Journal of Advances in Biology & Biotechnology
8(2): 1-7, 2016; Article no.JABB.26715
ISSN: 2394-1081

SCIENCEDOMAIN *international*
www.sciencedomain.org

Computational Analysis of Genomic Regions of Human Insulin Receptor Gene

Shahid Raza[1*] and Hira Mubeen[1]

[1]Department of Biotechnology, University of South Asia, Lahore, Pakistan.

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JABB/2016/26715
Editor(s):
(1) Laura Pastorino, Laboratory of Nanobioscience and Medical Informatic, Dept. Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genoa, Italy.
Reviewers:
(1) Eman Mohammed Faruk, Benha University, Egypt.
(2) Neslihan Toyran, Dokuz Eylül University, Turkey.
(3) Rajendra Nath, King George's Medical University, Lucknow (UP), India.
Complete Peer review History: http://www.sciencedomain.org/review-history/15599

Original Research Article

Received 29th April 2016
Accepted 21st June 2016
Published 31st July 2016

ABSTRACT

Insulin plays an essential role in metabolic control of multiple cellular processes. A number of DNA sequences that regulate transcription of the insulin gene have already been identified. The regulation of carbohydrates and fats can be operated by insulin specific receptor molecules. Previously, insulin was found to be a polypeptide in 1928 with its amino acid sequence identified in 1952. In fact it was considered as a dipeptide, containing A and B chains respectively, linked by disulphide bridges. The insulin receptor plays a critical role both in directing insulin to specific target tissues and in initiating the response of these tissues to the hormone. Hence, the study was performed to analyze the insulin gene receptor associated with regulation of various mechanism and also with several insulin deficiency syndromes. Identification of protein domains from PROSITE, Pfam and UniPROt reveals various conserved regions associated with insulin receptor gene (INSR). For this, various bioinformatics tools and software's were used.

Keywords: Disulphide bridges; receptor; Pfam; transcription; PROSITE.

**Corresponding author: E-mail: dr.raza03@gmail.com;*

1. INTRODUCTION

Insulin is an important hormone produced by specialized beta cells in the pancreas. It is mainly involved in regulation of carbohydrates and fats metabolism. It maintains normal blood glucose levels by facilitating cellular glucose uptake. Also, it supports the cell division and growth through its mitogenic effects. It was observed that the blood glucose levels fall below a certain level and the body begins to use stored glucose as an energy source through process of glycogenolysis. It helps in breakdown of glycogen stored in the liver and muscles into glucose. This glucose can than used as an energy source for many processes.

1.1 Insulin linked Receptors

Certain receptors are linked with insulin gene. The insulin receptor is known to be a transmembrane receptor activated by insulin itself. A number of different receptors belongs to tyrosine kinase receptor family [1]. The insulin receptor plays a necessary role in the regulation of glucose homeostasis [2,3]. Hence, the insulin receptor is encoded by a single gene *INSR*. Resistance against several genes exist. According to some studies, insulin receptor is a tetramer of 2 alpha and 2 beta subunits. The alpha and beta subunits are coded by a single gene and are joined by disulfide bonds [4].

The insulin interacts with target cells by specific receptor molecules located on surface of plasma membrane [5,6]. This specific receptor plays an important role to initiate a response of these tissues to hormones [7,8]. Hence, it was noticed that insulin receptors are present on surface of all cells [9,10]. Many examples were found. The most important is glucocorticoids which enhance transcription of the insulin receptor gene whereas comparatively insulin down-regulates its own receptor [11,12].

It was also found that, receptor tyrosine kinase helps to mediate the pleiotropic actions of insulin. Moreover, binding of insulin leads to phosphorylation of several intracellular substrates, including the insulin receptor substrates and many other signaling intermediates. Each of these phosphorylated proteins can serve as docking proteins for other signaling proteins that contain Src-homology-2 domains.

The study was performed to analyze the human insulin gene receptor. The selected gene coding sequence was obtained with translates protein sequence. The genomic regions were analyzed with various splice variants showing several protein domains. This in silico study was done by using multiple bioinformatics tools.

