondary structures.

### *Homology modeling*

The templates used for the structure prediction of HIV proteins have maximum sequence similarity and identity obtained after using the BLAST program on Protein Data Bank. Table 4 shows the PDB IDs of the targeted templates of the respective HIV proteins with maximum sequence similarity and identity. For some proteins like HIV nucleocapsid, more than one template was used due to fewer sequence identities with the query sequences. MODELLER software predicted the 3D structures of all the HIV structural and accessory proteins. Figure 1 shows the 3D modeled structures of HIV structural and accessory proteins.

### *3D structure validation of modeled proteins*

The 3D modeled structures of HIV structural and accessory proteins were validated by Ramachandran plot, local and global quality estimation scores, QMEAN scores, and Z-scores. Figure 2 shows the Ramachandran plot 3D modeled structures of HIV structural and accessory proteins. In the Ramachandran plot, the distribution of torsion angles of all the HIV structural and accessory proteins was analyzed to be present in favorable regions of the graphs. Table 5 displays the parameters of the Ramachandran plot, such as residues in the most favored region, residues additionally allowed region, residues generously allowed region, residues in the disallowed region, the total number of non-glycine and non-proline residues, numbers of end residues (excluding glycine and proline), numbers of glycine residues, numbers of proline residues, and the total number of residues in HIV structural and accessory proteins. QMEAN online web server calculated the local, global and normalized QMEAN4 scores of all the HIV 3D modeled protein structures. Figure 3 presents the local quality estimation graphs of HIV structural and accessory proteins. Figure 4 exhibits the global quality estimation scores of all the 3D-modeled HIV structural and accessory proteins. Figure 5 indicates the normalization of QMEAN4 scores of the 3D modeled structures of HIV structural and accessory proteins. The QMEAN scores below  $-4.0$  are considered a low-quality of predicted structure. Figure 6 shows the plots of Z-scores of all the 3D modeled structures of HIV proteins by comparing

**Table 1.** HIV structural and accessory proteins with amino acid sequence and UniprotKB ID.

GENES	PROTEINS	SIZE <sup>a</sup>	UNIPROTKB ID	AMINO ACIDS	SEQUENCE
<i>gag</i>	Capsid	p24	S5N8H7	231	PIVQNLRGQMVHQAIASPRTLNNAVVKVVEEKAFSPEVPMFSA LSEGATPQDLNMLNTVGGHQAAMQMLKETINEEAAEWDR HPVHAGPIAPGQMRPRGSDIAGTTSTLQEQIGWMTNPP VGEIYKRWILGLNKIVRMYSPSILDIRQGPKEPFRDYVDR KTLRAEQASQEVKNWMTETLLVQANPDCKTILKALGPAAT EEMMTACQGVGGPGHKARVL
	Matrix	p17	Q97730	132	MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASREL ERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELKSLFNTVA TLYCVHQKIEVKDTKEALEKVEEQNKSKKKAQQAAGTGN SSQVSQNY
	Nucleocapsid	p7	Q283U0	161	MTLYLVPPLDSADKELPALASKAGVTLLEIEFLHELWPHLSGG QIVIAALNANNLAILNRHMSTLLVELPVAVMAVPGASYRSDWN MIAHALPSEDWITLSNKMLKSGLLANDTVQGEKRSAGAEPLSP NVYTDALSRLGIATAHAIPVEPEQPFVDEVSA
	p6	p6	PDB ID: 2C55	52	LQSRPEPTAPPEESFRFGEETTPSQKQEPIDKELYPLASLRS LFGSDPSSQ
<i>pol</i>	Reverse transcriptase	p51/66	Q9WJQ2	259	PISPIEVPVKLPGMDGPKVKQWPLTEEKIKALVEICTEMEKE GKISKIGPENPYNTPVFAIKKDKSTRWRKLVDFRELNKRTQDF WEVQLGIPHPAGLKKKRSVTVLDVGDAYFSVPLDKFRKYTA FTIPSINNETPGIRYQYNVLPQGWKGSFAIFQSSMTKILEPFRK QNPDIYQYMDLLYVGSDEIGQHRTKIEELRQHLLKWGFTT PDKKHQKEPFLWVGYEHPDKWTVQPIVLEKDSWTVNDIQK
	Invertase	p32	Q76353	288	FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCD KCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVLVAVHVASG YIEAEVIPAETGQETAYFLLKLAGRWPVKTVHTDNGSNFTST TVKAACWWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQVR DQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQT KELQKQITKIQNFVYRDRSDPVWKGPAKLLWKGEGAVVIQ DNSDIKVVPRRKAIIIRDYDGKQMGAGDDCVASRQDED
	Protease	p10	O90777	99	PQVTLWQRPIVITIKIGGQLKEALLDGTGADDTVLEEMSLPGKW KPKMIGGIGGFIVRQYDQVSIICGHKAIGTVLIGTPVNIIGR NLLTQLGCTLNF
<i>env</i>	Gp120		Q9IZE4	455	VPVWRDADTTLFCASDAKSHVTEAHNVWATHACVPTDNP QEIHLNVTENFNMWKNMVEQMVEDVISLWEQSLKPCVK LTPLCVTLNCTNANLTNANLTNANNITNVENITDEVNCSFNV TDLRDKQKQVHALFYRLDIVQINSKNSDYRLINCNTSVIKQ ACPKISFDPIPIHYCTPAGYAILKCNKDFNGTGPKNVSSVQ CTHGKIPVSTQLLNGLSLAEEEEIIRSENLTNNVKTIVHLNKS VEINCTRPSNNTRTSITIGPGQVFYRTGDIIGDIRKVSCELNGT KWNEVLKQVKEKLEHFNKNISFQPPSGGDLEITMHHFSCR GEFFYCNTTQLFNNTYSNGTITLPCIKIKIINMWQGVQAM APPISGRINCLSNITGLLLTRDGNNGTNETFRPGGNIKDNW RSELYKCKVVQIEPLGIAPTRAKRRVVEREKK
	Gp41		PDB ID: 1AIK	72	XSGIVQQQNNLLRAIEAQHLLQLTVWGIKQLQARILXWMEW DREINNYTSLIHSLEESQNNQEQEKNEQELL
<i>vpr</i>	Virus protein r	p15	P12520	96	MEQAPEDQGPQREPYNEWTLLELEELKSEAVRHFPRIWLHN LGQHIYETYGDWAGVEAIIIRILQQLLFIHFRIGCRHSRIGVTR QRRARNGASRS
<i>vif</i>	Viral infectivity factor	p23	P12504	192	MENRWQVMIVWQVDRMRINTWKRLVKHHMYISRKAKDWFY RHHYESTNPKISSEVHIPLGDAKLVTITYWGLHTGERDWHLGQ GVSIEWRKKRYSTQVDPDLADQLIHLHYDFCFSEAIRNTILGR IVSPRCEYQAGHNKVGSLQYLALAALIKPKQIKPPLPSVRKLTE DRWNKPKQTKGHRGSHTMNGH
<i>vpu</i>	Virus protein unique	p16	P05923	81	MQPIQIAIAALVVAIIIVVSVIIEYRKILRQRKIDRLIDRLIERA EDSGNESEGEISALVEMGVEMGHHPWDIDDL
<i>rev</i>	RNA splicing regulator	p19	P04618	116	MAGRSGDSDEELIRTVRLIKLLYQSNPPNPEGTRQARRNRR RRWRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDC NEDCGTSGTQGVGSPQILVESPTVLESSTKE

