Comparative analysis of transcription factors of insulin signaling
(Insulin receptor substrate family)

Sampoornam Balakrishnan, Satyavani Kumpatla and Vijay Viswanathan*
M V Hospital for Diabetes and Diabetes Research Centre, Chennai, 600 013 India

Received 11 November 2008; revised 23 March 2009; accepted 8 June 2009

The aim of this study was to identify the role of insulin receptor substrates (IRS) in insulin signaling by doing sequence based comparative analysis. The sequences of IRS family (1, 2, 4, 5 & 6) were submitted to predict protein motif scan and clustal W for motif, domain identification and multiple sequence alignment. All the 5 IRS had some common motifs like amidation, Asn glycosylation, myristoylation, cAMP, casein kinase II, protein kinase C phosphorylation sites and functional domains like PTB and PH. Some functional motifs like tyrosine phosphorylation, Rho kinase α binding, OGFR, PSGP were identified in IRS-1, RGD motifs in IRS-1 and 4, XYPPX motif and ECM protein domain in IRS-2, NLS motif and extension like domain in IRS-2 and 4, ascorbate cytosolic peroxidase, class II peroxidase superfamily domain, NHL and 2 octapeptide repeats and ATP_GTP_ attachment sites in IRS-4. These domains and motifs may have inducing and inhibitory effects on insulin signaling.

Keywords: signaling pathway, insulin signaling, IRS, Rho kinase α, upregulation, downstream signaling

Introduction
In type 2 diabetes, the body becomes resistant to the effects of insulin presumably because of defects in the insulin signaling pathway. The insulin receptor is composed of two extracellular α subunits and two trans membrane β subunits linked together by disulphide bonds. Binding of insulin to the α subunit induces α residues present in the β subunit. These residues are recognized by phosphotyrosine-binding (PTB) domains of adaptor proteins such as members of the insulin receptor substrate (IRS) family. Much of the specificity of signal transduction depends on the recruitment of several signaling components such as protein kinases and G-protein GTPases into short lived active complexes in response to an activating signal such as a insulin binding to its receptor.

Receptor activation leads to phosphorylation of key tyrosine residues on IRS proteins, some of which are recognized by other members of the signaling pathway in the downstream, finally resulting in the storage of glucose as glycogen.

The IRS family consists of 6 members viz., IRS-1, IRS-2, IRS-3, IRS-4, IRS-5 and IRS-6. Among the 6 IRS, knockdown of IRS-1 results in up-regulation (the process by which a cell increases the number of cellular components such as RNA or protein in response to an external variable) of gluconeogenic enzymes such as glucose 6 phosphatase and phosphoenol pyruvate carboxy kinase, as well as marked increase in hepatic nuclear factor-4α. Decreased IRS-1 was associated with a decrease in the gluco kinase expression and a trend towards increased blood glucose.

Knockdown of IRS-2 results in the up-regulation of lipogenic transcription factor, SREBP-1c (sterol regulatory element binding protein 1c), which plays a major role in the effect of insulin on the transcription of hepatic genes such as gluco kinase and fatty acid synthase. The IRS-5 and IRS-6 are referred to as Dok 4 and Dok 5. The aim of this study was to identify the role of IRS family in insulin signaling by doing sequence based comparative analysis.

Materials and Methods

Tools used
2. ClustalW.

Methods
IRS sequences were retrieved after cross-references with databases like NCBL, SWISS-PROT, TrEMBL and UNIPROT. Among the 6 IRS, IRS3 is only 50%
of the length of IRS-1, -2 and -4. These three IRS were more or less of equal length. The sequence length of IRS-1 was 1242, IRS-2 was 1338 and IRS-4 was 1257. The length of IRS-5 and IRS-6 was, however, 326 and 306 amino acids sequences long, respectively.

Functional motifs of all IRS were identified by submitting the sequence query of all the 5 IRS to Motif scan and Predict protein tools. The sequences were aligned with Clustal W.

