Protein Sequence Analysis and Homology Modeling of the Human T-cell Leukemia

Uma Kumari[1], Tanishka Uttam[2]
[1]*Bioinformatics Scientist (Bioinformatics Project and Research Institute), Noida, India*
[2]*Project Trainee at Bioinformatics Project and Research Institute, Noida, India*

Abstract- The human deals with many diseases emerging from many sources, but only a few of them are curable. Diseases caused by viruses as sources are the most difficult ones to deal with. HTLV-1(Human T cell lymphotropic/leukaemia type-1) comes under retrovirus classification. In most cases, no sure cure is present till nowadays. The current status the estimated at 5 million to 10 million. The causative agents of both adult T cell leukaemia (ATL) and HTLV-1 Associated Myelopathy Spastic Paraparesis (HAM/TSP). The main geographical regions are the Southwestern part of Japan, Sub-Saharan Africa and South America. The initial stage of infection is unknown but certain studies show that the virus attack $CD4^+$ T cells, $CD34^+$, B lymphocytes, T lymphocytes and hematopoietic progenitor cells etc. Through an In-silicon approach for designing the drug against this protein expressed by HTLV-1 cells using bioinformatics database and software. Designing of ligands for substrate (HTLV-1) was performed. It was identified that the drug Allopathy Glatiramer, Plant-based drug Oleandrin and antibodies Doxycycline showed partially able to cure the T cell leukaemia virus replication protein structure.

Keywords: HTLV-1; T cell leukemia; carcinogenesis; miRNAs, Homology modelling, Molecular Docking, Structural Analysis

1. INTRODUCTION

1.1. T cell leukaemia virus

Human T– Cell leukaemia is infectious to human T lymphatic cells, belonging to the retrovirus family. There are two types – HTLV – 1 (type – 1) and HTLV - 2 (type – 2) (*Human T-Lymphotropic Virus Type 1*, n.d.). Type 1 (HTLV-1) viruses are mostly recorded cases causing T cell leukaemia diseases. HTLV-1 infection generally remains asymptomatic at the initial stages of infection. Its uncontrolled pace causes demyelinating diseases called HTLV1 in which the myelin sheath of the neuron got damaged and ultimately leads the paralysis, loss of movement, and sensing in the individual. ATL causes the proliferation and accumulation of infected cells carrying genes responsible for diseases that get integrated into the host genome forming the initial stage provirus form of the virus. ATL is divided into five different stages: asymptomatic, pre-leukaemia, chronic, smouldering, lymphoma and acute. HAM/TSP is characterized by progressive, spasticity, hyperreflexia, sensory disturbances and urinary incontinence. The patients who are suffering from this disease show symptoms like multiple sclerosis, and the signs and symptoms wane. Two main lymphocytes are the most studied $CD4^+$ and $CD8^+$ cells. $CD4^+$ T lymphocytes represent a primary cellular target for HTLV-1 to proliferate in vitro and express proinflammatory cytokines which cause traffic from the PB into the central system. $CD8^+$ cells represent the additional cell population for the viral reservoir (Nagai et al., 2001b).CD8+T and B lymphocyte dendritic cells etc. HAM/TSP is characterized by progressive, spasticity, hyperreflexia, sensory disturbances and urinary incontinence. The patients who are suffering from this disease show symptoms like multiple sclerosis, and the signs and symptoms wane. Cells such as astrocytes, dendritic cells and microglial cells are derived from PB monocytes which are infected with HTLV-1 cells in vivo and in vitro.

1.2. Genetic arrangement

The HTLV-1 has a single-stranded RNA genome of 9 kb, which encodes structural and enzymatic proteins gag, env, and pol. The virus consists of RNA stranded into a double DNA structure and incorporated into the host cell. The HTLV-1 long terminal repeats located at 5’ and 3 and ends of the viral genome have promoter and regulatory elements and are distributed into three main regions U3, R and U5 regions. The unique region at the 3’end designated pX region codes for two main

regulatory proteins named Tax and Rex and only mwith the promoter and other regulatory elements such as U3, R and U5 regions end and 5' end of the viral genome having the promoter and other regulatory elements such as U3, R and U5 regions which are responsible for provirus transcription and mRNA termination and polyadenylation signals.The presence of 21 base pairs of nucleotides repeats TRE (Tax responsive elements) used for transcription of the *Tax* gene. The full-length mRNA is responsible for many processes such as gag proteins (p55) cleaved by viral proteases to yield matrix (MA, p19) protein.The domain organization of Tax is depicted as having 48 amino acids and the amino-terminal domain interacts with a cellular transcription factor, cyclic AMP responsive element binding protein (CREB) and serum response factor overlap which interact with the and are responsible for HTLV-1 LTR trans- activation gen during cellular transformation. The Tax interacts with CREB and coactivators p300 CBP on three 21bp repeats in HTLV-1 LTR report to form a stable formation of dimers which bound to the DNA. (Giebler et al., 1997).The proteases encode the by-reading fraSeveral spliced products are derived from the X region's open reading frame (ORF) I, II, III, and IV. Rex-mediated regulation is required to balance the spliced and unsliced mRNA necessary for the production of infectious viruses PTHrP mRNA is overexpressed in ATL cell types which is responsible for indolent subtypes and causes renal dysfunction. The peripheral blood smear condensed chromatin with a convoluted or polylobate nucleus called a flower cell or cloverleaf path genomic for ATL. (Jeang et al., 1990).There is an amalgam and interaction between many proteins and transcription factors. Some important genes and factors are listed below the Table 1.

Protein	Function
Regulatoray factors –U3, R and U5	Proviral transcriptional and termination and polyadenylation signals
TRE (Tax responsive elements)	transcription activation
Gag protein/ p55	Formation of capsid, nucleocapsid
env protein	Formation of envelope and involvement in mRNA formation.
Tax	Enhancer repeats sequence, responsible for proliferation gene during cellular transformation.
Rex	Viral replication and expression of post –transcription stage of replication.
$p13^{II}$ and $p30^{II}$	transcriptional of cellular genes and potentially relevant for HTLV-1 replication or pathogenicity
$p12^{I}$	Host cell activation

1.3. Transmission and Symptoms

HTLV-1 primarily through direct contact breastfeeding, semen and cervical secretion, injection of the drug by the same syringe or blood transfer lead to the easy transmission of the virus to the individuals. In general terms, there are no early symptoms of detecting the leukaemia virus. Lower back pain and bowel and bladder dysfunction is been reported in HAM/TSP diagnosis one while the person who is suffering from ATL shows lymphadenopathy, hepatosplenomegaly, and hypercalcemia (increased amount of calcium).The cancer symptoms like infective dermatitis, bronchiectasis, rheumatoid arthritis and fibromyalgia.