2. MATERIALS AND METHODS

Sequence analysis of INSR gene depicted coding region of gene. Transcripts of INSR gene was analysed which are mainly involved in formation of functional proteins. Protein domain shows the conserved area of protein within the protein family.

2.1 INSR Nucleotide and Protein Sequence Retrival

NCBI's is used to study gene homologs, orthologues, protein domains and for gene expression studies. Nucleotide and protein sequence was obtained by using NCBI data repository.

2.2 Chromosme Map

Chromosome map was obtained for INSR gene from ensemble software. The map shows the exact loction of INSR.

2.3 Genomic Regions

The complete genomic regions were analysed which shows all the variations with exons. This was done by using bioinformatics approach.

2.4 Transcript Identification

Trancript map was analysed for INSR gene. This was done by using ensemble software. Ensemble assigns separate ID’s to all identified transcripts with complete details. Transcripts on total genomic regions are analysed with annotated gene regions and splice variants.

2.5 Analysis of Splice Variants for Identification of Protein Domains

The transcripts were analysed for identification of putative domains with the proetin sequence. The Conserved Domain Database (NCBI) is a resource for the annotation of functional units in proteins. Study of these domain models will

help to predict 3D structure and function of protein.

3. RESULTS

Insulin is a globular protein having two chains including the alpha chain which consists of 21 residues and the Beta chain which consists of 30 residues. Insulin is located on the short arm of chromosome 11 and synthesized in the β cells of the pancreatic islets of Langherhans. The analysis of INSR gene shown various transcripts, 77 orthologues, 13 paralogs and associated with 1 Ensemble protein familily with 27 distinguishing phenotypes. The results are shown alongwith respective findings.

3.1 INSR Gene Chromosome Map

The INSR gene was located on chromosome 19. The Fig. 1 shows the complete map of chromosome.

3.2 Nucleotide Sequence of INSR Gene

The nucleotide sequence of human INSR gene was retrieved from NCBI-Genbank. Fig. 2 shows the complete coding sequence of selected gene.

3.3 Translated Sequence of INSR Gene

The nucleotide sequence of gene was translated to analyze the protein coded by INSR gene.

