DATABASES

RILM: A Web-Based Resource to Aid Comparative and Functional Analysis of the Insulin and IGF-1 Receptor Family

Acely Garza-Garcia,[1] Dhaval S. Patel,[2] David Gems,[2] and Paul C. Driscoll[1,3]*

[1]Division of Molecular Structure, National Institute for Medical Research, MRC, London, United Kingdom; [2]Centre for Research on Aging, Department of Biology, University College London, London, United Kingdom; [3]Department of Biochemistry and Molecular Biology, University College London, London, United Kingdom

Communicated by Bruce Gottlieb

The metazoan receptors for insulin (INSR), insulin-like growth factor 1 (IGF1R), and other insulin-like molecules are transmembrane tyrosine kinases involved in the regulation of cell size, cell proliferation, development, signaling of nutritional and environmental conditions, and aging. Historically, mutations in the human insulin receptor have been studied because such changes often lead to severe insulin resistance. More recently, amino acid sequence alterations in the insulin receptor-like receptors of *Drosophila melanogaster* and *Caenorhabditis elegans*, as well as in the mouse insulin receptor have been the focus of attention. These modifications can have profound effects on growth, body size, metabolism, and aging. To integrate the many findings on insulin/IGF1 receptor structure and function across species we have created "Receptors for Insulin and Insulin-like Molecules" (RILM), a curated computer-based resource that displays residue-by-residue information on sequence homology, three-dimensional structure, structure/function annotation, and documented mutations. The resource includes data obtained from sequence and structure analysis tools, primary database resources, and published reports. The information is integrated via a structure-based multiple sequence alignment of diverse members of the family. RILM was designed to provide easy access to multiple data types that could prove useful in the analysis of the effect of mutations on protein structure and ligand binding within this receptor family. RILM is available at www.biochem.ucl.ac.uk/RILM. Hum Mutat 0, 1–9, 2007. **© 2007 Wiley-Liss, Inc.**

KEY WORDS: insulin receptor; INSR; IGF1R; INSRR; DAF2; dInR; structure; sequence alignment

INTRODUCTION

The vertebrate receptors for insulin (Ins) and insulin-like growth factor 1 (IGF1) have been extensively studied due to their critical roles in energy metabolism and cellular development. Impairment of human Ins receptor (INSR; MIM# 147670) function causes extreme insulin-resistance, which is a syndrome characterized by dysmorphic abnormalities, growth deficiency, and hyperinsulinemia, a condition that causes major damage to the skin and ovaries. The severity of the condition is determined by the degree of functional impairment of the receptor [Longo et al., 2002], which in turn depends on the nature of the mutation and whether one or both alleles are affected. Total loss of function of INSR is lethal shortly after birth [Hone et al., 1995; Joshi et al., 1996; Accili et al., 1996], while patients that suffer from a moderate loss of function often reach adulthood. In general, homozygotes and compound heterozygotes for detrimental mutations have more severe phenotypes. It is estimated that 0.1% of the human population is a heterozygous carrier of a deleterious mutation in INSR [Taylor et al., 1992].

The human receptor for IGF1 (IGF1R; MIM# 147370) shares 56% identity with INSR. Naturally-occurring viable mutations in the IGF1R appear to be less common than for INSR. In the few documented cases, IGF1R function loss was reported to cause intrauterine and postnatal growth retardation, and dysmorphic features [Tamura et al., 1993; Kawashima et al., 2005]. In animal models, IGF1R-null mutants have severe intrauterine growth deficiency and die soon after birth [Liu et al., 1993]. In addition to its fundamental role in somatic cell development, IGF1R signaling is required for the establishment and survival of cancerous tumors. IGF1R is necessary for cell transformation by various oncogenes [Sell et al., 1994; Valentinis et al., 1997], can rescue cells from apoptosis [Werner and Le Roith, 1997], and aids in cell mobility [Stracke et al., 1989]. Moreover, downregulation of IGF1R function in cancerous cells can promote apoptosis and inhibit transformation [Bahr and Groner, 2004], making IGF1R a

Received 4 August 2006; accepted revised manuscript 9 January 2007.

*Correspondence to: Paul C. Driscoll, Professor, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.
E-mail: pdrisco@nimr.mrc.ac.uk

Grant sponsor: Mexican National Council for Science and Technology (CONACyT); Grant sponsor: Biotechnology and Biological Sciences Research Council; Grant sponsor: Wellcome Trust; Grant sponsor: Medical Research Council.
DOI 10.1002/humu.20491
Published online in Wiley InterScience (www.interscience.wiley.com).

promising target for the treatment of cancer [Baserga, 2005]. Mammals possess a third receptor of the Ins receptor family, the insulin receptor-related receptor, the function of which remains to be discovered [Shier and Watt, 1989; Hanze et al., 1999]. The human insulin receptor-related receptor (INSRR; MIM# 147671) shares 51% identity with IGF1R and 49% identity with INSR. In contrast to the three paralogous proteins found in vertebrates, invertebrates appear to possess only one receptor of the Ins receptor family (marine sponges have multiple receptor tyrosine kinases with homology of the cytoplasmic domains of Ins receptor family, but with completely different extracellular domains [Skorokhod et al., 1999]). This suggests that the different classes of vertebrate receptors are the result of the duplications of a gene in the common ancestor of vertebrate species. Consistent with this hypothesis, although INSR and IGF1R have distinct physiological roles, they display a moderate degree of functional redundancy [Nakae et al., 2001].