(Continued)

**Table 1.** (Continued)

GENES	PROTEINS	SIZE <sup>a</sup>	UNIPROTKB ID	AMINO ACIDS	SEQUENCE
<i>tat</i>	Transactivator protein	p14	P04608	86	MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITK ALGISYGRKKRRRQRRRAHQNSQTHQASLSKQPTSQPRGDP TGPKE
<i>nef</i>	Negative regulating factor	p27	P04601	206	MGGKWSKSSVIGWPTVRERMRRRAEPAADRVAASRDLEKH GAITSSNTAATNAACAWLEAQEEEEVGFVTPQVPLRPMTYK AAVDLSHFLKEKGGLEGLIHSQRRQDILDWYHTQGYFPDW QNYTPGPGVRYPLTFGWICYKLVPEPKIEEANKGENTSLLH PVSLHGMDPPEREVLEWRFDLSRLAFHHVARELHPEYFKNC
<i>vpx</i>	Virus protein x	p15	P18099	113	MTDPRERVPPGNSGEETIGEAFEWLERTIEALNREAVNHLPR ELIFQVWQRSWRYWHDEQGMASASYTKYRYLCLMQKAIFTHF KRGCTCWGEDMGREGLEDQPPPPPPGLV

<sup>a</sup>Numbers correspond to proteins (p) size in 1000 Da.

**Table 2.** Physicochemical characteristics of HIV structural and accessory proteins.

SR. NO.	HIV PROTEINS	MOL. WT.	THEORETICAL PI	EC (M <sup>-1</sup> CM <sup>-1</sup> , AT 280 NM)	GRAVY	II	AI	R <sup>+</sup>	R <sup>-</sup>
1	Capsid	25593.45	6.26	33585	-0.325	44.03	83.2	22	24
2	Matrix	14715.75	9.39	15595	-0.739	43.36	78.33	22	16
3	Nucleocapsid	17299.86	4.73	20970	0.161	51.03	111.55	9	19
4	p6	5807.34	4.48	1490	-1.065	107.9	48.85	5	9
5	Reverse transcriptase	30194.88	9.12	57410	-0.7	42.48	77.07	40	34
6	Invertase	32198.77	7.75	50795	-0.404	35.34	82.95	38	37
7	Protease	10724.67	8.81	12615	0.166	36.56	114.14	10	8
8	Gp120	51088.18	8.53	53035	-0.415	38.22	82.64	49	42
9	Gp41	8576.08	4.93	17990	-0.592	64.11	109.72	5	9
10	Virus protein r	11378.88	8.02	20970	-0.691	51.37	84.38	13	12
11	Viral infectivity factor	22699.12	9.93	56045	-0.734	37.35	77.66	31	17
12	Virus protein unique	9159.72	4.69	12490	0.358	33.1	140.86	8	14
13	RNA splicing regulator	13074.73	9.29	8605	-0.874	86.74	78.97	18	15
14	Transactivator protein	9837.29	9.88	8855	-1.213	54.76	34.07	17	5
15	Negative regulating factor	23469.46	5.99	49055	-0.602	41.93	70.1	24	29
16	Virus protein x	13207.9	5.36	33585	-0.814	42.25	56.11	13	17

the NMR and X-ray crystallographic structures. The Z-scores of HIV capsid, matrix, nucleocapsid, p6, reverse transcriptase, invertase, protease, gp120, gp41, virus protein R, viral infectivity factor, virus protein unique, RNA splicing regulator, transactivator protein, negative regulating factor, and virus protein X were -7.02, -6.2, -0.74, -1.22, -8.37, -7.27, -4.97, -8.73, -2.32, -3.24, -5.26, 0.13, -2.75, -1.06, -4.4, and -3.32, respectively.

## Discussion and Conclusions

Computational protein structure modeling has significantly increased the determination of the 3D structure of proteins as

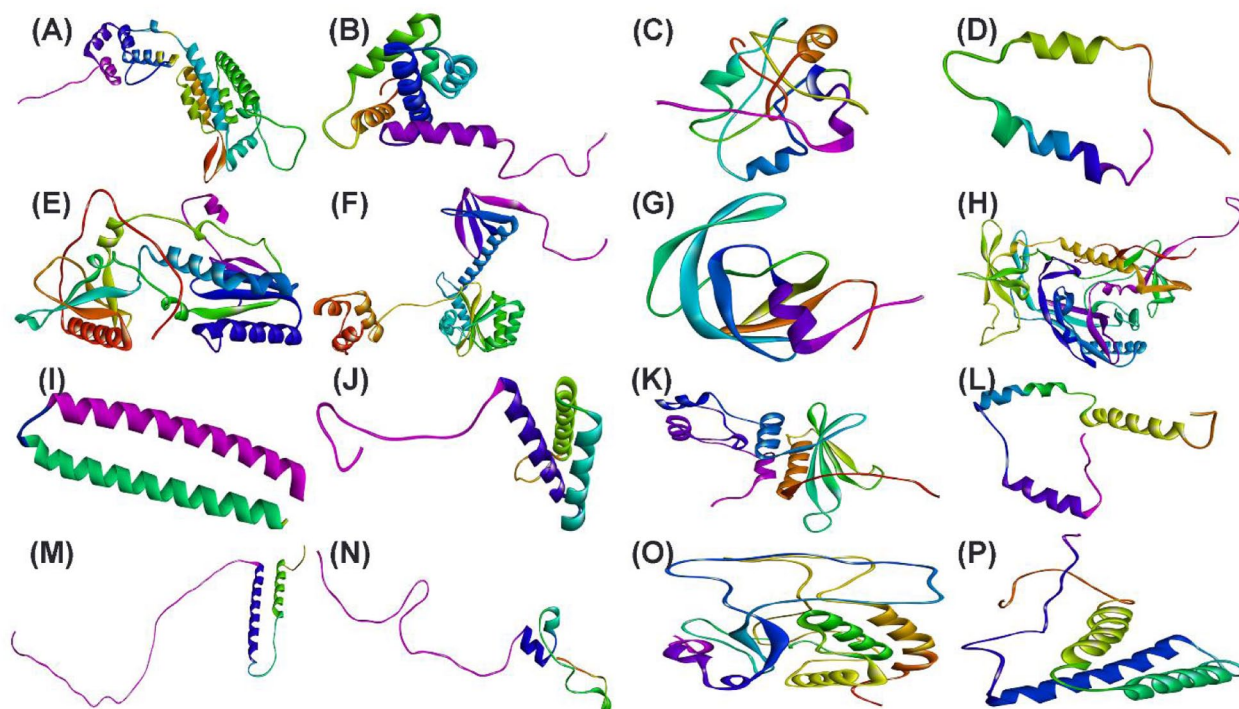
x-ray crystallization and NMR spectroscopy are time-consuming processes and have many difficulties, like purification, crystallization, and low-resolution 3D structures of proteins.<sup>22</sup> Along with structure determination, homology modeling has many other applications, such as structure-based drug designing, mutations analysis, active sites identification, novel ligands designing, substrate-specific modeling, protein-protein molecular docking, molecular simulation, structural refinement at the molecular level, future planning in computational experimentations.<sup>39-43</sup> HIV is an infectious disease that attacks the host's immune cells and weakens the immune system to fight other diseases. HIV encodes 16 distinct proteins, and their 3D

