The EMBO Journal vol.5 no.10 pp.2503-2512, 1986

Insulin-like growth factor I receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity

Axel Ullrich, Alane Gray, Albert W.Tam, Teresa Yang-Feng[1], Makoto Tsubokawa, Carol Collins, William Henzel, Thomas Le Bon[2], Satish Kathuria[2], Ellson Chen, Steven Jacobs[3], Uta Francke[1], J.Ramachandran and Yoko Fujita-Yamaguchi[2]

Departments of Developmental Biology and Molecular Biology, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, [1]Department of Human Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, [2]Department of Molecular Genetics, Beckman Research Institute of the City of Hope, 1450 E.Duarte Road, Duarte, CA 91010, and [3]Department of Molecular Biology, The Wellcome Research Laboratories, Research Triangle Park, NC 27709, USA

Communicated by J.Schlessinger

To identify structural characteristics of the closely related cell surface receptors for insulin and IGF-I that define their distinct physiological roles, we determined the complete primary structure of the human IGF-I receptor from cloned cDNA. The deduced sequence predicts a 1367 amino acid receptor precursor, including a 30-residue signal peptide, which is removed during translocation of the nascent polypeptide chain. The 1337 residue, unmodified proreceptor polypeptide has a predicted M_r of 151 869, which compares with the 180 000 M_r IGF-I receptor precursor. In analogy with the 152 784 M_r insulin receptor precursor, cleavage of the Arg-Lys-Arg-Arg sequence at position 707 of the IGF-I receptor precursor will generate α (80 423 M_r) and β (70 866 M_r) subunits, which compare with ~135 000 M_r (α) and 90 000 M_r (β) fully glycosylated subunits.

Key words: IGF-I receptor/insulin receptor/tyrosine kinase/signal transduction

Introduction

Insulin-like growth factor I (IGF-I), a 70 amino acid polypeptide with extensive structural homology to insulin (49%) and IGF-II (61%) exerts its biological effect by binding to a specific receptor on the surface of target cells. The IGF-I receptor is similar to but distinct from the insulin receptor (Froesch *et al.*, 1985; Rechler and Nissley, 1985). Like the insulin receptor, the IGF-I receptor is a membrane glycoprotein of M_r 300 000-350 000, consisting of two α subunits (M_r ~135 000) and two β subunits (M_r ~90 000) that are connected by disulfide bonds to form the functional $\beta-\alpha-\alpha-\beta$ heterotetrameric receptor complex (Chernausek *et al.*, 1981; Bhaumick *et al.*, 1981; Massague and Czech, 1982; Kull *et al.*, 1983; Rechler and Nissley, 1985). In analogy with the insulin receptor (Ronnett *et al.*, 1984), IGF-I receptor α and β subunits are thought to be encoded within a single 180 000 mol. wt receptor precursor (Jacobs *et al.*, 1983a) that is glycosylated, dimerized, and proteolytically processed to yield the mature $\alpha_2-\beta_2$ form of the receptor. Upon binding to the extracellular domain, IGF-I stimulates an intracellular, tyrosine-specific protein kinase activity which leads to β subunit autophosphorylation (Jacobs *et al.*, 1983b; Rubin *et al.*, 1983; Zick *et al.*, 1984) and presumably phosphorylation of cytoplasmic components of an IGF-I-specific signal transfer cascade.

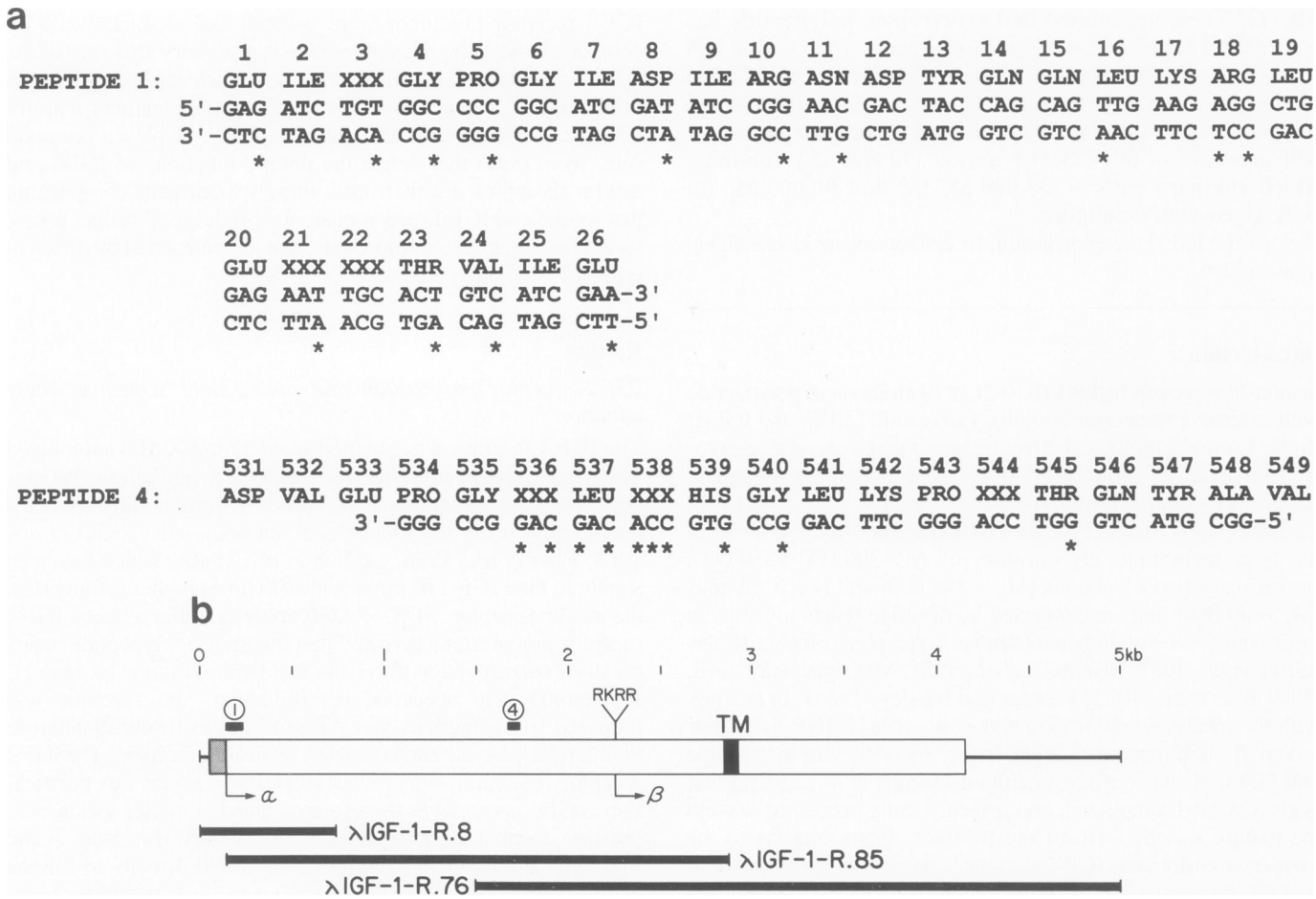
Despite functional and structural similarities, the receptors for IGF-I and insulin are thought to play different biological roles during mammalian development and mature life. Whereas insulin plays a key role in regulation of a variety of metabolic processes, the IGFs appear to be more potent in promoting growth. Multiple biological effects have been attributed to these hormones; however, identification of the specific biological function of the two hormones was complicated by the fact that insulin and IGFs can interact with each other's receptors in a concentration-dependent manner. Insulin binds with high affinity to the insulin receptor, low affinity to the IGF-I receptor, and very poorly to the IGF-II receptor. The converse is true for IGF-II, whereas IGF-I binds with the highest affinity to its own receptor, has slightly lower affinity for the IGF-II receptor, and still lower affinity for the insulin receptor. To enhance our understanding of the molecular mechanisms that evoke specific hormonal effects in cellular systems, and to attempt to identify structural features underlying these processes, cDNA clones encoding the human IGF-I receptor precursor were isolated and characterized. By comparison with the recently established primary structure of the closely related human insulin receptor (Ullrich *et al.*, 1985; Ebina *et al.*, 1985), we are now able to distinguish features that are common to both receptors, and thus likely to play a common role, from those that define the unique functions of IGF-I and insulin receptors. Furthermore, these experiments demonstrate that insulin and IGF-I receptors are the products of distinct genes, located on separate chromosomes, that are controlled by different types of regulatory signals.