1.4 Treatment, Prevention and Diagnosis

There is no efficient cure possible for these diseases nor any biological marker of predicting the development of quantities the risk of diseases although the levels of HTLV-1 provirus load. Certaintechniques like Antiviral therapy, Chemotherapy, Allogeneic stem cell transplantation, and Monoclonal antibody (Brentuximab vedotin, Alemtuzumab, Daclizumab and Mogamulizumab).An ongoing approach for building the derived B cell epitopes for induction of neutralizing antibodies for preventing and therapy against HTLV-1 infection. Prevention the can be taken such as Cessation of breastfeeding, Breast freeze-thaw methods and Antibody screening among blood donors). Diagnosis through blood tests computed tomography,biopsy and molecular testing).

2. MATERIAL AND METHODS

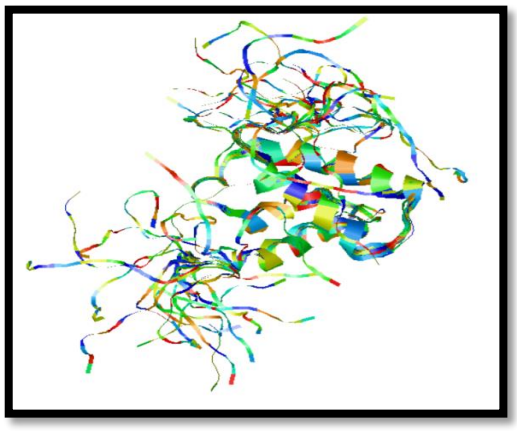
The structure of the protein was taken from the PDB server and 3D structure visualization was done with

the help of RasMol with different selection parameters such as temperature, shapely, group and structure. Pymol tools then further are used for visualizing the secondary structure and presence of any ligands site. The two-drug data against the protein structure were taken from Drug Bank Database which is allopathy Glatiramer and antibody Doxycycline. Through literature the third drug which is plant-based Oleandrin was taken. SDF file chemical structure of the ligands (Drug) was taken from PubChem Database. Chain A, Matrix protein p19 protein was selected for Multiple Sequence Alignment under NCBI Database and conserved Sequence was analysis. Panorama view, Rasmol view, BLOSUM 46 view and Phylogenetic tree done by COBALT tools build within NCBI database and identification of ORF reading Frame was done by ORF Finder. Through SmartBlast within NCBI database taxonomy position of our sequence against the landmark database (NC_001436.1) was observed. The Pathway of the molecular interaction, reaction and relation networks. T cell leukaemia virus was prepared by the KEGG Pathways Database. The Ligands and Receptor Structure was prepared by the help of auto dock bioinformatics tools and their respective Grid Score was analysis and prepared structure was visualized in PyMol

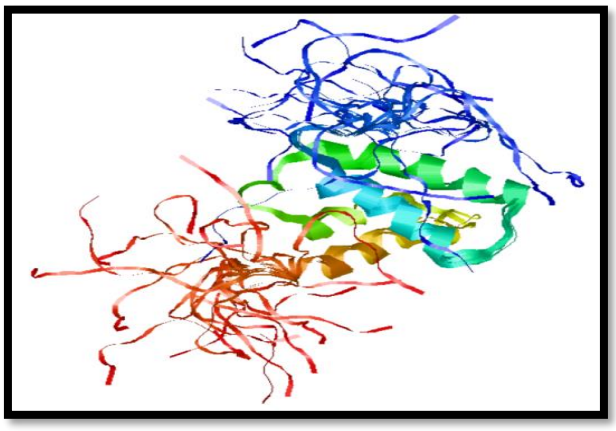
3. RESULTS

T cell leukaemia virus type -1 is classified under retrovirus. The protein structure is designed with the help of the NMR spectroscopy technique. The protein structure consists of 105 amino acid units. The virus is expressed and replicated only in Homo sapiens organisms. The structure consists of a linear arrangement of chain A protein. The structure consists of alpha helix and loops. The major core of Gag 19. BLAST run of the protein show a very low least E value. The query coverage lies range of 78%. conserve domain sequence belongs to the Gag_p19 superfamily which is an important inner protein layer of viral cell structure. The range of superfamily ranges from amino acids 18^{th} to 92^{th}. Paranormal sequenced view of the query is been represented by multiple sequence alignment tools COBALT. The query is highly mutable in nature and represented by a grey colour. The presence of a mismatch sequence is founded and represented by blue region in. Rasmol feature present in the COBALT. The phylogenetic tree represented features COBALT show that p19MA belong to the virus. The PyMol analysis shows that there is the presence of two main domain S1 and S2. Through Auto docking, the structure of the protein was treated with three ligands. Glatiramer ligand docks protein structure at N terminal at first amino acids methionine. Oleandrin ligand is able to dock protein structure at Glycine and Doxycycline dock at glycine and arginine amino acids. Since the spacing and dimension of the grid box dimension does not > -10 value although the spacing value lies less than 2 so the drug can partially contribute for curing this disease The binding of the ligand was at starting N terminal of the protein structure which also provides the advantage for quick repression of the protein to perform its function. The residue amino acid were mainly glycine, methionine and arginine.

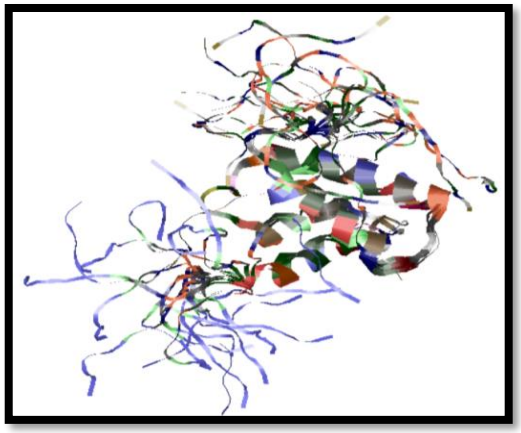
1.RasMol



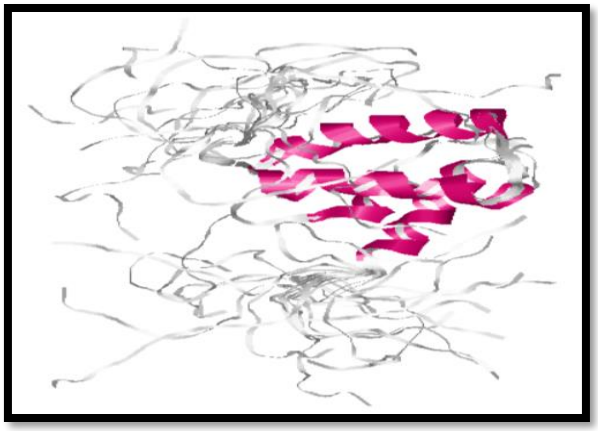
1 a)



1 b)



