

## Original Article

# Comparative Analysis and Molecular Structure of the Protease Molecule from Human Lymphotropic Virus Type-1 (HTLV-1)

Tarban N<sup>1</sup>, Habibi Rezaei M<sup>2</sup>, Shafifar M<sup>3</sup>, Mohammad hosayni M<sup>4</sup>, Rezaee S A<sup>5</sup>, Jazayeri S M<sup>2</sup>, Norouzi M<sup>2\*</sup>

1. Department of Biology, Kish International Campus, University of Tehran, Kish, Iran
2. School of Biology, College of Science, University of Tehran, Tehran, Iran
3. Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
4. Department of immunology, shahed university, school of medicine, Tehran, Iran
5. Immunology Research Center, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

## Abstract

**Background and Aims:** Human lymphotropic virus type-1 (HTLV-1) causes various diseases in human such as adult T-cell leukemia/lymphoma (ATLL) and HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM / TSP). The main goal of this study was to compare Iranian protease subtypes structure of this virus (HTLV-1) to samples collected from other part of world in order to understand their differences.

**Materials and Methods:** During 1394 -1395, 10 blood samples taken from HTLV-1 virus infected individuals. Samples were tested using polymerase chain reaction (PCR) process. The obtained products were sequenced and phylogenetic analysis was performed. The second and third structures of these sequences were determined using a specialized websites.

**Results:** The Iranian samples were completely exposed in to the cluster of Cosmopolitan subtype. The result of first structure alignment of protease protein in different subtypes of the virus, revealed some differences in the gene of interest. The conversion of the first structure to second and third structures, pairwise and multiple alignment showed no significant difference in the protease protein conformation.

**Conclusions:** The comparison of HTLV-1 virus protease protein from Iran and other sequences in the world which were obtained from GenBank showed no significant dissimilarity. This dissimilarity helps to design a plan for production of the drug. Therefore, future studies can target a part of the protein and generalize a treatment for all subtypes circulating. This comparison has beneficiary effect in making the right medication that inhibit the subtypes of the virus emerging during the course of treatment.

**Keywords:** Human lymphotropic virus type -1 (HTLV-1), protease, homology modeling.

## Introduction

Human lymphotropic virus type-1 (HTLV-1) is a complex virus. It is classified into Retroviridae family and Deltaretrovirus genus (1, 2). This virus has a

virion with icosahedral capsid and its genome contain two versions of single strand RNA with positive polarity (3). The virus genome's is linear (3) with a length of 9,032 bp and include areas such as gag, pro, pol, and env(1, 2). HTLV-1 is capable to make long period of infection in individuals (4, 5). Because of long-term and latent infection the symptoms of disease in about 5% of carrier (2, 3, 6) which appear as Adult T-cell leukemia/lymphoma (ATL) and HTLV-1 associated myelopathy/tropical spastic paraparesis (HAM/TSP) (1, 2,

\* **Corresponding author:** Mehdi Norouzi, Department of Virology, School of Public Health, Tehran University of Medical Sciences, Tehran  
mnorouzi@tums.ac.ir

7, 8). Nowadays, about 10-20 million people are infected with this virus worldwide (2, 9, 10). But the universal pattern of the disease distribution is not uniform (5). Infection with HTLV-1 is endemic in some parts of the world such as North America (4) Central Africa, Caribbean islands, and Northeastern Iran, in particular the state of Khorasan (9, 10). Khorasan Razavi especially Mashhad, Neyshabour and Sabzevar are known as endemic regions of infection with HTLV-1 and its prevalence in the whole population is about 2-4 percent (3, 5, 10, 11). The primary methods of transmission are unprotected sexual intercourse, breast-feeding by infected mothers, and blood transfusions (2, 9). Treatment of HTLV-1 is important because of two aspects; first, the infection with the virus is asymptomatic and could endanger public health. Second, the virulence, oncogenicity, and inflammatory reaction of this virus is significant (12). Protease protein produces structural and enzymatic important proteins for the replication of the virus. It is also responsible for the processing of end products of the coding part of Gag means Matrix, Capsid (CA) and Nucleocapsid (NC) (1, 13). This protein also produced by Ribosomal frame shifting (1, 14). The active site of HTLV-1 protease in recognition substrate is different from other retroviruses and even the sequences that are very similar to its substrate sequences, which could not transform to the product (1). Because of the importance of protease protein of HTLV-1 virus in replication (14), growth, and maturation (2, 15), this protein could be a suitable target for the treatment of the diseases caused by this virus (1). To date, no study has been performed to understand the difference between this protein in different strains of the virus HTLV-1. In this study, the gene bank and a conformity sequence software were utilized to study the difference of the protease protein in HTLV-1 virus subtypes and to determine the structure for comparing the protease protein in samples from Iran with the samples from other parts of the world. We believe the unique function of protease and its construction calls