Results

IGF-I receptor purification and amino acid sequence determination

The IGF-I receptor was purified from Triton X-100 solubilized human placental membranes by wheat germ agglutinin (WGA)-Sepharose chromatography, followed by immuno-affinity chromatography using the monoclonal antibody αIR3 (Kull *et al.*, 1983; Flier *et al.*, 1986; Le Bon *et al.*, 1986), which has been shown to bind IGF-I receptor with 100 times higher affinity than the insulin receptor. SDS-PAGE analysis of the reduced IGF-I receptor preparations revealed that the major polypeptide bands obtained correspond to the α and β subunits (Figure 1a, lane 1). For amino acid sequence determination, the receptor was prepared from fractions that did not bind to insulin-Sepharose to minimize possible contamination by insulin receptor. The IGF-I receptor α subunit, which was more stable under our purification conditions and was therefore obtained at higher yields, was purified from a polyacrylamide gel (Figure 1a, lane 2 and Materials and methods) and either subjected directly to Edman degradation sequence analysis or first to lysyl peptidase treatment and fractionation of the peptides by reversed-phase h.p.l.c.

Fig. 1. IGF-I receptor purification and protein sequence analysis. **(a)** Silver stained SDS−PAGE of the purified IGF-I receptor (2 μg, **lane 1**) and electrophoretically purified α subunit **(lane 2)**. The positions of the M_r standards (myosin, β-galactosidase, phosphorylase b, bovine serum albumin and ovalbumin) are indicated. **(b)** Amino-terminal amino acid sequence (peptide 1) and peptide sequences from purified IGF-I receptor α subunit. Parentheses indicate uncertainty in assignment; X's indicate unknown amino acids.



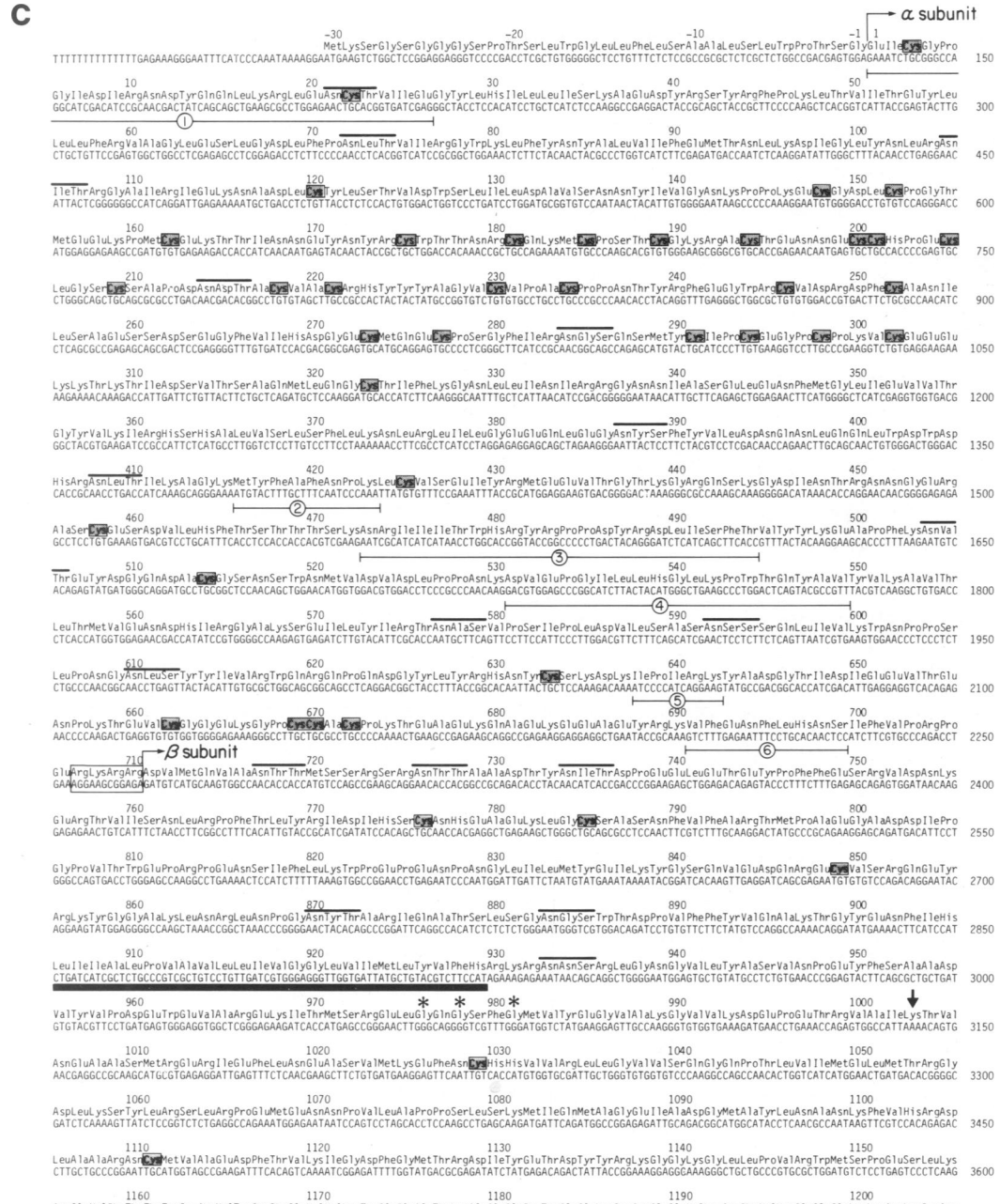


Fig. 2. IGF-I receptor cDNA clones, nucleotide sequence and predicted amino acid sequence. **(a)** Nucleotide sequence of synthetic oligonucleotide probes deduced from peptide amino acid sequences. Asterisks indicate differences between predicted and actual nucleotide sequence. **(b)** Schematic diagram of the cDNA structure of IGF-I receptor. Overlapping cDNA clones used in sequence determination are shown below. Translated sequences are boxed, with the signal sequence shaded. Also indicated are the putative precursor processing site (RKRR), the transmembrane domain (TM) and the positions of the synthetic oligonucleotide probes derived from peptides 1 and 4. **(c)** Nucleotide and predicted amino acid sequence. Amino acids of the proreceptor are numbered above starting at Glu 1, and are preceded by a 30-residue signal sequence; nucleotides are numbered to the right. Experimentally determined peptide sequences are underlined and numbered (Figure 1b); potential N-linked glycosylation sites are overlined; cysteine residues are shaded; and the transmembrane domain is heavily underlined. The potential ATP binding site is indicated by asterisks over Gly 976, 978 and 981, and by an arrow over Lys 1003. The putative precursor processing site is boxed.