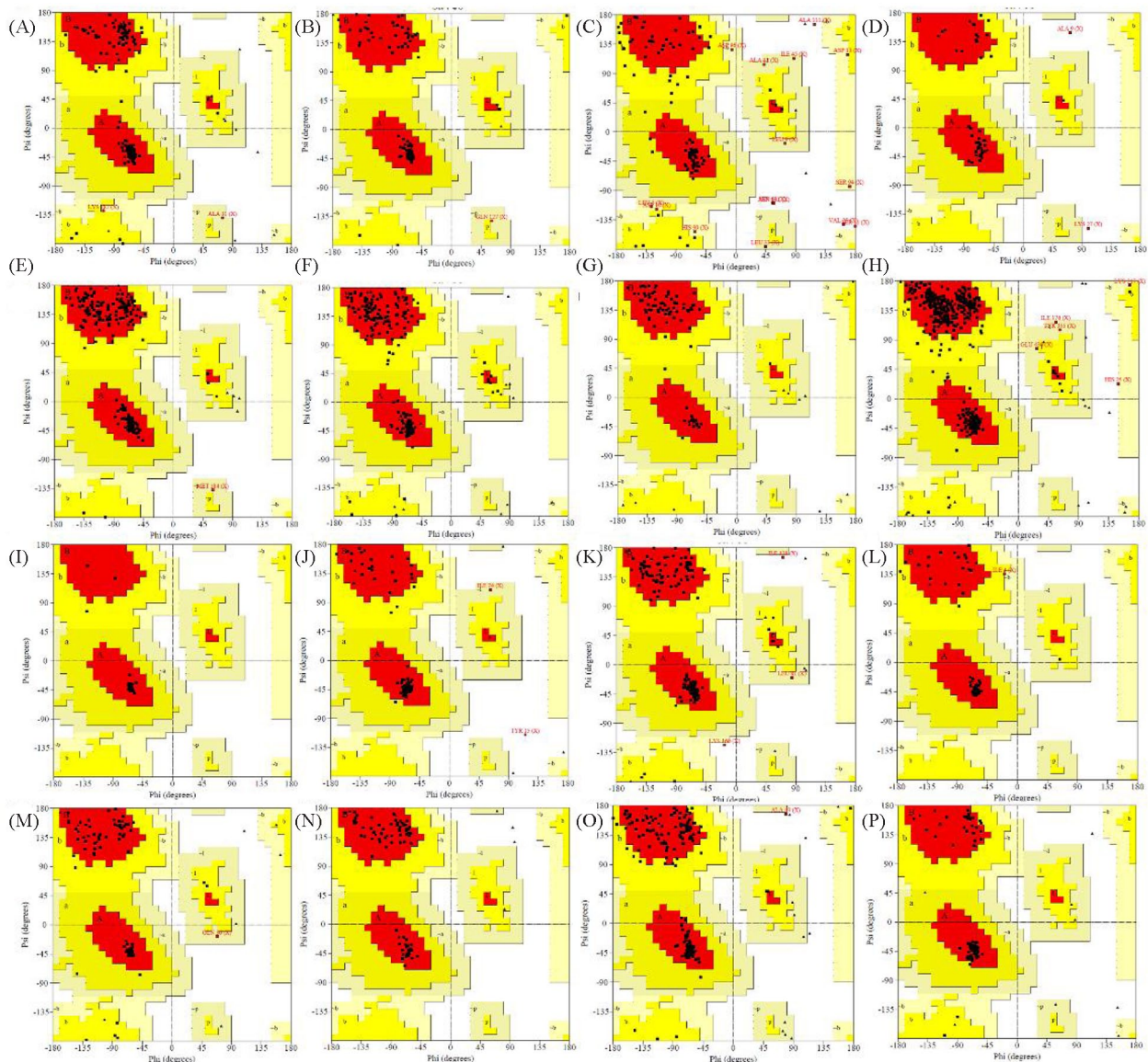
**Table 3.** Prediction of different parameters of secondary structures of HIV gp120 and gp41 proteins.

SR. NO.	HIV PROTEINS	ALPHA HELIX (%)	310 HELIX (%)	PI HELIX (%)	BETA BRIDGE (%)	EXTENDED STRAND (%)	BETA TURN (%)	BEND REGION (%)	RANDOM COIL (%)	AMBIGUOUS STATES (%)	OTHER STATES (%)
1	Capsid	54.11	0	0	0	5.63	4.33	0	35.93	0	0
2	Matrix	65.91	0	0	0	6.82	4.55	0	22.73	0	0
3	Nucleocapsid	37.89	0	0	0	13.04	8.7	0	40.37	0	0
4	p6	15.38	0	0	0	1.92	1.92	0	80.77	0	0
5	Reverse transcriptase	25.1	0	0	0	21.62	6.18	0	47.1	0	0
6	Invertase	38.54	0	0	0	15.62	5.9	0	39.93	0	0
7	Protease	12.12	0	0	0	42.42	6.06	0	39.39	0	0
8	Gp120	22.2	0	0	0	29.67	3.96	0	44.18	0	0
9	Gp41	83.33	0	0	0	2.78	2.78	0	11.11	0	0
10	Virus protein r	46.88	0	0	0	12.5	6.25	0	34.38	0	0
11	Viral infectivity factor	33.33	0	0	0	18.75	2.6	0	45.31	0	0
12	Virus protein unique	65.43	0	0	0	8.64	2.47	0	23.46	0	0
13	RNA splicing regulator	30.17	0	0	0	12.07	1.72	0	56.03	0	0
14	Transactivator protein	0	0	0	0	13.95	9.3	0	76.74	0	0
15	Negative regulating factor	42.23	0	0	0	8.25	3.88	0	45.63	0	0
16	Virus protein x	47.79	0	0	0	3.54	5.31	0	43.36	0	0

**Table 4.** List of the templates used for homology modeling HIV structural and accessory proteins with maximum sequence similarity and identity.

