Int J Pharm Bio Sci 2013 Oct; 4(4): (P) 902 - 909

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

IN SILICO ANALYSIS OF INTERLEUKIN-2: A COMPARATIVE STUDY BETWEEN *HOMO SAPIENS* AND *MUS MUSCULUS*

NEHA SINGH, AKRITI JAIN AND ANAND ANBARASU*

School of BioSciences and Technology, VIT University, Vellore-632014, Tamil Nadu, India.

ABSTRACT

Homo sapiens and *Mus musculus* genomes have the similarity ranges from 70% to 90% with an average of 85%. Interleukin (IL) has 17 different families among which interleukin 2 (IL2) is present in both *Homo sapiens* and *Mus musculus*. IL2 activates the receptor named IL2-R which further proliferate lymphocytes such as T-cell and B-cell. IL2 is a glycosylated polypeptide lymphokine. It requires cleavage of the signal sequence for its activity. IL2 is used for immunotherapy to treat cancer and its suppression avoids autoimmunity in patients with transplantations. IL2 of *Homo sapiens* and *Mus musculus* are studied and analysed using simulation and structural analysis methods. The results are compared to obtain their similarity profile. This result can further be applied in the study of inflammatory disease related aspects in future.

KEYWORDS : Interleukin 2, Physio-chemical properties, Topology mapping, Solvent accessible area, Stabilizing Centre and Stabilizing Residue

ANAND ANBARASU
School of BioSciences and Technology, VIT University,
Vellore-632014, Tamil Nadu, India.

INTRODUCTION

Episodic lymphopenia is a disease rather a syndrome linked with many immunosuppressive diseases like Schimke immunoosseous dyplasia. In this syndrome, episodic and profound depression in T-cell functions with loss in immunological response[1]. T-cell are responsible elements for cell-mediated immune response to the foreign materials of human body. These cells have various functions like activation of cytotoxicity and immunological memory[2]. T-cell proliferation and activation is particularly dependent on interleukin-2 molecules binding to the interleukin receptors. In such a manner, IL-2 plays essential role in contributing in maintaining tolerance and memory of immunological responses[3]. This implies that any changes or defects in the IL-2 molecule may have resulted in serious implications in the form of the episodic lymphopenia syndrome. IL-2 is one of the members of cytokine family including 21 interleukins. IL-2 is four α helix cytokine which is produced by activated CD4+ T cells. This molecule is synthesised in regulation with mRNA levels. IL-2 molecules binds to receptor complexes of three subunits of IL-2Rα (CD25), IL-2Rβ (CD122), and common γ-chain (γ_c;[2] CD132) which have high affinity for the molecule [4]. It is a protein which regulates the activities of various immunological cells by binding to IL-2 receptors. IL-2 is necessary for the growth, proliferation, and differentiation of T cells to become 'effector' T cells. IL-2 is normally produced by T cells during an immune response [5, 6]. IL-2 and the above mechanism are present in many mammals like *Bovine*, *Mus musculus*, *Homo sapiens* and others also. The similarity between the genomes of *Homo sapiens* and *Mus musculus* ranges from about 70% to 90% with an average of 85%. Both the genomes have an average gene density of 1 gene per 100,000 bases and estimated gene number approximately 25,000[7, 8]. In this study, the IL-2 genes of both the species are studied and analysed using structural analysis and docking results to prepare a similarity profile between them. The secondary structure of IL-2 protein of *Mus musculus* is modelled and its properties are predicted to compare with the existing IL-2 of *Homo sapiens*.

MATERIALS AND METHODS

(i) Collection of data

The FASTA sequence of IL2 of *Homo sapiens* and *Mus musculus* were retrieved with their accession id from UniProt database[10]. The 3-Dimensional (3D) structure of *Homo sapiens* IL2 was obtained from Protein Data Bank (PDB id 1PW6) [11]. The hypothetical structure of *Mus musculus* IL2 was generated in Swiss-Model Workspace [12].

