



INVESTIGATION OF ROLE AND CONSERVATION OF INSULIN RECEPTOR SUBSTRATE 2 PROTEIN

Meenu Sharma¹, C.S. Kapoor^{2*}

Department of Biotechnology¹, Department of Environmental Sciences², (PIBS), Pacific University of Higher Education and Research, Udaipur-313003, Rajasthan, India.

*Corresponding Author Email: drcskapoor@yahoo.co.in

ABSTRACT

Insulin receptor substrate (IRS) proteins have indispensable role signalling via insulin and insulin receptors interactions. The IRS 2 is one of the members of IRS family. In this study we explored the phylogenetic conservation and evolution of IRS 2 protein while involving statistical analysis and algorithms. IRS 2 is not only an alternate means to fight off diabetes but also play a crucial role in various types of cancers. To have a better understanding of evolution of IRS 2 and its direct and indirect role in other pathways, it's very important gain useful information of closely related species. Numerous methods including statistical analysis were employed to find its probability of occurrence. The study provides a rapid way to study evolution and make confident predictions to speed up the future research.

KEY WORDS

Diabetes, Evolution, Insulin receptor substrate 2, IRS 2, Phylogeny

INTRODUCTION:

Insulin receptor proteins, mainly IRS 1 and IRS 2 are cytoplasmic protein entities that play vital role in downstream signalling events mediated by the insulin receptor (IR) and the insulin like growth factor receptor1 (IGF1R) [1]. Insulin receptor substrate (IRS) proteins are believed to be involved in type2 diabetes as they arbitrate pleiotropic signals initiated by receptors for insulin and other cytokines. Studies have revealed that disruption of IRS 1 in animal experiments resulted in retarded growth and to our surprise insulin secretion increases to overcome the mild resistance developed to insulin [2]. In our earlier study we worked on the events involved in the insulin and insulin receptor interaction along with evolution and conservation of IRS 1 [3]. No doubt IRS proteins exhibit a high degree of homology and both IRS 1 and IRS 2 are capable of recruiting and trigger downstream signalling by activating phosphatidylinositol-3 kinase (PI3K) [1]. However, it is

now established that IRS 2, a member of insulin receptor substrate protein family is an alternate path for exciting insulin receptors [4]

Intracellular action of insulin hormone is initiated by the binding of insulin to its unambiguous cell surface receptors, the insulin receptor (IR) [5, 6]. The IR is a hetero-tetrameric protein having two alpha and two beta subunits that are connected by disulfide bridges. The alpha subunits are present extra-cellularly for the binding of insulin hormone and two beta subunits that are present within the cell membrane, i.e., it is a transmembrane subunit [7, 8]. The insulin binds to the extracellular alpha subunit of insulin receptor and results in its conformational change. This conformational change of insulin receptor is responsible for the activation of intracellular tyrosine kinase domain present in the beta subunits of insulin receptor [9, 10]. With the activation of tyrosine kinase domain, auto-phosphorylation of key tyrosine residues (Tyr-1158, Tyr-1162, and Tyr1163 as in homo sapiens) takes place in

the intracellular part of beta subunit of insulin receptor [11]

Any disturbance in IRS-2 has two-fold effects, firstly it impairs the function of pancreatic beta cells and secondly it impairs the peripheral insulin signalling. In animal study involving IRS-2-deficient mice shows the increasing deterioration in glucose homeostasis because of the development of insulin resistance in skeletal muscle and liver cells in addition of absence of presence of beta-cell compensation for this insulin resistance. It is concluded that disruption with the IRS 2 may be responsible for the pathophysiology of type 2 diabetes in humans [2].

MATERIALS AND METHODS:

Data Collection

The sequence data for IRS 2 is collected from the National Center for Biotechnology Information Database (NCBI) [12]. For the purpose amino acid sequence of insulin receptor substrate 2 peptide was obtained from NCBI against ID - NP_003740.2 for organism type – *Homo sapiens*. The functional protein sequences (in FASTA format) were gathered from the NCBI database and cross referred with expasy for annotation and are further analyzed using various online and offline bioinformatics tools.