SR. NO.	HIV PROTEINS	TEMPLATE PDB ID	SEQUENCE SIMILARITY	SEQUENCE IDENTITY (%)
1	Capsid	5JPA	1.426	97
2	Matrix	7JXR	1.333	90
3	Nucleocapsid	2CC0, 5I10, 3N6T, 3ZIE, 4M1B, 1UCD, 4RPO, 5TZD, 2EKC, 3DX5	0.399, 0.375, 0.277, 0.311, 0.271, 0.386, 0.088, 0.299, 0.162, 0.126	26, 25, 25, 23, 22, 20, 19, 19, 18, 15
4	p6	2C55	1.441	100
5	Reverse transcriptase	4G1Q, 3ISN,	1.548, 1.548	97, 97
6	Invertase	6PUY, 6T6E, 1WJA	1.516, 1.578, 1.555	100, 100, 100
7	Protease	2HS1, 3KA2, 3KA2	1.371, 1.425, 1.404	85, 84, 83
8	Gp120	5FCU, 3DNL, 7LX2	1.496, 1.505, 1.304	91, 89, 78
9	Gp41	3O3X, 5Y14, 3VH7, 3WV7	1.315, 1.271, 1.271, 1.138	97, 97, 97, 75
10	Virus protein r	5JK7	1.581	100
11	Viral infectivity factor	4N9F	1.606	100
12	Virus protein unique	2N28	1.378	94
13	RNA splicing regulator	2X7L	1.45	94
14	Transactivator protein	3MI9	1.653	99
15	Negative regulating factor	6CRI, 4EN2	1.595, 1.595	98, 98
16	Virus protein x	4CC9	1.315	66

**Figure 1.** Modeled 3D Structure of HIV Proteins: (A) capsid (p24), (B) matrix (p17), (C) nucleocapsid (p7), (D) p6, (E) reverse transcriptase (p51/66), (F) invertase (p32), (G) protease (p10), (H) GP120 (I) GP41, (J) virus protein R (p15), (K) viral infectivity factor (p23), (L) virus protein unique (p16), (M) RNA splicing regulator (p19), (N) transactivator protein (p14), (O) negative regulating factor (p27), and (P) virus protein x (p15).



**Figure 2.** Ramachandran plots validate the 3D modeled structures of HIV structural and accessory proteins. Ramachandran plot of: (A) capsid (p24), (B) matrix (p17), (C) nucleocapsid (p7), (D) p6, (E) reverse transcriptase (p51/66), (F) invertase (p32), (G) protease (p10), (H) GP120 (I) GP41, (J) virus protein R (p15), (K) viral infectivity factor (p23), (L) virus protein unique (p16), (M) RNA splicing regulator (p19), (N) transactivator protein (p14), (O) negative regulating factor (p27), and (P) virus protein x (p15) proteins. Phi (X-axis) and Psi (Y-axis) represent backbone conformation angles of amino acid residues.

structure determination is substantial to understanding their functions in the viral life cycle.<sup>44,45</sup> This study identifies the computational 3D modeled of HIV structural and accessory proteins to better understand the structural basis of HIV proteins interacting with host cells and viral replication.

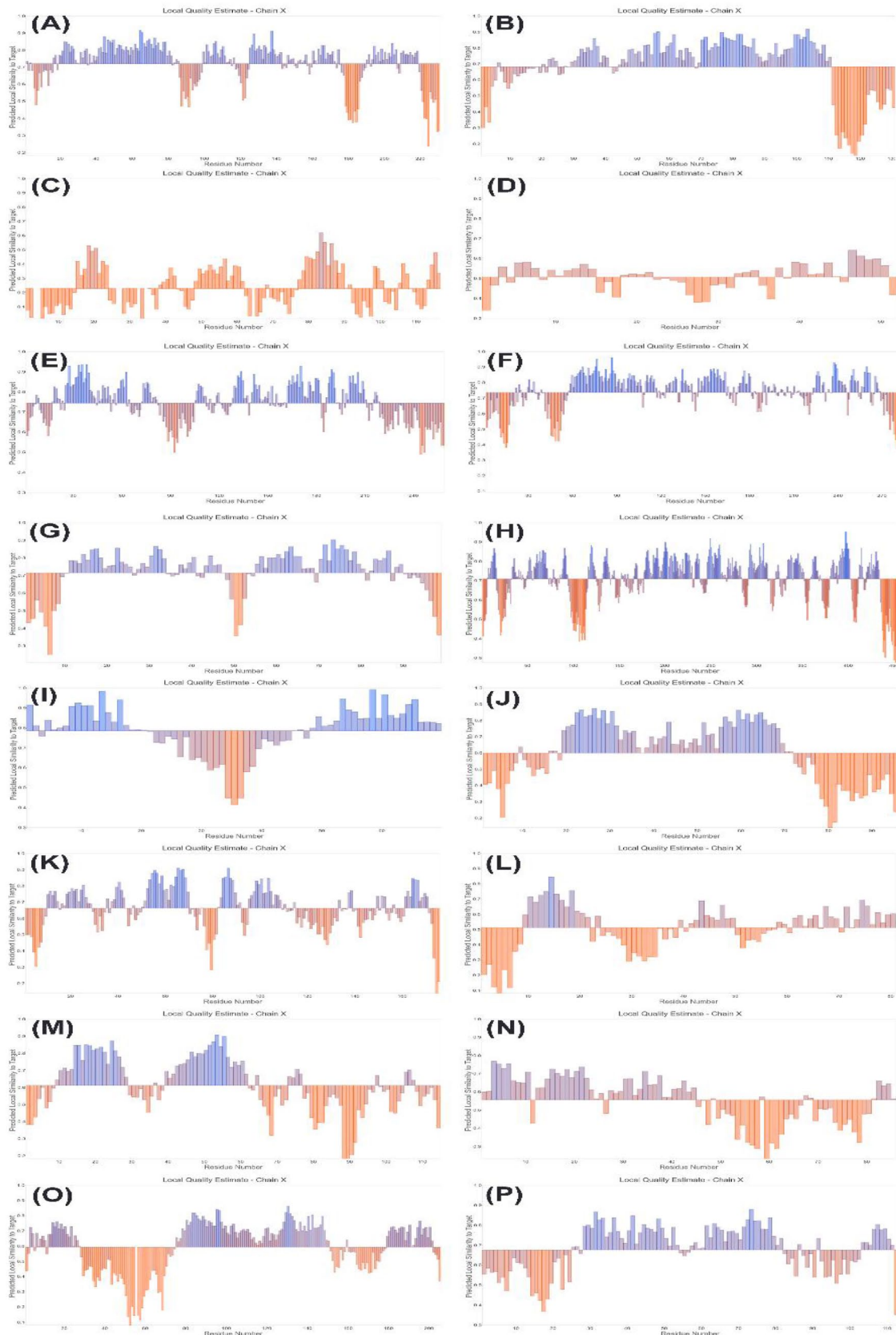
The sequences of HIV capsid (231 bp), matrix (132 bp), nucleocapsid (161 bp), p6 (52 bp), gp120 (455 bp), gp41 (72 bp), reverse transcriptase (259 bp), invertase (288 bp), protease (99 bp), virus protein r (96 bp), viral infectivity factor (192 bp), virus protein unique (81 bp), RNA splicing regulator (116 bp), transactivator protein (86 bp), negative regulating factor (206 bp), and virus protein x (113 bp) proteins were downloaded from UniPort. All the UniPort KB IDs with protein sequences of HIV proteins are mentioned in Table 1. The