(ii) Amino acid composition and Secondary Structure Topology mapping

Statistical Analysis of Protein sequence (SAPS) calculates different protein structure properties [13]. We used SAPS to analyze amino acid composition of the two IL2 sequences. SCRATCH protein predictor server is a freely available bioinformatics server to predict protein tertiary structure and structure features [14]. We calculate certain properties such as disulphide bond topology, antigenicity propensity, and domains through SCRATCH protein predictor.

(iii)Physio-chemical properties, Conserved Residues and Hydrophobic Segment Prediction

Physical and chemical properties of a given protein were evaluated through ProtParam tool of ExPASy [15]. We calculated and compared physio-chemical properties such as molecular weight, number of amino acid, extinction coefficient, grand average of hydropathicity, etc for both the IL2 sequence of *Homo sapiens* and *Mus musculus*. ConSurf server allowed the users to visualize the hydrophobic and hydrophilic surface properties using the concept of molecular hydrophobicity potential (MHP) [16]. It found the functional and structural residue of the given sequences. We used ConSurf server to find the detailed information including position, score, and sequence name of IL2 sequence for the species. We used ProtScale to analyze

the hydrophobicity of the protein by plotting hydropathy score against the position [17].

(iv) Solvent accessible area

Solvent accessible area is the specific area of the protein which was accessible to the solvent. We used ASAView tool to find the solvent accessible area which displayed the graphical output based on DSSP program [18]. The 3D structure of IL2 of both *Homo sapiens* and *Mus musculus* were used to generate these areas.

(v) Stabilizing Centre and Stabilizing Residue

Stabilization centres are the elements of a protein which stabilizes the structure by preventing it from decaying and elongating their interactions [19]. We used the program SCide to predict stabilization centre [20]. We submitted 3D structure of IL2 to SCide server and produced the output in text format. Protein residue stabilization depends on various non-covalent interactions. We used SRide server to find the stabilizing residues [21].

Stabilization centres and stabilizing residues present in the sequence were calculated by using the surrounding hydrophobicity, long-range order, stabilization centre and conservation score.

RESULTS AND DISCUSSION

(i) Amino acid composition and Secondary Structure Topology mapping

Statistical analysis of Protein Sequence (SAPS) shows the amino acid composition. We find out that IL2 of *Homo sapiens* and *Mus musculus* contains similar low and high amino acid which are non-polar hydrophobic. Tryptophan (W) has the lowest percentile in the sequence, i.e., it is 0.7% in *Homo sapiens* and 0.6% in *Mus musculus*. Leucine (L) has the highest percentile in the sequence, i.e., 27% in *Homo sapiens* and 25% in *Mus musculus*. Disulphide bonds play a very important role in stabilizing the protein structure and cysteines form the disulphide bonds due to the presence of sulphur in its structure [22]. Here, Cysteine residues are present at the two positions in both the species. The results of the SCRATCH server shows that Cystine residues are present at 78^{th} and 125^{th} position in *Homo sapiens* sequence, and 140^{th} and 160^{th} in *Mus musculus* sequence. At this residue position, water molecules content is less. The antigenicity is the ability of the chemical structure to bind with the certain groups which have the adaptive immunity [23]. The probability of antigenicity in *Homo sapiens* is 0.695013 and in *Mus musculus* is 0.917846. This shows *Mus musculus* have the higher antigenicity than *Homo sapiens*. Domains are the conserved regions in the protein sequence that evolve, function, and exit independently [23]. There are 1 domain in *Homo sapiens* and 2 domains in *Mus musculus*. Due to the presence of more domains in *Mus musculus*, there is more probability to alter the sequence at the genetic level than in *Homo sapiens*.

(ii) Physical-chemical properties, Conserved residues, and Hydrophobic Segment Prediction

Comparative properties of IL2 of both *Homo sapiens* and *Mus musculus* are given in Table 1. We find that the total number of negatively charged residue is more in *Mus musculus* whereas total number of positively charged residue is more in *Homo sapiens*. Instability index is the property of a protein which estimates the stability of the protein in the test tube [24]. The instability index of both the sequence is above 40 which shows these protein sequences are unstable. Aliphatic index shows the thermostability of the globular protein [25]. The aliphatic index of *Homo sapiens* is more than *Mus musculus* which implies that *Homo sapiens* IL2 globular protein is more thermodynamically stable than *Mus musculus* IL2 globular protein.