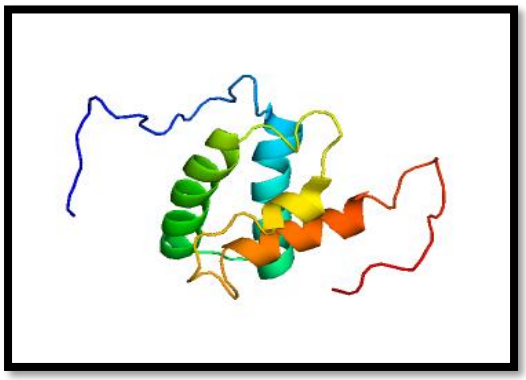
1 c)



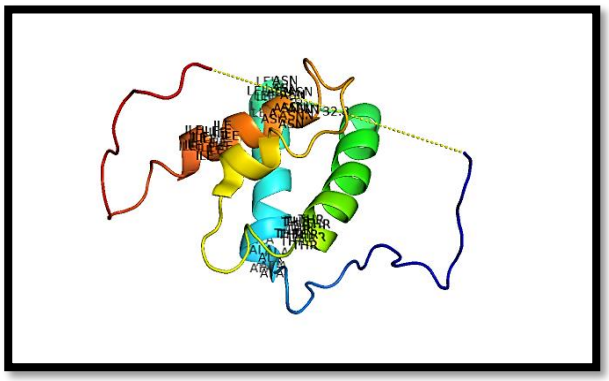
1 d)

Figure 1 - 3D presentation in software RasMol of 7M1W protein 1a) Temperature, 1b) Group, 1c)Shaply and 1d)Structure. 1a) The structure's main temperature colour is in blue which indicate that lower values of the structure. 1b) The N terminal of the proteins ae colored in blue and C terminal are coloured red . There is presence of alpha helix loop and tertiary structure are been observed. 1c) The shapely represent various amino acids present in which the N terminal consits of Methionine and Cysteine residue, while the C terminal mostly consists from the end portion Histine amino acid. 1d) The structure colour scheme indicates magenta which denotes alpha helices and white-colour denote all other residues.

1. PyMol



2a)



2 b)

Figure 2- a) Dimensional cartoon of the structure 7M1W show in PyMol software. b) The structure of the protein consists of an alpha helix, motif and multiple loops. N- Terminal (Blue region) and C terminal (Red) are apart from each other at a distance of 32.8 cm. the structure of the protein is made of two main domains.

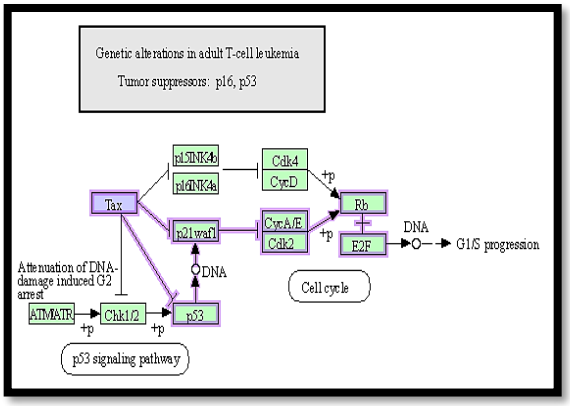
Pubchem 2D SDF file for chemical molecule Glatiramer, 3b) Oleandrin and 3c) Doxycycline

Multiple Sequenced Alignment of our Query Sequence Gag19 gene and 4 b) Conserved domain sequence of the query.

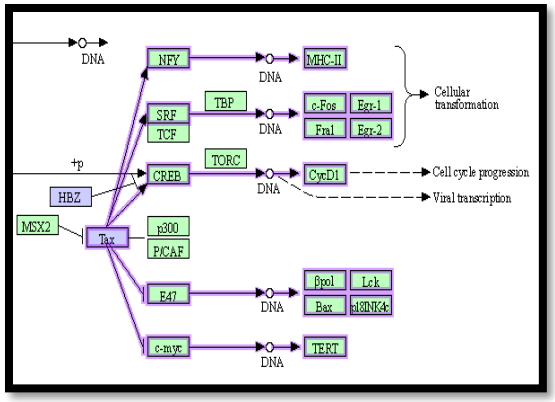
Auto docking T-Cell Leukemia Virus 3D structural viewed in Autodock.

Observation 3 Dimensional properties of the molecule 7M1W T- cell leukaemia virus protein gag using MMDB (Molecular Modelling Database) database. The structure of the protein was quantified with the help of the NMR (Nuclear Magnetic Resonance) technique. The PDB ID is 7M1W and the MMDB ID is 206805.

B.KEGG Pathways

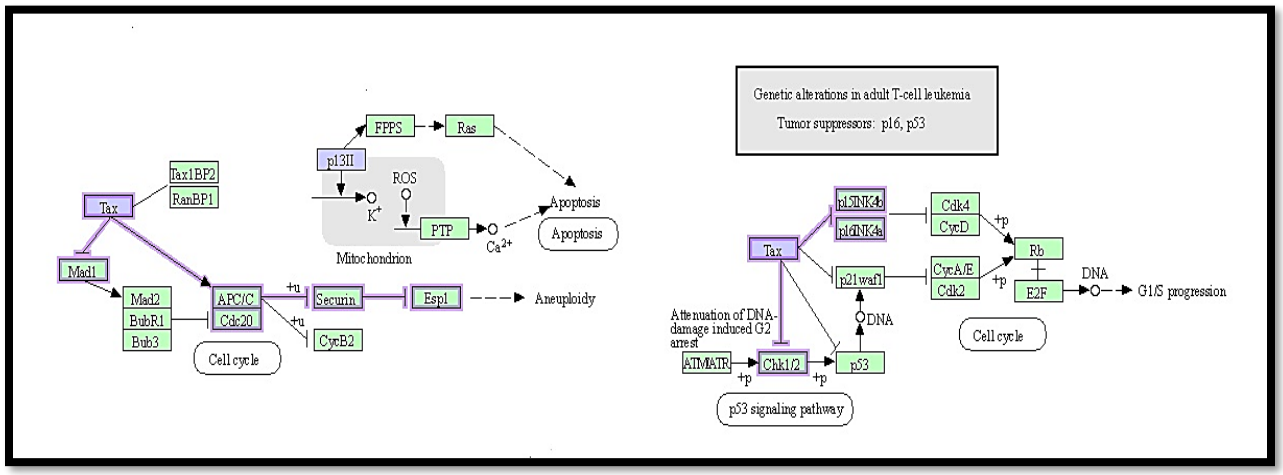


B.1)



B.2)

B.3)



B.4)