Insilico alignment of IRS 2 sequence using align bl2seq

To obtain similar sequences for a better understanding of evolution and conservation IRS 2 undergo bl2seq (NCBI protein blast). The resultant parameters obtained from bl2seq protein blast are recorded as Score, Query cover, E-value, Identities, Positives and Gaps.

Sequence conservation, Multiple sequences alignment and generation of scores

The resultant 20 sequences are then subjected to NCBI-Cobalt and MEGA-ClustalW for multiple sequence alignment [13] with a gap opening penalty of 10 and a gap extension penalty of 0.1 for pair wise alignment and with gap opening penalty of 10 and gap extension penalty of 0.2 for multiple alignment with PAM matrix and 4 gap separation distance.

Phylogenetic Tree Construction

The sequences used in the last step for analysis conservation are now subjected to phylogenetic tree construction. The evolutionary history was inferred using the unweighted pair-group method using arithmetic averages (UPGMA) method. The

phylogenetic tree is constructed with branch lengths in dimension same as the dimensions that are used to draw phylogenetic tree from evolutionary distance. The evolutionary distances were computed as number of amino acid substitutions per site using the Poisson correction method and all positions containing gaps and missing data were eliminated

Analysis using MEGA7

Composition distance, pairwise distance and gamma parameters were obtained for the resulting 20 blast hits for each receptor [14]. The software calculates the composition distance for a given pair of sequences as a measure of the difference in amino acid composition.

The estimated value of the shape parameter for the discrete gamma distribution was calculated by maximum likelihood method and substitution pattern and rates were estimated under the Jones-Taylor-Thornton model (+G). Then to model evolutionary rate differences among five different categories we used discrete gamma distribution [+G]. For estimating maximum log values, a tree topology was computed using MEGA7.

Tajima's Neutrality Test was carried out [15, 16]. All positions containing gaps and missing data were eliminated and number of sequences, total number of sites, number of segregating sites and nucleotide diversity was calculated using divergence ratio.

RESULTS AND DISCUSSION:

NCBI blast is carried out to obtain a set of sequences that are more likely to be similar to homo sapiens IRS 2 protein sequence. For this bl2seq sequence alignment was carried out for IRS2. The matrix used for the purpose is BLOSUM62 with gap penalty of 11.1 against non-redundant protein sequences. Table 1 summarizes the important parameters considered for sending the blast query for searching hits. The resulting blast hits are filtered. The results that are partial, hypothetical and predicted sequences are neglected during this process. The selected top 20 hits are then used for further analysis in next steps. Refer Table 2 that enlists the organism IDs, expectation value, identities, positives, gaps and query coverage for In-silico alignment of IRS2 sequence. In this study we found that IRS 2 protein is evolutionary conserved with varied level of similarity with various organism species. Query coverage for these results is also 97% that provides us that most of the IRS 2 homo sapiens sequence is included in the blast results.

Table 1: Search parameters that are used in the blast query.

Algorithm	blastp (protein-protein BLAST)
Matrix used	BLOSUM62
Hit list size	100
Gap penalty	11.1
Expect	10
Word size	6
Window size	40
Genetic code	1
Databases	Non-redundant protein sequences

Table 2: In silico alignment of IRS2 sequence using align bl2seq.