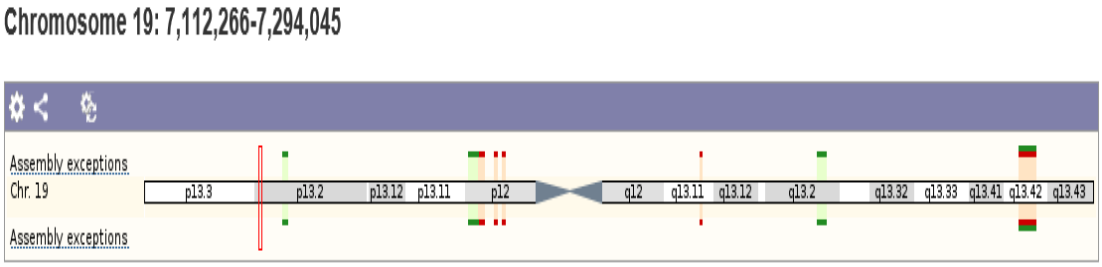

Fig. 1. Shows the chromosome map obtained from ensemble data repository

ATGGGCACCGGGGGCCGGCGGGGGGCGGCGGCCGCGCCGCTGCTGGTGGCGGTGGCCGCGCTGCTACTGGGCGCCGCGGGCCACCTGTA
CCCCGGAGAGGTGTGTCCCGGCATGGATATCCGGAACAACCTCACTAGGTTGCATGAGCTGGAGAATTGCTCTGTCATCGAAGGACACTT
GCAGATACTCTTGATGTTCAAAACGAGGCCCGAAGATTTCGAGACCTCAGTTTCCCCAAACTCATCATGATCACTGATTACTTGCTGCTCTTC
CGGGTCTATGGGCTCGAGAGCCTGAAGGACCTGTTCCCCAACCTCACGGTCATCCGGGGATCACGACTGTTCTTTAACTACGCGCTGGTCAT
CTTCGAAGGTTCACCTCAAGGAACTCGGCCTCTACAACCTGATGAACATCACCCGGGGTTCTGTCCGCATCGAGAAGAACATGAGCTCTGT
TACTTGGCCACTATCGACTGGTCCCGTATCCTGGATTCCGTGGAGGATAATCACATCGTGTTGAACAAAGATGACAACGAGGAGTGTGGA
GACATCTGTCCGGGTACCGCGAAGGGCAAGACCAACTGCCCCGCCACCGTCATCAACGGGCAGTTTGTCGAACGATGTTGGACTCATAGTC
ACTGCCAGAAAGTTTGCCCGACCATCTGTAAGTCACACGGCTGCACCGCCGAAGGCCTCTGTTGCCACAGCGAGTGCCTGGGCAACTGTTC
TCAGCCCGACGACCCCACCAAGTGCGTGGCCTGCCGCAACTTCTACCTGGACGGCAGGTGTGTGGAGACCTGCCCGCCCCCGTACTACCACT
TCCAGGACTGGCGCTGTGTGAACTTCAGCTTCTGCCAGGACCTGCACCACAAATGCAAGAACTCGCGGAGGCAGGGCTGCCACCAATACG
TCATTCACAACAACAAGTGCATCCCTGAGTGTCCCTCCGGGTACACGATGAATTCCAGCAACTTGCTGTGCACCCCATGCCTGGGTCCCTGT
CCCAAGGTGTGCCACCTCCTAGAAGGCGAGAAGACCATCGACTCGGTGACGTCTGCCCAGGAGCTCCGAGGATGCACCGTCATCAACGGG
AGTCTGATCATCAACATTCGAGGAGGCAACAATCTGGCAGCTGAGCTAGAAGCCAACCTCGGCCTCATTGAAGAAATTTCAGGGTATCTAA
AAATCCGCCGATCCTACGCTCTGGTGTCACTTTCCTTCTTCCGGAAGTTACGTCTGATTCGAGGAGAGACCTTGGAAATTGGGAACTACTCC
TTCTATGCCTTGGACAACCAGAACCTAAGGCAGCTCTGGGACTGGAGCAAACACAACCTCACCACCACTCAGGGGAAACTCTTCTTCCACTA
TAACCCCAAACTCTGCTTGTCAGAAATCCACAAGATGGAAGAAGTTTCAGGAACCAAGGGGCGCCAGGAGAGAAACGACATTGCCCTGAA
GACCAATGGGGACAAGGCATCCTGTGAAAATGAGTTACTTAAATTTTCTTACATTCGGACATCTTTTGACAAGATCTTGCTGAGATGGGAG
CCGTACTGGCCCCCCGACTTCCGAGACCTCTTGGGGTTCATGCTGTTCTACAAAGAGGCCCCTTATCAGAATGTGACGGAGTTCGATGGGC
AGGATGCGTGTGGTTCCAACAGTTGGACGGTGGTAGACATTGACCCACCCCTGAGGTCCAACGACCCCAAATCACAGAACCACCCAGGGT
GGCTGATGCGGGGTCTCAAGCCCTGGACCCAGTATGCCATCTTTGTGAAGACCCTGGTCACCTTTTCGGATGAACGCCGGACCTATGGGGC
CAAGAGTGACATCATTTATGTCCAGACAGATGCCACCAACCCCTCTGTGCCCCTGGATCCAATCTCAGTGTCTAACTCATCATCCCAGATTAT
TCTGAAGTGGAAACCACCCTCCGACCCCAATGGCAACATCACCCACTACCTGGTTTTCTGGGAGAGGCAGGCGGAAGACAGTGAGCTGTTC
GAGCTGGATTATTGCCTCAAAGGGCTGAAGCTGCCCTCGAGGACCTGGTCTCCACCATTCGAGTCTGAAGATTCTCAGAAGCACAACCAGA
GTGAGTATGAGGATTCGGCCGGCGAATGCTGCTCCTGTCCAAAGACAGACTCTCAGATCCTGAAGGAGCTGGAGGAGTCCTCGTTTAGGA
AGACGTTTGAGGATTACCTGCACAACGTGGTTTTCGTCCCCAGAGAGAGTCACTGACCTCATGCGCATGTGCTGGCAATTCAACCCCAAGA
TGAGGCCAACCTTCCTGGAGATTGTCAACCTGCTCAAGGACGACCTGCACCCCAGCTTTCCAGAGGTGTCGTTCTTCCACAGCGAGGAGAA
CAAGGCTCCCGAGAGTGAGGAGCTGGAGATGGAGTTTGAGGACATGGAGAATGTGCCCCTGGACCGTTCCTCGCACTGTCAGAGGGAGG
AGGCGGGGGGCCGGGATGGAGGGTCCTCGCTGGGTTTCAAGCGGAGCTACGAGGAACACATCCCTTACACACACATGAACGGAGGCAAG
AAAAACGGGCGGATTCTGACCTTGCCTCGGTCCAATCCTTCCTAA