Results

The motifs identified were: amidation site, Asn-glycosylation site, cAMP phosphorylation site, casein kinase II phosphorylation site, myristoylation site and protein kinase c phosphorylation site. These motifs have been commonly found in all the 5 IRS. Common motifs like ASN glycosylation were observed in all the members of IRS except IRS-6. Some features of the sequences are Ala rich, Arg, Gly, Asn, Pro, Ser rich regions, which are commonly found in IRS-1, -2 and -4. The phospho tyrosine domain, and pleckstrin homology regions were observed in all the 5 IRS. Apart from these motifs some other specific motifs were also found in IRS family with respect to their molecular role in insulin signaling. Multiple sequence alignment with Clustal W, showed that only 15% homology existed between the sequences but yet they belonged to a single IRS family. The details of motif information are given in Table 1. The role of IRS family members in insulin signaling mechanism is shown in Fig. 1.

INS activates INSR, in turn inhibited by IDE. IRS-1 gets activated by INSR in turn activates PI3K, GluT4, G6P and PEPCK, inhibited by ROK α- whereas IRS-4 inhibits PI3K, GluT4 and INSR, in turn inhibited by TNF α via PP4. IRS-2 activated by INSR activates SREBP1C 4, GK and FAS and gets activated by IL4 and JAK.

IRS-1

IRS-1 is the major substrate of the insulin receptor. It is thought to interact with the receptor itself, a process which is completely dependent upon receptor tyrosine kinase activity. When phosphorylated by the insulin receptor, IRS-1 binds specifically to cellular proteins that contain SH2 domains such as phosphatidylinositol 3-kinase p85 subunit6. IRS-1 is phosphorylated on multiple tyrosine residues by ligand-activated insulin receptors. These tyrosine phosphorylation sites serve to dock several