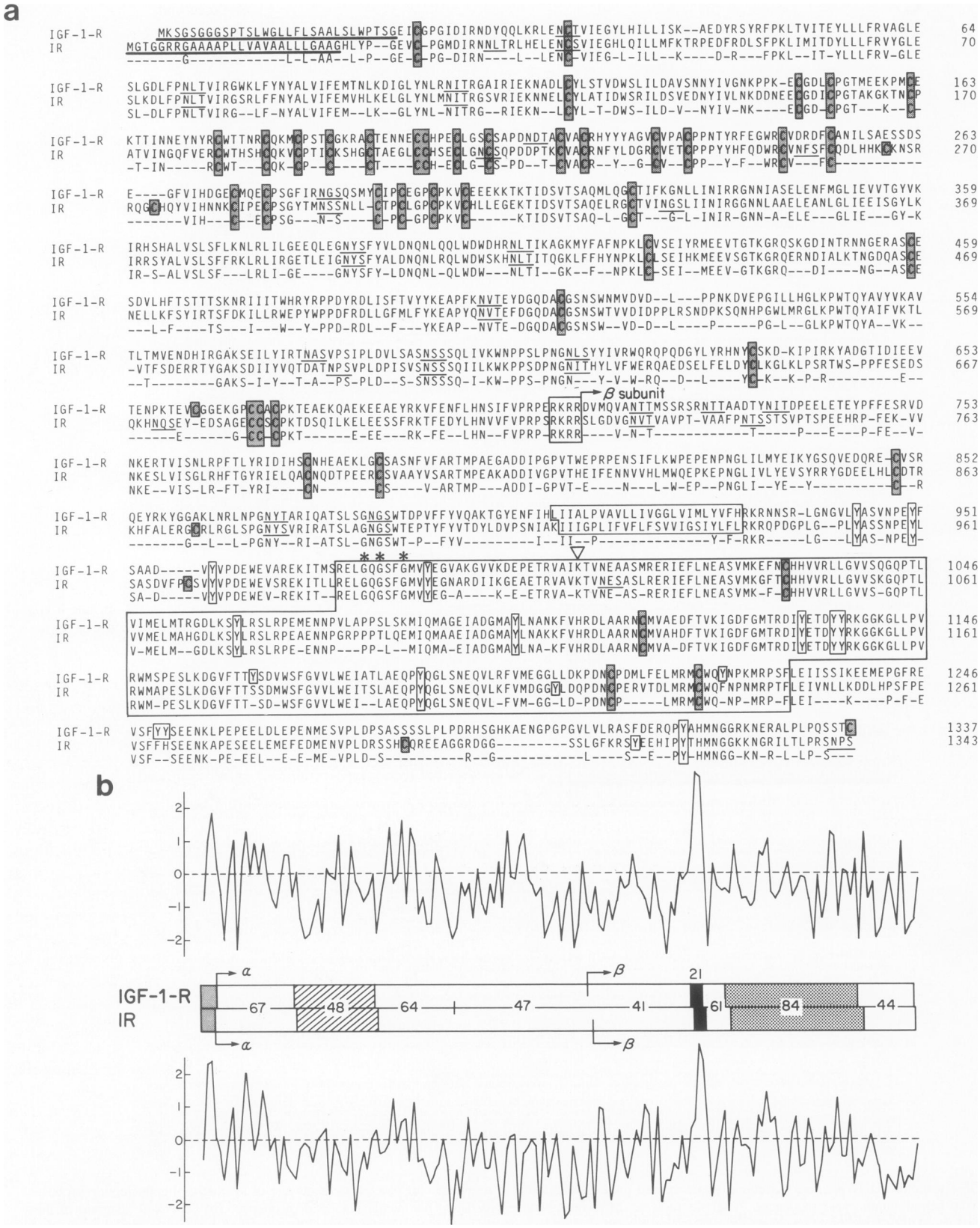


Fig. 3. Comparison of IGF-I and insulin receptors. **(a)** Comparison of primary amino acid sequences of IGF-I receptor (IGF-I-R) and insulin receptor (IR). Identical residues are indicated on the third line, gaps having been introduced to optimize alignment. The signal sequences are heavily underlined, cysteine residues shaded, tyrosines in the cytoplasmic domain are boxed, and potential N-glycosylation sites are lightly underlined. Asterisks and open triangles indicate residues involved in ATP binding. Boxed regions include, in order, the putative precursor processing site, the transmembrane domain and the tyrosine kinase domain. **(b)** Hydropathy analysis (Kyte and Doolittle, 1982) of IGF-I receptor and insulin receptor precursor sequences. Receptor domains are schematically represented and the percent homologous residues indicated. The signal sequence is shown by fine shading, cysteine-rich domain by cross-hatching, transmembrane domain by a black bar, and tyrosine kinase domain by coarse shading.

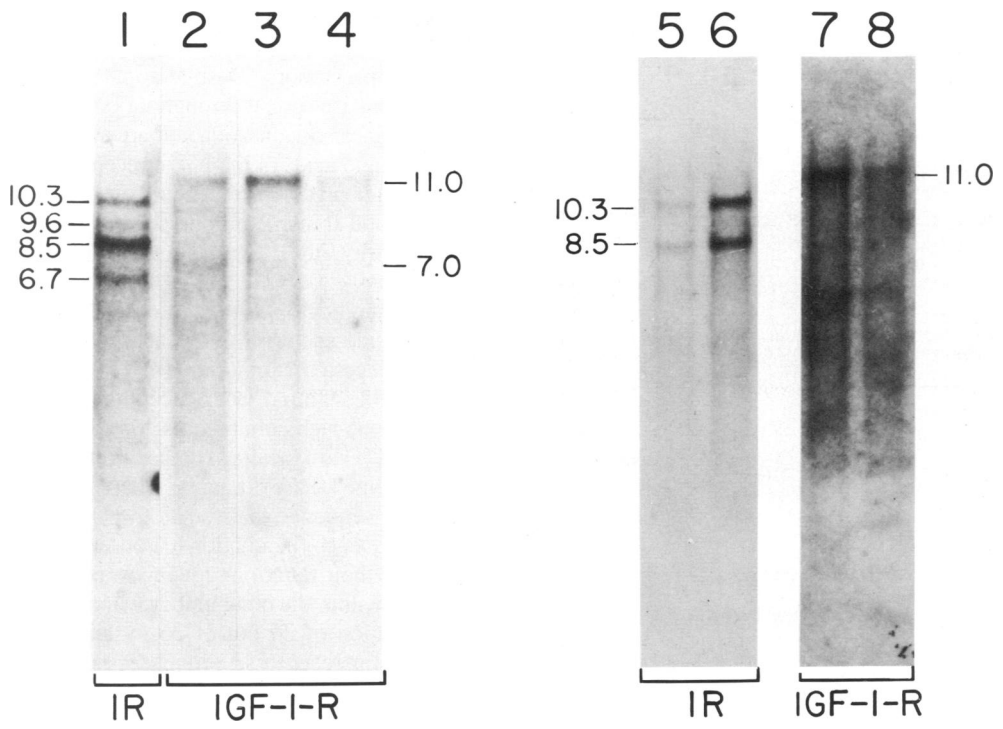


Fig. 4. Northern blot analysis of human, rat and mouse mRNAs. **Lanes 1** and **2:** human term placental mRNA. **Lane 3:** BRL-3A2. **Lane 4:** rat term placental mRNA. **Lanes 5** and **7:** mouse 3T3-LI fibroblasts. **Lanes 6** and **8:** 3T3-LI adipocytes. cDNA probes are indicated below lanes. IR, human insulin receptor; IGF-I-R, human IGF-I receptor. Exposure times: **lane 1**, 24 h; **lanes 2–4**, 5 days; **lanes 5–8**, 4 days. Sizes are in kb and were deduced from commercial, synthetic RNAs (RNA Ladder, BRL) ranging in size from 0.3 to 9.5 kb.

The amino-terminal sequence and five peptide sequences were determined, as shown in Figure 1b.

Isolation and characterization of IGF-I receptor cDNA

Long, single oligonucleotide hybridization probes were designed on the basis of partial sequences 1 and 4 (Figure 1b). Codon choices were based on corresponding nucleotide sequences of the insulin proreceptor (Ullrich *et al.*, 1985) as well as codon usage frequencies of human genes (Grantham *et al.*, 1981). Gaps in the sequence were filled with codons found in the insulin receptor homologous sequence.

A 78-nucleotide synthetic probe derived from the amino-terminal α subunit sequence (Figure 2a) was used initially to screen 2×10^6 clones of a human term placental cDNA library in λgt10 (Ullrich *et al.*, 1985). Clone λIGF-1-R.8 (Figure 2b) contained the probe sequence. The 730 bp *Eco*RI insert of λIGF-1-R.8 and the synthetic probe derived from peptide 4 (Figure 2a) were used for subsequent screening of the same placental library to yield overlapping clones λIGF-1-R.85 and λIGF-1-R.76, of 2.8 kb and 3.5 kb, respectively (Figure 2b).

Complete nucleotide sequence analysis of the cloned cDNA resulted in deduction of the complete primary sequence of the human IGF-1-receptor precursor (Figure 2c). The 4989-nucleotide sequence contains an open reading frame of 4101 nucleotides, which begins with an ATG codon that is flanked by sequences that meet the requirements for an initiation codon as defined by Kozak (1984). The 45-nucleotide sequence preceding this potential initiation codon includes a purine-rich portion and a 5′-terminal oligo T sequence and is likely to be part of a 5′-untranslated sequence, although no in-frame stop codon can be identified. At the 3′ end, the open reading frame is flanked by a TGA signal for translation termination and 840 nucleotides of 3′-untranslated sequence. No polyadenylation signal (AATAAA) or poly(A) tail is found at the 3′ end of the λIGF-1-R.76 cDNA insert, suggesting that the 3′-untranslated region of our sequence is incomplete.