Fig. 2. Shows the INSR nucleotide sequence obtained from Genbank-NCBI

3.4 Genomic Regions and Transcripts

The analysis of total genomic content represents whole coding area with predicted exons and gene variants. Small variations due to single nucleotide change covers maximum region. Many other genetic variations are also found which have been found to be distributed within the genome. Total 147 different variants were found in the selected gene. The intron spanning regions were found dispersed over coding region of genome.

3.5 INSR Gene Transcripts with Ensemble ID

Transcript analysis shown 7 different transcripts of different sizes. Out of which 3 were found to be protein coding. The Fig. 5 shows different transcripts with their ensemble ID.

3.6 INSR Reported Protein

Analysis of INSR shows three proteins within the protein family. Three separate ID's were assigned by ensemble software for their identity. The annotated sequence shows to subunits including alpha and beta.

3.7 Splice Variants

The INSR was assigned a specific ID by ensemble software. ENST00000302850 INSR-001 was assigned to INSR gene. Analysis shows different domains from PROSITE, Pfam, Prints and some other protein database. The residue positions are indicated in the Fig. 7. The location of each domain and receptor molecules highlights the particular catalytic domain associated with each family.

"MGTGGRRGAAAAPLLVAVAALLLGAAGHLYPGEVCPGMDIRNNLTRLHELENCSVIEGHLQILLMFKTRPEDFRDLSF
PKLIMITDYLLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLYNLMNITRGSVRIEKNNELCYLATID
WSRILDSVEDNHIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVERCWTHSHCQKVCPTICKSHGCTAEGLCCHSE
CLGNCSQPDDPTKCVACRNFYLDGRCVETCPPPYYHFQDWRCVNFSFCQDLHHKCKNSRRQGCHQYVIHNNKCIPECP
SGYTMNSSNLLCTPCLGPCPKVCHLLEGEKTIDSVTSAQELRGCTVINGSLIINIRGGNNLAAELEANLGLIEEISGYLKIR
RSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLTTTQGKLFFHYNPKLCLSEIHKMEEVSGTKG
RQERNDIALKTNGDKASCENELLKFSYIRTSFDKILLRWEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSW
TVVDIDPPLRSNDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLDPISVSNSSS
QIILKWKPPSDPNGNITHYLVFWERQAEDSELFELDYCLKGLKLPSRTWSPPFESEDSQKHNQSEYEDSAGECCSCPKTD
SQILKELEESSFRKTFEDYLHNVVFVPRKTSSGTGAEDPRPSRKRRSLGDVGNVTVAVPTVAAFPNTSSTSVPTSPEEHRP
FEKVVNKESLVISGLRHFTGYRIELQACNQDTPEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVVHLMWQEPKE
PNGLIVLYEVSYRRYGDEELHLCVSRKHFALERGCRLRGLSPGNYSVRIRATSLAGNGSWTEPTYFYVTDYLDVPSNIA
KIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGPLGPLYASSNPEYLSASDVFPCSVYVPDEWEVSREKITLLRELGQGSFGMVY
EGNARDIIKGEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMAHGDLKSYLR
SLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMVAHDFTVKIGDFGMTRDIYETDYYRKG
GKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLWEITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLM
RMCWQFNPKMRPTFLEIVNLLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDMENVPLDRSSHCQREEAGGRDGGSS
LGFKRSYEEHIPYTHMNGGKKNGRILTLPRSNPS"