![Fig. 1—A schematic diagram indicating the role of IRS and other members in insulin signaling in Type 2 diabetes (INS-insulin, INSR- insulin receptor, IRS1,2,4-insulin receptor substrates 1,2,4,IL4- interleukin 4, JAK- jak kinase, SREBP1C- transcription factor sterol regulatory element binding protein 1c, GK- Glucokinase, FAS- fatty acid synthase, TNF α - Tumour necrosis factor α, PP4-protein phosphatase 4, PI3K- Inositol 3 phosphate kinase, GluT4- Glucose transporter 4, ROK α- Rho associated kinase α, G6P- Glucose-6-phosphatase, PEPCK-phosphoenol pyruvate carboxy kinase).](image-url)
<table>
<thead>
<tr>
<th>Motif id</th>
<th>Description</th>
<th>IRS-1</th>
<th>IRS-2</th>
<th>IRS-4</th>
<th>IRS-5</th>
<th>IRS-6</th>
<th>Investigator</th>
<th>Putative role</th>
</tr>
</thead>
<tbody>
<tr>
<td>prf:IRS_PTBD</td>
<td>Phosphotyrosine domain</td>
<td>160</td>
<td>264</td>
<td>194</td>
<td>298</td>
<td>231</td>
<td>335</td>
<td>Kavanaugh et al\textsuperscript{a}</td>
</tr>
<tr>
<td>prf:PH_DOMAIN</td>
<td>Pleckstrin homology</td>
<td>12</td>
<td>115</td>
<td>30</td>
<td>144</td>
<td>78</td>
<td>199</td>
<td>Baltimore et al\textsuperscript{a}</td>
</tr>
<tr>
<td>pfam:fs:Phage_fiber</td>
<td>ROK-alpha binding site</td>
<td>350</td>
<td>363</td>
<td>30e</td>
<td>335</td>
<td>231</td>
<td>335</td>
<td>Sahar Farah et al\textsuperscript{a}</td>
</tr>
<tr>
<td>pfam:ls:OGFr_III</td>
<td>Opioid growth factor receptor</td>
<td>495</td>
<td>513</td>
<td>30e</td>
<td>144</td>
<td>78</td>
<td>199</td>
<td>Zagon et al\textsuperscript{a}</td>
</tr>
<tr>
<td>freq_pat:RGD</td>
<td>Cell attachment sequence</td>
<td>986</td>
<td>986</td>
<td>986</td>
<td>1198</td>
<td>1200</td>
<td>Gougos A, Letarte M\textsuperscript{a}</td>
<td>Key recognition structure in cellular adhesion</td>
</tr>
<tr>
<td>freq_pat:TYR_PHOSPHO_SITE</td>
<td>Tyr kinase phosphorylation site</td>
<td>457</td>
<td>465</td>
<td>1221</td>
<td>1229</td>
<td>410</td>
<td>424</td>
<td>Patschinsky et al\textsuperscript{a}</td>
</tr>
<tr>
<td>prf:NLS_BP</td>
<td>Bipartite nuclear localization signal</td>
<td>-</td>
<td>557</td>
<td>573</td>
<td>40</td>
<td>424</td>
<td>Kalderon et al\textsuperscript{a}</td>
<td>Targets the (pro)ein members of the signaling cascade to the nucleus through the cytosolic nuclear transport receptors.</td>
</tr>
<tr>
<td>pfam:fs:Amelogenin</td>
<td>extracellular matrix (ECM) protein</td>
<td>-</td>
<td>908</td>
<td>921</td>
<td>-</td>
<td>-</td>
<td>Nakahori et al\textsuperscript{a}</td>
<td>Involved in cell adhesion.</td>
</tr>
<tr>
<td>pfam:ls:Extensin_2</td>
<td>Extensin-like region</td>
<td>-</td>
<td>707</td>
<td>1130</td>
<td>425</td>
<td>907</td>
<td>Lamport\textsuperscript{14}</td>
<td>Responsible for cell wall self-assembly. (may be it play a role in assembly of IRS2 on the membrane.</td>
</tr>
<tr>
<td>pfam:ls:PSGP</td>
<td>Polysialoglycoprotein</td>
<td>1037</td>
<td>1049</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Schlossman\textsuperscript{11}</td>
<td>Binds to the poly sialo glyc an chain.</td>
</tr>
<tr>
<td>freq_pat:ATP_GTP_A</td>
<td>ATP-GTP attachment site</td>
<td>-</td>
<td>-</td>
<td>602</td>
<td>609</td>
<td>642</td>
<td>649</td>
<td>New motif Viswanathan et al, current study.</td>
</tr>
<tr>
<td>pfam:ls:XYPPX</td>
<td>XYPPX repeat</td>
<td>-</td>
<td>1263</td>
<td>1267</td>
<td>-</td>
<td>-</td>
<td>Viswanathan et al, current study.</td>
<td>New motif Viswanathan et al, current study.</td>
</tr>
</tbody>
</table>
that leads to inhibition of both insulin-induced IRS-1 tyrosine phosphorylation and phosphatidylinositol 3-kinase (PI3-kinase) activation.

Opioid growth factor receptor, (OGFR), is also known as [Met(5)]-enkephalin. It was found between 495 and 513. The encoded unbound receptor for OGF has been localized to the outer nuclear envelope, where it binds OGF and is translocated into the nucleus.

Cell attachment sequence which was 3 residues long and found between 986 and 988, contained an RGD sequence and is a likely candidate for supporting integrin-mediated cell attachment.

Tyr kinase phosphorylation site lies in motifs between 457-465 (GEEELSNY) and 1221-1229 (RSSEDPGL). Substrates of tyrosine protein kinases are generally characterised by a lysine or an arginine seven residues to the N-terminal side of the phosphorylated tyrosine. An acidic residue (Asp or Glu) is often found at either three or four residues to the N-terminal side of the tyrosine.

Polysialoglycoprotein motif (PSGP) lies between 1037 and 1049. This family represents a series of 13 residue repeats found in the apo poly sialo glycoprotein, consists of tandem repeats of a glycol-tridecapeptide, Asp-Asp-Ala-Thr*-Ser*-Glu-Ala-Ala-Thr*-Gly-Pro-Ser-Gly (*denotes the attachment site of a polysialoglycan chain).