S.No	Organism / Accession No	Score	Query cover	E-value	Identities	Positives	Gaps
1	gi 38683860 ref NP_003740.2 ;gi 62298062 sp Q9Y4H2.2 IRS2_HUMAN;gi 12247740 gb AAG50013.1	2644	100	0	100	100	0
2	gi 14537854 gb AAK66750.1 AF385932_1	2637	99	0	99.851	99.85	1
3	gi 829983074 ref XP_012607460.1	1988	100	0	88.749	90.6	14
4	gi 524954869 ref XP_005077412.1	1902	97	0	86.191	88.92	11
5	gi 124487073 ref NP_001074681.1 ;gi 341940841 sp P81122.2 IRS2_MOUSE;gi 223461918 gb AAI47581.1	1899	100	0	85.469	88.38	11
6	gi 1195735440 ref XP_021026522.1	1896	100	0	85.469	88.3	11
7	gi 148690106 gb EDL22053.1	1895	98	0	85.66	88.62	11
8	gi 537265949 gb ERE91019.1	1882	94	0	86.912	89.58	12
9	gi 274323811 ref NP_001162104.1 ;gi 149057571 gb EDM08814.1	1868	100	0	85.863	88.91	12
10	gi 1195523702 ref XP_021075109.1	1867	100	0	85.48	88.31	12
11	gi 1168093 gb AAB35237.1	1857	100	0	84.948	87.85	11
12	gi 1147386523 ref XP_020041910.1	1848	97	0	86.071	88.72	10

13	gi 1212201875 ref XP_021551623.1	1844	97	0	88.249	89.84	11
14	gi 1211409608 ref XP_021493195.1	1766	97	0	84.226	87.32	14
15	gi 532062622 ref XP_005317187.1	1630	85	0	86.964	89.33	7
16	gi 512993983 ref XP_004858614.1	1593	97	0	77.235	81.22	22
17	gi 1190444331 ref XP_020825718.1	1518	97	0	73.007	79.01	23
18	gi 1190444329 ref XP_020825717.1	1517	97	0	72.933	79.01	23
19	gi 1210014038 ref XP_021401280.1	1144	97	0	59.145	64.45	45
20	gi 1210014036 ref XP_021401279.1	1142	97	0	59.114	64.43	45

Phylogenetic tree is constructed for the filtered 20 IRS 2 sequences. The statistical method used for the purpose is maximum likelihood method and it involved 100 bootstrap replications to obtain a best inferred tree. In the process poisson method for amino acid substitution is used. The parameters used for the tree construction are listed in table 3.

Table 3: Analysis preferences selected for phylogenetic tree construction

Analysis	
Analysis	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
Phylogeny Test	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	100
Substitution Model	
Substitutions Type	Amino acid
Model/Method	Poisson model
Rates and Patterns	
Rates among Sites	Uniform rates
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Tree Inference Options	
ML Heuristic Method	Nearest Neighbour Interchange (NNI)
Initial Tree for ML	Make initial tree automatically (Default NJ/BioNJ)
Branch Swap Filter	None
System Resource Usage	
Number of Threads	1

Here we found that most of the IRS 2 sequence is conserved among group of species, namely, *Homo sapiens*, *Mus musculus*, *Mus caroli*, *Mus pahari*, *Rattus norvegicus*, *Meriones unguiculatus*, *Mesocricetus*

auratus, *Cricetulus griseus*, *Castor canadensis*, *Microcebus murinus*, *Neomonachus schauinslandi*, *Ictidomys tridecemlineatus*, *Heterocephalus glaber*, X1 *Phascolarctos cinereus*, X2 *Lonchura striata domestica*, X1 *Lonchura striata domestica*

The work provides the reason behind using animal insulin as a substitute for insulin for diabetic human population.

The evolutionary history was inferred by using the Maximum Likelihood Estimation (MLE) based on the Poisson correction model [17]. The tree with the highest log likelihood (-7143.0947) is shown. Initial tree(s) are

obtained automatically with the application of Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The phylogenetic tree is build with branch lengths measured in terms of number of substitutions per site. The analysis involved 20 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 894 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [14].