Table 1
Physical Chemical properties of Homo sapiens and Mus musculus IL2 sequence.

S. No.	Property	*Homo sapiens*	*Mus musculus*
1.	Number of amino acid	153	169
2.	Molecular weight	17627.7	19400.0
3.	Theoretical pI	7.76	4.88
4.	Total number of negatively charge residue	15	20
5.	Total number of positively charge residue	16	14
6.	Extinction Coefficient (assuming all cysteine residues appear as half cystines)	11710	10220
7.	Extinction Coefficient (assuming no cysteine residue appear as half life)	11460	9970
8.	Instability index	47.71	64.97
9.	Aliphatic index	108.37	89.41

Residue position and function of each residue of *Homo sapiens IL2* and *Mus musculus* are depicted in Fig1(a) and Fig1(b). The result shows that the residues between 1 and 25 and at 90th position have the same function in both the sequences. There are 11 structural and exposed residues in IL2 of *Homo sapiens* and *Mus musculus*. There are 21 functional and buried residues in *Homo sapiens* and 22 residues in *Mus musculus*. This shows that there is almost 25% residue similarity in both the sequences.

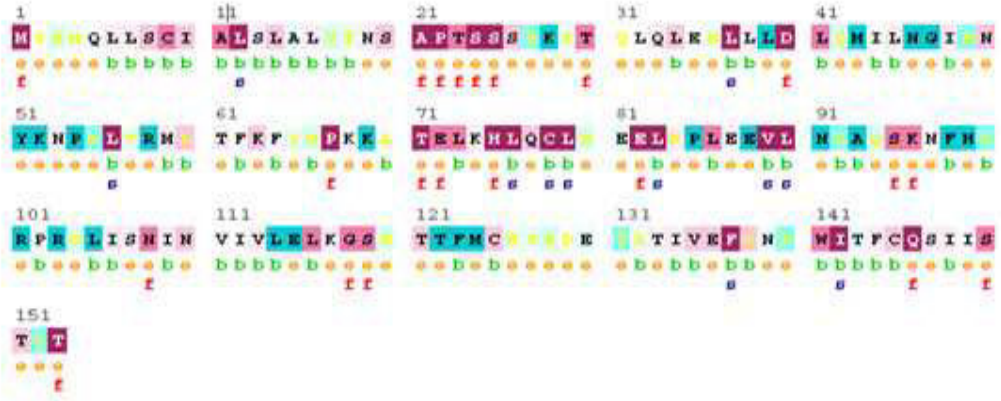

Fig1(a) ConSurf Result for IL2 *Homo sapeins*

Fig1(b) ConSurf Result for IL2 *Mus musculus*

a) The conservation scale:

Variable Average Conserved
b) e - An exposed residue according to the neural-network algorithm.
c) b - A buried residue according to the neural-network algorithm.
d) f - A predicted functional residue (highly conserved and exposed).
e) s - A predicted structural residue (highly conserved and buried).
f) X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Fig2 shows the hydropathy plot of IL2-*Homo sapiens*. The highest score ranges from 2.5 and 3 whereas the lowest score ranges from -2 and -2.5. The plot shows a wide range of hydropathy score for the IL2 sequence. Fig3 shows the hydropathy score of IL2-*Mus musculus*. The highest score ranges from 2 to 3 and the lowest score ranges from -3 to -4. The hydropathy score is found to be confined within the range of -2 to 1. The maximum number of score is in this range. The highest score of IL2 *Homo sapiens* and *Mus musculus* is in the same range with similar position. Hence, both sequences have maximum and similar hydrophobic and hydrophilic properties at position of amino acid within the range 1 to 20. The average hydropathy score, also known as GRAVY (Grand average of hypropathicity) is calculated through ProtParam tool. The GRAVY score for IL2 *Homo sapiens* is -0.007 and for IL2 *Mus musculus* is -0.343. This shows that IL2 of *Homo sapiens* have more hydropathy score than *Mus musculus* which implies that *Homo sapiens* have more hydrophobic and hydrophilic property than *Mus musculus*.