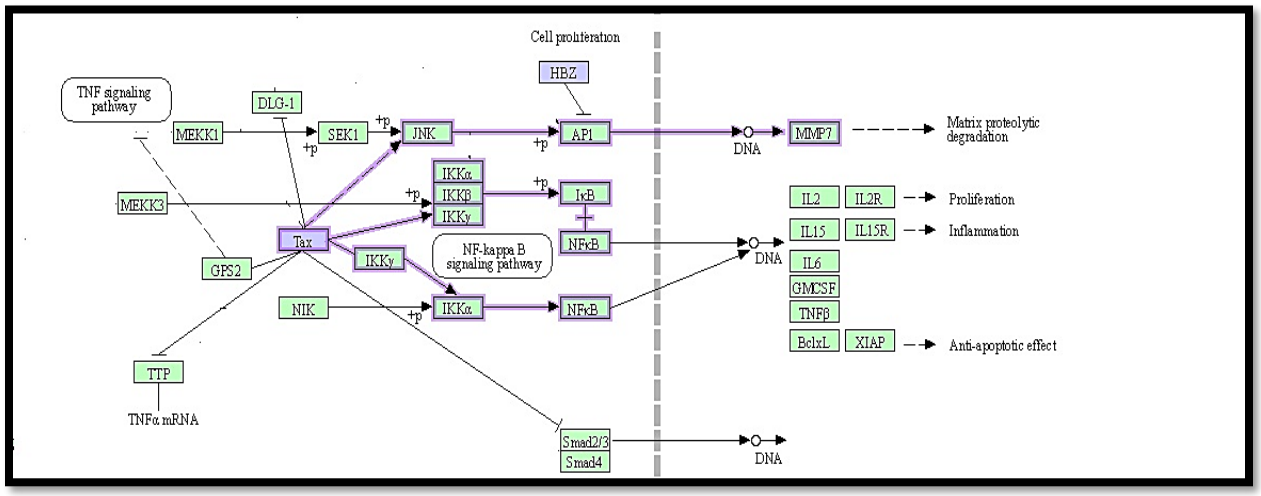


Figure- 3.B1) Human T cell Leukemia Virus p53 Pathways, B.2) Transcription expressing of the virus B.3) Cell Cycle of the Leukemia Virus and B.4) TNF Signalling pathway

Description of the T-cell leukaemia virus matric protein gag19 protein. The structure is expressed and replicated only in the human sapiens organisms. The protein consists of mainly three secondary structures and has linear form of conformation. The main major core region begins from the First amino acids structure and ends at ninety-nine amino acids structure.

Figure 4) –BLAST alignment of query sequence (T cell leukaemia virus gag protein) with first 10 sequence present in the database. The query cover is around 79-78% and expressed in Human sapiens.

4a)

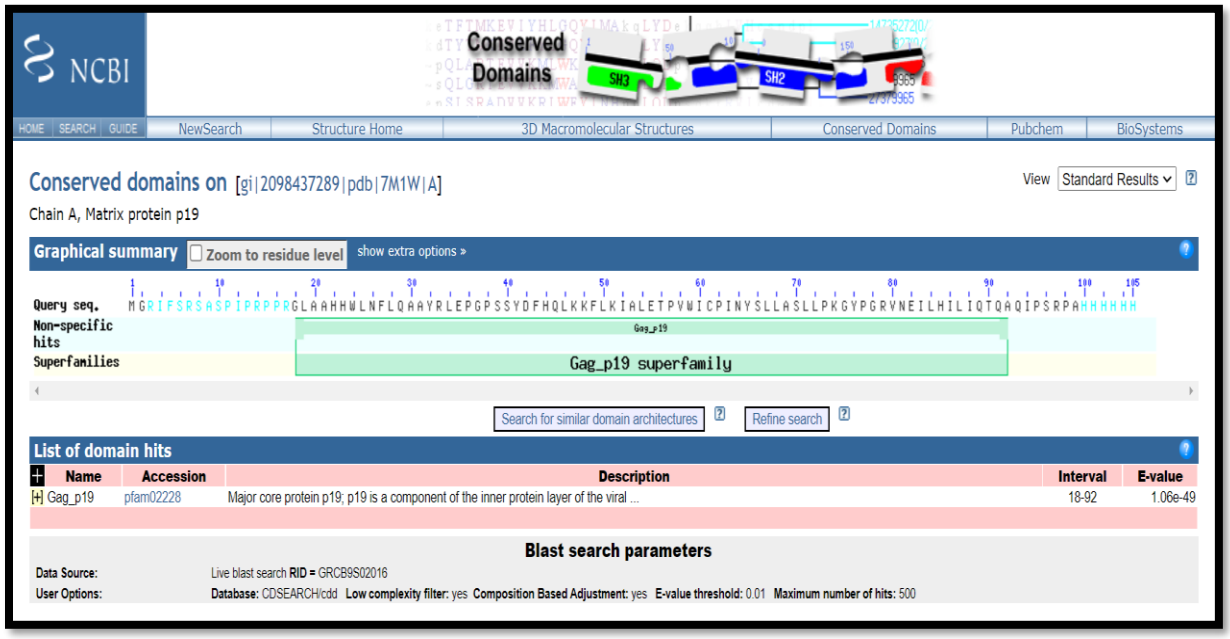


Figure 5) – Conserved Domain Sequenced of T cell leukaemia virus coding protein. The protein is consist of Gag_p19 superfamily protein which is a component of the inner protein layer of the virus cell. The blue colour region represents the non-specific region which can vary from species to species.

Graphical representation of the BLAST results showing poor alignment pairing and hence high rate of mutation replicability.