Fig. 3. Shows the translated protein sequence obtained from Just bio software

Fig. 4. Shows the genomic region with variants and exons obtained from ensemble

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq
INSR-002	ENST00000341500	8954	1370aa	Protein coding	CCDS42487	P06213	NM_001079817 NP_001073285
INSR-001	ENST00000302850	4721	1382aa	Protein coding	CCDS12176	P06213	NM_000208 NP_000199
INSR-004	ENST00000600492	653	152aa	Protein coding	-	M0R3E6	-
INSR-003	ENST00000598216	2961	No protein	Retained intron	-	-	-
INSR-005	ENST00000601099	570	No protein	Retained intron	-	-	-
INSR-006	ENST00000597211	532	No protein	Retained intron	-	-	-
INSR-007	ENST00000593970	419	No protein	Retained intron	-	-	-

Fig. 5. Shows the INSR transcripts obtained from ensemble software

Family ID	Consensus annotation	Other INSR proteins in this family
PTHR24416_SF140 (1 gene)	INSULIN RECEPTOR IR EC_2.7.10.1 CD220 ANTIGEN [CONTAINS INSULIN RECEPTOR SUBUNIT ALPHA; INSULIN RECEPTOR SUBUNIT BETA]	• ENSP00000342838 (INSR-002) • ENSP00000303830 (INSR-001) • ENSP00000473170 (INSR-004)