for an applicable medicine for all strains circulating in Iran and other parts of the world.

### Methods

#### **Sampling and confirmation of the disease.**

During 1394-1395, 10 individuals with lymphotropic virus type one (HTLV-1) were selected from the Ghaem Hospital patients pool in Mashhad. After the confirmation of a lymphotropic virus infection in each 10 individual samples, the research group collected a thorough medical history of each individual patient. Initial examination and serological tests were performed on the samples of peripheral whole blood from each selected patients.

#### **Isolation of peripheral blood mononuclear cells - (PBMCs).**

Isolation of peripheral blood mononuclear cells (PBMCs) was performed according to the manufacturer's instructions using ficole (SIGMA-Germany). Subsequently, DNA was extracted from isolated cells using High Pure Viral Nucleic Acid Kit of Roche manufacturing.

**Get the reference sequences.** The National Center of Biotechnology Information (NCBI) GenBank was used as reference for the genome sequence of the virus strains HTLV-1 for comparison of the strains isolated in other countries.

#### **Alignment and drawing preliminary phylogenetic tree of the reference sequences.**

BioEdit alignment software was used to compare the full genome of all viral strains. After alignment of obtained sequences, the first phylogenetic tree was drawn using neighbor joining (NJ) logic, 1000 bootstrapping and kimura two Parameter substitution model by MEGA software. In this study, we compared 26 sequences related to HTIV-1 and also HTLV-2, HTLV-3, HTLV-4 and HIV (results not shown). The aim of drawing control tree (first tree) was to determine the position of obtained sequences.

#### **Proliferation of samples and Duplication of sequences.**

After extracting the protease region of the sequence, appropriate primers for amplification of the full protease region were

designed using the Gen Runner as a reference software.

In the Polymerase Chain Reaction step (PCR), the QIAGEN, a German-made kit was used. The following material were used in the test: 31  $\mu$ l H<sub>2</sub>O, 10  $\mu$ l PCR buffer, 1  $\mu$ l leading primer, 1  $\mu$ l reverse primer, 2  $\mu$ l dNTP, 2  $\mu$ l Taq enzyme, 3  $\mu$ l samples, in the final volume 50  $\mu$ l was prepared. The following temperatures were used in the amplification steps: denaturation step at 94°C for 30 seconds, the annealing step at 54°C for 35 seconds, and the expansion phase at 72°C and 25 seconds.

#### **Determination and alignment of sequences.**

After purification, the product was obtained by the PCR test, the protease region sequences of Iranian samples was determined under sequencing process using the vivantis-America kit. The sequencing process was done in the Virology Department of Medical School of Tehran University using Genetic Analyzer machinery by ABI Hitachi 3130. After alignment with BioEdit software (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>), minor editing was done on the obtained sequences.

#### **Drawing the secondary phylogenetic tree of reference and Iran sequences.**

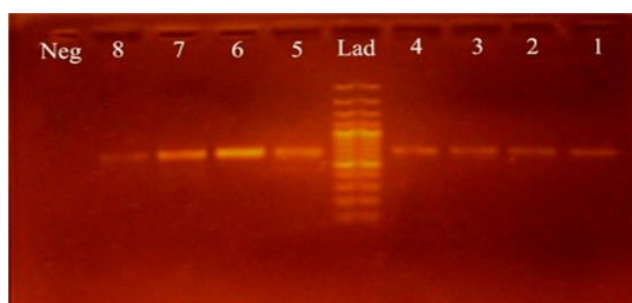
After constructing the first tree by full genome and determining the reference region, the sequences of protease region of these subtypes was determined. The obtained sequences of Iranian samples took place under alignment process. In order to draw the second phylogenetic tree based on the protease gene, in addition to the reference protease gene sequences, sequences of 10 samples from Iran were used. The phylogenetic tree was based on a protease gene using MEGA 6.6 software and was drawn using the logic of Maximum Likelihood with 1000 bootstrapping and kimura two Parameter substitution model.