COBALT

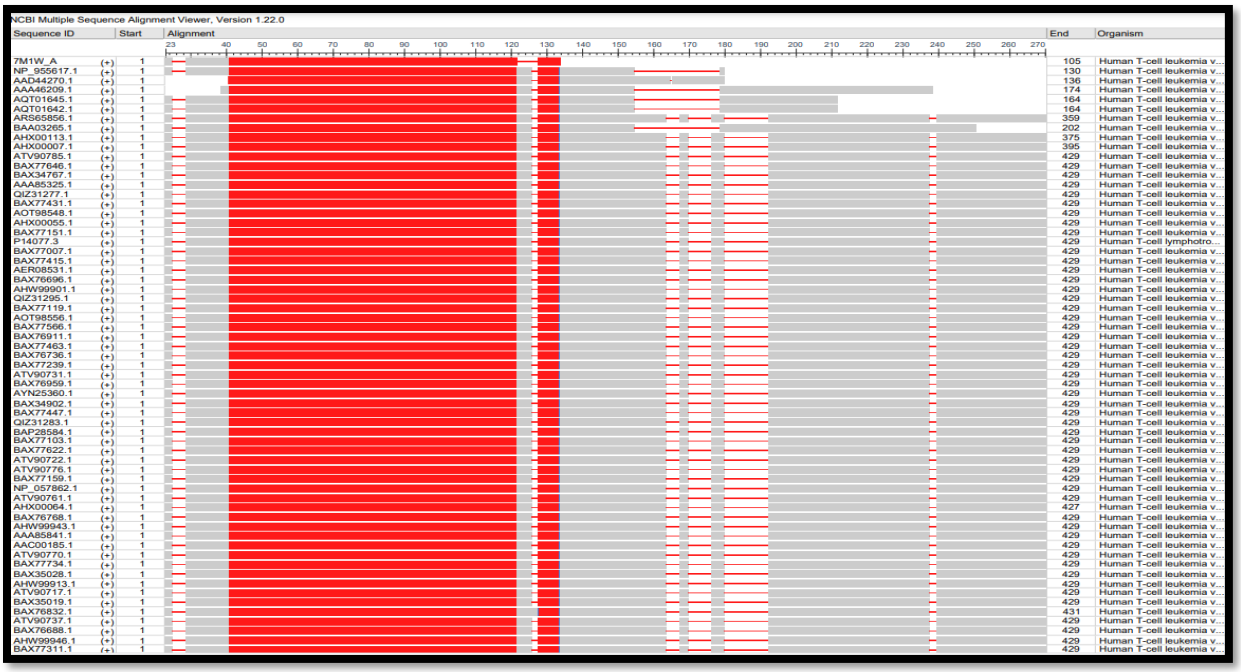


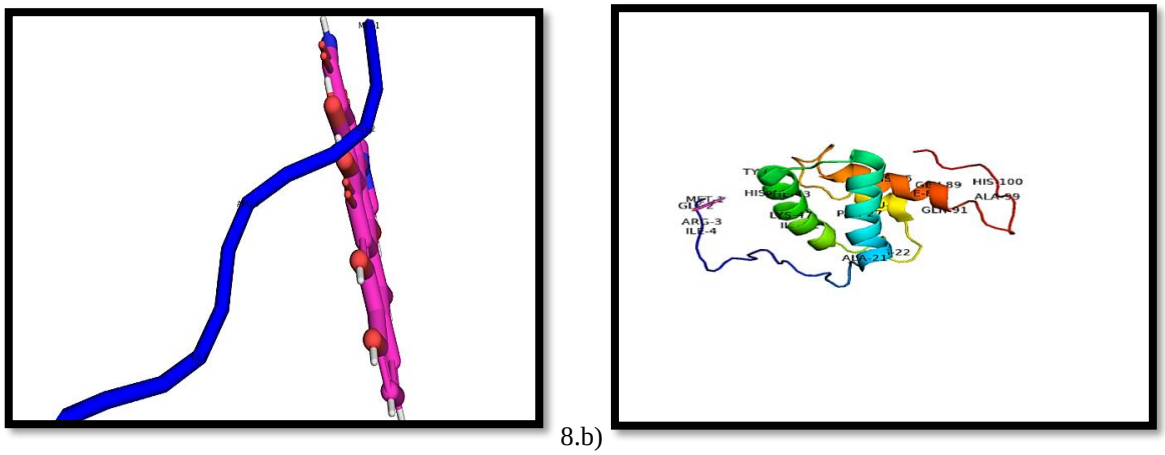
Figure 6 – Multiple Sequence Alignment using COBALT tools of the query sequence. A paranormal view of the query sequence is been observed.

A) ORF Finder ORF Finder tools represented various ORF. The six frame translation represent the various start and stop codon of different ORF. Figure - E.2) Various ORF frames are presented. + Strands represented different frames and – strands represent the complementary strands.

B) Smart BLAST

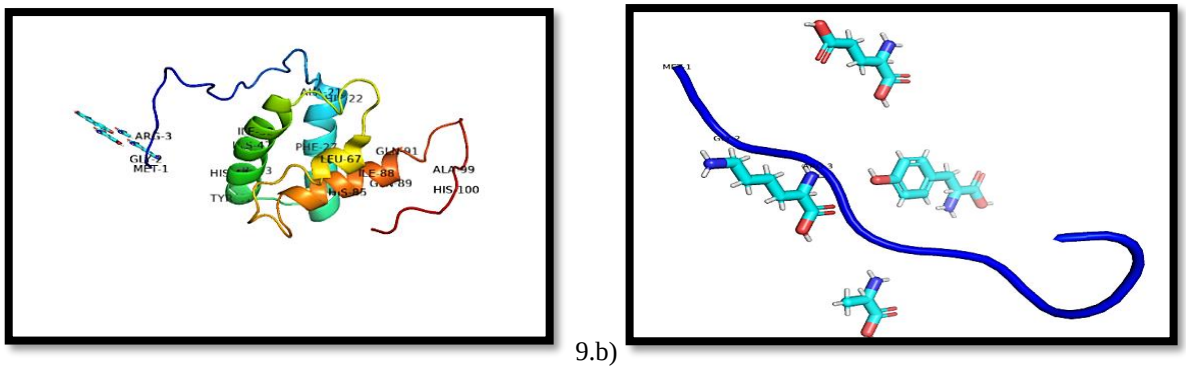
Figure 7 - Smart BLAST of ORF 2 belongs to the vertebrate. There are few matches representing the green region.

G) Auto Docking



8.a) 8.b)

Figure 8.a) Glatiramer (ligand) binding at the N terminal of the structure 8.b)glatiramer with a full structural view of T cell leukaemia virus structure

