Identification of Potential Drug Targets Among Hypothetical Proteins of HTLV-1 Through Integrated In Silico Approaches

Sagor Das, Md. Tamzid Islam, Sanzida Afrin Oishy
Department of Biomedical Engineering,
Military Institute of Science and Technology (MIST), Dhaka, Bangladesh
Emails: sagor.bme8@gmail.com, tamzid.bme8@gmail.com, oishy.bme8@gmail.com

Abstract

Human T-cell Leukemia Virus type 1 (HTLV-1), the etiological agent of HTLV-1 Associated Myelopathy Tropical Spastic Paraparesis (HAM/TSP), is a neurodegenerative retrovirus that causes progressive spinal cord inflammation and disability. Despite its significant impact on affected populations in endemic regions, HAM/TSP remains underexplored, especially in the context of computational biology. In this study, we employed a comprehensive in silico approach to investigate hypothetical proteins (HPs) encoded by the HTLV-1 genome that may contribute to its neurotropism and pathogenicity. Through genome and proteome analysis of HTLV-1 strain ATK-1, we identified and annotated 18 hypothetical proteins using physicochemical characterization, subcellular localization, transmembrane topology, antigenicity prediction, and virulence profiling. Functional annotations revealed that several HPs may act as viral regulators, RNA-binding proteins, or immune evasion molecules. Notably, HP ATK1_RS00135 demonstrated strong virulent potential, cytoplasmic localization, and predicted interaction with host immune signalling pathways. Furthermore, molecular docking and dynamics simulation suggested that antiviral compounds such as Ribavirin and Zidovudine exhibit stable binding with HP ATK1_RS00090, indicating its potential as a repurposable drug target. This computational analysis of HTLV-1 HPs not only expands our understanding of its molecular pathogenesis but also presents novel targets for drug development and vaccine design against HAM/TSP.

Keywords: *HTLV-1, Human T-cell Leukemia Virus type 1, HAM/TSP, Hypothetical Proteins, In Silico Analysis, Drug Target Identification, Molecular Docking, Protein Functional Annotation, Biomedical Engineering, Computational Biology.*

Introduction

Human T-cell Leukemia Virus type 1 (HTLV-1) is a retrovirus and the etiological agent responsible for HTLV-1 Associated Myelopathy / Tropical Spastic Paraparesis (HAM/TSP), a chronic neurodegenerative disease characterized by progressive inflammation and demyelination of the spinal cord ***[1].*** HAM/TSP leads to spastic paraparesis, sensory disturbances, and bladder dysfunction, significantly impairing the quality of life of affected individuals ***[2].*** Globally, HTLV-1 infection is endemic in regions such as Japan, the Caribbean, South America, and parts of Africa, with an estimated 5–10 million infected individuals worldwide ***[3].*** Despite its considerable public health impact, HAM/TSP remains underdiagnosed and poorly understood, particularly at the molecular level.*[7]*

HTLV-1 primarily infects CD4+ T-cells, leading to persistent infection and immune dysregulation that contribute to disease progression ***[4].*** While several viral proteins have been characterized for their roles in viral replication and pathogenesis, approximately 20–30% of the HTLV-1 genome encodes hypothetical proteins (HPs) with unknown functions ***[5].*** These uncharacterized proteins may play critical roles in viral persistence, immune evasion, or neurotoxicity but remain largely unexplored due to limited biochemical data.

