Primary Structure of a Putative Receptor for a Ligand of the Insulin Family*

(Received for publication, June 21, 1989)

Peter Shier$ and Valerie M. Watt

From the Department of Physiology, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Nucleotide sequence analysis of human and guinea pig genomic DNA encoding a new member of the insulin receptor (IR) family revealed that the predicted primary structure of this IR-related protein is as similar to the IR and insulin-like growth factor (IGF) I receptor as the IR and IGF-IR are to each other. The conservation of this IR-related sequence among mammals and with the IR and IGF-IR suggests that this IR-related protein is a novel receptor for insulin, IGF-I, IGF-II, or an as yet unidentified peptide hormone or growth factor belonging to the insulin family.

Although insulin and one structurally related insulin-like growth factor (IGF), IGF-I, are known to be primary regulators of metabolism and growth, the role of IGF-II remains unclear (1-5). Precise determination of the physiological functions of each of these peptides is complicated, however, by the ability of three known receptors to bind these peptides with overlapping specificities. Two of these receptors, the insulin receptor (IR) and the IGF-I or type I IGF receptor, belong to a family that also includes several more distantly related tyrosine kinases: the protein products of tkh and of the proto-oncogenes c-ros, met, and trk (4, 5). The IR and IGF-IR exhibit a high degree of overall similarity (6-8); extracellular a-subunits containing the ligand-binding region are disulfide-bonded to b-subunits which span the membrane and contain a cytoplasmic tyrosine kinase activated by ligand binding. Surprisingly, the IR or IGF II receptor lacks any homology to the IR and IGF-IR as well as any known mechanism of signal transduction (9, 10). This IGF-II receptor is the same molecule as the cation-independent mannose 6-phosphate (Man-6-P) receptor involved in lysosomal enzyme targeting (11) and shares antigenic properties with an IGF-binding protein present in serum (9). This IGF-II receptor is unlikely to mediate IGF-II action, at least in the chicken; the avian Man-6-P receptor lacks the high affinity binding site for IGF.

* This work was supported in part by Medical Research Council of Canada Grant MA8786. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J05046 and J05047.

$ Funded by grants from the Quebec Diabetes Association and Canadian Diabetes Association.

1 The abbreviations used are: IGF, insulin-like growth factor; IGF-IR, insulin-like growth factor I receptor; IGF-II, insulin-like growth factor II receptor; IR, insulin receptor; IRR, insulin receptor-related receptor; kb, kilobase(s); Man-6-P, cation-independent mannose 6-phosphate.

II (12). IGF-II also binds with high affinity to the receptor encoded by the human IGF-IR cDNA, although in most reports IGF-II has been shown to bind the type 1 IGFRI with lower affinity than IGF-I (13). These results raise the possibility that there is another receptor in the IR family whose binding specificity differs from those of IR and IGF-IR (3, 13). Previous reports of anomalous binding of insulin and the IGFs have also been explained by hypothesizing the existence of additional receptors belonging to the IR family (14-19). In this study, we have defined the primary sequence of a new receptor structurally similar to the IR and IGF-IR that it suggests this insulin receptor-related receptor (IRR) mediates the action of a ligand(s) that is identical or very similar to insulin, IGF-I, or IGF-II.

EXPERIMENTAL PROCEDURES

Southern Blot Analysis—Genomic DNA (10 μg) was digested with EcoRI and separated on a 1% agarose gel before transfer to nitrocel-lulose (20, 21). Nitrocellulose filters were hybridized in 30% formamide at 42 °C (22) with DNA probes labeled with [32P]dCTP (23) and then washed in 15 mM sodium chloride, 1.5 mM sodium citrate, and 0.1% sodium dodecyl sulfate at 42 °C. Library Screening, Cloning, and Sequencing—Part of the human IR gene was isolated by screening a library of genomic DNA (24) using synthetic oligonucleotides based on the human IR cDNA sequence (241-260 and 261-280 base pairs (6)). DNA encoding part of exon 2 from the human IR (an Alul/SrdI fragment encoding L25-L125) was used as a probe to isolate IRR genomic DNA from a library of guinea pig genomic DNA. A guinea pig IRR DNA probe encoding part of exon 2 (PstI fragment encoding L29-Q133, Fig. 3) was used to isolate a phage containing all of the human IRR gene except exon 1 from a Xcharon 4A library (24). Insert DNAs from both human and guinea pig genes were subcloned into vectors pEMBL18 or pEMBL19 (25); deletions were created by restriction enzyme digestion; and single stranded DNA was sequenced by the dideoxy method using Klenow (26) and resequenced using modified T7 DNA polymerase and dITP to resolve ambiguities (27).