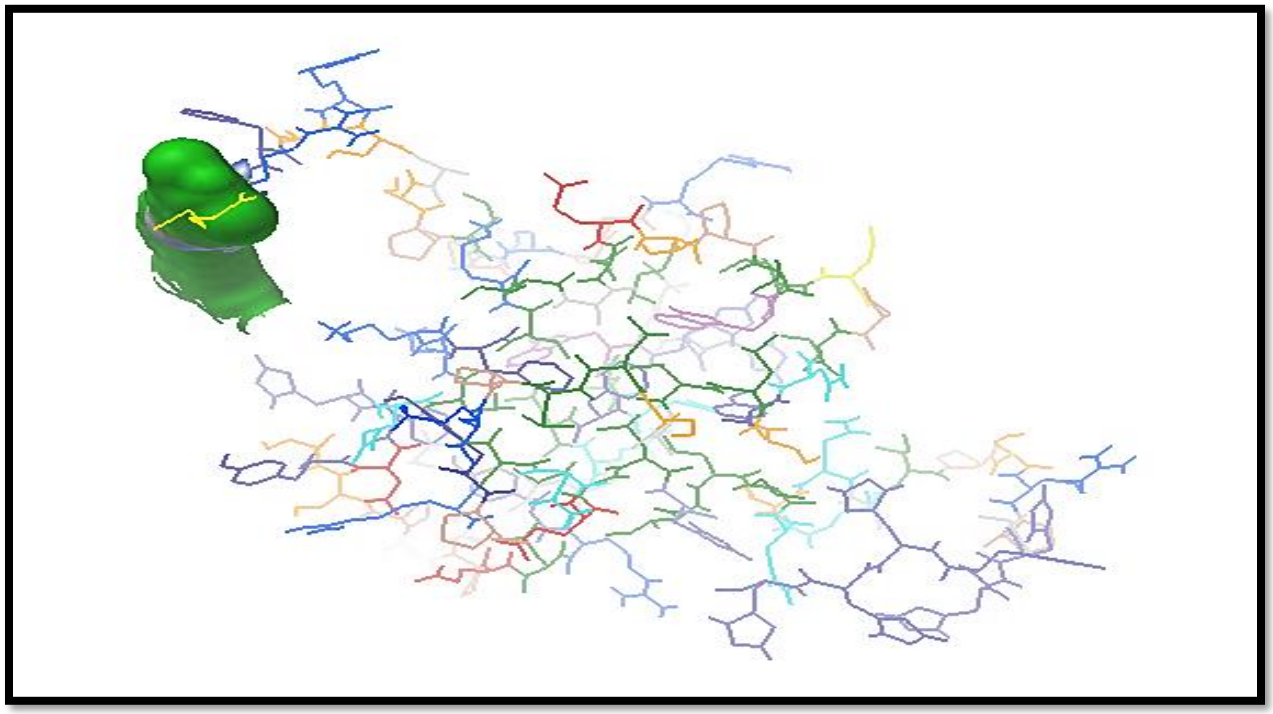


9.a) 9.b)

Figure 9 a) Oleandrin (ligand) binding at N terminal of the structure 8.b) Oleandrin with a full structural view of T cell leukaemia virus structure

Figure10.a) Doxycycline (ligand) binding at the N terminal of the structure 9.b) Doxycycline with a full structural view of T cell leukaemia virus structure.

11.a)



11 b)

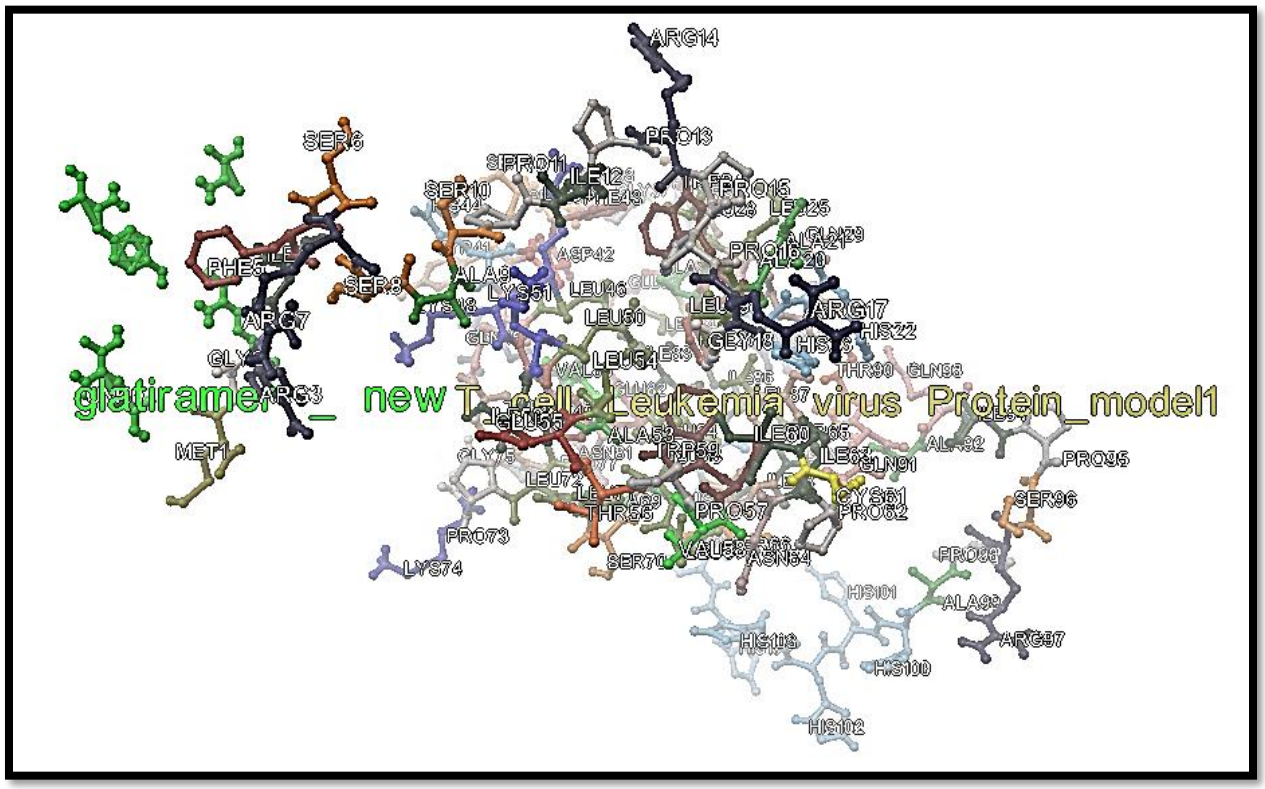
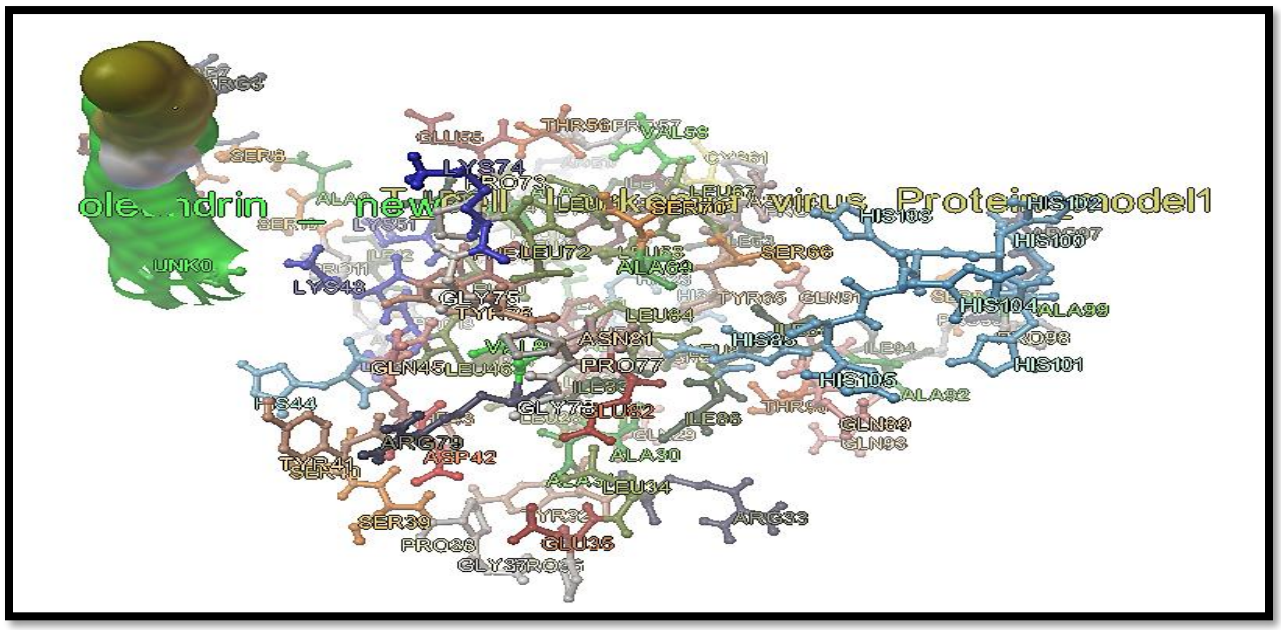