In the last decade, interest in the Ins receptor family has been further fuelled by the discovery that they have an important role in the determination of lifespan in a range of metazoan organisms [Kenyon, 2005]. As in vertebrates, in the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, mutations that lead to complete loss of function of the insulin/IGF1-like receptor (dInR and DAF-2, respectively) are lethal [Fernandez et al., 1995; Chen et al., 1996; Gems et al., 1998]. In *D. melanogaster*, heterozygote mutations that reduce but do not abolish dInR signaling result in developmental delay, diminished cell size and number, small body size, and female infertility [Chen et al., 1996; Brogiolo et al., 2001]. Surprisingly, mild reduction of function of Ins/IGF1-like receptors can also extend adult lifespan. This was first discovered for the *C. elegans* DAF-2 receptor [Kenyon et al., 1993; Kimura et al., 1997] and subsequently for *D. melanogaster* dInR [Tatar et al., 2001] and the mouse [Bluher et al., 2003]. Mice with either reduced levels of IGF1 receptor expression or lacking the fat-specific insulin receptor, live longer than wild-type animals [Bluher et al., 2003; Holzenberger et al., 2003].

These and other observations have inspired research into the effects of mutations in the Ins/IGF1-like signaling pathway in model organisms. For example, over 30 *C. elegans daf-2* alleles that increase lifespan have been documented [Gems et al., 1998; Muñoz and Riddle, 2003]. These long-lived mutants display various combinations of phenotypic traits such as resistance to thermal stress, reduced fertility and movement, and/or constitutive entry into the developmentally arrested dauer larva stage [Gems et al., 1998]. The molecular basis of phenotypic differences in *daf-2* mutants remains poorly understood, as do the effects of mutation of insulin/IGF1 receptors generally. However, given the level of amino acid sequence identity among the Ins receptor homologs throughout the Metazoa (e.g., INSR and DAF-2 share 29% sequence identity and 43% similarity), cross-phylum integration of structural and functional information should help in the understanding of the molecular effects of mutations in Ins receptor family members.

Analysis of evolutionary conservation of amino acid sequences has been useful in predicting which of a set of mutations might have a negative impact on protein function, and subsequently lead to disease [Ng and Henikoff, 2001; Ferrer-Costa et al., 2005; Mooney, 2005]. Residues that are conserved among a family of proteins are typically important for function, either because they are needed to preserve the three dimensional (3D) structure or because they have a functional role, such as being a part of the ligand-binding or enzymatic active sites, target sites for post-translational modification, or involved in the quaternary interactions with other polypeptides or between distinct structural modules of the intact protein. Additional information for the prediction or rationalization of the effect of residue substitutions can be obtained from the analysis of protein 3D structure. Amino acid substitutions that lead to a drastic change of physicochemical properties, such as size, charge, hydrophobicity, and secondary structure propensity, are usually damaging. The identification of mutations that could lead to loss of protein function or stability using 3D structure and sequence conservation information has been a field of intense research in recent years [Mooney, 2005]. There are now several excellent examples in the literature describing methods to identify such mutations in a systematic and quantitative manner [Chasman and Adams, 2001; Sunyaev et al., 2001; Ferrer-Costa et al., 2002; Saunders and Baker, 2002; Yue et al., 2006].

To facilitate the analysis of mutations in the Ins receptor family using all available structure and sequence data we have created the "Receptors for Insulin and Insulin-like Molecules" (RILM) database resource. The resource provides online utilities that: 1) make available to the research community a curated structure-based multiple sequence alignment of Ins receptor family members from a variety of metazoan phyla; 2) provide residue-specific information on sequence conservation and 3D structural properties that could help to rationalize the consequence of a substitution at a particular position; and 3) compile the information available in the literature that describes the phenotypic effects of mutations in the INSR, IGF1R, dInR, DAF-2 proteins, and relatives in other organisms; and 4) take account of recent advances in the experimental 3D structure determination of various component parts of the receptors. We demonstrate the use of the resource as a valuable tool in the interpretation of phenotypes associated with a range of *daf-2* alleles and hope other researchers, more widely involved in the investigation of this family of receptors, will benefit from its capabilities.

RILM WEB RESOURCE

The RILM resource is a client/server application accessible via a standard web browser at www.biochem.ucl.ac.uk/RILM. The resource comprises a collection of HTML files and purpose-built CGI scripts written in Perl that integrate data on sequence conservation, protein structure, and mutations within the Ins receptor family. The resource, the general structure of which is schematically represented in Figure 1, comprises the following core elements:

Per-Residue Structural Context Information

The mammalian members of the Ins receptor family are tetrameric transmembrane kinases comprised of a dimer of αβ heterodimers, where the α and β subunits are derived by posttranslational proteolysis from a single precursor polypeptide chain [Hedo et al., 1983; Olson et al., 1988; Knutson, 1991]. Each β subunit is covalently bound to an α subunit through a single disulphide bond, and α subunits are bound to each other by at least two, but probably four, disulphide bonds [Schaffer and Ljungqvist, 1992; Sparrow et al., 1997]. Each of the precursor chains comprises seven structural domains. Following the signal sequence at the N-terminus, there are two ~150-residue L domains, L1 and L2, separated by a ca. 150 residue cysteine-rich region (CRR). L2 is linked to three consecutive fibronectin type III (FnIII) domains, FnIII 1, FnIII 2 and FnIII 3. After the FnIII domains come the transmembrane and cytoplasmic juxta-