Primary structure of the IGF-I receptor precursor

The initiation methionine is the first amino acid of a 30-residue sequence which precedes the chemically identified amino-terminal glutamic acid residue. This displays structural features characteristic of signal peptides, necessary for transfer of the nascent polypeptide chain into the membrane of the endoplasmic reticulum. The IGF-I receptor signal sequence is unusually rich in polar residues such as threonine and serine (30%). Cleavage by signal peptidase occurs after a glycine residue to expose the amino-terminal IGF-I receptor α subunit glutamic acid residue; this establishes the preproreceptor organization to be NH_2-signal peptide-α subunit-β subunit-COOH. The α subunit region contains a single cysteine-rich (24 cys) region between residues 148 and 302, and 11 potential N-linked glycosylation sites. An Arg-Lys-Arg-Arg tetrapeptide after residue 706 marks the putative cleavage site of the $\alpha\beta$ proreceptor polypeptide, suggesting that the β subunit begins at Asp 711.

The β subunit portion is characterized by the presence of a single, 24 amino acid hydrophobic sequence (residues 906–929) that is likely to be the transmembrane domain, and is flanked at its carboxy terminus by a stretch of basic amino acids that may facilitate plasma membrane anchoring. Five potential N-linked glycosylation sites are found between residues 711 and 905 upstream from this hydrophobic sequence. This region probably represents β subunit extracellular sequences. Structural features characteristic of tyrosine kinase enzymatic domains are found downstream from the transmembrane sequence, thus the cytoplasmic domain is generated by the carboxy-terminal 407 amino acids of the β subunit.

Tyrosine kinase features are located within a 257-residue region

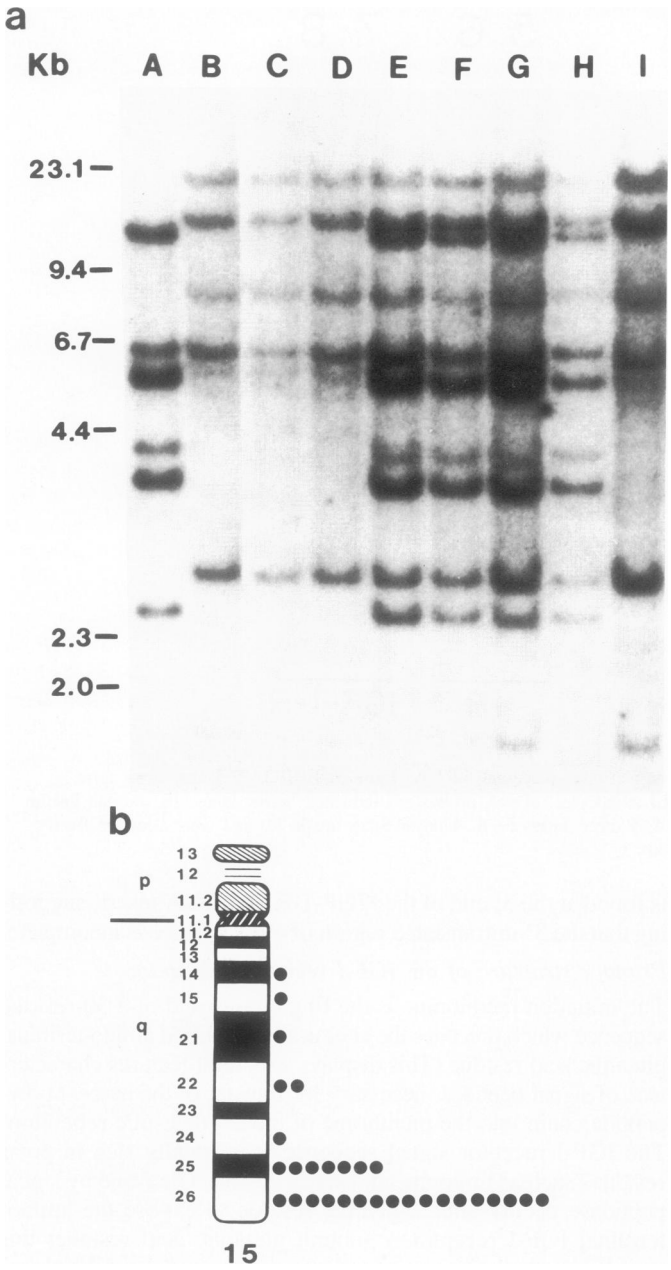


Fig. 5. Chromosome mapping of IGF-I receptor locus. **(a)** Autoradiogram obtained following hybridization of IGF-I receptor cDNA probe (λIGF-1-R.76 insert) with *Eco*RI-digested DNA from human × Chinese hamster hybrid cells and controls. **Lane A:** human diploid cell line; **lane B:** Chinese hamster Don/a23; **lanes E–H:** hybrid cell lines containing human chromosome 15; **lanes C, D** and **I:** hybrid cell lines missing human chromosome 15. **(b)** Silver grain distribution along chromosome 15 (ideogram from ISCN, 1985).

between amino acids 973 and 1229 and include a potential ATP binding site (Gly 976 to Gly 981 and Lys 1003). The entire unmodified IGF-I proreceptor polypeptide chain having a predicted M_r of 151 869 can thus be subdivided into the 80 423 dalton α subunit that lacks sequences with membrane-spanning characteristics, and a 70 866 dalton β subunit that contains tyrosine kinase-homologous sequences and appears to anchor the intact receptor complex in the plasma membrane.

Comparison with insulin receptor

Figure 3 shows an alignment of insulin and IGF-I receptor precursor amino acid sequences (a) and a comparison of their hydropathy profiles (b). This analysis demonstrates the close similarities in the overall organization and primary structures of the two receptor precursors. Despite some minor insertions and/or deletions that are spread throughout both sequences, the precursors show colinear organization and are approximately the same length (1337 amino acids of the IGF-I receptor versus 1343 for the insulin receptor precursor excluding the signal peptide). While signal peptide and transmembrane sequences show the lowest degree of similarity, their overall hydrophobic character is well conserved. As expected, the highest homology (84%) is found in the β subunit cytoplasmic region, defining the enzymatic domain for tyrosine-specific kinase activity. Interestingly, the second highest degree of similarity (64–67%) between the two receptor precursors is found in extracellular α subunit regions flanking the cysteine-rich subdomains. Similar levels of sequence homology (61%) are localized in the interface between transmembrane and tyrosine kinase domains, where conservation of a basic Arg-Lys-Arg sequence that flanks the C terminus of the transmembrane sequence, in addition to conservation of three tyrosine residues and their flanking sequences, is found. Sequence similarity is lower within the α subunit cysteine-rich domain (48%), despite conservation of 24 out of 26 cysteine residues between IGF-I and insulin receptors. Sequences surrounding these conserved cysteine residues are very hydrophilic in both proteins, suggesting a tightly structured conformation with a hydrophilic surface. In the two receptors, α subunit C-terminal and β precursor extracellular sequences are ~44% homologous, and include conserved cysteine residues (7 out of 8) as well as the precursor cleavage sequence (Arg-Lys-Arg-Arg). Fifteen out of sixteen potential N-linked glycosylation sites in the extracellular domains of α and β subunits of the IGF-I receptor are found at nearly identical locations in the insulin receptor, which contains three additional Asn-X-Thr/Ser consensus sequences. Within the cytoplasmic domain, the C-terminal hydrophilic tails of IGF-I and insulin receptors display the most divergent sequences, although they are similar in length (108 versus 99, respectively). Conserved structural features in the cytoplasmic domain of IGF-I receptor include four out of five cysteines and 11 out of 15 tyrosine residues. As can be seen in Figure 3b, the hydropathy profiles of the two receptor precursors are very similar, yet the insulin receptor is slightly more hydrophilic in most regions.