Fig. 6. Shows the protein family for INSR obtained from annotation reports

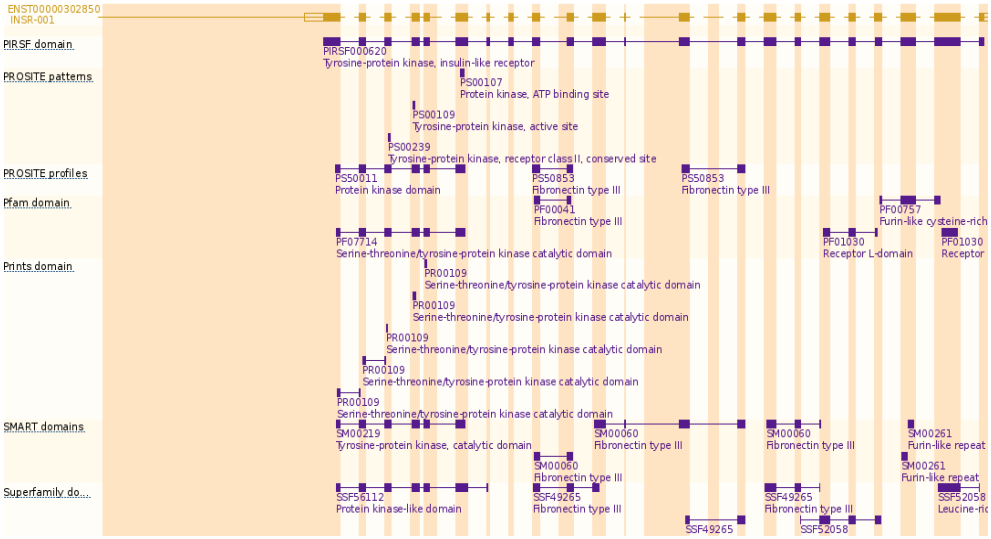

Fig. 7. Shows the splice variants obtained from ensemble software

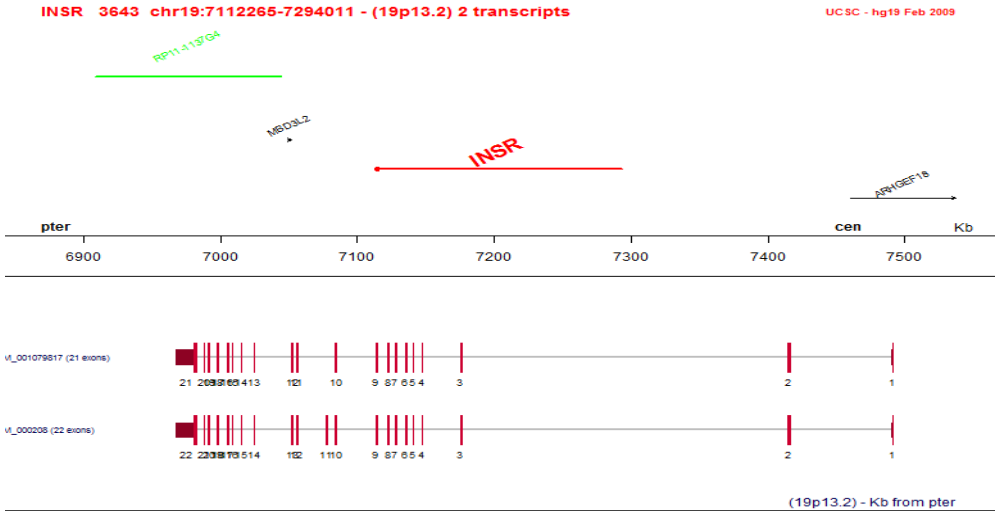

Fig. 8. Shows the internal view of INSR

4. DISCUSSION AND CONCLUSION

Insulin regulates the metabolic homeostasis and is mainly secreted from pancreatic beta cells in response to some nutrient stimulation. According to some previous studies, Insulin gene expression phenomena begins early in the embryonic development of the pancreas and supposed to be tightly-regulated throughout adult life. Hence, the mouse genome contains two insulin genes, *Ins1* (GeneID: 16333) and *Ins2* (GeneID: 16334). The *Ins2* gene has greater structural and functional similarity to other mammalian insulin genes, like homo sapiens [13].

The entire 1,370-amino acid sequence of the insulin receptor from a cDNA clone was studied by some scientist [14]. It was found that the precursor starts with a 27-amino acid signal sequence, followed by the receptor alpha subunit, a processing enzyme cleavage site, then the beta subunit containing a single 23-amino acid transmembrane sequence. Moreover, the differences in in molecular mass, carbohydrate composition, and antigenicity between the insulin receptor alpha subunit in liver, muscle and mainly the adipose tissue [15].