RESULTS AND DISCUSSION

To determine if IR DNA could be used to identify previously unknown receptors of the IR family, Southern blots of mammalian genomic DNA were probed with DNA encoding part of the extracellular ligand-binding region of the human IR. DNA from exon 2 which encodes an N-terminal region of the α-subunit (28) hybridized at reduced stringency to multiple fragments in guinea pig, human, and rat genomic DNA (Fig. 1A). The single fragment in each species which encodes the IR was identified by washing at 60 °C to eliminate hybridization to less related sequences (data not shown). The size estimated for this EcoRI fragment in human DNA (∼18 kb, Fig. 1A) is similar to that of an EcoRI fragment containing exon 2 of the human IR isolated by Seino et al. (28). One other hybridizing fragment in each genome (Fig. 1A) probably encodes the IGF-IR since there is 75% identity between DNA encoding the human IR and IGF-IR in the region used as probe (6-8). At least one hybridizing fragment in each genome, however, could not be attributed to these known genes.

Since analysis of genomic DNA by Southern blotting suggested that IR DNA could identify novel, homologous DNA sequences, we used the human IR DNA as probe to isolate IR-related sequences from a guinea pig genomic library. The region of this guinea pig DNA that was homologous to the human IR probe selectively detected EcoRI fragments in
guinea pig, human, and rat genomic DNAs (Fig. 1B) that had hybridized only weakly to the human IR probe (Fig. 1A). We used this IRR DNA from guinea pig to isolate the analogous human IRR gene. The novel identity of this IRR DNA was confirmed by localizing the human gene to chromosome 1. In contrast, genes encoding other members of the IR family are known to be located on human chromosomes 6, 7, 15, and 19 (c-ros, met, IGF-IR, and IR, respectively (8, 29-31)).

Nucleotide sequence analysis of the guinea pig and human genomic DNA encoding the IRR revealed an overall gene organization similar to that of the human IR. The guinea pig IRR gene consists of 22 exons (Fig. 2). All of these exons, with the exception of exon 1 which encodes the signal peptide and the first three amino acids of the α-subunit, were also present in our human IRR isolate (Fig. 2). This intron/exon organization of both IRR genes is identical to that of the human IR gene (28). Indeed, of the 42 identifiable intron/exon boundaries in the IRR and IR genes, 36 are at analogous positions and the remaining 6 accommodate deletions or insertions of no more than 5 amino acids. Although the relative sizes of introns within each gene are similar, the IRR genes in both species are approximately 10-fold smaller (≈15 kb in the guinea pig; ≈12 kb without the signal exon in the human) than the human IR gene (>120 kb (28)). The identical exon organization of both IRR genes (Fig. 2) and the human IR gene (28) revealed that the IRR gene does not exhibit the common pseudogene structure arising from reverse transcription of a processed RNA (32). Although we have not yet detected IRR transcripts using Northern analysis of poly(A) RNA from multiple tissues and cell lines including placenta, liver, muscle, and brain, the absence of in-frame stop codons in all human and guinea pig IRR exons even though the amino acid sequence of the IRR has diverged significantly from those of the IR and IGF-IR (Fig. 3) also indicates that the IRR gene is not a pseudogene.

Alignment of the predicted amino acid sequences of the human and guinea pig IRR with those of the human IR and IGF-IR revealed colinear organization of these preproreceptors over their entire length (Fig. 3). The processing of the IRR to its mature form, therefore, is likely similar to that of the IR and IGF-IR. A 26-residue sequence beginning at a consensus initiator methionine (GCCATGG compared to consensus ACCATGG (33)) of guinea pig prepro-IRR (Fig. 3) indicates that the IRR precursor which contains 1271 amino acids with a calculated M₉ = 141,000 would result in an α-subunit of 716 amino acids with M₉ = 79,000 and a β-subunit of 551 amino acids with M₉ = 61,000. The predicted subunit sizes of the IR and IGF-IR are similar: 84 and 80 kDa for the α-subunits; 70 and 71 kDa for the β-subunits, respectively (6, 8). Within the extracellular regions, 8 of 10 potential N-linked glycosylation sites conserved between the guinea pig and human IRR are also present in both the human IR and IGF-IR (Fig. 3). The IRR may therefore exhibit a high level of glycosylation similar to that of the IR and IGF-IR (6-8).