Figure 11.a) Interaction of the ligand (Glatiramer) with T cell leukaemia virus protein structure.11.b) 3D presentation of the interaction between ligand (Glatiramer) with T cell leukaemia virus protein structure) shapely presentation of the interaction of the ligand with target protein structure with labelled residue amino acids.

12.a)



Interaction of the ligand (Oleandrin) with T cell leukaemia virus protein structure. 12.a) 3D presentation of the interaction between ligand (Oleandrin) with T cell leukaemia virus protein structure) shapely presentation of the interaction of the ligand with target protein structure with labelled residue amino acids.

13.a)

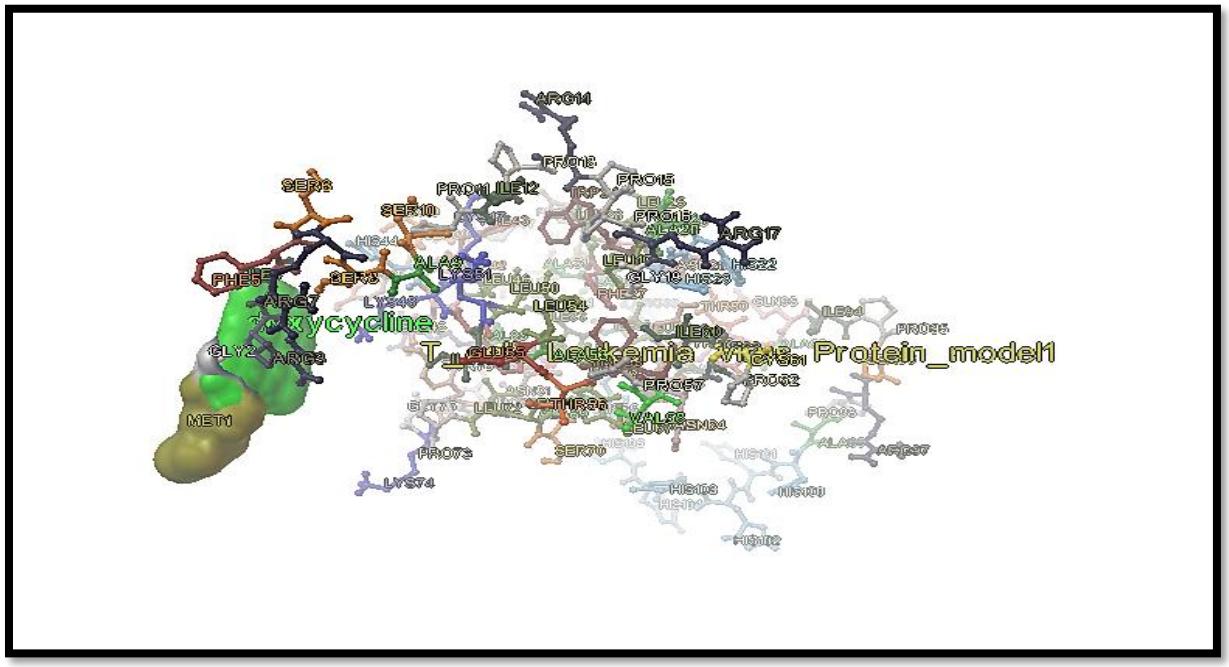


Figure 13 a.Interaction of the ligand (Doxycycline) with T cell leukaemia virus protein structure.G.6a) 3D presentation of the interaction between ligand (Doxycycline) with T cell leukaemia virus protein structure) shapely presentation of the interaction of the ligand with target protein structure with labelled residue amino acids.

Grid dimension

Serial number	Ligand	X coordinate	Y coordinate	Z coordinate	Spacing
1	Glatiramer	15.437	-16.922	10.902	0.375
2	Oleandrin	15.437	-16.922	10.902	0.356
3	Doxycycline	15.437	-16.922	10.902	0.339

Figure 13. b – Grid Dimension of the ligands against T-cell leukaemia virus.

CONCLUSION

Human T cell leukaemia virus has been classified under retrovirus of type -1. It causes cancer termed ATL (Adult T cell Lymphocyte). The diseases have been expressed and replicate in Homo sapiens organisms. The disease has been transferred through blood, breast milk and sexual intercourse. To previous literature and database like Drug Bank and PubChem one allopathic drug Glatiramer, one plant-based phytochemical substance named Oleandrin and one monoclonal antibody name doxycycline was taken for the in- silica drug design against of the protein expressed by the T cell leukaemia virus. The protein is linear structure consists of Alpha helix and loops. Through various bioinformatics tools and technique, the drug can binds to the at the first starting amino acids methionine, Glycine and Arginine. The drug can be used for a partial treatment of the diseases.

REFERENCE

[1] Bading, H., Beutler, C., Beimling, P., Gerdes, J., Stein, H., & Moelling, K. (1987). Monoclonal antibodies against the viral and human cellular myb gene product. *Haematology and Blood Transfusion*, *31*, 488–492. https://doi.org/10.1007/978-3-642-72624-8_104

[2] Bhat, R. V., Axt, K. J., Fosnaugh, J. S., Smith, K. J., Johnson, K. A., Hill, D. E., Kinzler, K. W., & Baraban, J. M. (1996). Expression of the APC tumor suppressor protein in oligodendroglia. *Glia*, *17*(2), 169–174. https://doi.org/10.1002/(SICI)1098-1136 (199606)17: 2<169: AID-GLIA8>3.0.CO;2-Y

[3] Durer, C., & Babiker, H. M. (2021). Adult T cell Leukemia. In *StatPearls [Internet]*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK558968/

[4] Giebler, H. A., Loring, J. E., van Orden, K., Colgin, M. A., Garrus, J. E., Escudero, K. W., Brauweiler, A., & Nyborg, J. K. (1997). Anchoring of CREB binding protein to the human T-cell leukemia virus type 1 promoter: A molecular mechanism of Tax transactivation. *Molecular and Cellular Biology*, *17*(9), 5156–5164.