primary structures were determined by using HIV protein sequences in the ProtParam tool of the ExPASy server. In this study, the theoretical pI of gp120 was calculated as 8.53, which was very close to the predicted pI of MN-rgp120 (8.7) and A244-rgp120 (8.4).<sup>46</sup> Similarly, the pI of gp41 was calculated as 4.93, which is acidic, as reported in a previous study.<sup>47</sup> Further, the HIV capsid, nucleocapsid, p6, gp41, virus protein unique, negative regulating factor, and virus protein x showed acidic pI values due to more negative (acidic) amino acid residues (Table 2). While HIV matrix, reverse transcriptase, invertase, protease, gp120, virus protein r, viral infectivity factor, RNA splicing regulator, and transactivator protein had more positive (basic) amino acids, so that showing pI values in a basic range (Table 2).

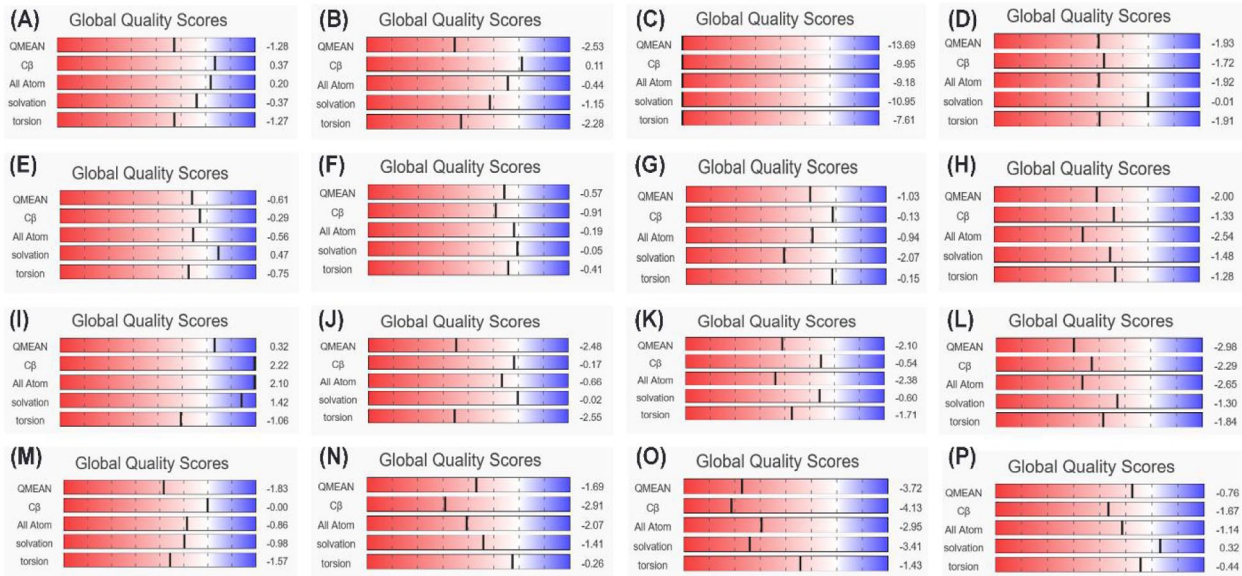


**Table 5.** Ramachandran plot statistics of Chi1, Chi2, and BcLPMO proteins.

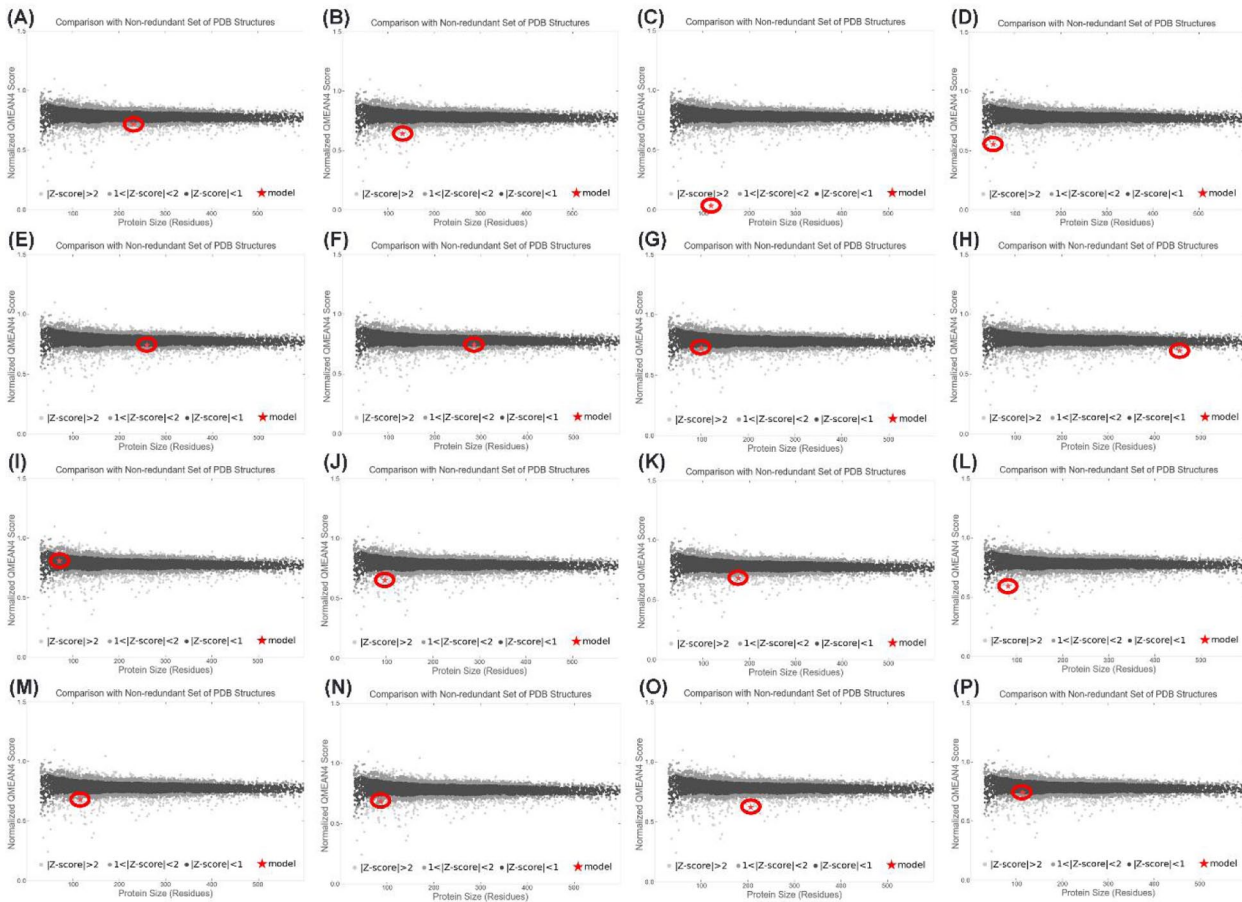
RAMACHANDRAN PLOT PARAMETERS	CAPSID	MATRIX	NUCLEOCAPSID	P6	REVERSE TRANSCRIPTASE	INVERTASE	PROTEASE	GP120	GP41	VIRUS PROTEIN R	VIRAL INFECTIVITY FACTOR	VIRUS PROTEIN UNIQUE	RNA SPLICING REGULATOR	TRANSACTIVATOR PROTEIN	NEGATIVE REGULATING FACTOR	VIRUS PROTEIN X