Current therapeutic options for HAM/TSP are limited to symptomatic management, with no effective antiviral treatment to halt disease progression ***[6].*** This therapeutic gap underscores the urgent need to identify novel molecular targets within the HTLV-1 proteome. In this context, in silico approaches provide a powerful platform to predict functions and druggability of HPs, guiding future experimental validation and drug development efforts ***[7].***

In this study, we conducted a comprehensive bioinformatics analysis of HTLV-1 hypothetical proteins derived from the core viral genome to predict their physicochemical properties, subcellular localization, virulence potential, and protein-protein interactions. Furthermore, molecular docking and dynamics simulations were employed to evaluate possible drug candidates against prioritized HPs.*[10]* Our integrative computational approach aims to uncover novel drug targets within the HTLV-1 proteome, paving the way for improved therapeutic interventions against HAM/TSP.*[8]*

Table 1: Characteristic Overview of HTLV-1 Infection

Feature	Details
Causative Agent	*Human T-cell Leukemia Virus type 1 (HTLV-1)*
Type	Viral, retrovirus
Transmission	Blood, sexual contact, breastfeeding, contaminated needles
Primary Symptoms	Muscle stiffness, leg weakness, back pain, bladder issues

Target System	Central Nervous System (especially the spinal cord)
Common Regions	Japan, South America, Caribbean, parts of Africa
Chronic?	Yes – progressive and lifelong
Treatment	No cure; only symptomatic treatment available

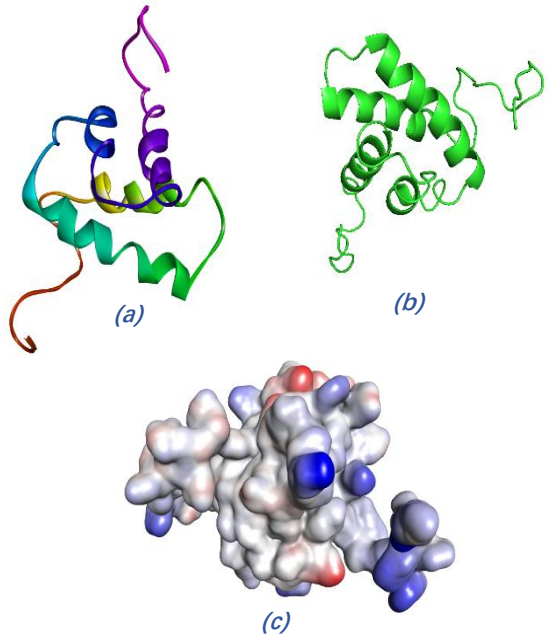


Figure 1:Structural representation of a hypothetical protein of HTLV-1. (a) Predicted 3D structure of the protein visualized in cartoon model showing predominant alpha-helices.
(b) Rainbow-colored ribbon model displaying secondary structure features including helices and loops.
(c) Electrostatic surface representation highlighting the distribution of charged and hydrophobic regions.

TABLE 2:THE FASTA FORMAT OF THE HTLV-1 PROTEIN

>7M1W_1|Chain A|Matrix protein p19|Human T-cell leukemia virus type I (11908)
MGRIFSRSASPIPRPPRGLAAHHWLNFLQAAYRLEPG PSSYDFHQLKKFLKIALETPVWICPINYSLLASLLPKGYP GRVNEILHILIQTQAQIPSRPAHHHHHH

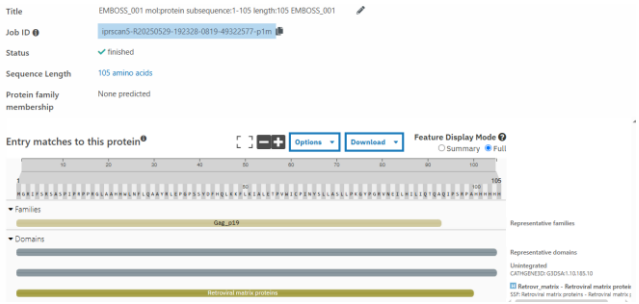


Figure 2: Domain and Motif

User-provided sequence:
10 20 30 40 50 60
MGRIFSRSAS PIPRPPRGLA AHHWLNFLQA AYRLEPGPSS YDFHQLKKFL KIALETPVWI

70 80 90 100
CPINYSLLAS LLPKGYPGRV NEILHILIQT QAQIPSRPAH HHHHH

Number of amino acids: 105
Molecular weight: 12005.98
Theoretical pI: 10.21
Theoretical pI/Mw: 10.21 / 12005.98