**Determination of the first and second structure of the protease protein based on the reference and the Iranian sequences.** The translation of the samples and reference (obtained from NCBI) sequences of the first protein structure was done using the BioEdit

software. One of the similar structures was chosen based on geographical location, then the first and second structures were determined using Phyre2 ([www.sbg.bio.ic.ac.uk/phyre2](http://www.sbg.bio.ic.ac.uk/phyre2)). Subsequently, the determined second structure was designated using the handheld adaptation process.

Tertiary structural determination and alignment of protease proteins to study their structural differences was obtained using Phyre2 from the VADAR website (<http://vadar.wishartlab.com/index.html>).

Subsequently, the Ramachandran diagram was drawn. The amount of tertiary structure similarity and alignment of selected sequences were analyzed using the Chimera Protein structure visualizer software (<http://www.rbvi.ucsf.edu/chimera>). The results is emblematic of the similarity between the sequences.



**Fig. 1.** The result of PCR – the protease bands of clinical samples is shown. All of the fragments are on 700 bp band. According to gel image, the 9th well is negative control and the sample between the 4th and 5th wells contains the DNA Ladder.

## **Results**

**Primer Design and protease gene amplification.** The fragments with the length of 700 bp in received clinical samples were obtained using the following designed primers and PCR method (Figure 1). TRN-FORWARD: 5'-CAACTCACTGGAAGCGRGAC -3' TRN-REVERSE: 5'-GRTTCRATATGGCCTGCCTC-3'



Fig. 2. Phylogenetic tree of the protease sequence of HTLV-1 virus

**Obtaining the reference sequences and draw phylogenetic tree.** 26 full sequences of references were obtained. The first phylogenetic tree was drawn (results not shown) and determined the position of the each reference sequence. Subsequently, repetitive sequences were removed and the protease gene was determined. HTLV-2 protease sequences were considered as outside groups. Consequently, by drawing this tree, it was found that the viruses circulating in Iran are closest to the Cosmopolitan type. The phylogenetic results showed a strong relationship between Iran's samples to other sequences, thus, the majority of target gene sequences were similar (Figure 2).

The alignment data showed little differences in the gene. After alignment and drawing the similarities of the first structure matrix, there was high similarity in all 26 amino acid sequences of reference. Then a number of sequences were chosen based on the greatest difference and also according to the results of phylogenetic and geographic location. These sequences were KF242506, NC001436, AF042071, AY563953, AF259264, L03562, and HQ606137. Also of the analyzed sequences from Iran, sample number 8 was selected randomly (due to lack of similarity between samples) to comparison with these sequences (Figure 3).