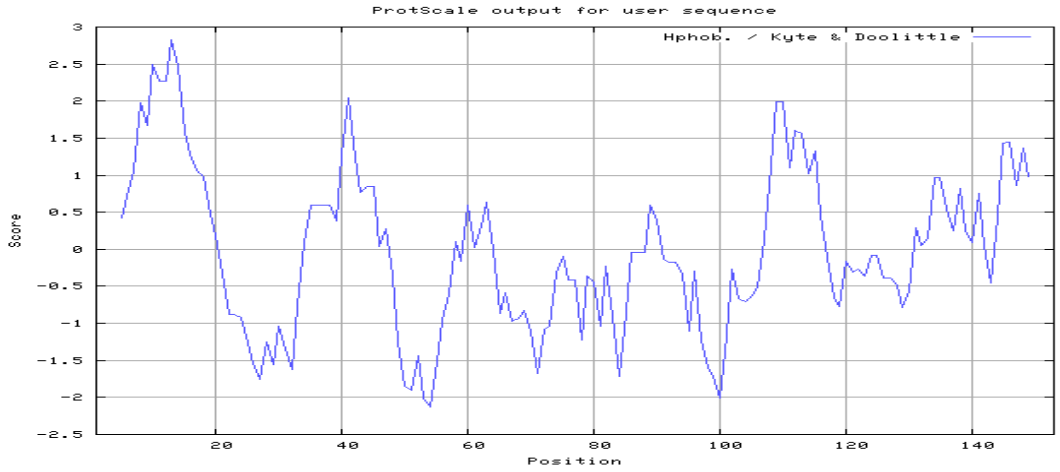

Figure 2
ProtScale result for IL2 of Homo sapeins

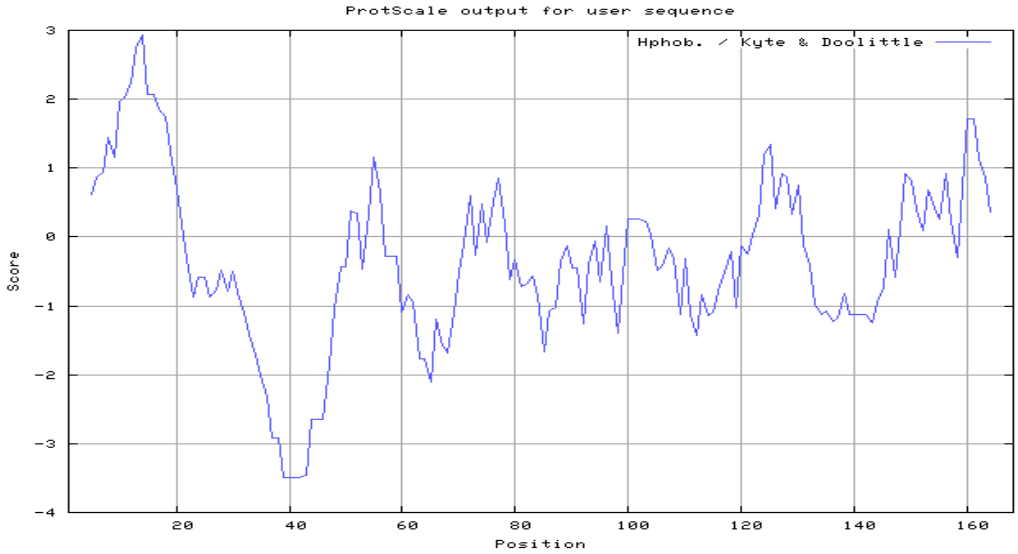

Figure 3
ProtScale result for IL2 of Mus musculus

(iii) Solvent accessible area

ASAview tool result shows that *Mus musculus* IL-2 protein has a larger number of residues in the interior of the spiral plot. Thus, the protein is more tightly packed than the *Homo sapiens* IL-2 protein which has comparatively fewer residues in the interior. *Mus musculus* IL-2 protein also has relatively more residues on the outer radii and larger number of circles than the *Homo sapiens* IL-2 protein hence, the *Mus musculus* IL-2 protein has higher solvent accessibility and consequently, it posses greater number of possible sites. Charged residues on the surface will fall on the outermost ring of the spiral and hence these plots automatically suggest potential binding sites of the protein [26]. The relative solvent accessible surface area can be used to predict the magnitude of binding-induced conformational changes in protein-protein interactions [27].

(iv) Stabilizing residue and Stabilizing Centre

It is interesting to note that 23.07% of residues in *Mus musculus* contribute to stabilizing centre whereas in *Homo sapiens*, it is only 18.3% of residues. Leucine residues at 14^{th} and 16^{th} positions contribute to stabilization centres in both *Homo sapiens* and *Mus musculus* proteins. In *Mus musculus* IL2 proteins, cysteine residues are present at 140^{th} and 160^{th} positions which contribute to the stabilization centres by forming the disulfide bonds. These stabilization residues are involved in long range interactions. More number of long range contacts refer to stability which implicates that *Mus musculus* IL2 protein is more stable than the *Homo sapiens* IL2 protein. From the prediction of stabilization residues, there is only 1 stabilization residue present in *Mus musculus* IL-2 protein whereas in *Homo sapiens* IL-2 protein, 