Residues in the most favored region	181	109	60	34	211	232	75	374	65	77	147	69	83	70	155	86
Residues additionally allowed region	12	6	28	4	7	18	4	27	1	4	8	3	9	0	18	4
Residues generously allowed region	2	1	10	0	1	0	0	4	0	1	1	1	1	0	0	0
Residues in the disallowed region	0	0	5	2	0	0	0	1	0	1	2	0	0	0	1	0
Total number of non-glycine and non-proline residues	195	116	103	40	219	250	79	406	66	83	158	73	93	70	174	90
Numbers of end residues (Excluding Glycine and Proline)	1	1	1	2	1	2	1	2	2	2	2	2	2	2	2	2
Numbers of Glycine residues	17	11	6	2	14	23	13	24	2	7	8	4	10	5	15	10
Numbers of Proline residues	18	3	7	8	25	10	6	22	0	4	8	2	10	9	15	11
Total number of residues	231	131	117	52	259	285	99	454	70	96	176	81	115	86	206	113



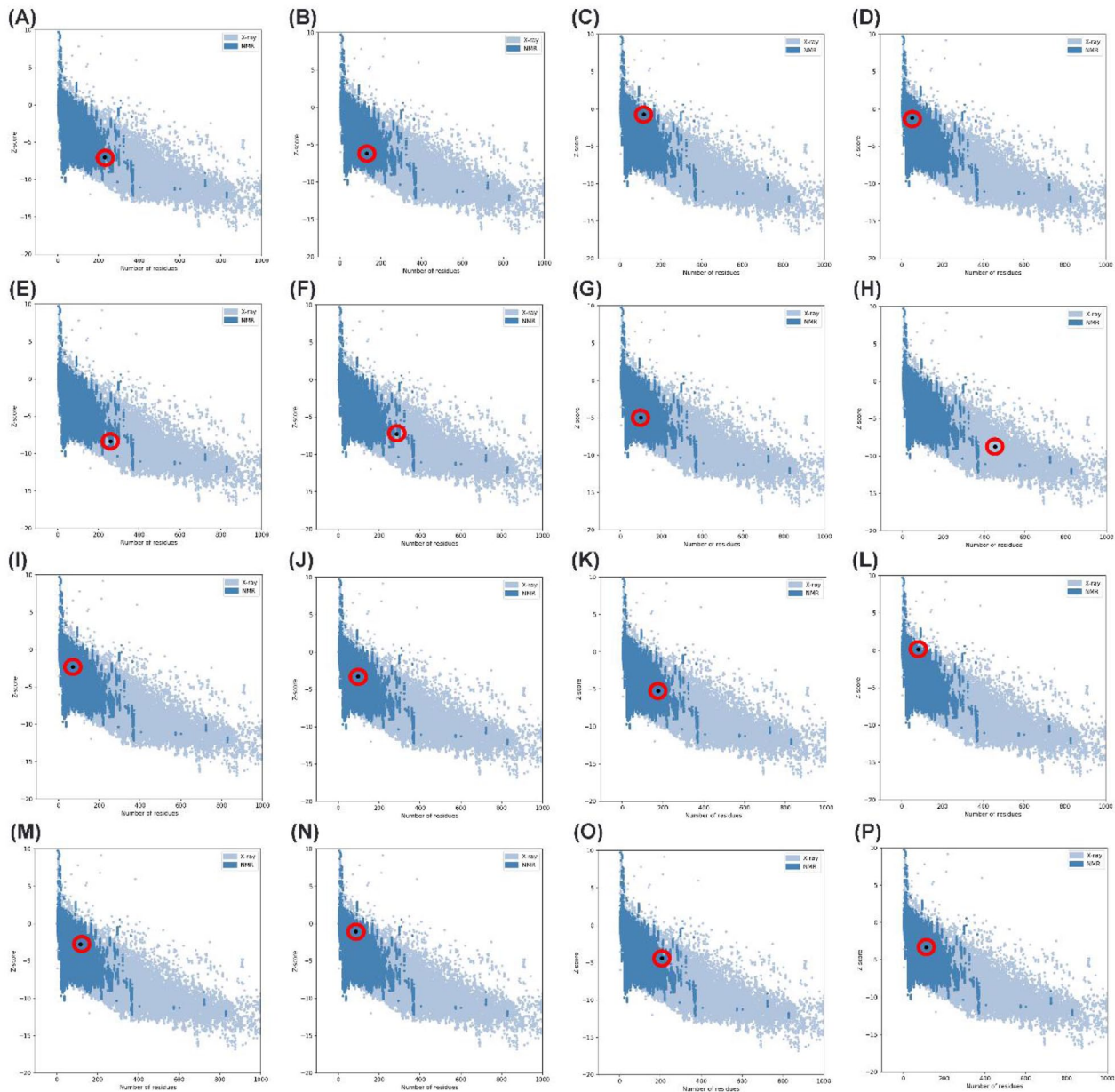
**Figure 3.** Local quality estimation graphs comparing the 3D models of expected similarity to the native structures of: (A) capsid (p24), (B) matrix (p17), (C) nucleocapsid (p7), (D) p6, (E) reverse transcriptase (p51/66), (F) integrase (p32), (G) protease (p10), (H) GP120 (p160), (I) GP41, (J) virus protein R (p15), (K) viral infectivity factor (p23), (L) virus protein unique (p16), (M) RNA splicing regulator (p19), (N) transactivator protein (p14), (O) negative regulating factor (p27), and (P) virus protein x (p15) proteins of HIV.