RS-2
Apart from PTB and PH domains like IRS-1, IRS-2 has the following specific motifs: Bipartite nuclear localization signal, Extra cellular matrix (ECM) protein Extensin-like region XYPPX repeat.

Bipartite nuclear localization signal lies in the region 557 and 573. (RRVSGDAAQDLDDLDRGL). A nuclear localizing sequence (NLS) is an amino acid sequence, which acts like a ‘tag’ on the exposed surface of a protein. This sequence is used to target the protein to the cell nucleus through the nuclear pore complex and to direct a newly synthesized protein into the nucleus via its recognition by cytosolic nuclear transport receptors. Typically, this signal consists of one or more short sequences of positively charged lysines or arginines.

Amelogenin, the ECM protein motif lies between 908 and 921. ECM, LPPEPKSPGEYN, is a hydrophobic, proline rich and cell adhesion protein.

There are four classes of cell adhesion proteins, namely, SAMs, CAMs, cadherins and selectins. The SAM (surface adherent material) consists of

<table>
<thead>
<tr>
<th>Table</th>
<th>Specific motifs of IRS family — Cond</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRS-2</td>
<td>759 766</td>
</tr>
<tr>
<td>IRS-5</td>
<td>1226 1333</td>
</tr>
<tr>
<td>IRS-6</td>
<td>499 828</td>
</tr>
<tr>
<td>IRS-7</td>
<td>1053 1257</td>
</tr>
<tr>
<td>IRS-3</td>
<td>1039 1051</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>Motif</th>
<th>Description</th>
<th>Motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oecapptide</td>
<td>plm_k4Oecapptide</td>
<td>Ascorbate cysteine peroxidase</td>
<td>peroxo-35_APCYCT</td>
</tr>
<tr>
<td></td>
<td>PKAPDNK</td>
<td>Class II peroxidase superfamily</td>
<td>peroxo-40_CLASSUF</td>
</tr>
<tr>
<td></td>
<td>PEREDSN</td>
<td>NII repeat</td>
<td>prENII</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Putative role</th>
<th>Possible role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigator</td>
<td>Function is unknown.</td>
</tr>
<tr>
<td>New motif</td>
<td>New motif</td>
</tr>
<tr>
<td>Vasiathanan et al., current study</td>
<td>Vasiathanan et al., current study</td>
</tr>
<tr>
<td>Vasiathanan et al., current study</td>
<td>Vasiathanan et al., current study</td>
</tr>
<tr>
<td>Vasiathanan et al., current study</td>
<td>New motif</td>
</tr>
<tr>
<td>Vasiathanan et al., current study</td>
<td>New motif</td>
</tr>
</tbody>
</table>
extracellular matrix proteins that bind to membrane intercalated receptors (integrins) and require divalent cations for activity. Amelogenin has several characteristics of the SAM in that amelogenin is an extracellular matrix molecule, is not membrane-intercalated, and requires divalent cations for activity.

Since amelogenin does not contain an RGD sequence or several other sequences known to bind to integrins, the cell-surface receptor for amelogenin may not be an integrin. While the SAMs often have a binding site for collagen, amelogenin apparently does not contain a binding site for collagen. In contrast, amelogenin does bind to hydroxyapatite. Amelogenin not only supports cell adhesion but also promotes cell spreading. It is important to note that the RGD containing peptide of fibronectin can support cell attachment but requires a heparin-binding domain of fibronectin to mediate cell-spread. Based on lack of binding to a heparin-Sepharose affinity column amelogenin apparently does not interact with heparin under physiological conditions. Thus, amelogenin has both similarities to and marked differences from members of the SAM class of cell adhesion proteins.

Extensin-like region was found between 707 and 1030 (423 residues). Extensins are homologous hydroxyproline-rich glycoproteins (HRGPs) found in the extracellular matrix. The key to the role of HRGPs in cell wall self-assembly and cell extension lies in their chemistry, which is dependent on extensive post-translational modifications (PTMs) such as hydroxylation, glycosylation, and cross-linking. Repetitive peptide motifs characterise HRGPs.