IGF-I receptor transcripts

It has previously been shown that a cDNA encoding the human insulin precursor hybridizes with 4–6 mRNAs in term placenta (Ullrich *et al.*, 1985; Ebina *et al.*, 1985). Given the structural similarities and ligand overlap of the insulin and IGF-I receptors, it was possible that one or more of these mRNA species represented cross-hybridization with an IGF-I receptor transcript. To address this question, Northern blot hybridizations were carried out using poly(A)-containing RNA preparations from various tissues, and the hybridization patterns obtained with both IGF-I and insulin receptor probes were compared. As can be seen in Figure 4, IGF-I and insulin receptor cDNAs displayed different patterns of hybridization with human RNA. While insulin receptor cDNA hybridized to four major RNAs of 10.3, 9.6, 8.5 and 6.7 kb (Figure 4, lane 1), IGF-I cDNA hybridization yielded fainter signals of 11 and 7 kb (Figure 4, lane 2). The hybridization signals detected reflect about 10- to 20-fold higher levels of insulin receptor mRNAs than IGF-I receptor mRNAs in human term placenta, which is different from the relative amounts of these proteins in this tissue (about 5-fold; Fujita-Yamaguchi *et al.*, 1983; Le Bon *et al.*, 1986).

Table I. Correlation of human chromosomes and human IGF-I receptor sequences in human × rodent hybrid cell lines

	Hybridization/ chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
Concordant hybrids	+/+	2	1	4	2	2	4	2	4	2	1	2	3	4	3	6	2	1	4	3	3	5	5	2	0
	−/−	4	4	3	7	7	3	6	4	10	8	5	7	4	1	10	4	7	5	4	6	2	4	3	8
Discordant hybrids	+/−	4	5	2	3	4	2	4	2	4	5	4	3	2	2	0	3	5	2	3	3	1	1	1	6
	−/+	2	5	6	3	2	6	3	5	0	2	4	3	5	8	0	5	3	5	6	4	7	6	3	2
Discordant hybrids		6	10	8	6	6	8	7	7	4	7	8	6	7	10	0	8	8	7	9	7	8	7	4	8
Informative hybrids[a]		12	15	15	15	15	15	15	15	16	16	15	16	15	14	16	14	16	16	16	16	15	16	9	16

[a]Data for chromosomes rearranged or present at a frequency of less than 10% were excluded.

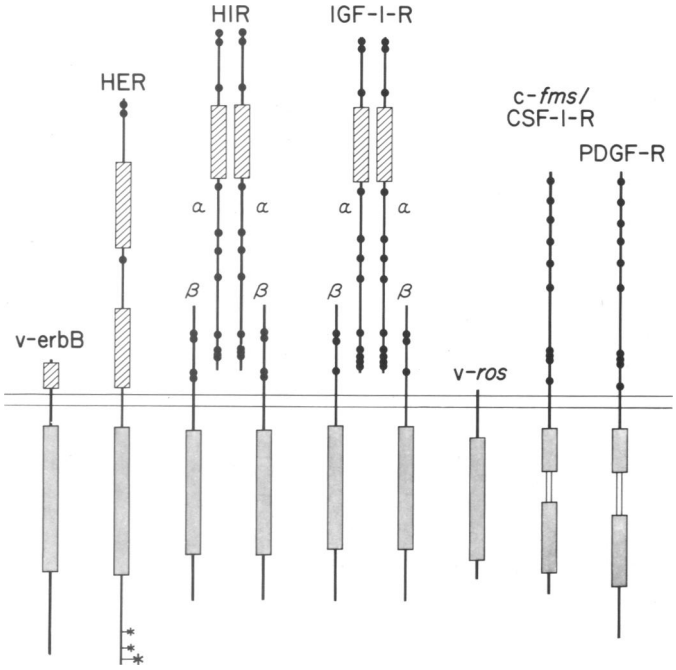


Fig. 6. Schematic comparison of IGF-I receptor with other cell surface receptors and oncogene products. Cysteine-rich domains are shown as cross-hatched boxes; other cysteine residues in extracellular domain as filled circles. Tyrosine kinase domains are shaded boxes, with insertions in c-*fms*/CSF-1-R and PDGF-R shown as an open box. Asterisks indicate sites for *in vitro* autophosphorylation in HER. HER, human epidermal growth factor receptor; CSF-1-R, colony stimulating factor-1 receptor; PDGF-R, platelet-derived growth factor receptor.

Rat tissues contained a single, 11-kb IGF-I receptor mRNA compared with 10.3- and 8.5-kb insulin receptor transcripts (not shown). Buffalo rat liver cells (BRL-3A2) were found to express high levels of IGF-I receptor mRNA (Figure 4, lane 3) and no detectable insulin receptor mRNA (not shown). Massague and Czech (1982) have reported that IGF-I receptors as well as insulin receptors increase during the differentiation of 3T3-LI fibroblasts to adipocytes. To determine if this increase reflected increased IGF-I receptor mRNA levels, the different 3T3-LI cell types were analyzed by Northern hybridization. Interestingly, reduced levels of 11-kb mRNA were found in the mouse adipocytes when compared with fibroblasts (Figure 4, lanes 7 and 8). In contrast, substantially higher insulin receptor mRNA levels were detected in a parallel analysis of the same mRNA preparations (Figure 4, lanes 5 and 6, and Ullrich *et al.*, 1985), in agreement with previous reports (Rubin *et al.*, 1978; Karlsson *et al.*, 1979; Massague and Czech, 1982).

Chromosome mapping

Southern blot analysis of DNA from 16 human × rodent cell hybrids were used to localize human IGF-I receptor sequences to chromosome 15. ^{32}P-labelled human IGF-I receptor cDNA was hybridized to Southern filters of *Eco*RI-digested DNA from hybrids and parental cell lines. Six human-specific fragments of 13, 6.6, 5.6, 4.2, 3.9 and 2.6 kb were observed (Figure 5, lane A). Seven hybridizing bands (21, 14, 8.3, 6.6, 5.8, 2.8 and 1.3 kb) were present in DNA from the Chinese hamster parental cell line (Figure 5, lane B). Furthermore, five hybridizing bands (21, 10.5, 8, 4.2 and 3.5 kb) were detected in rat DNA, and nine (15.5, 13, 8.6, 7.8, 4.4, 3.8, 3.2, 2.6 and 2.2 kb) in mouse 3T3 DNA (data not shown). In Chinese hamster × human hybrid DNAs, all five human-specific *Eco*RI fragments that were distinguishable from hamster fragments were present in hybrids containing human chromosome 15 (Figure 5, lanes E, F, G and H), and all five were absent in hybrids lacking chromosome 15 (Figure 5, lanes C, D and I). The results of the discordancy analysis for 16 hybrids are summarized in Table I. The 14 Chinese hamster × human hybrids were from six different series: one was a mouse × human and one a rat × human hybrid. Chromosome 15 shows perfect concordance with the human IGF-I receptor sequences, whereas all other chromosomes are excluded by at least four discordant hybrids.

Regional assignment of the human IGF-I receptor gene was carried out by *in situ* hybridization of ^{3}H-labelled IGF-I receptor cDNA probe to human chromosome spreads. Twenty-three out of 75 (30.7%) cells exhibited silver grains at bands q25→q26 of chromosome 15. Of the 174 total grains scored, 24 (13.8%) were found over this specific region with most grains at the distal band 15q26 (Figure 5b). No other chromosomal site was labelled above background. The *in situ* hybridization data confirm the mapping of the IGF-I receptor gene to chromosome 15 made by somatic cell hybrid analysis and further refine its map position to the distal band of the long arm of chromosome 15.

Discussion

Since binding of insulin and IGF-I to their respective cell surface receptors triggers distinct cellular responses, the insulin and IGF-I receptors must differ not only in the structure of their ligand binding pockets, but also in regions necessary for their unique physiological functions. Both of these receptors are tyrosine-specific protein kinases, and their distinct roles in differentiation and cellular metabolism suggests that they utilize different pathways of signal transduction involving phosphorylation of different classes of substrates.