[5] Gonçalves, D. U., Proietti, F. A., Ribas, J. G. R., Araújo, M. G., Pinheiro, S. R., Guedes, A. C., & Carneiro-Proietti, A. B. F. (2010). Epidemiology, Treatment, and Prevention of Human T-Cell Leukemia Virus Type 1-Associated Diseases. *Clinical Microbiology Reviews*, *23*(3), 577–589. https://doi.org/10.1128/CMR.00063-09

[6] Hanly, S. M., Rimsky, L. T., Malim, M. H., Kim, J. H., Hauber, J., Duc Dodon, M., Le, S. Y., Maizel, J. V., Cullen, B. R., & Greene, W. C. (1989). Comparative analysis of the HTLV-I Rex and HIV-1 Rev trans-regulatory proteins and their RNA response elements. *Genes & Development*, *3*(10), 1534–1544. https://doi.org/10.1101/gad.3.10.1534

[7] Hidaka, M., Inoue, J., Yoshida, M., & Seiki, M. (1988). Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. *The EMBO Journal*, *7*(2), 519–523.

[8] *Human T-lymphotropic virus type 1*. (n.d.). Retrieved July 17, 2022, from https://www.who.int/news-room/fact-sheets/detail/human-t-lymphotropic-virus-type-1

[9] Jacobson, S., Shida, H., McFarlin, D. E., Fauci, A. S., & Koenig, S. (1990). Circulating CD8+ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. *Nature*, *348*(6298), 245–248. https://doi.org/10.1038/348245a0

[10] Jeang, K. T., Widen, S. G., Semmes IV, O. J., & Wilson, S. H. (1990). HTLV-I trans-activator protein, tax, is a trans-repressor of the human β-polymerase gene. *Science*, *247*(4946), 1082–1084. https://doi.org/10.1126/science.2309119

[11] Johnson, E. S., Ma, P. C. M., Ota, I. M., & Varshavsky, A. (1995). A Proteolytic Pathway That Recognizes Ubiquitin as a Degradation Signal. *Journal of Biological Chemistry*, *270*(29),

17442–17456. https://doi.org/10.1074/jbc.270.29.17442

[12] Johnson, J. M., Harrod, R., & Franchini, G. (2001). Molecular biology and pathogenesis of the human T-cell leukaemia/lymphotropic virus Type-1 (HTLV-1). *International Journal of Experimental Pathology*, *82*(3), 135–147. https://doi.org/10.1046/j.1365-2613.2001.00191.x

[13] Kalyanaraman, V. S., Sarngadharan, M. G., Robert-Guroff, M., Miyoshi, I., Golde, D., & Gallo, R. C. (1982). A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. *Science (New York, N.Y.)*, *218*(4572), 571–573. https://doi.org/10.1126/science.6981847

[14] Li, Q., Milo, R., Panitch, H., Swoveland, P., & Bever, C. T. (1998). Glatiramer acetate blocks the activation of THP-1 cells by interferon-γ. *European Journal of Pharmacology*, *342*(2), 303–310. https://doi.org/10.1016/S0014-2999(97)01509-4

[15] Milo, R., & Panitch, H. (1999). Glatiramer Acetate or Interferon-β for Multiple Sclerosis? *CNS Drugs*, *11*(4), 289–306. https://doi.org/10.2165/00023210-199911040-00005

[16] Mulloy, J. C., Migone, T. S., Ross, T. M., Ton, N., Green, P. L., Leonard, W. J., & Franchini, G. (1998). Human and simian T-cell leukemia viruses type 2 (HTLV-2 and STLV-2(pan-p)) transform T cells independently of Jak/STAT activation. *Journal of Virology*, *72*(5), 4408–4412. https://doi.org/10.1128/JVI.72.5.4408-4412.1998

[17] Nagaya, H. (1985). Induction of antigen-specific suppressor cells in patients with hay fever receiving immunotherapy. *The Journal of Allergy and Clinical Immunology*, *75*(3), 388–394. https://doi.org/10.1016/0091-6749(85) 90077-6

[18] Neusiedler, J. (2011). *To translate or to degrade? The role of INT6 in histone mRNA translation and Nonsense Mediated mRNA Decay*.

[19] O'Brien, K., Gran, B., & Rostami, A. (2009, December 17). *T-cell based immunotherapy in experimental autoimmune encephalomyelitis and multiple sclerosis* (London, UK). Http://Dx.Doi.Org/10.2217/Imt.09.61; Future Medicine Ltd London, UK. https://doi.org/10.2217/imt.09.61

[20] Panfil, A. R., Martinez, M. P., Ratner, L., & Green, P. L. (2016). Human T-cell Leukemia Virus-associated Malignancy. *Current Opinion in Virology*, *20*, 40–46. https://doi.org/10.1016/j.coviro.2016.08.009

[21] Patel, R. S., & Parmar, M. (2022). Doxycycline Hyclate. In *StatPearls [Internet]*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK555888/

[22] Pfab, C., Schnobrich, L., Eldnasoury, S., Gessner, A., & El-Najjar, N. (2021). Repurposing of Antimicrobial Agents for Cancer Therapy: What Do We Know? *Cancers*, *13*(13), 3193. https://doi.org/10.3390/cancers13133193

[23] Pillai, V., Ortega, S. B., Wang, C. K., & Karandikar, N. J. (2007). Transient regulatory T-cells: A state attained by all activated human T-cells. *Clinical Immunology (Orlando, Fla.)*, *123*(1), 18–29. https://doi.org/10.1016/j.clim.2006.10.014

[24] Ratner, L., Rauch, D., Abel, H., Caruso, B., Noy, A., Barta, S. K., Parekh, S., Ramos, J. C., Ambinder, R., Phillips, A., Harding, J., Baydoun, H. H., Cheng, X., & Jacobson, S. (2016). Dose-adjusted EPOCH chemotherapy with bortezomib and raltegravir for human T-cell leukemia virus-associated adult T-cell leukemia lymphoma. *Blood Cancer Journal*, *6*, e408. https://doi.org/10.1038/bcj.2016.21

[25] Timbrell, J. A. (2009). *Principles of biochemical toxicology*. New York: Informa Healthcare. http : //archive.org/details/principlesbioche00timb_973

[26] Uma kumari, NavjotKaurVirk, Identification of new potential drug for lung adenocarcinoma causing protein RMB10 using computer aided drug design approach, JournalIJBTR, Volume 12, Issue 2, Pages, 1-8, TJPRC Pvt.Ltd

[27] Uma Kumari "Insilico analysis and computer aided drug designing approach for mutant cancer gene (IJBTR) DEC 2021(Impact factor 6.6, ICV 61.5 ,NASS RATING 3.8)