As in the IR and IGF-IR, the α- and β-subunits of the IRR are also likely to be disulfide-linked forming a biologically active α₂β₄ heterotetrameric receptor complex. The Cys residues suggested to be involved in this linkage of α- and β-subunits of the IR and IGF-IR (8) are also present in the IRR between Cys₁⁰⁸ and Cys₁⁴⁸ (Fig. 3). These α/β-subunit heterotetramers could anchor in the membrane via a single 22-amino acid hydrophobic sequence (Val¹⁰⁶ to Phe¹¹⁷) of the β-

---

**Fig. 1.** Genes encoding members of the IR family detected by Southern blot analysis of mammalian genomic DNA. Genomic DNA isolated from guinea pig (lane 1), human (lane 2), or Sprague-Dawley rat (lane 3) was hybridized with part of exon 2 from human IR genomic DNA (A) or guinea pig IRR genomic DNA (B). DNA fragments encoding the IRR and IR are indicated by open and closed triangles, respectively.

**Fig. 2.** Partial restriction maps of genomic DNA encoding the human and guinea pig IRR. Exons are indicated by black boxes. Exon boundaries were determined by the presence of consensus splice donor and acceptor sites at identical positions in both species and comparison of the primary translation product of the IRR genes with the human IR and IGF-IR amino acid sequences. Arrows in the predicted preproreceptor indicate the putative signal peptide and precursor processing sites; $\alpha$, the signal peptide; $\beta$, the Cys-rich domain; $\gamma$, the transmembrane domain; and $\delta$, the tyrosine kinase domain. B, BamHI; E, EcoRI; S, SstI.
Putative Receptor for a Ligand of the Insulin Family

Fig. 3. Predicted amino acid sequences of human and guinea pig IRR aligned with those of closely related members of the IR family. Open boxes indicate residues of human IRR (hIRR) identical to those of human IR (hIR), human IGF-IR (hIGF-IR), or human c-ros protein (hc-ros). Only differences between guinea pig IRR (gpIRR) and human IRR are shown. Numbers indicate amino acid residues of each proreceptor and for human IRR were assigned assuming that the unidentified initial amino acids of the a-subunit are identical in number to those in the guinea pig IRR. Cys residues are indicated with black boxes; Tyr residues in the cytoplasmic region are bold. The signal peptide and transmembrane regions are underlined with solid lines.

Consensus N-glycosylation sites (NXS/T) of human IRR are outlined with solid lines; consensus sequences for tyrosine kinase activity with dashed lines. Intron positions are indicated by triangles; the potential ATP binding site (GXGXXG and K) by asterisks.

Characteristics of ligand binding and/or signal transduction may also be determined by a small region encoded by exon 11 immediately N-terminal to the putative αβ cleavage site of the IRR (Fig. 3). This exon is differentially spliced from human IR RNA in specific tissues and is absent from the placental human IGF-IR sequence. Signal transduction following ligand binding will lead to activation of the tyrosine kinase present in the cytoplasmic region of the β-subunit. Structural features characteristic of tyrosine kinases are located between amino acid residues Leu1160 and Thr1164 and include a potential ATP binding site (Gly-X-Gly-X-X-Gly and Lys1169) as well as all other residues conserved among kinases. Tyrosine specificity of this putative kinase is supported by the presence in the IRR of the two protein tyrosine kinase consensus sequences (Asp1099 to Asn1104, Pro1120 to Glu1126) (4). Within the tyrosine kinase domains, the four Cys residues and six of seven Tyr residues conserved among kinases (4). Tyrosine kinase activity is also present in the IRR. Among these are the autophosphorylation sites implicated in the regulation of IR kinase activity (Tyr1162, 1163) in human IR.
binding characteristics for insulin or IGF-I. Alternatively, the IRR may mediate the biological action of IGF-II while the novel IGF-IR/MAN-6-P receptor acts only as a binding protein without the ability to function in signal transduction. This DNA encoding the IRR may thus facilitate determination of the physiological role of IGF-II.

Acknowledgments—We thank J. Spencer for the gift of the guinea pig genomic library; M. Shales for computer assistance; S. Runciman for technical assistance; and C. J. Ingles, J. E. Kudlow, A. Pawson, J. Segall, H. F. Willard, and C. C. Yip for helpful discussion and/or critical review of the manuscript.