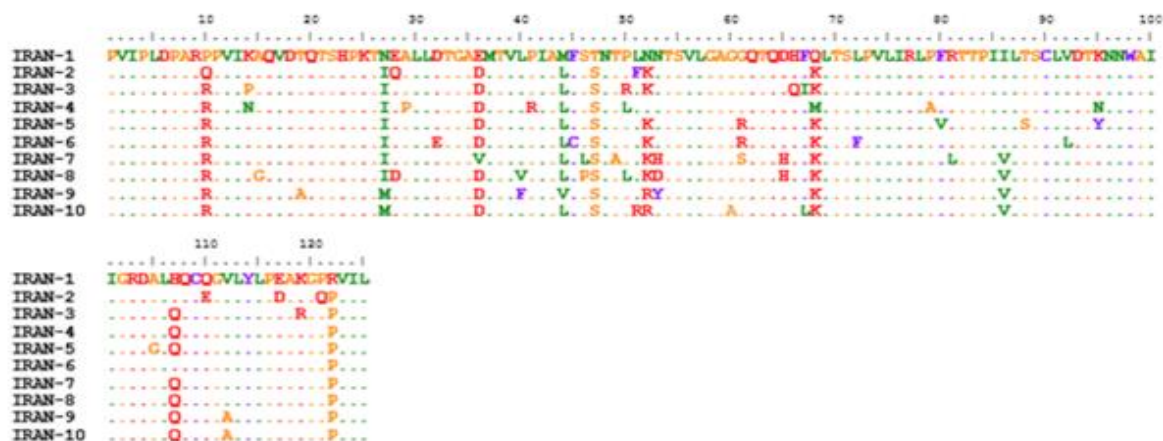


Fig. 3. Results of the protease gene alignment of the Iran sequences

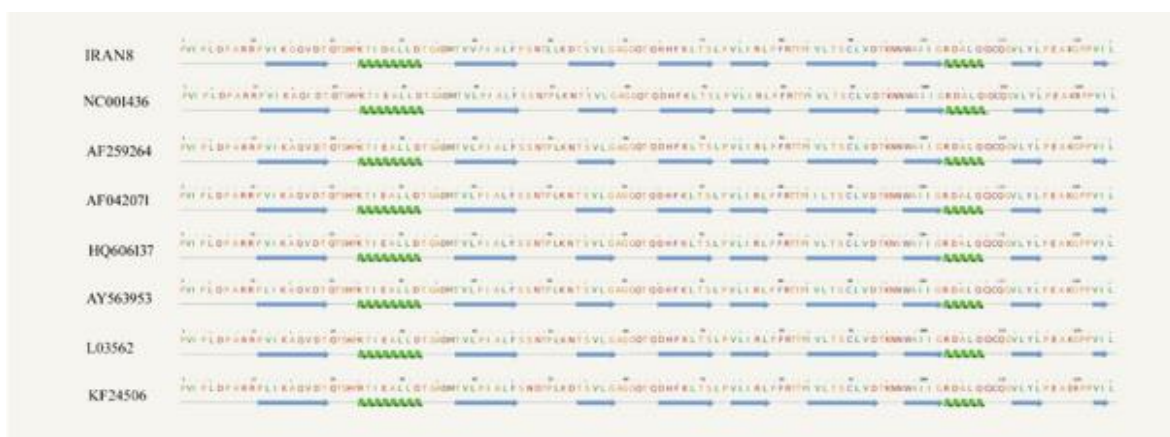


Fig. 4. Alignment result by Phyre2.

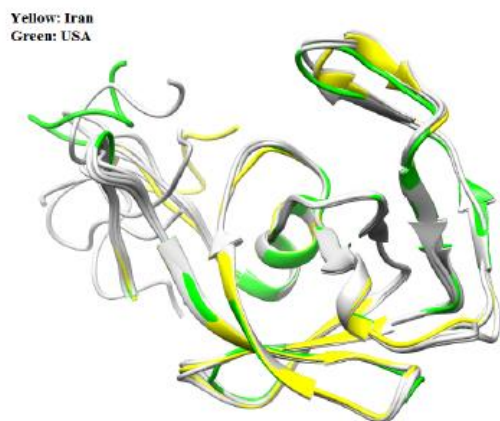
**The second and third alignment of the selected sequences.** The secondary structure of protease protein was examined by pohyre2 and ESPriptsoftwares that was explained as follow (Figure 4).

In the above figures, the allpha helices, beta-sheets regions, and the rate of difference between sequences are visible. Therefore, it can be concluded that there are many similarities in the secondary structure of the Iranian sequence and other sequences relative to each other.

In the next step, the similarity and alignment of tertiary structure of selected sequences were analyzed by the Chimera software that the results indicate the similarity between sequences (Figure 5).