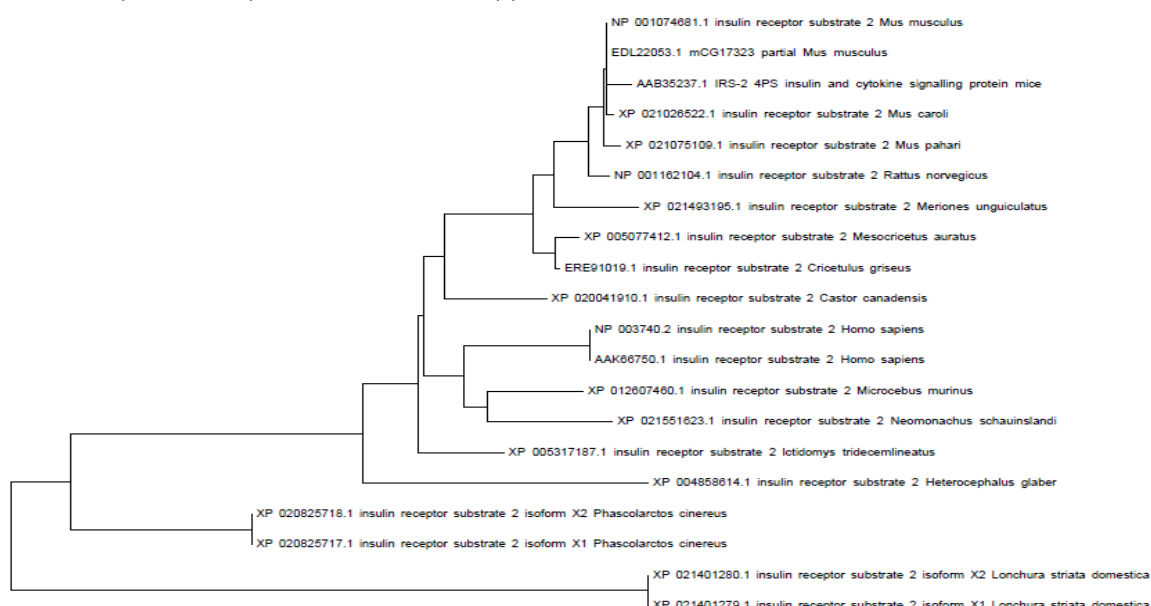


Fig 1. Molecular Phylogenetic analysis by Maximum Likelihood method

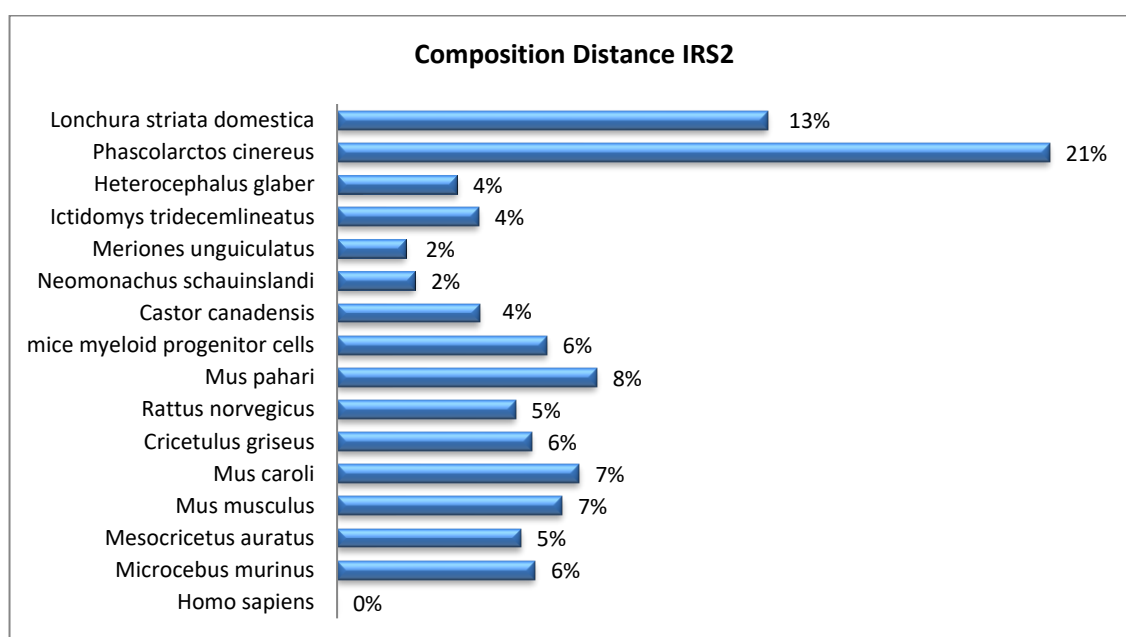


Fig 2. Bar chart showing composition distance of homo sapiens against various organism hits of IRS2 sequence.