8 stabilizing residues are present. In *Mus musculus*, Thr127 has highest conservation score of 7 with surrounding hydrophobicity of 27.35 while in *Homo sapiens* IL-2 protein, Leu41 and Asp40 have highest conservation scores of 9 along with high values of surrounding hydrophobicity (Hp) of 27.3 and 27.87 respectively. The results also shows that all the stabilizing residues have a conservation score greater than 6, indicating the high conservation among other evolutionarily related protein sequences. This result correlates with the fact that the residues essential for the preservation of the protein architecture are well conserved by evolution [28].

CONCLUSION

In the present study, it has been shown that the IL-2 protein molecules of the species *Homo sapiens* and *Mus musculus* posses similarity in their structural and compositional features. Although IL-2 protein is more prone to alterations at genetic level, both the molecules showed conservation of residues and protein architecture during the evolutionary process. Besides having 25% residue similarity between the sequences, IL-2 of *Homo sapiens* possesed comparatively more hydropathicity. IL-2 protein of *Mus musculus* is more stable and more tightly packed than that of *Homo sapiens* protein. This study also shows that IL-2 in *Mus musculus* has more antigenicity than that of *Homo sapiens*. Hence, it can be recommended that during the diagnosis and clinical trials of episodic lymphopenia, this similarity profile should be considered while preparation and testing of drug potency between the *Mus musculus* and *Homo sapiens* species.

ACKNOWEDGEMENT

We thank the management of VIT University for providing the necessary facilities to carry out this research project.

REFERENCES

1. Ananthanarayan and Paniker, Ananthanarayan and Paniker's textbook of microbiology, 7^{th} edition, page 156.
2. Shankhajit De, Saroj Singhmura and Ajoy Kumar Ghos, Overview of renal cell carcinoma, International Journal of