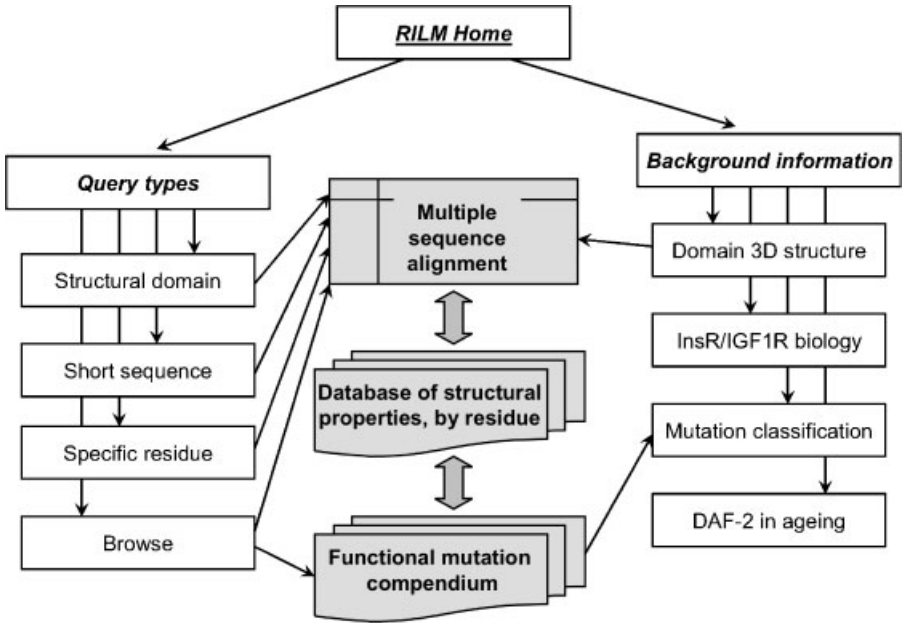

FIGURE 1. **Schematic representation of the structure of the RILM web resource, indicating the seamlessly connected (thick double-headed arrows) core data elements in the center (gray), modes of entry to the multiple sequence alignment and derived structural and functional data at left, and background textual information at right.**

membrane regions, and then a tyrosine kinase domain. At the extreme C-terminus of the chain there is a region with no apparent regular 3D structure that is implicated in the recruitment of phosphotyrosine binding proteins. There is a ~130-residue-long insert of unknown 3D structure within the FnIII 2 domain that, in vertebrates and some invertebrate receptors, contains a conserved stretch of basic residues that is known to be the proteolytic cleavage site which divides the α- and β-chains. Thus in the mature receptor, the α-subunit is composed of L1, CRR, L2, FnIII 1, and the first 3 β-strands of the FnIII 2 domain. The β-subunit is formed by the remaining 4 β-strands of the FnIII 2 domain, FnIII 3, the transmembrane and juxtamembrane regions, the tyrosine kinase domain, and the C-terminal region.

No quaternary structure of any transmembrane *holo*-Ins/IGF1 receptor is known at atomic resolution. However, high quality (resolution better than 2.7 Å) 3D crystal structures of a variety of constructs encompassing most of the INSR and IGF1R domains have been solved: namely, the tyrosine kinase domains of INSR [Hubbard et al., 1994; Hubbard, 1997] and IGF1R [Pautsch et al., 2001; Munshi et al., 2003] in both their *apo*- and activated states; and a large fragment of the α subunit of both IGF1R [Garrett et al., 1998] and INSR [Lou et al., 2006] comprising the tandem L1-CRR-L2 domains. Recently the intermediate resolution (3.8 Å) 3D structure of the complete dimeric ectodomain of INSR in complex with monoclonal antibody F_{ab} fragments was reported [McKern et al., 2006]. This last structure provides the clearest picture of the likely physiological arrangement of the component domains in the extracellular portion of a member of this family of receptors, at least in the *apo*-form. The crystal structure appears to supersede the predicted quaternary structure of the whole tetramer that was modeled based on low resolution electron microscopy data [Luo et al., 1999; Ottensmeyer et al., 2000] and whose validity had already been questioned [De Meyts and Whittaker, 2002]. In the new structure (Protein Databank Code 2DTG), the domains are arranged linearly, bending over between the L2 and FnIII 1 domains, hence assuming an overall V-shape [McKern et al., 2006]. This configuration is supported by other electron microscopy data, in which projections agreeing with crystallographic envelope were found [Tulloch et al., 1999]. In spite of substantial research in this area, further structural investigations will be required to fully elucidate the mechanism of ligand-induced tyrosine kinase activation and effector protein recruitment.

In RILM we have compiled a set of key descriptors of the structural environment (secondary structure, side chain accessibility, backbone and side chain contacts, and hydrogen bonding) of each of the residues in the experimentally determined 3D structures of INSR and IGF1R components and each of these sets of descriptors is associated with the corresponding column of the multiple sequence alignment.

Multiple Sequence Alignment

The core component of the RILM resource is a multiple sequence alignment of 45 members of the Ins receptor family. Sequences were selected from model organisms of five metazoan phyla including three previously unavailable nematode DAF-2 sequences from *C. briggsae*, *C. remanei*, and the long-lived parasitic nematode *Brugia malayi* (D.S. Patel, N. Nadkarni, and D. Gems, unpublished results). Sequences were obtained from UniProt (www.ebi.uniprot.org) by text search and from GenBank (www.ncbi.nlm.nih.gov/Genbank) and Ensembl (www.ensembl.org) using BLAST [Altschul et al., 1990]. We computed two alternative alignments for each domain: a structure-based multiple sequence alignment using the program Expresso [Armougom et al., 2006] followed by manual refinement, and an alignment based on the corresponding profile hidden-Markov models (HMMs) of Pfam (www.sanger.ac.uk/Software/Pfam) using HMMER (http://hmmer.janelia.org). We found that both alignments are very similar for the L1, CRR, L2, and the kinase domain. However, the structure-based alignments of the three FnIII domains have a better average quality score (computed with

Clustal X [Thompson et al., 1997]). Thus, the curated structure-based alignment appears to be of better quality than those generated automatically from the profile HMMs of Pfam. The alignment HTML files as they appear in RILM were created with Jalview [Clamp et al., 2004]. A Javascript routine was written for the user interface to the sequence alignment that displays the sequence name and residue number corresponding to the current cursor position (illustrated in Fig. 2).