In order to identify structural features of the insulin and IGF-I receptors that define their distinct physiological roles, the com-

plete primary structure of the IGF-I receptor has been characterized. By first isolating sufficient quantities of the receptor glycoprotein from human placental membranes using an IGF-I receptor-specific monoclonal antibody (Le Bon *et al.*, 1986), partial amino acid sequence information was obtained and thus overlapping cDNA clones containing coding sequences for the entire precursor of the receptor were identified. The assembled cDNA sequence predicts a 1367 amino acid receptor precursor, including a 30 amino acid signal peptide, which is removed during translocation of the nascent polypeptide chain. The 1337 residue, unmodified proreceptor polypeptide has a predicted M_r of 151 869, which compares with the 180 000 M_r IGF-I receptor precursor (Jacobs *et al.*, 1983a). In analogy with the 152 784 M_r insulin receptor precursor, cleavage of the Arg-Lys-Arg-Arg sequence at position 707 of the IGF-I receptor precursor will generate α (80 423 M_r) and β (70 866 M_r) subunits, which compare with ~135 000 M_r (α) and 90 000 M_r (β) fully glycosylated subunits.

Extensive similarity with the insulin receptor is not only found at the level of precursor and subunit size, but also at the level of structural topology and primary sequence. Only one characteristic transmembrane domain can be identified in the IGF-I receptor $\alpha\beta$ precursor. Thus, like the insulin receptor, the ~350 000 M_r heterotetrameric IGF-I receptor complex spans the plasma membrane via two β subunit membrane-spanning domains, leaving 195 amino acid portions of the β subunits protruding from the cell surface to which entire extracellular α subunits are attached by disulfide bonds (Ullrich *et al.*, 1985; Grunfeld *et al.*, 1985).

The ligand binding pockets of IGF-I and insulin receptors are formed by the extracellular α subunits and possibly some extracellular portions of the β subunits. Differences in receptor ligand specificities are likely to be dictated by sequence differences within this region, and indeed lower homology was found in the extracellular cysteine-rich domains (48%), C-terminal one-third of the α subunits (47%) and N-terminal portion of the β subunits (41%). These regions are more hydrophilic than the remainder of the extracellular domain sequences, and are likely to be exposed on the surface of this domain and to function there to define ligand specificity.

Conservation of cysteine residue spacing in the single cysteine-rich domain of the IGF-I receptor α subunit indicates a common evolutionary origin with the closely related insulin receptor as well as the epidermal growth factor (EGF) receptor (Ullrich *et al.*, 1985) and the *neu* oncogene-related putative hormone receptor HER2 (Coussens *et al.*, 1985) (Figure 6). Divergence of sequences flanking these conserved cysteine residues in receptor ligand-binding domains suggests that they may define ligand specificity within a common, compact structural unit formed by a network of disulfide bridges. Disturbance of a similar structure in the low density lipoprotein receptor has been shown to abolish ligand binding (Yamamoto *et al.*, 1984).

Most cysteine residues outside this Cys-cluster are also conserved between IGF-I and insulin receptors and may crosslink α and β subunits to form the biologically active $\alpha_2-\beta_2$ heterotetrameric receptor complex, as suggested in Figure 6. In addition, the locations of most N-linked glycosylation sites are conserved between IGF-I receptor and insulin receptor α and β subunits. The size difference between unmodified and fully glycosylated subunits indicates that most N-linked glycosylation sites are modified.

Binding of IGF-I to the extracellular portion of its cell surface receptor in some way initiates a cytoplasmic signal cascade that includes autophosphorylation of tyrosine residues within the intracellular portion of the insulin receptor β subunit. Signal transduction through the plasma membrane may involve intramolecular signal transfer through the transmembrane domain or ligand-induced, intermolecular receptor association. In the case of IGF-I and insulin receptors, the heterotetrameric complex structure may represent a stabilized receptor dimer, in which a conformation change in one $\alpha-\beta$ subunit extracellular domain could trigger both β cytoplasmic domains into an activated conformation. Although direct transfer of ligand binding information via a single transmembrane domain cannot be excluded, available data support the aggregation of intermolecular model of transmembrane signal transfer (Heffetz and Zick, 1986).

The extensive homology of sequences downstream from the putative ATP-binding consensus sequence with members of the *src* family of tyrosine kinase oncogenes, as well as the hormone receptors of this gene family, indicates that they represent the domain encoding the tyrosine kinase enzymatic activity. Despite its overall homology with the insulin receptor kinase domain (84%), it includes three discrete regions of sequence divergence (Figure 3a) following residues 986, 1072 and 1208. The nonapeptide Val-Leu-Ala-Pro-Pro-Ser-Leu-Ser-Lys, which begins at residue 1073, is present at the exact position in the IGF-I receptor where 70- and 100-residue heterologous insertion sequences are found in c-*fms*/CSF-1 receptor and platelet-derived growth factor receptor structures (Sherr *et al.*, 1985; Coussens *et al.*, 1986; Yarden *et al.*, 1986), respectively (Figure 6). The presence of such a highly heterogeneous sequence within otherwise highly conserved tyrosine kinase domains of gene family members appears highly significant and indicates a possible function of this subdomain in definition of specific receptor function. This possibility is emphasized by the fact that even receptors as closely related as the IGF-I and insulin receptors diverge at this point. This insertion may therefore reflect a structural determinant that defines unique tyrosine kinase substrate interactions.

The sequence homology between IGF-I and insulin receptor cytoplasmic domains ceases abruptly at isoleucine 1233 and remains at a remarkly low level (44% overall) up to the carboxyl terminus. Carboxy-terminal domains of other tyrosine kinase receptors have been characterized to be equally sequence heterogeneous, retaining in all cases, however, a hydrophilic character as shown in Figure 3b for IGF-I and insulin receptors. This domain contains major and minor *in vitro* autophosphorylation tyrosine target sites in the EGF receptor which are conserved in the putative receptor HER2/*neu*. As suggested by sequence heterogeneity, this carboxy-terminal receptor domain may, in conjunction with the nonapeptide sequence at position 1073–1081 and the divergent membrane-proximal region between residues 933 and 955, be responsible for receptor-specific, ligand-induced, intracellular signal generation.

Several cell surface receptors are substrates for protein kinase C phosphorylation. In the human EGF receptor, the specific residue phosphorylated (Thr 654) is adjacent to the membrane-spanning region of the receptor and is surrounded by basic residues (Hunter *et al.*, 1984). This threonine residue is conserved in HER2/*neu* (Thr 685) (Coussens *et al.*, 1985), and an identically situated serine residue, also flanked by basic residues, is the target of protein kinase C in $pp60^{v\text{-}src}$ (Gould *et al.*, 1985). Although it has not yet been demonstrated that IGF-I or insulin receptors are protein kinase C substrates, both receptors are phosphorylated predominantly on serine residues when intact cells are treated with phorbol esters, known activators of protein kinase C (Jacobs *et al.*, 1983c; Takayama *et al.*, 1984; Jacobs and Cuatrecasas,

1986). It is thus noteworthy that the IGF-I receptor has a serine residue (935), flanked by basic residues, that is located near the membrane-spanning domain and could serve as a protein kinase C target. Since serine 935 is not conserved in the insulin receptor, the similar receptor phosphopeptide maps obtained after phorbol ester treatment (Jacobs and Cuatrecasas, 1986) must not reflect phosphorylation at this site.

The finding that the v-*erbB* and v-*fms* oncogenes of acute transforming retroviruses were derived from growth factor receptor proto-oncogenes by carboxy-terminal sequence truncation further emphasizes the functional importance of this receptor domain, and raises the question as to whether the IGF-I receptor gene represents a potential proto-oncogene. The transforming gene of the avian sarcoma virus UR-2 (v-*ros*) (Neckameyer and Wang, 1985) shares sequence homology with tyrosine kinase family members and was found to be most similar to the insulin receptor (Ullrich *et al.*, 1985). V-*ros* is equally homologous to IGF-I and insulin receptors (Figure 6), which are in fact more similar to each other. Comparison of the IGF-I receptor with all currently known oncogene sequences did not yield any new potential receptor–oncogene relationship.

Analysis of IGF-I receptor mRNA size and expression revealed 7- and 11-kb mRNAs in human tissues, of which a 5-kb cDNA copy was cloned. The two IGF-I receptor mRNAs identified in human term placenta are derived from a single gene, as determined by use of 3′-untranslated sequence probes on Southern blot genomic analysis which revealed a single hybridizing band after cleavage with *Eco*RI and *Pst*I (not shown). It is unknown whether both mRNAs code for the same IGF-I receptor precursor sequence or whether structural variants exist, like the two insulin receptor species of Ullrich *et al.* (1985) and Ebina *et al.* (1985) (insulin receptor variants A and B; Tam,A., Gray,A. and Ullrich,A., in preparation), which differ by a 12 amino acid insertion at the C terminus of the α subunit. The two human mRNAs may include 3- and 7-kb untranslated sequences, respectively. The rat and mouse tissues examined only express an 11-kb mRNA.