XYPPX repeat was in region 1263 and 1267. This repeat is found in a wide variety of proteins and generally consists of the motif XYPPX where X can be any amino acid. The family includes annexin VII, the carboxy tail of certain rhodopsins This family also includes plaque matrix proteins, however, this motif was embedded in a ten residue repeat. The molecular function of this repeat is unknown.

**IRS-4**

IRS-4, like other IRS, contains one PTB and PH domain and cell attachment sequence, nuclear localization signal and extensin like region. Cell attachment sequence was found in the region 1198 and 1200; Bipartite nuclear localization signal is in between 410 and 424; Extensin-like region was observed in between 425 and 907; Octapeptide sequence in between 759-766 and 1226-1233; Ascorbate cytosolic peroxidase in region 499-828; Class II peroxidase superfamily region in 1053-1257; NHL repeat was found between 1039 and 1051; and ATP_GTP attachment site in between 602-609 and 642-649.

The NHL repeat, named after NCL-1, HT2A and Lin-41, was found largely in a large number of eukaryotic and prokaryotic proteins. For example, the repeat was found in a variety of enzymes of the copper type II, ascorbate-dependent mono oxygenase family, which catalyses the C-terminus α-amidation of biological peptides.

Class II peroxidase superfamily region is in 1053-1257 region; in plants this enzyme activity depolymerizes cellulose and lignin. *Caenorhabditis elegans* gene ncl-1 encodes a zinc finger protein and may be a repressor of ribosome synthesis, RNA polymerase I and III transcription, and an inhibitor of cell growth. Loss of functional mutations in *ncl-1*, shown to result in enlarged nucleoli, increased rates of rRNA and 5S RNA transcription and enlarged cells.

The mammalian 5-HT2A receptor is a subtype of the 5-HT2 receptor which belongs to the serotonin receptor family and is a G protein coupled receptor (GPCR). This is the main excitatory receptor subtype among the GPCRs for serotonin (5-HT), although 5-HT2A may also have an inhibitory effect on certain areas such as the visual cortex and the orbitofrontal cortex. Lin-41 – abnormal cell lineage family member 41.

PKAPDTNK, PEREDSDN Octapeptide motif was located in the region between 759 and 766, and also located in 1226 and 1233. The function of these two motifs are still unclear.

The ATP_GTP attachment site (P-loop) contains nucleotide triphosphate hydrolase activity. Conserved sequences for this motif are GDGERGKS and GTGGKDGKS. The mammalian 5-HT2A receptor which belongs to the serotonin receptor family and is a G protein coupled receptor (GPCR). This is the main excitatory receptor subtype among the GPCRs for serotonin (5-HT), although 5-HT2A may also have an inhibitory effect on certain areas such as the visual cortex and the orbitofrontal cortex. Lin-41 – abnormal cell lineage family member 41.

Multiple sequence alignment with ClustalW showed that only 15% homology exists between sequences (202 residues). The cladogram obtained showed that IRS-1 was originated first, and then IRS-2 and IRS-4 were evolved from IRS-1.

**IRS-5**

The sequence length of this protein is 326, it has some common motifs like PTB, PH, ASN glycosylation, cAMP phosphorylation, CK2 phosphorylation, PKC phosphorylation, myristoylation, and tyrosine phosphorylation sites like other IRS.
IRS-6

The sequence length of this protein is 306, it has some common motifs like PTB, PH, cAMP phosphorylation, CK2phosphorylation, PKC phosphorylation, myristoylation, and tyrosine phosphorylation sites like other IRS.

Discussion

IRS-1

The PTB and PH domains of IRS-1 bind to INSR and get phosphorylated by tyrosine kinase activity of INSR on tyrosine phosphorylation sites, in turn activates inositol 3 phosphate kinase, which in turn translocates GluT4 to plasma membrane, essential for glucose uptake by cells, called + regulation. In contrast to this activity, the IRS-1 has Rho kinase α (ROK) binding site, which undergoes phosphorylation and dephosphorylation by insulin, negatively regulates IRS-1, inhibits glucose uptake by cells causing increase in blood glucose probably by insulin resistance.