Pairwise Distance IRS2

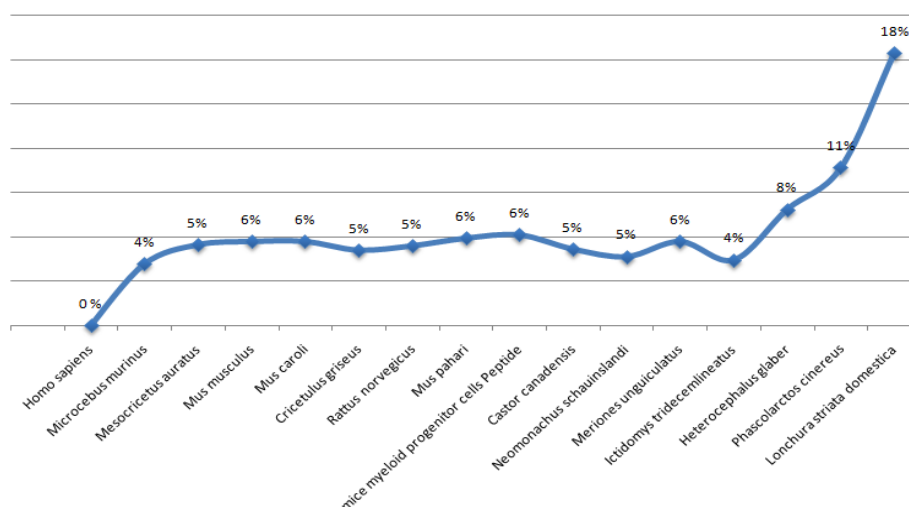


Fig 3. Line chart showing pairwise distance of homo sapiens against various organism hits of IRS2 sequence.

In figure 2, Bar chart shows the graphical representation of composition distance obtained for homo sapiens against various organism hits of IRS 2 sequence including homo sapiens. In this analysis of IRS 2 sequence of homo sapiens maximum composition distance is obtained in case of *Phascolarctos cinereus* (21%) followed by *Lonchura striata domestica* (13%) and minimum composition distance is obtained against *Meriones unguiculatus* and *Neomonarchus schauinslandi* (2% each) followed by *Heterocephalus glaber*, *Ictidomys tridecemlineatus* and *Castor canadensis* (4% each). This simply means that IRS 2 homo sapiens sequence is more related to *Meriones unguiculatus* and *Neomonarchus schauinslandi* as compared to other sequences under study.

Similarly, figure 3, displays the pairwise distance of homo sapiens against various organism hits of IRS2 sequence using MEGA7 which results in maximum pairwise distance with *Lonchura striata domestica* (18%) and minimum pairwise distance with *Microcebus murinus*, *Ictidomys tridecemlineatus* (4% each) and *Mesocricetus auratus*, *Cricetulus griseus*, *Rattus norvegicus*, *Castor canadensis*, *Neomonachus schauinslandi* (5% each). Here we can say that homo

sapiens IRS 2 is more related to *Microcebus murinus*, *Ictidomys tridecemlineatus* along with and *Mesocricetus auratus*, *Cricetulus griseus*, *Rattus norvegicus*, *Castor canadensis*, *Neomonachus schauinslandi* as compared to *Lonchura striata domestica*.