Although closely related to the insulin receptor, the IGF-I receptor gene maps to a distinct chromosomal locus. The IGF-I receptor gene maps to 15q25→26, whereas the insulin receptor gene was recently localized on chromosome 19 band p13.3→p13.2 (Yang-Feng *et al.*, 1985b). Chromosomal assignment of *IGF1R* to 15q25→26 coincides with the map position of the cellular proto-oncogene c-*fes* (Harper *et al.*, 1983). C-*fes* was originally thought to be involved in a specific translocation [t(15;17)(q22; q21)] found in 70% of patients with acute promyelocytic leukemia; however, it now appears that c-*fes* is not associated with this form of leukemia (LeBeau and Rowley, 1986). Chromosomes 15 and 19 contain several loci for genes that are functionally and/or evolutionarily related: *CYP2*, cytochrome P450-dioxin inducible on 15 (Hildebrand *et al.*, 1985) and *CYP1*, cytochrome P450- phenobarbital inducible on 19 (Phillips *et al.*, 1985); *MANA*, α-mannosidase A on 15 (Champion *et al.*, 1978) and *MANB*, α-mannosidase B on 19 (Ingram *et al.*, 1977); *MPI*, mannose phosphate isomerase on 15 (Chern and Croce, 1975) and for polio, Echo II and Baboon M7 viruses on 19 (Siddique *et al.*, 1985; Gerald and Bruns, 1978; Brown *et al.*, 1979). Thus, it is possible that the *IGF1R* locus on 15q and the *INSR* locus on 19p are part of an ancestrally related cluster of genes.

During differentiation of mouse 3T3-LI fibroblasts into adipocytes, we found that the levels of the 11-kb IGF-I receptor mRNA decreased, in direct contrast with the insulin receptor mRNAs, which dramatically increased. This clearly demonstrates that the insulin and IGF-I receptors are subject to different mechanisms of gene regulation, consistent with the distinct functions of their products.

The availability of complete cDNAs encoding human insulin and IGF-I receptors will enable, for the first time, a molecular dissection of domains that mediate the distinct physiological functions of these highly homologous receptors, which play central roles in human development and metabolism.

Materials and methods

Purification and amino acid sequence determination of the IGF-I receptor

The IGF-I receptor was purified from Triton X-100 solubilized human placental membranes by WGA-Sepharose chromatography followed by immuno-affinity chromatography using αIR-3, a monoclonal antibody directed against the IGF-I receptor (Kull *et al.*, 1983), as previously described (Le Bon *et al.*, 1986). Amino acid sequence determination was carried out on receptor that was prepared from large pooled fractions of the flow-through portion of insulin–Sepharose chromatography, which had been accumulated and stored frozen. Briefly, the flow-through fraction prepared from 20–25 placentae as described for purification of the insulin receptor (Fujita-Yamaguchi *et al.*, 1983) was applied to a 2 × 6.4 cm αIR-3 column (~0.5 mg antibody/ml of packed gel) which had been equilibrated with 50 mM Tris–HCl buffer, pH 7.4. The column was washed extensively with this buffer, and then eluted with 50 mM acetate buffer, pH 5 and 1 M glycine buffer, pH 2.2. All buffers used contained 1 M NaCl, 0.1% Triton X-100, 0.1 mM phenylmethylsulfonyl fluoride, and 2 mM *N*-α-benzoyl-L-arginine ethyl ester. Fractions (2 ml) were collected in tubes containing 3 ml of 0.5 M Tris to neutralize the eluates. The pH 2.2 fractions were pooled and concentrated using a PM-10 Diaflo ultrafiltration membrane (Amicon Corp.). The amount of the IGF-I receptor recovered was ~200 μg.

The IGF-I receptor preparations obtained by affinity chromatogaphy on αIR-3-Sepharose columns were precipitated by the methanol–chloroform procedure (Wessel and Flugge, 1984). The precipitate was redissolved in 50 mM Tris–HCl, pH 6.8, containing 2% SDS and 20 mM dithiothreitol and heated at 60°C for 30 min. This material was subjected to SDS–PAGE in 7% polyacrylamide gels in the presence of 0.1 mM thioglycolate. After electrophoresis, a guide strip of the gel was stained with Coomassie blue and the bands corresponding to M_r = 120 000 and ~90 000 were electro-eluted as described (Hunkapiller *et al.*, 1983). The purified α and β subunits obtained by electro-elution were precipitated by the methanol–chloroform procedure. A part of the isolated α subunit (130 pmol) was subjected to amino acid sequence analysis, and the rest (~300 pmol) was digested by lysyl peptidase. The digests were separated by reversed-phase h.p.l.c. on a 2 mm × 10 cm C4 column by elution with a linear gradient of 0.1% trifluoroacetic acid–70% I-propanol. The peptide peaks which seemed to be pure were collected, concentrated, and subjected to amino acid sequence determination on the Applied Biosystem Protein Sequencer 470B equipped with PTH Analyzer 120A.

cDNA cloning and library screening

Preparation of human term placenta poly(A)$^+$ mRNA (Cathala *et al.*, 1983), cDNA cloning in λgt10 (Huynh *et al.*, 1985), and screening of a cDNA library of 2 × 10^6 recombinant phages was carried out as previously described (Ullrich *et al.*, 1984a). Two complementary unique 78-mers (Figure 2a) were synthesized (Crea and Horn, 1980), radioactively labelled with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (Amersham) and used for plaque hybridization in 30% formamide and otherwise under conditions described earlier (Ullrich *et al.*, 1984b). Subsequent plaque hybridizations were carried out in parallel with a human insulin receptor fragment under high stringency conditions (50% formamide) to eliminate cross-hybridizing insulin receptor clones.

Northern blot analysis

Formaldehyde agarose gels (1.2%) were used according to the procedure of Lehrach *et al.* (1977). Total poly(A)$^+$ RNA (5 μg per lane) from various tissues was analyzed using a 2.8-kb *Eco*RI fragment as a radioactively labelled (Taylor *et al.*, 1976) hybridization probe. Nitrocellulose filters were washed and exposed to X-ray film at −70°C using an intensifying screen (Cronex Lightning Plus, Dupont).

Chromosomal mapping

The 2.8-kb cDNA *Eco*RI insert of λIGF-1-R.85 was used for Southern blot analysis of *Eco*RI-digested human × Chinese hamster hybrid cell DNAs and for *in situ* hybridization to human metaphase and prometaphase chromosomes as described in detail elsewhere (Harper and Saunders, 1981; Yang-Feng *et al.*, 1985a).

Acknowledgements

We thank Mark Vasser and Parkash Jhurani for preparing oligonucleotide probes, and Jeanne Arch for typing the manuscript. We are especially grateful to Dr Suzanne Pfeffer for critically reviewing and editing this paper. This work was supported by Research Grants AM29770 and AM34427 to Y.F.Y. and CA-16417 to J.R. from the National Institutes of Health.

References

Bhaumick,B., Bala,R.M. and Hollenberg,M.D. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 4279–4283.

Brown,S., Oie,H.K., Gazda,A.F., Minna,J.D. and Francke,U. (1979) *Cell*, **18**, 135–143.

Cathala,G., Savouret,J-F., Mendez,B., West,B.L., Karin,M., Martial,J.A. and Baxter,J.D. (1983) *DNA*, **2**, 329–335.

Champion,M.J., Brown,J.A. and Shows,T.P. (1978) *Cytogenet. Cell Genet.*, **22**, 498–502.

Chern,C.J. and Croce,C.M. (1975) *Cytogenet. Cell Genet.*, **15**, 299–305.

Chernausek,S.D., Jacobs,S. and Van Wyk,J.J. (1981) *Biochemistry*, **20**, 7345–7350.

Coussens,L., Yang-Feng,T.L., Liao,Y-C., Chen,E., Gray,A., McGrath,J., Seeburg,P.H., Libermann,T.A., Schlessinger,J., Francke,U., Levinson,A. and Ullrich,A. (1985) *Science*, **230**, 1132–1139.