IRS-1 lacks NLS sequence for nuclear membrane attachment instead it has OGFR motif which localizes IRS-1 to nuclear membrane. The presence of RGD and PSGP motifs indicates that IRS-1 is initially localized on cell surface but after activation by INSR moves to nuclear region.

IRS-2

The PTB and PH domains of IRS-2 like IRS-1 bind to INSR and gets activated, but IRS-2 does not contain motif for tyrosine phosphorylation, instead has XYPPX repeat, may be it gets phosphorylated (This point needs further study, because the action of this motif in insulin signaling is still unknown) by INSR and participates in lipid mobilization into the cells.

It has Amelogenin, an extracellular matrix (ECM) protein motif, and extensin like motif indicates involvement in cell-cell and cell-matrix adhesion needed for participation in signaling cascades. The presence of NLS motif clearly indicates that after activation at ECM, it gets translocated to nucleus for inducing transcription of proteins or enzymes involved in lipid metabolism. IRS-2 is up-regulated by IL4 through its tyrosine phosphorylated α receptor. IL 4 signaling in turn is activated by JAK kinase.

IRS-4

Like IRS-1 and IRS-2, IRS-4 also has PTB and PH domains, which bind to INSR but lacks tyrosine phosphorylation and XYPPX motifs like other IRS. The outcome of INSR binding needs to be studied further. It may be involved in internalization of INSR. But somehow it decreases the concentrations of IRS-1 and IRS-2, inhibits the action of inositol 3 phosphate kinase and translocation of GluT4 to the plasma membrane and increases plasma glucose concentration.

IRS-4 has one NLS and extensin motif indicating the participation in signaling cascade and it gets activated at the ECM and translocated to the nucleus and may inhibit the transcription of RNA polymerase I and III and inhibit ribosome biosynthesis. Due to inhibition of transcription of RNA polymerases and ribosomes it may interfere in insulin signaling.

The ATP_GTP attachment site has nucleotide triphosphate hydrolase activity and it may be responsible for inhibition of RNA pol I and III transcription.

The ATP_GTP attachment site has nucleotide triphosphate hydrolase activity and it may be responsible for inhibition of RNA pol I and III transcription.

PKAPDTNK, PEREDSDN Octapeptide motif was located in the region between 759 and 766, and also located in the region of 1226-1233. The function of these two motifs is still unclear, and needs to be further investigated.

IRS-4 is subjected to regulation by TNF-α and that protein phosphatase 4 mediates TNF-α-induced degradation of IRS-4. This area of cascading event needs further attention.

When IRS-4 was bound to INSR, through its class II peroxidase superfamily domain, it may be involved in depolymerisation of glycan part of insulin receptor and ATP_GTP attachment site and NHL domain may be involved in inhibition of transcription further required for downstream insulin signaling thereby increasing the plasma glucose concentration. Further, insulin gets degraded in 71 min after release into the circulation and this adds to the pathophysiology of type 2 diabetes.

Conclusion

This study indicated that the class II peroxidase motif and ATP_GTP attachment site and NHL motif of IRS-4 and ROK binding motif of IRS-1 may have inhibitory role in insulin signaling and they may be responsible for insulin resistance. Further attention is needed for the identification of role of XYPPX motif of IRS-2 and detailed mechanism of action of ascorbate cytosolic peroxidase and one class II peroxidase superfamily domain and NHL motif and octapeptide motifs of IRS-4. The molecular mechanisms of
in ositol 3 phosphate kinase and IDE needs to be studied further and may direct the research community involved in identification of new drug targets for type 2 diabetes and its treatment into a new direction.

References

15. Interpro database: InterPro is an integrated documentation resource for protein families, domains and sites.
17. Prosite database is an integrated documentation resource for protein families, domains and motifs.
20. 5-HT2A receptor (Wikipedia, the free encyclopedia): http://www.wikipedia.com