We use several categories of rates with equal probability to approximate the gamma distribution for each category. The mean of each category is then used to represent all the rates falling in the category. Maximum Likelihood estimate of gamma parameter for Site rates was also calculated using MEGA7. Refer table 4 for IRS 2 Maximum Likelihood estimate of gamma parameter for site Rates. In this table maxima and minima were displayed for various Gamma categories for #5 positions. The line graph in figure 4 shows the relative plot of minima and maxima of five Gamma Categories using Maximum Likelihood Estimation of Gamma Parameters for site rates for IRS 2.

Refer table 5 that displays the maxima and minima for IRS2 MLE of Gamma parameter for Site Rates obtained using MEGA7. The line graph in figure 4 shows the relative plot of minima and maxima of five Gamma Categories using MLE of Gamma Parameters for Site rates for IRS2 sequence.

Table 4: IRS2 Maximum Likelihood Estimate of Gamma Parameter for Site Rates

Value Type	Gamma Categories				
	#1	#2	#3	#4	#5
max	0.32	0.301	0.318	0.973	3.631
min	0	0	0	0	0.011

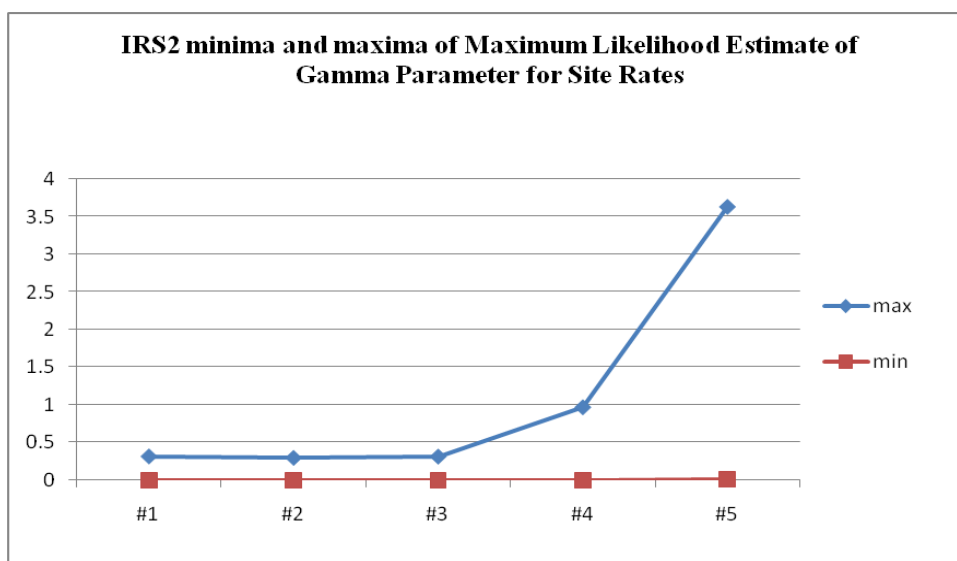


Fig 4. Line graph showing relative plot of minima and maxima of five Gamma Categories using Maximum Likelihood Estimate of Gamma Parameters for Site rates for IRS2

Table 5: Neutrality test statistics for IRS1 sequences

Peptide Name	m	S	p_s	Θ	π	D
IRS 2	20	334	0.373602	0.105307	0.121830	0.654357

Abbreviations: m = number of sequences, n = total number of sites, S = Number of segregating sites, $p_s = S/n$, $\Theta = p_s/a_1$, π = nucleotide diversity, and D is the Tajima test statistic

We also conducted Tajima's test of neutrality [14, 15] that compares the number of segregating sites per site. It's an important statistic that is widely used in population genetics. The site is considered independently segregating if one can find two or more nucleotides at that site in comparison to m number of sequences under study. We found that during simulations the D value as a small positive number which is not 0 but very close to zero but never converges to zero. This is because of finite number of samples.

When Tajima's D is greater than 0, it suggests that either there is a recent population bottleneck or there is presence of some form of balancing selection. A positive Tajima's D signifies low levels of low and high frequency polymorphisms, indicating a decrease in population size and/or balancing selection. In such population, rare alleles are present at high frequencies.