Table 3: Amino Acid composition

Amino Acid	Symbol	Count	Percentage (%)
Alanine	A	9	8.6%
Arginine	R	7	6.7%
Asparagine	N	3	2.9%
Aspartic Acid	D	1	1.0%
Cysteine	C	1	1.0%
Glutamine	Q	5	4.8%
Glutamic Acid	E	3	2.9%
Glycine	G	5	4.8%
Histidine	H	10	9.5%
Isoleucine	I	9	8.6%
Leucine	L	13	12.4%
Lysine	K	4	3.8%
Methionine	M	1	1.0%
Phenylalanine	F	4	3.8%
Proline	P	12	11.4%
Serine	S	8	7.6%
Threonine	T	2	1.9%
Tryptophan	W	2	1.9%
Tyrosine	Y	4	3.8%
Valine	V	2	1.9%
Pyrrolysine	O	0	0.0%
Selenocysteine	U	0	0.0%
Ambiguous	B, Z, X	0	0.0%

Total number of negatively charged residues (Asp + Glu): 4
Total number of positively charged residues (Arg + Lys): 11

Table 4: Atomic composition

Element	Symbol	Atom Count
Carbon	C	555
Hydrogen	H	852
Nitrogen	N	160
Oxygen	O	136
Sulfur	S	2

Formula: $C_{555}H_{852}N_{160}O_{136}S_2$
Total number of atoms: 1705
Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.
Ext. coefficient 16960
Abs 0.1% (=1 g/l) 1.413, assuming all pairs of Cys residues form cystines
Ext. coefficient 16960
Abs 0.1% (=1 g/l) 1.413, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).
Instability index:
The instability index (II) is computed to be 46.31
This classifies the protein as unstable.

Aliphatic index: 95.81
Grand average of hydropathicity (GRAVY):-0.257

Table 5: Physicochemical Properties of the Protein

Property	Value	Notes
Number of residues	105	Total number of amino acids
Molecular weight	12,005.83 g/mol	Indicates the mass of the protein
Extinction coefficient	16,500 $M^{-1}cm^{-1}$	At 280 nm, assuming presence of Trp/Tyr
Isoelectric point (pI)	10.59	Protein is positively charged below this pH
Net charge at pH 7	+7.9	Strongly basic protein at physiological pH
Estimated solubility	Poor water solubility	May aggregate or precipitate in solution

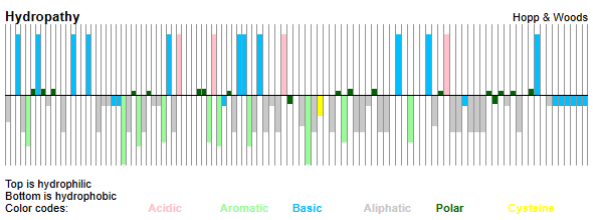


Figure 3: Hydropathy graph

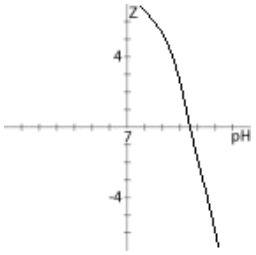


Figure 4: Net Charge Vs pH

CELLO RESULTS
SeqID: 7M1W_1|Chain A|Matrix protein p19|Human T-cell leukemia virus type I (11908)

Table 6: Subcellular Localization Analysis Report

Analysis Method	Predicted Localization	Reliability Score
Amino Acid Composition	Nuclear	0.378
N-peptide Composition	Mitochondrial	0.566
Partitioned Sequence Composition	Nuclear	0.461
Physicochemical Composition	Mitochondrial	0.763 (highest confidence)
Neighboring Sequence Composition	Nuclear	0.442

Table 7:Top Predicted Localizations

Subcellular Location	Score	Interpretation
Mitochondrial	2.138	Most likely localization
Nuclear	1.577	Also, highly probable
Extracellular	0.400	Less likely
Cytoplasmic	0.281	Unlikely
Plasma Membrane	0.229	Unlikely

Prediction of 7M1W_1|Chain
ID 7M1W_1|Chain
FT TOPO_DOM 1 105 CYTOPLASMIC.