It has been our experience that the existence of different residue numbering schemes for this family of receptors complicates the cross-referencing of the information in the literature and the database deposits. Traditionally, for INSR and IGF1R, residues have been numbered according to the mature receptor primary sequence, while in the electronic databases the numbering for sequence of the pre-proreceptor is usually used. Another inconsistency arises from the fact that INSR exists in two isoforms

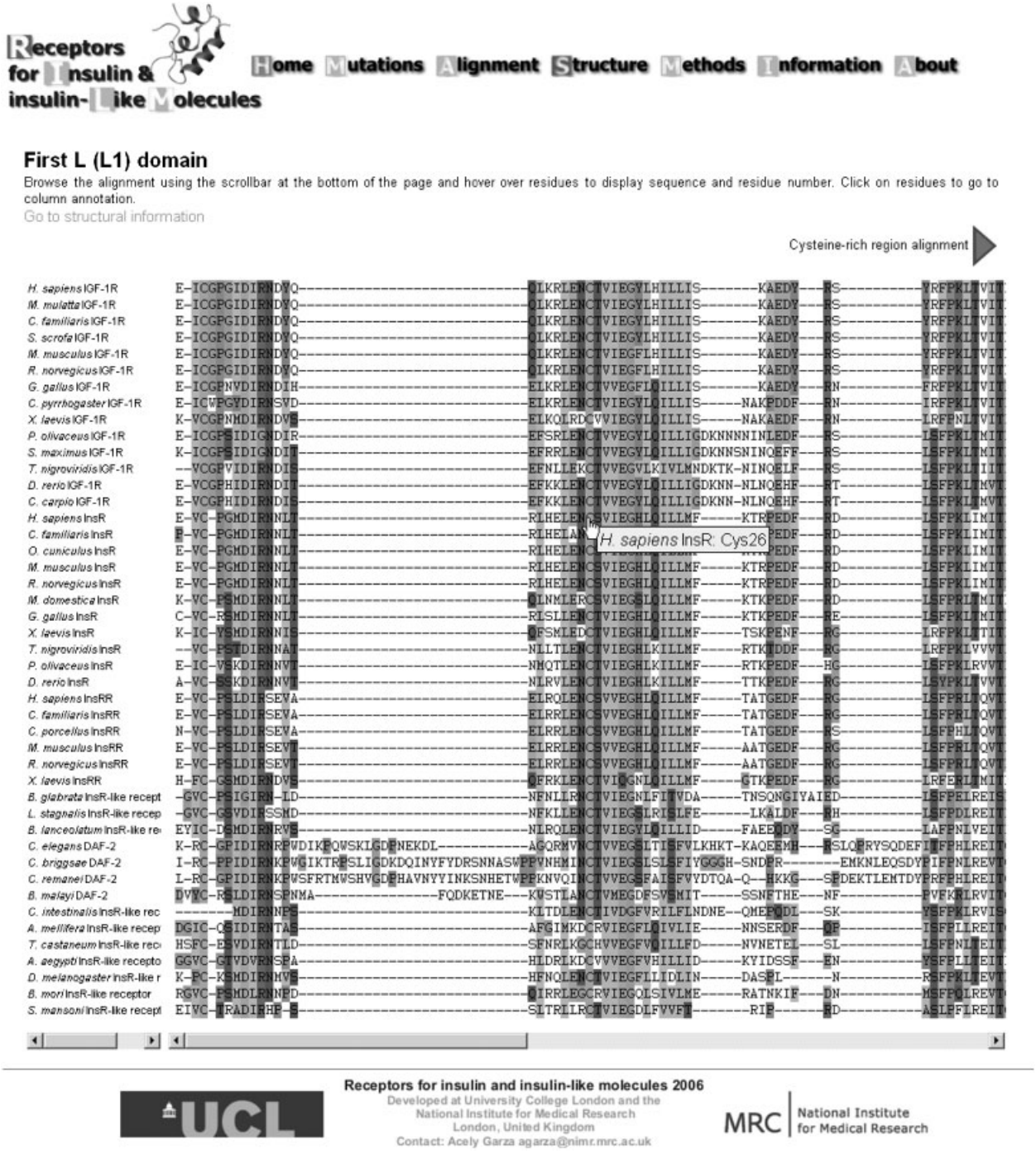

FIGURE 2. **Screen snapshot of the cross-phyla multiple sequence alignment at the core of RILM. The example shown is for the L1 domain. In the online web resource each residue is given a background color according to conventional categorization of the physicochemical properties of the side chain. The cursor bar at bottom allows one to easily move along the sequence within a given domain. Jumping between domains is enabled by navigational tools at the top (here the right arrow provides a link to the CRR domain that follows L1). The snapshot illustrates the amino acid residue identity at the current cursor position (here Cys26 of INSR).**

as a result of alternative splicing of exon 11 [Seino and Bell, 1989]. The long isoform (usually denoted B), is 12 residues longer than the short isoform (denoted A), which lacks residues 718 to 729. To be consistent with the literature, the entries for RILM of INSR and IGF1R have been numbered according to the mature receptor sequence, and in the case of INSR to the long isoform. For all other sequences, residues are numbered according to the precursor form prior to posttranslational modification. The sequences of the *D. melanogaster* and the various *Caenorhabditis* receptors possess long additional sequences of unknown function at the N-terminus that precedes the signal peptide. SignalP 3.0 [Bendtsen et al., 2004] predicts that the signal peptide cleavage site is between residues 140 and 141 for DAF-2 and between residues 285 and 286 for dInR.

Literature-Derived Biological/Functional Annotation

RILM collates published data, on a per-residue basis, from experimental studies of the Ins receptor family. For each residue, a manually annotated section of the resource contains details of its naturally occurring missense and nonsense mutations, mutagenesis studies, posttranslational modifications (phosphorylation, glycosylation, or disulfide bond formation) and any experimental information that illuminates its structural or functional role, e.g., in ligand binding. This literature-derived information is displayed for all equivalent residues in each column of the alignment to allow straightforward comparison of the family members. The full compendium of known mutations can also be accessed directly from the *Mutation* link in the header of each page of the resource (see below). RILM currently includes 241 entries for known mutations in INSR, 24 entries for IGF1R, 11 entries for DAF-2, and five entries for dInR.

The RILM resource also includes brief descriptive articles on general Ins receptor family topics, with background information, relevant literature references and links to related online resources. These pages have been built using wiki-based software [Michaud, 2006], and are therefore open to be remotely edited by users. The current article topics include the INSR/IGF1R signal transduction pathway, the various syndromes associated with mutations in INSR and IGF1R, and the role of DAF-2 in development and aging in nematodes.