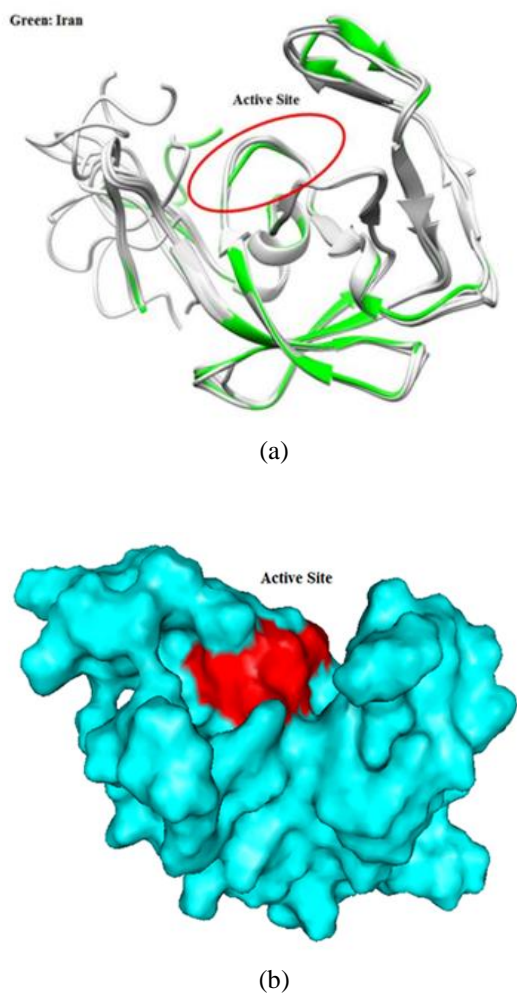
Fig. 5. Structural alignment of 8 selected sequences and comparison of the third structural in the Iranian selected sequences and reference sequences.



**Fig. 6.** Third structural alignment of the 8 selected sequences.

Each of the used colors in the above figure represents the structure of a sequence. Tm score and RMSD numbers were obtained using the Chimera Protein structure visualizer site (<http://www.rbvi.ucsf.edu/chimera>) which indicates the degree of Iranian sequences similarity with the reference sequences, shown in the table below (Table1).

<b>Table 1: Tm score and RMSD Numbers</b>		
<b>Reference (USA)</b>		
	<b>RMSD</b>	<b>Tm-score</b>
<b>Iran</b>	1.03	0.91
<b>Australia</b>	1.27	0.95
<b>China</b>	1.19	0.91
<b>Germany</b>	1.59	0.92
<b>Brazil</b>	0.78	0.91
<b>Canada</b>	1.88	0.92
<b>Japan</b>	0.88	0.92



**Fig. 7.** (a) View of protease active site (b) View the protein's surface.

The figure below displays the Iranian number 8 sequence and reference sequence (NC001436) The American portions were shown using the colored and white sequences in structural alignment in order to enhance the conception of the similarity in Iranian and reference sequences (Figure 6).

The following figures illustrate the active site of the protease protein in the structure. The first figure displayed the second structures of protease protein, and the active site position is showed by a red ring. The structure of the Iranian sample sequence is marked in green and other sequences are in white. The second figure displays the third structure of the protease protein as a surface shape and the active site of protein is illustrated in red [Figure7: (a), (b)].

## Discussion

Due to the relative success of anti-protease drugs against the HIV-1, protease protein can be used as an important target for the treatment of HTLV-1 infections. We are planning to run the first comparative study of protease protein structures subtypes of the HTLV-1 virus.

Increasing prevalence of the virus around the world necessitate to develop a drug to alleviate

this disease. This study was based on whether the protease sequences in all strains of the virus are identical or significantly different. Another goal of the study was to determine if a drug designed on the basis of HTLV-1 protease structure can be useful for treatment of related virus strains.

After receiving the complete genome sequence of the virus subtypes from different parts of the world, drawing the phylogenetic tree, and determining the affinity of these strains with each other, the protease gene was extracted from the Iranian samples. Next, by phylogenetic analysis, the affinity among protease protein of all strains of HTLV-1 worldwide and the sequences of 10 clinical samples collected from Mashhad Ghaem hospital were analyzed. The analytical result indicates a high similarity and affinity between all reference protease sequences and Iranian samples, with an exception of Australia. The result of the study confirmed the findings of Arthur Paiva and colleagues who reported the transmission of the virus between the continents of Africa, Europe, Asia and America (16). The phylogenetic tree obtained from this study showed of high similarity between Iran (Mashhad) and German strains .

The length of the branches of phylogeny tree base on protease protein and compression of Iranian samples with other parts of the world indicates that in this protein there is a high affinity among the samples. This affinity decreases in the Malaysian subtype according to phylogenetic tree clusters that relate to the protease gene and the similarity matrix between genes confirmed for the full genome tree of this virus. From each tree cluster, one candidate is selected and then put under comparison analysis where homology modeling takes place .