REFERENCES
1. Froesch, E. R., Schmid, C., Schwander, J., and Zapf, J. (1986) Annu. Rev. Biochem. 55, 47-449.
2. Rechler, M. M., and Niesley, S. P. (1985) Annu. Rev. Physiol. 47, 425-442.
3. Rosenthal, M. (1988) Science 239, 1269-1271.
4. Hanks, S. K., Quinn, A. M., and Hunter, T. (1988) Science 241, 42-52.
5. Han-Neriah, Y., and Baskin, A. R. (1988) Nature 333, 672-676.
6. Kaidbey, A., Ellis, J., Jarmagak, K., Edery, M., Grass, L., Clausen, E., Ou, J., Masiarz, F., Kan, Y. W., Goldfine, I. D., Roth, R. A., and Rutter, W. J. (1985) Cell 40, 747-758.
7. Ulrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzelli, L. M., Dull, T. J., Gray, A., Cossens, L., Liao, Y.-C., Tsukobowa, M., Mason, A. E., Seeburg, P. H., Grunfeld, C., Rosen, O. M., and Ramachandran, J. (1985) Nature 313, 705-707.
8. Ullrich, A., Gray, A., Tam, A. W., Yang-Feng, T., Tsukobowa, M., Collins, C., Kaelin, W., Le, Q. T., and Cooper, C. A. (1985) Nature 319, 1134-1136.
9. Uhle, N., Alon, C., M., Kyle, J. W., Grubb, J. H., and Sly, W. S. (1988) J. Biol. Chem. 263, 2553-2562.
10. Canfield, W. M., and Knorfield, S. (1985) J. Biol. Chem. 260, 7100-7103.
11. Short, J. E., Turner, J., Edman, J. C., Hari, J., Pierce, S. B., Stover, C., Rutter, W. J., and Roth, R. A. (1966) J. Biol. Chem. 261, 11466-11492.
12. Spath, V. R., Hall, K., Missaki, M., Frykland, L., Christiansen, N., and Wetterberg, L. (1987) J. Clin. Invest. 71, 1094-1014.
13. Jonas, H. A., and Harrison, L. C. (1986) J. Biol. Chem. 261, 2392-2394.
14. Jonas, H. A., Cox, A. J., and Harrison, L. C. (1989) Biochem. J. 257, 101-107.
15. Meier, P., Hintz, R. L., and Rosenfeld, R. G. (1986) J. Clin. Endocrinol. Metab. 63, 1400-1405.
16. Morgan, D. O., and Roth, R. A. (1986) Biochem. Biophys. Res. Commun. 134, 1541-1547.
17. Ito, K., and Kato, M. (1986) Biochem. Biophys. Res. Commun. 134, 1541-1547.
18. Burant, C. F., Treutelaar, M. K., Allen, K. D., Sens, D. A., and Buse, M. G. (1987) Biochem. Biophys. Res. Commun. 147, 100-107.
19. Wigler, M., Sweet, R., Grass, G. K., Wold, B., Pelicic, A., Lacy, E., Maniatis, T., Silverstein, S., and Axel, R. (1979) Cell 16, 777-785.
20. Southern, E. M. (1970) J. Mol. Biol. 38, 530-539.
21. Southern, E. M., Stern, S. R., and Marks, A. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3683-3687.
22. Roder, J. W. I., and Kunkel, M., Rhodes, R., and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
23. Warrant, R. M., Fritsch, E. F., Parker, R. C., and Maniatis, T. (1987) J. Cell Biol. 105, 1107-1114.
24. Allison, L. A., Moyle, M., Shales, M., and Ingles, C. J. (1985) Cell 42, 599-610.
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
26. Tabak, R. H., and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 4767-4771.
27. Seino, S., and Nishizawa, S. (1985) Proc. Natl. Acad. Sci. U.S.A. 86, 111-114.
28. Nagaoka, L., and Sugiyama, T. (1986) Cell 45, 721-722.
29. von Honege, G. (1965) J. Mol. Biol. 18, 99-105.
30. Yip, C. C., Hsu, H., Patel, R. G., Hawley, D. M., and Goldfine, I. D. (1986) Biochem. Biophys. Res. Commun. 157, 321-329.
31. Seino, S., and Bell, G. I. (1989) Biochem. Biophys. Res. Commun. 159, 312-316.
32. Ellis, L., Clausen, E., Morgan, D. O., Edery, M., Roth, R. A., and Rutter, W. J. (1986) Cell 45, 721-722.
33. Short, J. E., and Yip, C. C. (1989) EMBO J. 8, 435-440.
34. Yarden, Y., and Ullrich, A. (1988) Annu. Rev. Biochem. 57, 443-478.
35. Matsushima, H., Wang, L., and Shibuya, M. (1986) Mol. Cell. Biol. 6, 3000-3004.