The RILM User Interface

The available data for a particular position in the multiple sequence alignment can be accessed by browsing the multiple sequence alignment or by menu-based search. The user can move through a graphical representation of the multiple sequence alignment and click on a residue to display the position entry. Entries can also be retrieved by input of residue number and selection of the receptor name from a menu list, or by searching with a user-defined sequence of residues. The latter option is useful when there is uncertainty on the part of the user about the specific residue number, or when it is desired to see all matches of a particular amino acid in the receptor sequence.

Once the residue number and corresponding receptor have been requested, a script refers to the pregenerated multiple sequence alignment to identify equivalent residues in the other sequences; all of the available information for the equivalent residues at the selected position is then displayed. An example for the output of a residue-based query is shown in Figure 3. At the top of the output there is a sequence navigation section that can be used to move back and forth along the sequence of the selected receptor. The output displays information in three categories: alignment, structure, and annotation.

The Alignment section reports the residues found at that position in other species, the structural domain in which the selected position is located, and three scores: Residue conservation, computed with Scorecons [Valdar, 2002]; ClustalX quality score [Thompson et al., 1997], and probability of a mutation being tolerated at that position, calculated with SIFT [Ng and Henikoff, 2001].

The Structure section provides key descriptors of the local environment for each residue. The reported calculations are: local secondary structure context assessed with the Kabsch and Sander [1983] algorithm ; solvent accessibility computed with Naccess [Hubbard, 1996]; a listing of neighboring side-chain atoms computed with Molmol [Koradi et al., 1996]; residue contacts calculated using the contact utility of the MMTSB Tool Set [Feig et al., 2001]; and optimal hydrogen bonds predicted with What If [Hooft et al., 1996]. The selected residue can be visualized on the corresponding structure using a Jmol (www.jmol.org) applet.

The Annotation section comprises reports of published naturally occurring, induced, and site-directed nonsense and missense mutations, as well as posttranslational modifications and other remarks. The entries for human receptor mutations, where applicable and if known, include: the location and nature of the nucleotide and amino acid change; phenotype associated with the mutation; link to the corresponding entry for the disease in the OMIM knowledgebase; classification of the mutation according to the scheme devised by Taylor et al. [1992]; a textual description of the effects of the mutation; the relevant literature reference hyperlinked to an appropriate online catalog (e.g., PubMed-Entrez); and, when available, hyperlinks to three general purpose external mutation databases: SNPs3D [Yue et al., 2006], MutDB [Dantzer et al., 2005], and Polyphen [Ramensky et al., 2002]. The entries for DAF-2 mutations include: the name of the allele; the classification of the allele according to its phenotype [Gems et al., 1998]; and a textual description of the phenotype and the relevant reference. The Annotation section also displays remarks of any known structural and functional information regarding any of the equivalent residues in members of the receptor superfamily.

CASE STUDY

A driving force for the creation of the RILM resource was to attempt to decipher the molecular effects associated with the lifespan-increasing mutations in the *C. elegans daf-2* gene, using the available information for INSR and IGF1R. Quite unexpectedly, we found that 3 out of the 11 DAF-2 receptor missense mutations are in the same position as mutations in INSR associated with severe insulin resistance. Three of the mutations are substitutions to the same amino acid type in both receptors. Moreover, of the 11 *daf-2* missense mutations, nine are in highly conserved positions in the multiple sequence alignment, suggesting that they are important for receptor stability or function. A further two mutations are in positions that are not conserved across phyla, but their effect can be rationalized using the available structural information.

An example of our findings is the *daf-2(m596)* allele which corresponds to the DAF-2 mutation G547S [Scott et al., 2002]. The RILM entry for DAF-2 G547 shows that this position is fully conserved among vertebrate organisms, but not in invertebrates. In the structure of IGF1R, the equivalent residue (G356) is buried and at the site of a sharp change in the direction of the polypeptide backbone within the L2 domain. A mutation in the equivalent