Insulin plays a vital role in the regulation of vertebrate metabolism. The initial an important step in hormonal control involves the specific binding of insulin to the insulin receptor (INSR) [16-18]. However, it has been observed that binding of insulin to its receptor mainly triggers an activation of tyrosine kinase. Insulin increases sugar transport and initiates active protein synthesis. According to scientists, a conformational change of insulin occurs at the positions of two structural switches when insulin binds to its receptor. Also, studies shown the single insulin chain linked by a B29-A1 covalent bond is completely inactive [19]. However, process of insulin binding to its receptor is still unknown. The complex has not yet been characterized either by X-diffraction or by NMR. However, two insulin receptor mRNA transcripts resulting from alternative splicing of exon 11 in the receptor gene are expressed in a highly regulated tissue-specific fashion. The relative abundance of these 2 mRNA species in human tissues including the one containing exon 11 shows a much dominance in liver, whereas the isoform in which exon 11 has been spliced out showing a comparable predominance in leukocytes. Hence, it was concluded that identification of INSR gene transcripts and protein domains will help to analyze the receptor regulatory functions. And also to study various syndromes associated with insulin gene receptor dis functioning at genomic level. This will also help to overcome the gene silencing problems and to predict 3 D structure of proteins.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Chang X, Jorgensen AM, Bardrum P, Led JJ. Solution structures of the R6 human insulin hexamer. Biochemistry. 1997; 36(31):9409–22.
2. Wright James R, Yang Hua, Hyrtsenko Olga, Xu Bao-Yu, Yu Weiming, Pohajdak Bill. A review of piscine islet xenotransplantation using wild-type tilapia donors and the production of transgenic tilapia expressing a "humanized" tilapia insulin. Xenotransplantation. 2014;21(6): 485–495.
3. Sonksen P, Sonksen J. Insulin: understanding its action in health and disease. Br J Anaesth. 2000;85(1):69–79.
4. Rubin CS. Personal Communication. Bronx, NY. 12/8/1984.
5. Kahn CR. The molecular mechanisms of insulin action. Annu. Rev. Med. 1985;36:429-451.
6. Goldfine ID. The insulin receptor: Molecular biology and transmembrane signalling. Endocr. Rev. 1987;8:235-255.
7. White MF, Kahn CR. The insulin signaling system. J. Biol. Chem. 1994;269:1.
8. Brunetti A, Maddux BA, Wong KY, Goldfine ID. Muscle cell differentiation is associated with increased insulin receptor biosynthesis and messenger RNA levels. J. Clin. Invest. 1989;86:192-198.
9. Brunetti A, Goldfine ID. Differential effects of fibroblast growth factor on insulin receptor and muscle specific protein gene expression in BC3H-1 myocytes. Mol. Endocrinol. 1990;4:880-885.
10. Mamula PW, McDonald AR, Brunetti A, Okabayashi Y, Wong KY, Maddux BA, Logsdon C, Goldfine ID. Regulating insulin-receptor-gene expression by differentiation and hormones. Diabetes Care. 1990;13:288-301.
11. McDonald AR, Goldfine ID. Glucocorticoid regulation of insulin receptor gene

transcription in IM-9 cultured lymphocytes. J. Clin. Invest. 1988;81:499-504.
12. Rouiller DG, McElduff A, Hedo JA, Gorden P. Induction of the insulin proreceptor by hydrocortisone in cultured lymphocytes (IM-9 line). J. Clin. Invest. 1985;76:645-649.
13. Wentworth BM, Schaefer IM, Villa-Komaroff L, Chirgwin JM. Characterization of the two nonallelic genes encoding mouse preproinsulin. Journal of Molecular Evolution. 1986;23:305–312.
14. Ullrich A, Bell JR, Chen EY, Herrera R, Petruzzelli LM, Dull TJ, Gray A, Coussens L, Liao Y-C, Tsubokawa M, Mason A, Seeburg PH, Grunfeld C, Rosen OM, Ramachandran J. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature. 1985;313: 756-761.
15. Caro JF, Raju SM, Sinha MK, Goldfine ID, Dohm GL. Heterogeneity of human liver, muscle, and adipose tissue insulin receptor. Biochem. Biophys. Res. Commun. 1988;151:123-129.
16. Baker EN, Blundell TL, Cut field JF, Cut field SM, Dodson EJ, Dodson GG, Hodgkin DM, Hubbard RE, Isaacs NW, Reynolds CD, et al. The structure of 2Zn pig insulin crystals at 1.5 A resolution. Philos Trans R Soc Lond B Biol Sci. 1988;319:369–456.
17. De Meyts P. The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its elevance to mitogenic versus metabolic signalling. Diabetologia. 1994;37:S135–S148.
18. De Meyts P, Whittaker J. Structural biology of insulin and IGF1 receptors: Implications for drug design. Nat Rev Drug Discov. 2002;1:769–783.
19. Derewenda U, Derewenda Z, Dodson EJ, Dodson GG, Bing X, Markussen J. X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. J Mol Biol. 1991;220:425–433.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/15599