After determining the third structure of selected samples and observing any differences, the second and third structure were analyzed to each other consecutively. During the study, the selected structures were compared with the American reference sequences and Tm SCORE numbers, and its RMSD was obtained. The high Tm SCORE and lower RMSD indicates more affinity with

America's sequence. We found out that Brazil has the greatest and Germany has the least affinity with the reference sequences. Also the Iranian and Japanese sequences have the large scale affinity together compared to others. In another study, the active site of this enzyme had residues DTGAD (17). Some analysis shows that the first structure of the worldwide and Iranian sequences indicate that this region is conserved. After determining the secondary and tertiary structure and comparing of these structures, no difference was observed in the enzyme active site.

It is likely that the spatial structure of the protein tertiary structure is viewed in the second structure. Any difference in these structures needs a different treatment with combined medication. Observations of the slight differences between various subtypes of this virus need broader studies. In other words, more advances studies are necessary to ensure the power of adhesion of these drugs to very similar structures of protein. By comparing the HTLV-1 and HIV-1 protease structure and sequence, the enzymatic digestion and substrate binding sites were identified. Despite the high similarity of these two proteins, HIV-1 protease inhibitors have not any effect on HTLV-1 protease (6). Regarding this fact, there are not any other studies about the differences between various HTLV-1 virus subtypes and this study was done for the first time. This study, which proves the similarities in the structure of all protease subtypes has a substantial effect on designing and manufacturing the medicine that controls all viral subtypes .

The phylogenetic analysis performed in this study led to a study of molecular epidemiology of the virus. Dr. Rafatpanah's team have performed a phylogenetic study on the LTR region(18). The results of previous study are parallel to Dr. Rafatpanah's results. The protease structure determination of various sub-types of the HTLV-1 virus are practical not only in making of an effective drug for treatment but determining the phylogeny and phylodynamic of the HTLV-1 virus in Iran was a major step. The study also has a substantial effect in future studies in effective drug

treatment and confronting drug resistance as well. It seems that if the sequences of other Iranian cities samples were used, the results of the phylogenetic will be more reliable. But by considering the fact that there is not any significant difference in the second and third structure between the sequences of circulating subtypes in all over the world, using the sequences of other Iranian cities is not necessary because there is no difference in second and third structure. With an analysis of all the subtypes by crystallography, we came to this conclusion with 100% confidence in the Angstrom scale that this method is expensive and the processes of sample preparation is very complicated. The phylogenetic analysis indicated that the strains in Iran have more affinity with the French and German subtypes, which are sub-parallel to the tree designed by full genome sequences. After selecting a representative sequence of each cluster and obtaining the second and third structures, no significant differences between these structures were observed. This could be the main reason for the lack of need for polyvalent anti-protease drugs against infection with this virus. To ensure this result in future studies, these medicinal structures should dock with the representative sequences and be based on the similarity in the power of connection. The aim of this study was comparison of the various subtype's structures of the virus, which was accomplished successfully. This study achieved a scale which represent the percent of similarity between proteins by using the world's newest softwares technology .

### Acknowledgments

Thanks to the hepatitis laboratory personnel of the Virology Department, Tehran University of Medical Science, and especially Mrs. A. Khamse, who helped us in this study.