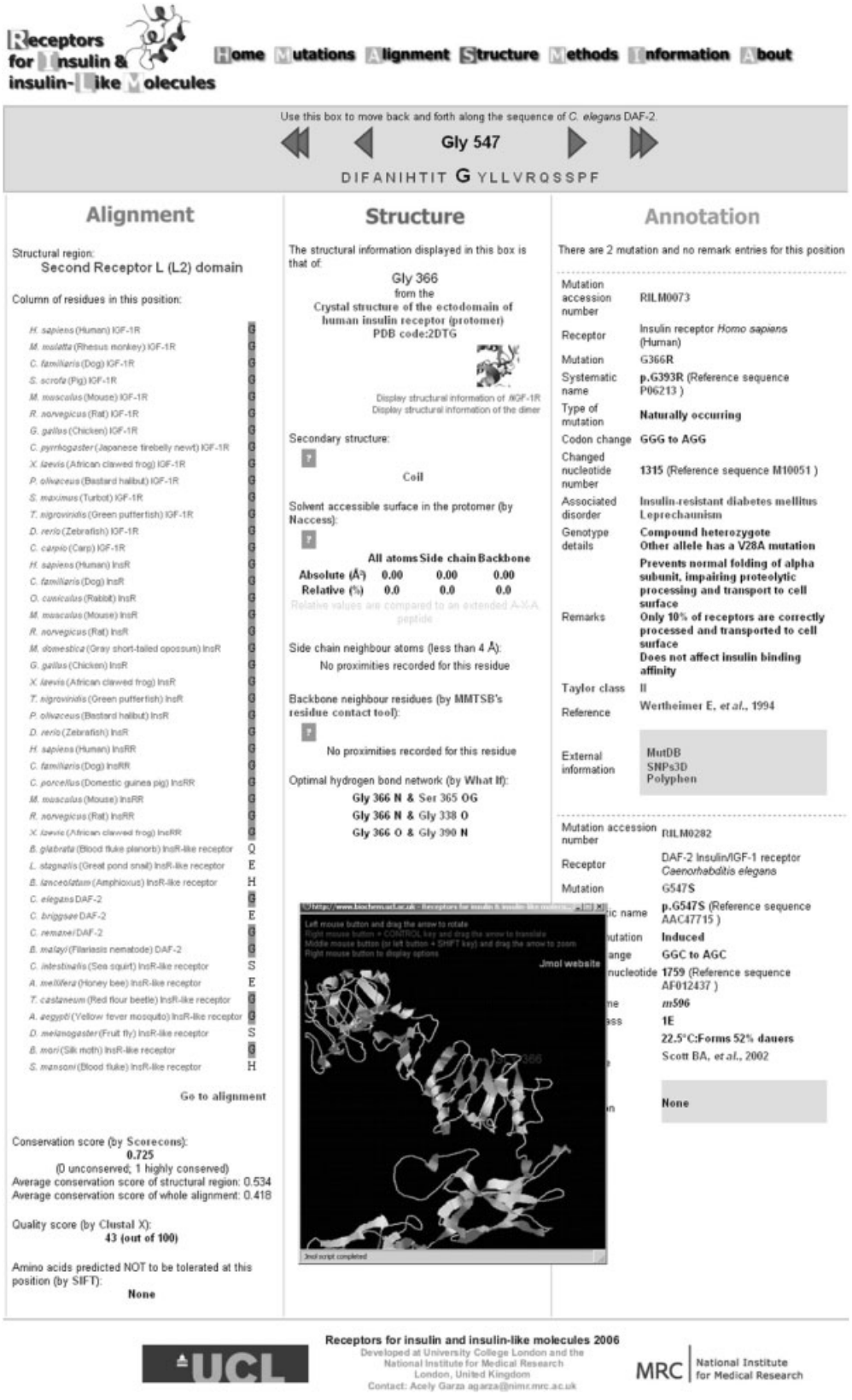

FIGURE 3. **Screen snapshot of the derived structural and functional data for a specific residue from the RILM web resource. The example shown is for Gly547 of *C. elegans daf-2*, which is described in more detail in the main text. The snapshot illustrates the navigational features of RILM at top. Browsing of the sequence alignment is facilitated by clicking on the "forward," "rewind," "fast-forward," and "fast-rewind" buttons, or by jumping between component domains. The Jmol applet at the bottom illustrates how the user is able to call up a view of the position of the selected residue in the most appropriate experimental 3D structure.**

residue of INSR (G366R) has been found in a patient who was a compound heterozygote and died shortly after birth [Barbetti et al., 1992]. Expression of the G366R mutant receptor in vitro has shown that it prevents normal processing of the pre-proreceptor and that only ~10% of mature receptor reaches the plasma membrane, but this mutation does not appear to affect ligand binding or autophosphorylation of the mature receptor [Wertheimer et al., 1994]. The fact that the Gly-to-Arg substitution is deleterious is readily explained, as Arg, being relatively voluminous and carrying a formal charge, would not fit in the protein packing and the fold would be severely disrupted. However, a Gly-to-Ser mutation in proteins is often benign because both amino acids have small side chain volumes. In this particular case, it is possible that the mutation affects the structure because Ser, having a side chain, is not suited to adopt the conformation needed to accommodate the sharp turn in the polypeptide fold. This idea is supported by the fact the Daf-c phenotype of *daf-2(m596)* is a relatively weak allele [Gems et al., 1998], which implies that the Gly-to-Ser mutation may have a milder effect on processing of the receptor than a Gly-to-Arg mutation. However, the fact that this position in the multiple sequence alignment is not conserved in invertebrates suggests that the local structure could also be different in this subset of the Ins receptor family.

As illustrated by this single example out of many documented in RILM, the compilation of cross-phylum structural and functional information brought together in the resource provides ready access to reasonable hypotheses to rationalize the phenotypes of missense mutant alleles of the Ins receptor family members and places them it in a cross-species context.

FUTURE DEVELOPMENTS

In the future we aim to incorporate into the mutation entries of RILM the curators' interpretation of the effect of documented INSR, IGF1R, DAF-2, and dInR mutations on protein structure/ligand-binding. We shall shortly add details of another 11 newly sequenced *daf-2* alleles (D.S. Patel, A. Garza-Garcia, M. Nanji, J.J. McElwee, P.C. Driscoll, and D. Gems, unpublished results). We are also planning to include the predictions of the effect of mutations on protein structure obtained from additional structure and sequence-based algorithms. The database will be updated to include any new refined 3D models of the intact INSR or IGF1R as they are placed in the public domain. Moreover we are aware of the community effort to unify the contents of human locus-specific databases [Claustres et al., 2002] and are currently working to make our data compatible with central databases and to follow the proposed guidelines.

CONCLUSIONS

In general, the analysis of protein structure and sequence conservation is a readily available source of valuable data for the understanding of the effect of mutations on protein function. However, it is not only necessary to make the data available, but to deliver it in a format such that researchers from a variety of scientific disciplines can easily benefit from it. RILM was born from a collaborative effort between geneticists and structural biologists in an attempt to achieve this goal. Ultimately, its use is to aid in better understanding insulin and IGF1 signaling, and the pathologies that they affect, including diabetes and aging. We are primarily concerned with making this resource useful to the research community and with keeping it up to date and therefore invite researchers that share our interest in this family of receptors to contact with comments or updates. To this end we have added a password-protected Wiki interface to the RILM resource, which can be accessed under the Structure and Information headings.

ACKNOWLEDGMENTS

A.G.G. was funded by the Mexican National Council for Science and Technology (CONACyT) during part of this project. D.S.P. and D.G. were funded by the Biotechnology and Biological Sciences Research Council, and the Wellcome Trust. A.G.G. and P.C.D. are supported by the Medical Research Council.

REFERENCES

Accili D, Drago J, Lee EJ, Johnson MD, Cool MH, Salvatore P, Asico LD, Jose PA, Taylor SI, Westphal H. 1996. Early neonatal death in mice homozygous for a null allele of the insulin receptor gene. Nat Genet 12: 106–109.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C. 2006. Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res 34:W604–W608.