Pharma and Biosciences, Vol.1, (3) (Jul-Sep.2010).
3. Thosmas R. Malek ,The biology of interleukin-2. Annual Review of Immunology, 26, 453-479, (2008).
4. Brad H. Nelson, IL-2, Regulatory T Cells, and Tolerance, The Journal of Immunology:172(7) , (2004).
5. Cantrell DA and Smith KA, The interleukin-2 T-cell system: a new cell growth model. Science 224 (4655): 1312–6. (1984).
6. Smith KA, Interleukin-2: inception, impact, and implications. Science 240 (4856): 1169–76. (1988).
7. Eric S. et al., International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 409: 860-921. (Feb 2001)
8. Waterston RH, Lindblad-Toh K, Lander ES, Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562. (Dec 2002).
9. Chiranjib Chakraborty and Alok Agarwal, Computational Analysis of c-Reactive Protein for Assessment of Molecular Dynamics and interaction Properties. Cell Biochemistry and Biophysics. Vol.1, Issue 12, (2013).
10. Michele Magrane and Uniprot Consortium, UniProt Knowledgebase: a hub of integrated protein data. Oxford University Press, (2011).
11. Westbrook et al., The Protein Data Bank. Biological Crystallography. 58, 899-907, (2002).
12. Konstantin Arnold, Lorenza Bordoli, Jurgen Kopp and Torsten Schwede, The SWISS-MODEL workspace : a web-based environment for protein structure homology modelling. Bioinformatics. 22, 195-201, (2006).
13. Brendel V, Bucher P, Nourbakhsh IR, Blaisdell BE and Karlin S, Methods and algorithms for statistical analysis of protein sequences. Proceedings of the National Academy of sciences of the United State of America. 89(6), 2002-2006.
14. J. Cheng, A. Z. Randall, M. J. Sweredoski and P. Baldi , SRACTCH: a protein structure and structural feature prediction server. Nucleic Acids Research. 33, W72-W76, (2005).
15. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi g, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, kuznetsov d, Leichti r, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, and Stockinger H, ExPASy: SIB bioinformatics resource portal, Nucleic Acids Res 40(W1): W597-W603, (2012).
16. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N., ConSurf 2010, calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research, 38, W529–W533, (2010).
17. Roseman, M. A., Hydrophilicity of polar amino acid sidechains is markedly reduced by flanking peptide bonds. Journal of Molecular Biology, 200, 513–522, (1988).
18. . Shandar Ahmad, Michael Gromiha, Hamed Fawareh and Akinori Sarai, *ASAView*: Database and tool for solvent accessibility representation in proteins, BMC Bioinformatics, (2004).
19. Doszta´nyi, Z, Fiser A, Simon I, Stabilizing centers in protein : identification, characterization and predictions. Journal of Molecular Biology. 272(4), 597-612, (1997).
20. Doszta´nyi Z., S. Magyar, C. S., Tusna´dy, G. E., & Simon, I., SCide, identification of stabilization centers in proteins. Bioinformatics, 19, 899–900, (2003).
21. Heitmann P, A model for sulfhydryl groups in proteins. Hydrophobic interactions of the cystein side chain in micelles, Eur. J. Biochem. 3 (3): 346–50, (January 1968).
22. Cruse, Julius M.; Lewis, Robert E., Atlas of Immunology. Boca Raton: CRC Press. ISBN 978-0849394898.(1998).
23. Eileen K. Clancy and Roy Duncan, Reovirus FAST Protein Transmembrane

Domains Function in a Modular, Primary Sequence-Independent Manner To Mediate Cell-Cell Membrane Fusion, (December 2008).
24. Guruprasad k, Reddy B.V.B., and Pandit M.W., Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng. 4, 155-161, (1990).
25. Ikai, A.j., Thermostability and aliphatic index of globular protein. Journal of biochemistry. 88, 1895-1898, (1980).
26. Magyar, C., Gromiha, M. M., Pujadas, G., Tusna´dy, G. E., & Simon, I., SRide, a server for identifying stabilizing residues in proteins. Nucleic Acids Research, 33, W303–W305, (2005).
27. Marsh JA, Teichmann SA, Relative Solvent Accessible Surface Area Predicts Protein Conformational Changes upon Binding, Structure, 19(6): 859–86, (2011).
28. M. Michael Gromiha, Gerard Pujadas, Csaba Magyar, Samuel Selvaraj, and Istvan Simon, Locating the Stabilizing Residues in alpha/beta Barrel Proteins Based on Hydrophobicity, Long-Range Interactions, and Sequence Conservation, PROTEINS: Structure, Function, and Bioinformatics, 55:316 – 329, (2004).