**Figure 4.** Global quality estimation scores showing the 3D models of entire structures of: (A) capsid (p24), (B) matrix (p17), (C) nucleocapsid (p7), (D) p6, (E) reverse transcriptase (p51/66), (F) invertase (p32), (G) protease (p10), (H) GP120 (I) GP41, (J) virus protein R (p15), (K) viral infectivity factor (p23), (L) virus protein unique (p16), (M) RNA splicing regulator (p19), (N) transactivator protein (p14), (O) negative regulating factor (p27), and (P) virus x (p15) proteins of HIV.



**Figure 5.** Normalized QMEAN4 score comparing with a non-redundant set of PDB of 3D model structures of: (A) capsid (p24), (B) matrix (p17), (C) nucleocapsid (p7), (D) p6, (E) reverse transcriptase (p51/66), (F) invertase (p32), (G) protease (p10), (H) GP120 (I) GP41, (J) virus protein R (p15), (K) viral infectivity factor (p23), (L) virus protein unique (p16), (M) RNA splicing regulator (p19), (N) transactivator protein (p14), (O) negative regulating factor (p27), and (P) virus protein x (p15) proteins of HIV.



**Figure 6.** Z-score plot obtained using the ProSA server showing the location of 3D model structures of HIV: (A) capsid (p24), (B) matrix (p17), (C) nucleocapsid (p7), (D) p6, (E) reverse transcriptase (p51/66), (F) invertase (p32), (G) protease (p10), (H) GP120 (I) GP41, (J) virus protein R (p15), (K) viral infectivity factor (p23), (L) virus protein unique (p16), (M) RNA splicing regulator (p19), (N) transactivator protein (p14), (O) negative regulating factor (p27), and (P) virus protein x (p15) proteins in X-ray and NMR structures.

The HIV proteins, except the transactivator protein, showed 4 secondary structures, that is, alpha helix, extended strand, beta-turn, and random coil, as shown in Table 3. The percentages of secondary structures in HIV proteins vary. Most proteins (nucleocapsid, p6, Reverse transcriptase, invertase, gp120, viral infectivity factor, RNA splicing regulator, transactivator protein, and negative regulating factor) contained more random coils in their secondary structures. While HIV capsid, matrix, gp41, virus protein r, virus protein unique, and virus protein x had more percentage of an alpha helix in their secondary structures. In HIV protease, extended strands were more present than other secondary structures.

For the 3D structure modeling of HIV proteins, templates with maximum sequence identities and similarities were used. Their PDB IDs are mentioned in Table 4. MODELLER software modeled the HIV proteins by using these templates (Figure 1). All the modeled HIV protein structures were validated by Ramachandran plots, local and global quality estimation scores, QMEAN scores, and Z-scores. The Ramachandran plots confirmed the quality of the 3D-modeled structure of the HIV proteins. All the amino acid residues of the capsid, matrix, reverse transcriptase, invertase, protease, gp41, virus protein unique, RNA splicing regulator, transactivator protein, and virus protein x were present in the most favored regions,

allowed regions, and generously allowed regions in the Ramachandran plots (Table 5 and Figure 2).

The Ramachandran plot of the 3D modeled structure of HIV-1 capsid protein showed 2 amino acids in the generously allowed region. However, the template (PDB: 5JPA) used to model the HIV-1 capsid protein showed only one amino acid in the generously allowed region of the Ramachandran plot. Similarly, Ramachandran plots of reverse transcriptase, invertase, and protease templates showed 3, 1, and 1 amino acid(s) in the generously allowed region, respectively. While Ramachandran's plot of modeled predicted structure of reverse transcriptase showed one amino acid in the generously allowed region. Nonetheless, nucleocapsid, p6, gp120, virus protein r, viral infectivity factor, and negative regulating factor had some amino acid residues in the disallowed region of the Ramachandran plots (Table 5 and Figure 2). Additional local quality estimation was used for the estimation of the quality of each residue of the protein. All the 3D modeled protein structures of HIV proteins showed local quality scores above 0.6, except nucleocapsid and p6 proteins (Figure 3). Each residue's local quality is calculated by comparing its model structure on the x-axis with its native structure on the y-axis. A local quality score below 0.6 is not considered a good 3D-modeled protein structure.<sup>20</sup> A global score determines the overall quality of an entire model. Depending on the alignment and template used to construct the model, the Global Model Quality Estimation can be expressed as a number between 0 and 1. In this number, the denominator is the target sequence coverage. The reliability of the system increases as the number rises.<sup>35-38</sup> The global quality score of the 3D modeled HIV proteins showed a QMEAN score above -4.0 except for nucleocapsid protein (QMEAN score -13.69) (Figures 4 and 5). Additionally, the degree of nativeness of 3D modeled structures of HIV proteins with native proteins of similar sizes was calculated by Z-scores.<sup>28</sup> All the 3D modeled structures of HIV proteins showed their native confirmation within the range of the plots, as shown in Figure 6. All the validation results of the HIV 3D modeled proteins confirmed the relatively good quality of predicted proteins which could be helpful in further studies like drug designing.

The present study focused on the HIV structural and accessory proteins 3D structural modeling by using MODELLER. The 3D structures of all the HIV proteins were modeled with maximum sequence similarity and identity protein templates. Further, the predicted structures were validated using a variety of online servers. The obtained 3D modeled structures of HIV proteins might be helpful in further studies to investigate the crystal structures with x-rays crystallography and NMR spectroscopy, proteins motifs and domains, physiochemical properties, new target positions, and drug designing. Moreover, this study predicted the unbounded 3D structures of HIV proteins. However, further studies can predict the bounded

structures of HIV proteins, such as HIV RT structure bound with nucleotide RT inhibitors (NRTIs) and nonnucleoside RT inhibitors (NNRTIs). HIV is a dangerous and infectious virus with no cure. Hence, further extensive investigation is still required to find possible treatment in the future.