Bahr C, Groner B. 2004. The insulin like growth factor-1 receptor (IGF-1R) as a drug target: novel approaches to cancer therapy. Growth Horm IGF Res 14:287–295.

Barbetti F, Gejman PV, Taylor SI, Raben N, Cama A, Bonora E, Pizzo P, Moghetti P, Muggeo M, Roth J. 1992. Detection of mutations in insulin receptor gene by denaturing gradient gel electrophoresis. Diabetes 41: 408–415.

Baserga R. 2005. The insulin-like growth factor-1 receptor as a target for cancer therapy. Expert Opin Ther Targets 9:753–768.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795.

Bluher M, Kahn BB, Kahn CR. 2003. Extended longevity in mice lacking the insulin receptor in adipose tissue. Science 299:572–574.

Brogiolo W, Stocker H, Ikeya T, Rintelen F, Fernandez R, Hafen E. 2001. An evolutionarily conserved function of the Drosophila insulin receptor and insulin-like peptides in growth control. Curr Biol 11:213–221.

Chasman D, Adams RM. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol 307:683–706.

Chen C, Jack J, Garofalo RS. 1996. The Drosophila insulin receptor is required for normal growth. Endocrinology 137:846–856.

Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. Bioinformatics 20:426–427.

Claustres M, Horaitis O, Vanevski M, Cotton RG. 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. Genome Res 12:680–688.

Dantzer J, Moad C, Heiland R, Mooney S. 2005. MutDB services: interactive structural analysis of mutation data. Nucleic Acids Res 33: W311–W314.

De Meyts P, Whittaker J. 2002. Structural biology of insulin and IGF1 receptors: implications for drug design. Nat Rev Drug Discov 1:769–783.

Feig M, Karanicolas J, Brookes III CL. 2001. MMTSB Tool Set (2001). http://mmtsb.scripps.edu

Fernandez R, Tabarini D, Azpiazu N, Frasch M, Schlessinger J. 1995. The Drosophila insulin receptor homolog: a gene essential for embryonic development encodes two receptor isoforms with different signaling potential. EMBO J 14:3373–3384.

Ferrer-Costa C, Orozco M, de la Cruz X. 2002. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J Mol Biol 315:771–786.

Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. 2005. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinformatics 21:3176–3178.

Garrett TP, McKern NM, Lou M, Frenkel MJ, Bentley JD, Lovrecz GO, Elleman TC, Cosgrove LJ, Ward CW. 1998. Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor. Nature 394:395–399.

Gems D, Sutton AJ, Sundermeyer ML, Albert PS, King KV, Edgley ML, Larsen PL, Riddle DL. 1998. Two pleiotropic classes of daf-2 mutation affect larval arrest, adult behavior, reproduction and longevity in *Caenorhabditis elegans*. Genetics 150:129–155.

Hanze J, Berthold A, Klammt J, Gallaher B, Siebler T, Kratzsch J, Elmlinger M, Kiess W. 1999. Cloning and sequencing of the complete cDNA encoding the human insulin receptor related receptor. Horm Metab Res 31: 77–79.

Hedo JA, Kahn CR, Hayashi M, Yamada KM, Kasuga M. 1983. Biosynthesis and glycosylation of the insulin receptor. Evidence for a single polypeptide precursor of the two major subunits. J Biol Chem 258: 10020–10026.

Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, Even PC, Cervera P, Le Bouc Y. 2003. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature 421:182–187.

Hone J, Accili D, Psiachou H, Alghband-Zadeh J, Mitton S, Wertheimer E, Sinclair L, Taylor SI. 1995. Homozygosity for a null allele of the insulin receptor gene in a patient with leprechaunism. Hum Mutat 6:17–22.

Hooft RW, Sander C, Vriend G. 1996. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 26: 363–376.

Hubbard SR, Wei L, Ellis L, Hendrickson WA. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 372: 746–754.

Hubbard SJ. 1996. Naccess: atomic solvent accessible area calculations. Ver. 2.1.1 Computer Program, Department of Biomolecular Sciences, UMIST, Manchester, UK. http://wolf.bms.umist.ac.uk/naccess/

Hubbard SR. 1997. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. EMBO J 16:5572–5581.

Joshi RL, Lamothe B, Cordonnier N, Mesbah K, Monthioux E, Jami J, Bucchini D. 1996. Targeted disruption of the insulin receptor gene in the mouse results in neonatal lethality. EMBO J 15:1542–1547.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637.

Kawashima Y, Kanzaki S, Yang F, Kinoshita T, Hanaki K, Nagaishi J, Ohtsuka Y, Hisatome I, Ninomoya H, Nanba E, Fukushima T, Takahashi S. 2005. Mutation at cleavage site of insulin-like growth factor receptor in a short-stature child born with intrauterine growth retardation. J Clin Endocrinol Metab 90:4679–4687.

Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. 1993. A *C. elegans* mutant that lives twice as long as wild type. Nature 366:461–464.

Kenyon C. 2005. The plasticity of aging: insights from long-lived mutants. Cell 120:449–460.

Kimura KD, Tissenbaum HA, Liu YX, Ruvkun G. 1997. daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science 277:942–946.

Knutson VP. 1991. Cellular trafficking and processing of the insulin receptor. FASEB J 5:2130–2138.

Koradi R, Billeter M, Wuthrich K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14: 51–32.

Liu JP, Baker J, Perkins AS, Robertson EJ, Efstratiadis A. 1993. Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type 1 IGF receptor (Igf1r). Cell 75:59–72.

Longo N, Wang Y, Smith SA, Langley SD, DiMeglio LA, Giannella-Neto D. 2002. Genotype-phenotype correlation in inherited severe insulin resistance. Hum Mol Genet 11:1465–1475.

Lou M, Garrett TP, McKern NM, Hoyne PA, Epa VC, Bentley JD, Lovrecz GO, Cosgrove LJ, Frenkel MJ, Ward CW. 2006. The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity. Proc Natl Acad Sci USA 103:12429–12434.