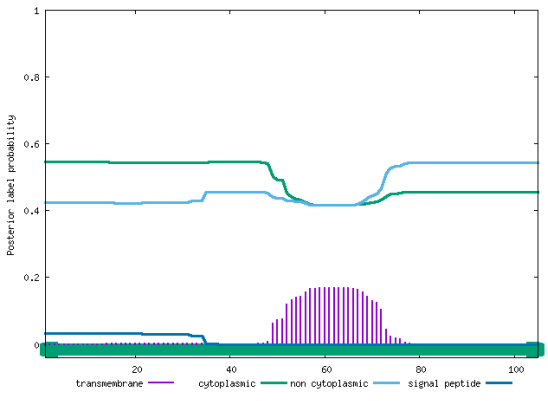


Figure 5: Phobius posterior probabilities for 7M1W_1|Chain

SecretomeP 1.0f Prediction Summary

In the context of SecretomeP 1.0f analysis, non-classically secreted proteins are characterized by a Neural Network (NN) score exceeding the standard threshold of 0.6, **without** the concurrent prediction of a signal peptide. Users are advised to consult the "Warning" column for indications of signal peptide presence as predicted by SignalP.

Prediction Results:

Table 8: Predicted Result

Name	NN-score	Odds by Prior	Weighted	Warning
7M1W_1_Chai	0.874	5.044	0.010	–

The protein **7M1W_1_Chain** exhibits an NN-score of **0.874**, which significantly exceeds the threshold of 0.6, indicating a strong likelihood of non-classical secretion. No signal peptide was predicted for this sequence, as reflected by the absence of a warning, thereby supporting its classification as a non-classically secreted protein.

Predicted proteins
7M1W_1_Chain
Prediction: Other

Table 9: Other Prediction

Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.0037	0.9963

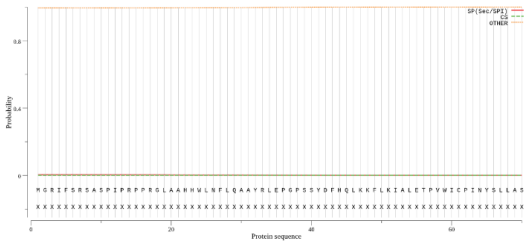


Figure 6: SignalP-5.0 prediction (Eukarya): 7M1W_1_chain

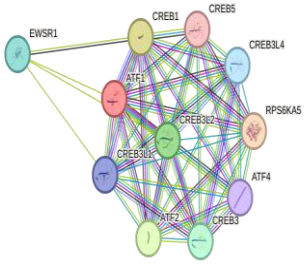


Figure 7: Protein-protein interaction of atf-1 protein

- Each node represents a protein. Each node corresponds to all protein products from a single gene locus.
- Edges represent functional associations, not just physical interactions and they likely work together in same biological process

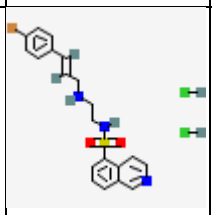
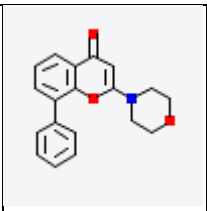
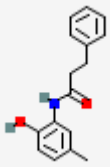
Table 10: Protein interaction chart legend

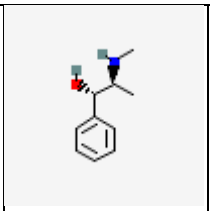
Gene Symbol	Protein Name	Function Summary
ALB	Serum albumin	Major plasma protein; binds water, ions, fatty acids, hormones, and drugs. Regulates osmotic pressure.
INSR	Insulin receptor	Tyrosine kinase receptor; mediates insulin's metabolic effects.
IRS1	Insulin receptor substrate 1	Mediates insulin signaling; activates PI3K when phosphorylated.
IGF1	Insulin-like growth factor I	Promotes cell growth; similar to insulin but more potent in growth functions.