Recent researches have established that dysfunction related to the IRS 2 contributes to the pathophysiology of human type 2 diabetes [18]. Some other researchers have also proved that IRS 2 is also capable of arbitrating anti-apoptosis with direct communication with various other cell surface receptors. IRS 2 has also been studied as a part of many vital biological processes that have

important role in cell proliferation, metabolism, clonogenicity and cell survival. To be specific IRS 2 not only play crucial role in regulation of diabetes, but also arbitrate antiapoptotic and mitogenic signaling from insulin receptor (IR), erythropoietin receptor (EPOR), insulin-like growth factor 1 (IGF1R), growth hormone (GH), thrombopoietin receptor (MPL), vascular endothelial growth factor receptor VEGFR (KDR), leptin LEP, interleukins and interferons (IF). Thus, stimulating growth factors cytokines that in the long run are responsible for the proliferation and survival of normal as well as the cancerous cells [19-28]

A study has also revealed that signalling through IRS1 and IRS2 also leads to distinct type of tumor cell outcomes in vitro and in vivo. In vitro, IRS1 regulates cell proliferation and growth and IRS2 regulates metabolism, survival and invasion. In vivo, Irs2 is a positive regulator of tumor metastasis, whereas Irs1 does not promote metastasis [1].

CONCLUSION:

It is now clear that IRS 2 protein is important not only as an alternate to regulate diabetes, but it also plays important role in regulating various cancers. Such

important roles make this protein draw interest and opens doors for future research. In present work we have make use of rapid and innovative approach to study evolutionary and functional relationship among insulin receptor substrate 2 protein on the basis of consensus to other animal species. Our study provides a fast method to compare and calculate evolutionary conservation rate. All this information will no doubt will speed up the future research related to IRS 2 protein.

REFERENCES:

- [1] Mercado-Matos, JR. A Mechanistic Investigation of Insulin Receptor Substrate 2 Function in Breast Cancer Progression. University of Massachusetts Medical School. GSBS Dissertations and Theses. Paper 918. DOI: 10.13028/M25D63 (2017).
- [2] Withers DJ1, Gutierrez JS, Towery H, Burks DJ, Ren JM, Previs S, Zhang Y, Bernal D, Pons S, Shulman GI, Bonner-Weir S, White MF, Disruption of IRS-2 causes type 2 diabetes in mice, *Nature*. Feb 26;391(6670):900-4 (1998)
- [3] Sharma M, Kapoor C. S., I-IR Interaction Exploration and Computational Investigation of Evolution of IRS1, *J. Applicable Chem.*, 7 (5):1319-1329 (2018)
- [4] Patti ME, Sun XJ, Bruening JC, Araki E, Lipes MA, White MF, Kahn CR, 4PS/insulin receptor substrate (IRS)-2 is the alternative substrate of the insulin receptor in IRS-1-deficient mice *J Biol Chem* Oct 20;270(42):24670-3 (1995) PMID 7559579
- [5] Pirola L, Johnston AM, Van Obberghen E. Modulation of insulin action. *Diabetologia*.47:170–84 (2004).
- [6] Myers MG Jr, White MF. Insulin signal transduction and the IRS proteins. *Annu Rev Pharmacol Toxicol*. 36:615–58 (1996)
- [7] De Meyts P, Whittaker J. Structural biology of insulin and IGF1 receptors: implications for drug design. *Nat Rev Drug Discov*. 1:769–83 (2002).
- [8] Hubbard SR. The insulin receptor: both a prototypical and atypical receptor tyrosine kinase. *Cold Spring Harb Perspect Biol*.5:a008946 (2013).
- [9] Ward CW, Lawrence MC. Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. *BioEssays*. 31:422–34 (2009).
- [10] Du Y, Wei T. Inputs and outputs of insulin receptor. *Protein Cell*. 5:203–13 (2014).
- [11] Wei L, Hubbard SR, Hendrickson WA, Ellis L. Expression, characterization, and crystallization of the catalytic core of the human insulin receptor proteintyrosine kinase domain. *J Biol Chem*.270:8122–30 (1995).
- [12] Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY.. Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res*, 35:D5–12 (2007).
- [13] Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD., Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*, 31:3497–500, (2003)
- [14] Kumar S., Stecher G, Tamura K., MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*, 33(7), 1870-1874 (2016). doi:10.1093/molbev/msw054
- [15] Tajima F, Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*, 123:585-595 (1989).
- [16] Tajima F, Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis. *Genetics*, 135(2), 599–607 (1993).
- [17] Zuckerkandl E. and Pauling L, Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York (1965).
- [18] Withers DJ, Gutierrez JS, Towery H, Burks DJ, Ren JM, Previs S, Zhang Y, Bernal D, Pons S, Shulman GI, Bonner-Weir S, White MF, Disruption of IRS-2 causes type 2 diabetes in mice. *Nature*. Feb 26;391(6670):900-4 (1998)
- [19] Argetsinger LS, Norstedt G, Billestrup N, White MF, Carter-Su C, Growth hormone, interferon-gamma, and leukemia inhibitory factor utilize insulin receptor substrate-2 in intracellular signaling *J Biol Chem* Nov 15;271(46):29415-21 (1996) PMID 8910607
- [20] Dearth RK, Cui X, Kim HJ, Kuaitse I, Lawrence NA, Zhang X, Divisova J, Britton OL, Mohsin S, Allred DC, Hadsell DL, Lee AV, Mammary tumorigenesis and metastasis caused by overexpression of insulin receptor substrate 1 (IRS-1) or IRS-2 *Mol Cell Biol* Dec;26(24):9302-14 (2006) PMID 17030631
- [21] Gibson SL, Ma Z, Shaw LM, Divergent roles for IRS-1 and IRS-2 in breast cancer metastasis *Cell Cycle* Mar 15;6(6):631-7 (2007) PMID 17361103
- [22] Johnston JA, Wang LM, Hanson EP, Sun XJ, White MF, Oakes SA, Pierce JH, O'Shea JJ, Interleukins 2, 4, 7, and 15 stimulate tyrosine phosphorylation of insulin receptor substrates 1 and 2 in T cells. Potential role of JAK kinases *J Biol Chem*. Dec 1;270(48):28527-30 (1995). PMID 7499365
- [23] Plataniias LC, Uddin S, Yetter A, Sun XJ, White MF, The type I interferon receptor mediates tyrosine phosphorylation of insulin receptor substrate 2 *J Biol Chem* Jan 5;271(1):278-82 (1996) PMID 8550573
- [24] Sun XJ, Wang LM, Zhang Y, Yenush L, Myers MG Jr, Glasheen E, Lane WS, Pierce JH, White MF, Role of IRS-2 in insulin and cytokine signalling *Nature* Sep 14;377(6545):173-7 (1995) PMID 7675087
- [25] Uddin S, Yenush L, Sun XJ, Sweet ME, White MF, Plataniias LC, Interferon-alpha engages the insulin

- receptor substrate-1 to associate with the phosphatidylinositol 3'-kinase *J Biol Chem* Jul 7;270(27):15938-41 (1995) PMID 7608146
- [26] Verdier F, Chrétien S, Billat C, Gisselbrecht S, Lacombe C, Mayeux P, Erythropoietin induces the tyrosine phosphorylation of insulin receptor substrate-2 An alternate pathway for erythropoietin-induced phosphatidylinositol 3-kinase activation *J Biol Chem*. Oct 17;272(42):26173-8 (1997). PMID 9334184
- [27] White MF, Kahn CR, The insulin signaling system, *J Biol Chem* Jan 7;269(1):1-4 (1994) PMID 8276779
- [28] Yenush L, White MF, The IRS-signalling system during insulin and cytokine action, *Bioessays* Jun;19(6):491-500 (1997) PMID 9204766

Received:04.08.18, Accepted: 07.09.18, Published:01.10.2018

***Corresponding Author:**

C.S. Kapoor*

Email: drcskapoor@yahoo.co.in