### References

1. SHUKER SB, MARIANI, V. L., HERGER, B. E. & DENNISON, K. J. Understanding HTLV-I Protease. *Chemistry & biology*, 10, 373-380. 2003.
2. Rafatpanah H, Farid R, Golanbar G, Jabbari Azad F. HTLV-I Infection: virus structure, immune response to the virus and genetic association studies in HTLV-I-infected individuals. *Iranian journal of allergy, asthma, and immunology*. 2006;5(4):153-66.
3. Ghezeldasht SA, Shirdel A, Assarehzadegan MA, Hassannia T, Rahimi H, Miri R, et al. Human T Lymphotropic Virus Type I (HTLV-I) Oncogenesis: Molecular Aspects of Virus and Host Interactions in Pathogenesis of Adult T cell Leukemia/Lymphoma (ATL). *Iranian journal of basic medical sciences*. 2013;16(3):179.
4. Boxus M, Willems L. Mechanisms of HTLV-1 persistence and transformation. *British journal of cancer*. 2009;101(9):1497-501.
5. Saffar S, Azadmanesh K, Golmohammadi T, Golkar M, Amminian M, Mirshahabi H, et al. Production of recombinant Human T Lymphotropic Virus type 1 Tax protein in Rosetta (DE3) bacterial host. *Journal of Paramedical Sciences*. 2010;1(4).
6. Kádas J, Boross P, Weber I, Bagossi P, Matúz K, Tozser J. C-terminal residues of mature human T-lymphotropic virus type 1 protease are critical for dimerization and catalytic activity. *Biochem J*. 2008;416:357-64.
7. Cook LB EM, Rowan AG, Asquith B. HTLV-1: Persistence and pathogenesis. *Virology* 435 (2013) 131–140. 2013.
8. Treviño A, Parra P, Bar-Magen T, Garrido C, de Mendoza C, Soriano V. Antiviral effect of raltegravir on HTLV-1 carriers. *Journal of antimicrobial chemotherapy*. 2011;dkr404.
9. Shoeibi A, Etemadi M, Ahmadi AM, Amini M, Boostani R. "HTLV-I Infection" Twenty-Year Research in Neurology Department of Mashhad University of Medical Sciences. *Iranian journal of basic medical sciences*. 2013;16(3):202.
10. Akbarin MM RH, HassanNia T, et al. Comparison of HTLV-I Proviral Load in Adult T Cell Leukemia/Lymphoma (ATL), HTLV-I-Associated Myelopathy (HAM-TSP) and Healthy Carriers. *Iran J Basic Med Sci*: 2013; 16:208-12. 2013.
11. Hosseni RF, Jabbari F, Shabestari M, Rezaee SR, Gharivani Y, Valizadeh N, et al. Human T Lymphotropic Virus Type I (HTLV-I) is a Risk Factor for Coronary Artery Disease. *Iranian journal of basic medical sciences*. 2013;16(3):217.
12. Bazzaz B, Rezaee SA. The Significance of HTLV-I in Molecular Oncology. MASHHAD UNIV MED SCIENCES VICE-CHANCELLOR



FOR RES CTR OFF IJBMS, DANESHGAH ST,  
PO BOX 9138813944-445, MASHHAD, 00000,  
IRAN; 2013.

13.J T. Comparative Studies on Retroviral  
Proteases: Substrate Specificity. *Viruses* 2010, 2,  
147-165. 2010.

14.Makoto Kobayashi ‘ YO, Tsuneo Asano’,  
Takaki Hayakawa’, Koichi Kato’. Atsushi  
Kakinuma’ and Masakazu Hatanaka’ , ‘.  
Purification and characterization of human T-cell  
leukemia virus type % protease produced in  
*Escherichia coli* Federation of European  
Biochemical Societies 00145793/91/S350  
1991;Volume 293, number 1.2. 106-I 10

15.Bagossi P, Sperka T, Fehér A, Kádas J,  
Zahuczky G, Miklóssy G, et al. Amino acid  
preferences for a critical substrate binding subsite  
of retroviral proteases in type 1 cleavage sites.  
*Journal of virology*. 2005;79(7):4213-8.

16.PAIVA AC, J. . Origin and prevalence of  
human T-lymphotropic virus type 1 (HTLV-1) and

type 2 (HTLV-2) among indigenous populations in  
the Americas. *Revista do Instituto de Medicina  
Tropical de São Paulo*, 57, 01-14. 2015.

17.TOZSER JW, I. T. . The protease of human T-  
cell leukemia virus type-1 is a potential therapeutic  
target. *Current pharmaceutical design*, 13, 1285-  
1294. 2007.

18.AZARPAZHOOH MR, HASANPOUR, K.,  
GHANBARI, M., REZAEE, S. R., MASHKANI,  
B., HEDAYATI-MOGHADDAM, M. R.,  
VALIZADEH, N., FARID HOSSEINI, R.,  
FOROGHIPOOR, M. & SOLTANIFAR, A. .  
Human T-lymphotropic virus type 1 prevalence in  
Northeastern Iran, Sabzevar: an epidemiologic-  
based study and phylogenetic analysis. *AIDS  
research and human retroviruses*, 28, 1095-1101.  
2012.