IGF1R	IGF1 receptor	Receptor tyrosine kinase for IGF1; important for cell survival and proliferation.
IRS2	Insulin receptor substrate 2	Similar role to IRS1; mediates insulin's cellular effects.
GCG	Glucagon	Counteracts insulin; raises blood glucose during hypoglycemia.
NTRK1	Neurotrophic receptor tyrosine kinase 1	Binds NGF; critical for neuron survival and differentiation.
EGF	Epidermal growth factor	Stimulates epithelial cell growth; involved in magnesium transport.

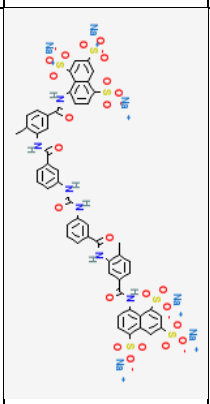
Inhibitors collected:

Table 11: Inhibitor's Chemical Structure and Toxicity

Inhibitor	Chemical Structure	Toxicity
H-89 Dihydrochloride		**Parameter** / **Value** AMES toxicity: No Max. tolerated dose (human): 0.017 hERG I inhibitor: No hERG II inhibitor: Yes Oral Rat Acute Toxicity (LD50): 2.322 Oral Rat Chronic Toxicity (LOAEL): 1.315 Hepatotoxicity: Yes Skin Sensitisation: No T.Pyriformis toxicity: 0.306 Minnow toxicity: -1.905
2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one		**Parameter** / **Value** AMES toxicity: Yes Max. tolerated dose (human): 0.149 hERG I inhibitor: No hERG II inhibitor: No Oral Rat Acute Toxicity (LD50): 2.307 Oral Rat Chronic Toxicity (LOAEL): 1.369 Hepatotoxicity: Yes Skin Sensitisation: No T.Pyriformis toxicity: 0.401 Minnow toxicity: -1.002
N-(2-hydroxy-5-methylphenyl)-3-phenylpropanamide		**Parameter** / **Value** AMES toxicity: No Max. tolerated dose (human): -0.033 hERG I inhibitor: No hERG II inhibitor: No Oral Rat Acute Toxicity (LD50): 2.257 Oral Rat Chronic Toxicity (LOAEL): 1.555 Hepatotoxicity: No Skin Sensitisation: No T.Pyriformis toxicity: 1.358 Minnow toxicity: 1.753

Pseudoephedrine		

Parameter	Value
AMES toxicity	No
Max. tolerated dose (human)	-0.359
hERG I inhibitor	No
hERG II inhibitor	No
Oral Rat Acute Toxicity (LD50)	2.942
Oral Rat Chronic Toxicity (LOAEL)	1.369
Hepatotoxicity	Yes
Skin Sensitisation	Yes
T.Pyriformis toxicity	-0.023
Minnow toxicity	1.887

Suramin Sodium		

Parameter	Value
AMES toxicity	No
Max. tolerated dose (human)	0.438
hERG I inhibitor	No
hERG II inhibitor	No
Oral Rat Acute Toxicity (LD50)	2.482
Oral Rat Chronic Toxicity (LOAEL)	5.052
Hepatotoxicity	No
Skin Sensitisation	No
T.Pyriformis toxicity	0.285
Minnow toxicity	2.736

Docking results:

Table 12: Result of Docking

Ligand	Score
H-89 Dihydrochloride	-6.6 (C5)
Suramin Sodium	-9.5(C1)
Pseudoephedrine	-4.1(C5)
2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one	-6.6(C4)
N-(2-hydroxy-5-methylphenyl)-3-phenylpropanamide	-5.9(C3)

Results

In this study the toxicity profiling of these compounds indicated that Suramin Sodium possessed the most favorable safety characteristics, exhibiting no AMES toxicity, hepatotoxicity, or hERG inhibition. In contrast, Pseudoephedrine showed the weakest binding affinity (−4.1 kcal/mol) and was associated with hepatotoxicity and skin sensitization risks, rendering it unsuitable for therapeutic application.