Luo RZ, Beniac DR, Fernandes A, Yip CC, Ottensmeyer FP. 1999. Quaternary structure of the insulin-insulin receptor complex. Science 285:1077–1080.

McKern NM, Lawrence MC, Streltsov VA, Lou MZ, Adams TE, Lovrecz GO, Elleman TC, Richards KM, Bentley JD, Pilling PA, Hoyne PA, Cartledge KA, Pham TM, Lewis JL, Sankovich SE, Stoichevska V, Da Silva E, Robinson CP, Frenkel MJ, Sparrow LG, Fernley RT, Epa VC, Ward CW. 2006. Structure of the insulin receptor ectodomain reveals a folded-over conformation. Nature 443:218–221.

Michaud P. 2006. Pmwiki website. Information and downloads for the PmWiki software. Available at: www.pmwiki.org. Last accessed: 26 January 2007.

Mooney S. 2005. Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Brief Bioinform 6:44–56.

Munshi S, Hall DL, Kornienko M, Darke PL, Kuo LC. 2003. Structure of apo, unactivated insulin-like growth factor-1 receptor kinase at 1.5 A resolution. Acta Crystallogr D Biol Crystallogr 59:1725–1730.

Muñoz MJ, Riddle DL. 2003. Positive selection of *Caenorhabditis elegans* mutants with increased stress resistance and longevity. Genetics 163: 171–180.

Nakae J, Kido Y, Accili D. 2001. Distinct and overlapping functions of insulin and IGF-I receptors. Endocr Rev 22:818–835.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Res 11:863–874.

Olson TS, Bamberger MJ, Lane MD. 1988. Post-translational changes in tertiary and quaternary structure of the insulin proreceptor. Correlation with acquisition of function. J Biol Chem 263:7342–7351.

Ottensmeyer FP, Beniac DR, Luo RZ, Yip CC. 2000. Mechanism of transmembrane signaling: insulin binding and the insulin receptor. Biochemistry 39:12103–12112.

Pautsch A, Zoephel A, Ahorn H, Spevak W, Hauptmann R, Nar H. 2001. Crystal structure of bisphosphorylated IGF-1 receptor kinase: insight into domain movements upon kinase activation. Structure 9:955–965.

Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900.

Saunders CT, Baker D. 2002. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. J Mol Biol 322: 891–901.

Schaffer L, Ljungqvist L. 1992. Identification of a disulfide bridge connecting the alpha-subunits of the extracellular domain of the insulin receptor. Biochem Biophys Res Commun 189:650–653.

Scott BA, Avidan MS, Crowder CM. 2002. Regulation of hypoxic death in *C. elegans* by the insulin/IGF receptor homolog DAF-2. Science 296: 2388–2391.

Seino S, Bell GI. 1989. Alternative splicing of human insulin receptor messenger RNA. Biochem Biophys Res Commun 159:312–316.

Sell C, Dumenil G, Deveaud C, Miura M, Coppola D, DeAngelis T, Rubin R, Efstratiadis A, Baserga R. 1994. Effect of a null mutation of the insulin-like growth factor I receptor gene on growth and transformation of mouse embryo fibroblasts. Mol Cell Biol 14:3604–3612.

Shier P, Watt VM. 1989. Primary structure of a putative receptor for a ligand of the insulin family. J Biol Chem 264:14605–14608.

Skorokhod A, Gamulin V, Gundacker D, Kavsan V, Muller IM, Muller WE. 1999. Origin of insulin receptor-like tyrosine kinases in marine sponges. Biol Bull 197:198–206.

Sparrow LG, McKern NM, Gorman JJ, Strike PM, Robinson CP, Bentley JD, Ward CW. 1997. The disulfide bonds in the C-terminal domains of the human insulin receptor ectodomain. J Biol Chem 272:29460–29467.

Stracke ML, Engel JD, Wilson LW, Rechler MM, Liotta LA, Schiffmann E. 1989. The type I insulin-like growth factor receptor is a motility receptor in human melanoma cells. J Biol Chem 264:21544–21549.

Sunyaev S, Ramensky V, Koch I, Lathe W, III, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. Hum Mol Genet 10: 591–597.

Tamura T, Tohma T, Ohta T, Soejima H, Harada N, Abe K, Niikawa N. 1993. Ring chromosome 15 involving deletion of the insulin-like growth factor 1 receptor gene in a patient with features of Silver-Russell syndrome. Clin Dysmorphol 2:106–113.

Tatar M, Kopelman A, Epstein D, Tu MP, Yin CM, Garofalo RS. 2001. A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. Science 292:107–110.

Taylor SI, Cama A, Accili D, Barbetti F, Quon MJ, de la Luz SM, Suzuki Y, Koller E, Levy-Toledano R, Wertheimer E. 1992. Mutations in the insulin receptor gene. Endocr Rev 13:566–595.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882.

Tulloch PA, Lawrence LJ, McKern NM, Robinson CP, Bentley JD, Cosgrove L, Ivancic N, Lovrecz GO, Siddle K, Ward CW. 1999. Single-molecule imaging of human insulin receptor ectodomain and its Fab complexes. J Struct Biol 125:11–18.

Valdar WS. 2002. Scoring residue conservation. Proteins 48: 227–241.

Valentinis B, Morrione A, Taylor SJ, Baserga R. 1997. Insulin-like growth factor I receptor signaling in transformation by src oncogenes. Mol Cell Biol 17:3744–3754.

Werner H, Le Roith D. 1997. The insulin-like growth factor-I receptor signaling pathways are important for tumorigenesis and inhibition of apoptosis. Crit Rev Oncog 8:71–92.

Wertheimer E, Barbetti F, Muggeo M, Roth J, Taylor SI. 1994. Two mutations in a conserved structural motif in the insulin receptor inhibit normal folding and intracellular transport of the receptor. J Biol Chem 269:7587–7592.

Yue P, Melamud E, Moult J. 2006. SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics 7:166.