## ORCID iDs

Amir Elalouf  <https://orcid.org/0000-0001-7950-0785>

## REFERENCES

- Alves NMP, de Moura RR, Bernardo LC, et al. In silico analysis of molecular interactions between HIV-1 glycoprotein gp120 and TNF receptors. *Infect Genet Evol.* 2021;92:104837.
- UNAIDS. Global HIV & AIDS statistics — fact sheet | UNAIDS. *UNAIDS.* 2021:2020–2022.
- Vidya Vijayan KK, Karthigeyan KP, Tripathi SP, Hanna LE. Pathophysiology of CD4+ T-Cell depletion in HIV-1 and HIV-2 infections. *Front Immunol.* 2017;8:580.
- Levy JA, Steele FR. HIV and the pathogenesis of aids. *Nat Med.* 1995;1:273.
- Checkley MA, Lutge BG, Freed EO. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J Mol Biol.* 2011;410:582–608.
- UniProt. ENV - envelope glycoprotein gp160 precursor - human immunodeficiency virus type 1 group M subtype B (isolate HXB2) (HIV-1) - env gene & protein.
- Vicenzi E, Poli G. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *Tissue Antigens.* 2013;81:61–71.
- Le Rouzic E, Benichou S. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology.* 2005;2:11.
- Del Cornò M, Donninelli G, Varano B, Da Sacco L, Masotti A, Gessani S. HIV-1 gp120 activates the STAT3/interleukin-6 axis in primary human monocyte-derived dendritic cells. *J Virol.* 2014;88:11045–11055.
- Ran X, Ao Z, Trajtmann A, et al. HIV-1 envelope glycoprotein stimulates viral transcription and increases the infectivity of the progeny virus through the manipulation of cellular machinery. *Sci Rep.* 2017;7:9487.
- Allan JS, Coligan JE, Barin F, et al. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-iii. *Science.* 1985;228:1091–1094.
- Kwong PD, Wyatt R, Majeed S, et al. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure.* 2000;8:1329–1339.
- Cardoso RM, Zwick MB, Stanfield RL, et al. Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity.* 2005;22:163–173.
- Caillat C, Guilligay D, Sulbaran G, Weissenhorn W. Neutralizing antibodies targeting HIV-1 gp41. *Viruses.* 2020;12:1210.
- Clark AJ, Gindin T, Zhang B, et al. Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of HIV-1. *J Mol Biol.* 2017;429:930–947.
- Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J Mol Biol.* 2019;431:2197–2212.
- Serra PA, Taveira N, Guedes RC. Computational modulation of the v3 region of glycoprotein gp125 of hiv-2. *Int J Mol Sci.* 2021;22:1–19.
- Andrianov AM, Anishchenko IV. Computational model of the hiv-1 subtype a v3 loop: study on the conformational mobility for structure-based anti-aids drug design. *J Biomol Struct Dyn.* 2009;27:179–193.
- Tripathy CS, Sahoo BC, Dash M, Sahoo D, Sahoo S, Kar B. In-silico structural modelling of cytochrome complex proteins of white turmeric (*Curcuma zedoaria*). *Plant Sci Today.* 2022;9:555–563.
- Santhoshkumar R, Yusuf A. In silico structural modeling and analysis of physicochemical properties of curcumin synthase (CURS1, CURS2, and CURS3) proteins of curcuma longa. *J Genet Eng Biotechnol.* 2020;18:24.
- Sailapathi A, Gunalan S, Somarathinam K, et al. Importance of homology modeling for predicting the structures of GPCRs. *Homol Mol Model Perspect Appl.* Published online November 19, 2021. doi:10.5772/intechopen.94402.
- Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: overview, current applications, and future perspectives. *Chem Biol Drug Des.* 2019;93:12–20.

23. Pal S, Mishra M, Sudhakar DR, Siddiqui MH. *In-silico* designing of a potent analogue against HIV-1 Nef protein and protease by predicting its interaction network with host cell proteins. *J Pharm Bioallied Sci.* 2013;5:66-73.
24. Singh S, Biswas S, Srivastava A, Mishra Y, Chaturvedi TP. In silico characterization and structural modeling of a homeobox protein MSX1 from homo sapiens. *Inform Med Unlocked.* 2021;22:100497.
25. Gurung AB. In silico structure modelling of SARS-CoV-2 Nsp13 helicase and Nsp14 and repurposing of FDA approved antiviral drugs as dual inhibitors. *Gene Rep.* 2020;21:100860.
26. Sakamoto K, Kayanuma M, Inagaki Y, Hashimoto T, Shigeta Y. In silico structural modeling and analysis of elongation factor-1 alpha and elongation factor-like protein. *ACS Omega.* 2019;4:7308-7316.
27. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics.* 2016;54:5.6.1-5.6.37.
28. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35:W407-W410.
29. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011;27:343-350.
30. 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.* 2003;31:3784-3788.
31. Geourjon C, Deléage G. Sopma: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics.* 1995;11:681-684.
32. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;33:W244-W248.
33. Rodriguez R, China G, Lopez N, Pons T, Vriend G. Homology modeling, model and software evaluation: three related resources. *Bioinformatics.* 1998;14:523-528.
34. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 2009;37:W510-W514.
35. Cao R, Wang Z, Wang Y, Cheng J. SMOQ: a tool for predicting the absolute residue-specific quality of a single protein model with support vector machines. *BMC Bioinformatics.* 2014;15:8.
36. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modeling of protein structures and complexes. *Nucleic Acids Res.* 2018;46:W296-W303.
37. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics.* 2020;36:1765-1771.
38. Bienert S, Waterhouse A, de Beer TA, et al. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* 2017;45:D313-D319.
39. Koseki Y, Aoki S. Computational medicinal chemistry for rational drug design: identification of novel chemical structures with potential anti-tuberculosis activity. *Curr Top Med Chem.* 2014;14:176-188.
40. Ahmed MZ, Hameed S, Ali M, Zaheer A. Computational analysis of the interaction of limonene with the fat mass and obesity-associated protein. *Sci J Inform.* 2021;8:154-160.
41. Tariq A, Rehman H, Mateen R, et al. A computer aided drug discovery based discovery of lead-like compounds against KDM5A for cancers using pharmacophore modeling and high-throughput virtual screening. *Proteins Struct Funct Bioinforma.* 2022;90:645-657.
42. Al-Karmalawy AA, Dahab MA, Metwaly AM, et al. Molecular docking and dynamics simulation revealed the potential inhibitory activity of ACEIs against SARS-CoV-2 targeting the hACE2 receptor. *Front Chem.* 2021;9:661230.
43. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules.* 2015;20:13384-13421.
44. Armstrong LA, Lange SM, Dee Cesare V, et al. Biochemical characterization of protease activity of nsp3 from SARS-CoV-2 and its inhibition by nanobodies. *PLoS One.* 2021;16:e0253364.
45. Estevez M, Li R, Paul B, et al. Identification and mapping of post-transcriptional modifications on the HIV-1 antisense transcript Ast in human cells. *RNA.* 2022;28:697-710.
46. Yu B, Morales JF, O'Rourke SM, Tatsuno GP, Berman PW. Glycoform and net charge heterogeneity in gp120 immunogens used in HIV vaccine trials. *PLoS One.* 2012;7:e43903.
47. Taylor SE, Schwarz G. The molecular area characteristics of the HIV-1 gp41-fusion peptide at the air/water interface. Effect of pH. *Biochim Biophys Acta.* 1997;1326:257-264.