Physicochemical analysis of one modeled protein (Matrix protein p19) revealed an instability index of 46.31 (indicative of instability), a basic isoelectric point (pI = 10.59), and low solubility, which may influence its behavior in a cellular environment. Structural features such as high alpha-helical content and predicted protein-protein interactions with key host signaling proteins (e.g., IGF1R, INSR) suggest functional relevance in viral replication or host manipulation.

Collectively, the results highlight as a prioritized druggable target with strong ligand-binding potential and favorable pharmacological interactions, laying the foundation for further experimental validation.

Conclusion

Based on the docking scores and toxicity profiles, Suramin Sodium emerges as the most promising inhibitor, demonstrating the strongest binding affinity (-9.5) and a favorable toxicity profile with no AMES toxicity, hepatotoxicity, or hERG inhibition. H-89 Dihydrochloride and 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one show moderate binding affinities (-6.6) but are associated with notable toxicities—H-89 with hepatotoxicity and hERG II inhibition, and the morpholinyl-benzopyran derivative with a positive AMES test, indicating potential mutagenicity. N-(2-hydroxy-5-methylphenyl)-3-phenylpropanamide has a slightly weaker binding score (-5.9) but stands out for its excellent safety profile, making it a potential candidate for further optimization. In contrast, Pseudoephedrine, with

the lowest docking score (-4.1) and concerns of hepatotoxicity and skin sensitization, appears least suitable as a therapeutic inhibitor.

References

[1] K. Matsuoka and K.-T. Jeang, "Human T-cell leukemia virus type 1 (HTLV-1) and leukemic transformation: Viral infectivity, oncogenesis, and cellular signal pathways," *Oncogene*, vol. 26, no. 33, pp. 6021–6030, 2007.

[2] A. Gessain and A. Cassar, "Epidemiological aspects and world distribution of HTLV-1 infection," *Front. Microbiol.*, vol. 3, p. 388, 2012.

[3] J. C. Patricio, R. A. T. Alves, and R. B. Leal, "HTLV-1 associated myelopathy/tropical spastic paraparesis: Pathogenesis and biomarkers," *Rev. Soc. Bras. Med. Trop.*, vol. 53, e20190469, 2020.

[4] A. Bateman, L. Coin, R. Durbin, R. Finn, V. Hollich, S. Griffiths-Jones, et al., "The Pfam protein families database," *Nucleic Acids Res.*, vol. 32, pp. D138–D141, 2004.

[5] R. S. Kumar, P. A. Harilal, and C. G. P. Doss, "In silico characterization and functional annotation of hypothetical proteins of Mycobacterium leprae," *3 Biotech*, vol. 6, no. 2, pp. 147–157, 2016.

[6] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, et al., "STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, Jan. 2019.

[7] D. J. McArthur and J. M. Bibb, "The annotation and prediction of hypothetical proteins," *Curr. Opin. Struct. Biol.*, vol. 18, no. 3, pp. 357–364, 2008.

[8] P. Daina, O. Michielin, and V. Zoete, "SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules," *Sci. Rep.*, vol. 7, p. 42717, Mar. 2017.

[9] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, Jan. 2010.

[10] M. L. Benson, R. J. Smith, T. L. Eaton, and P. A. FitzGerald, "Structure-function relationships in drug design: Molecular docking and pharmacophore modeling of viral proteins," *Biophys. Rev.*, vol. 3, no. 1, pp. 35–45, 2011.