Coussens,L., Van Beveren,C., Smith,D., Chen,E., Mitchell,R.L., Isacke,C.M., Verma,I.M. and Ullrich,A. (1986) *Nature*, **320**, 277–280.

Crea,R. and Horn,T. (1980) *Nucleic Acids Res.*, **8**, 2331–2348.

Ebina,Y., Ellis,L., Jarnagin,K., Edery,M., Graf,L., Clauser,E., Ou,J-H., Masiarz,F., Kan,Y.W., Goldfine,I.D., Roth,R.A. and Rutter,W.J. (1985) *Cell*, **40**, 747–758.

Flier,J.S., Usher,P. and Moses,A.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 664–668.

Froesch,E.R., Schmid,C., Schwander,J. and Zapf,J. (1985) *Annu. Rev. Physiol.*, **47**, 443–467.

Fujita-Yamaguchi,Y., Choi,S., Sakamoto,Y. and Itakura,K. (1983) *J. Biol. Chem.*, **258**, 5045–5049.

Gerald,P.S. and Bruns,G.A. (1978) In Summitt,R.L. and Bergsma,D. (eds), *Proceedings of the 1977 Birth Defects Symposium, Tennessee*. Alan R.Liss, Inc., New York, pp. 225–248.

Gould,K.L., Woodwet,J.R., Cooper,J.A., Buss,J.E., Shalloway,D. and Hunter, T. (1985) *Cell*, **42**, 849–857.

Grantham,R., Gautier,C., Gouy,M., Jacobzone,M. and Mercier,R. (1981) *Nucleic Acids Res.*, **9**, 43–74.

Grunfeld,C., Shigenaga,J.K. and Ramachandran,J. (1985) *Biochem. Biophys. Res. Commun.*, **133**, 389–396.

Harper,M.E. and Saunders,G.F. (1981) *Chromosoma*, **83**, 431–439.

Harper,M.E., Franchini,G., Love,J., Simon,M.I., Gallo,R.C. and Wong-Staal,F. (1983) *Nature*, **304**, 169–171.

Heffetz,D. and Zick,Y. (1986) *J. Biol. Chem.*, **261**, 889–894.

Hildebrand,C.E., Gonzales,F.J., McBride,O.W. and Nebert,D.W. (1985) *Nucleic Acids Res.*, **13**, 2009–2016.

Hunkapiller,M.W., Lujan,E., Ostrander,F. and Hood,L.E. (1983) In Hirs,C.H.W. and Timasheff,A. (eds), *Methods in Enzymology*. Vol. 91. Academic Press, New York, pp. 221–236.

Hunter,T., Ling,N. and Cooper,J.A. (1984) *Nature*, **311**, 480–483.

Huynh,T., Young,R. and Davis,R. (1985) In Glover,D. (ed.), *DNA Cloning: A Practical Approach*. IRL, Oxford, Vol. I, pp. 49–78.

ISCN (1985) *An International System for Human Cytogenetic Nomenclature*. S.Karger, Basel.

Ingram,P.H., Bruns,G.A., Regina,V.M., Eisenman,R.A. and Gerald,P.S. (1977) *Biochem. Genet.*, **15**, 455–476.

Jacobs,S. and Cuatrecasas,P. (1986) *J. Biol. Chem.*, **261**, 934–939.

Jacobs,S., Kull,F.C. and Cuatrecasas,P. (1983a) *Proc. Natl. Acad. Sci. USA*, **80**, 1228–1231.

Jacobs,S., Kull,F.C., Earp,H.S., Svoboda,M., Van Wyk,J.J. and Cuatrecasas,P. (1983b) *J. Biol. Chem.*, **258**, 9581–9584.

Jacobs,S., Sahyoun,N.E., Saltiel,A.R. and Cuatrecasas,P. (1983c) *Proc. Natl. Acad. Sci. USA*, **80**, 6211–6213.

Karlsson,F.A., Grunfeld,C., Kahn,C.R. and Roth,J. (1979) *Endocrinology*, **104**, 1383–1390.

Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.

Kull,F.C.,Jr, Jacobs,S., Su,F-Y., Svoboda,M.E., Van Wyk,J.J. and Cuatrecasas,P. (1983) *J. Biol. Chem.*, **258**, 6561–6566.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.

LeBeau,M.M. and Rowley,J.D. (1986) In Harris,H. and Hirschhorn,K. (eds), *Advances in Human Genetics*, Vol. 15. Plenum Press, New York, pp. 1–54.

LeBon,T.R., Jacobs,S., Cuatrecasas,P., Kathuria,S. and Fujita-Yamaguchi,Y. (1986) *J. Biol. Chem.*, **261**, 7685–7689.

Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry*, **16**, 4743–4751.

Massague,J. and Czech,M.P. (1982) *J. Biol. Chem.*, **257**, 5038–5045.

Neckameyer,W.S. and Wang,L.-H. (1985) *J. Virol.*, **53**, 879–884.

Phillips,I.R., Shephard,E.A., Povey,S., Davis,M.B., Kelsey,G., Monteiro,M., West,L.F. and Cowell,J. (1985) *Annu. Hum. Genet.*, **49**, 267–274.

Rechler,M.M. and Nissley,S.P. (1985) *Annu. Rev. Physiol.*, **47**, 425–442.

Ronnett,G.V., Knutson,V.P., Kohanski,R.A., Simpson,T.L. and Lane,M.D. (1984) *J. Biol. Chem.*, **259**, 4566–4575.

Rubin,C.S., Hirsch,A., Fung,C. and Rosen,O.M. (1978) *J. Biol. Chem.*, **253**, 7570–7578.

Rubin,J.B., Shia,M.A. and Pilch,P.F. (1983) *Nature*, **305**, 438–440.

Sherr,C.J., Rettenmier,C.W., Sacca,R., Roussel,M.F., Look,A.T. and Stanley, E.R. (1985) *Cell*, **41**, 665–676.

Siddique,T., Bartlett,R.J., McKinney,R., Hung,W-Y., Bruns,G., Wilfert,C. and Roses,A.D. (1985) *Cytogenet. Cell Genet.*, **40**, 745–756.

Takayama,S., White,M.F., Lauris,V. and Kahn,C.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7797–7801.

Taylor,J.M., Illmensee,R. and Summers,J. (1976) *Biochim. Biophys, Acta*, **442**, 324–330.

Ullrich,A., Coussens,L., Hayflick,J.S., Dull,T.J., Gray,A., Tam,A.W., Lee,J., Yarden,Y., Libermann,T.A., Schlessinger,J., Downward,J., Bye,J., Whittle, N., Waterfield,M.D. and Seeburg,P.H. (1984a) *Nature*, **309**, 418–425.

Ullrich,A., Berman,C.H., Dull,T.J., Gray,A. and Lee,J.M. (1984b) *EMBO J.*, **3**, 361–364.

Ullrich,A., Bell,J.R., Chen,E.Y., Herrera,R., Petruzzelli,L.M., Dull,T.J., Gray,A., Coussens,L., Liao,Y-C., Tsubokawa,M., Mason,A., Seeburg,P.H., Grunfeld,C., Rosen,O.M. and Ramachandran,J. (1985) *Nature*, **313**, 756–761.

Wessel,D. and Flugge,U.I. (1984) *Anal. Biochem.*, **138**, 141–143.

Yamamoto,T., Davis,C.G., Brown,M.S., Schneider,W.J., Casey,M.L., Goldstein,J.L. and Russell,D.W. (1984) *Cell*, **39**, 27–38.

Yang-Feng,T.L., Floyd-Smith,G., Nemer,M., Drouin,J. and Francke,U. (1985a) *Am. J. Hum. Genet.*, **37**, 1117–1128.

Yang-Feng,T.L., Ullrich,A. and Francke,U. (1985b) *Science*, **228**, 728–731.

Yarden,Y., Escobedo,J.A., Kuang,W-J., Yang-Feng,T.L., Daniel,T.O., Tremble, P.M., Chen,E.Y., Ando,M.E., Harkins,R.A., Francke,U., Fried,V.A., Ullrich,A. and Williams,L.T. (1986) *Nature*, in press.

Zick,Y., Sasaki,N., Rees-Jones,R.W., Grunberger,G., Nissley,S.P. and Rechler, M.M. (1984) *Biochem. Biophys. Res. Commun.*, **119**, 6–13.

Received on 18 July 1986