A third fibronectin type III domain in the extracellular region of the insulin receptor family

Cristina Marino-Buslje[a,b], Kenji Mizuguchi[a], Kenneth Siddle[b], Tom L. Blundell[a,*]

[a] *Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK*
[b] *Department of Clinical Biochemistry, University of Cambridge, Addenbrookes Hospital, Cambridge CB2 2QR, UK*

Received 5 November 1998

Abstract The insulin receptor family consists of the homologous tyrosine kinase receptors, insulin receptor (IR), insulin-like growth factor 1 receptor (IGF1R) and insulin receptor-related receptor. The three-dimensional structures of the tyrosine kinase domain of the IR and the first three extracellular domains (L1, Cys-rich and L2) of the IGF1R are known. Here we present evidence that the connecting domain of the IR family is a member of the fibronectin type III (FnIII) superfamily. Structure-based alignment of FnIII domains reveals several key residues that are also conserved in the sequence of the connecting domain. The alignment of the connecting domain with FnIII domains is in good agreement with secondary structure prediction. A model of the connecting domain shows a hydrophobic core formed by the conserved residues and is consistent with previously known biochemical data. This suggests that the IR family contains three FnIII domains in tandem in the extracellular juxtamembrane region.

Key words: Insulin receptor; Fibronectin type III; Signal transduction; Comparative modeling

1. Introduction

The insulin receptor family comprises the homologous insulin receptor (IR), insulin-like growth factor 1 receptor (IGF1R) and insulin receptor-related receptor (IRR). These integral membrane glycoproteins are each synthesised as a single polypeptide chain and then proteolytically processed to produce two chains, α and β, which are linked by a disulphide bond. The active receptor is a dimer α2β2, bound by two disulphide bonds [1]. The α-chain is extracellular, whereas the β-chain consists of a short extracellular region, a single transmembrane segment and an intracellular tyrosine kinase domain involved in signal transduction. The ligand binding regions are contained in the extracellular α-chain. Ligand binding affinity and specificity are central to receptor activation, regulation and function.

The extracellular region consists of six domains. The three-dimensional (3D) structure of the first three domains, L1, Cys-rich and L2, of the IGF1R has been defined by X-ray analysis [2,3], as well as the structure of the intracellular tyrosine kinase domain of IR [4]. Little is known, however, about the structure of the rest of the extracellular region, where three other domains have been identified: a connecting domain (residues 468–588) and two repeats of fibronectin type III domains (FnIII) (residues 593–820 and 821–929; residue numbers are those of human IR throughout this paper) [5,6].

In spite of the extensive similarity in amino acid sequence, domain structure and signalling mechanism between the receptors, the residues involved in ligand binding are different in IR and IGF1R. Affinity labelling experiments and studies with chimeric IRs indicate that at least three separate regions in the α-chain of IR (residues 1–120, 459–524 and 713–718) are involved in insulin binding [7,8]. More recent studies, using alanine scanning mutagenesis and analogue binding properties, have shown that the N-terminal 120 amino acids and residues 704–715 of the α-chain play a major role [9,10], although the fragment consisting of residues 1–462 of the IGF1R does not bind ligand on its own [3].

The region of the connecting domain can be deleted without seriously compromising insulin binding [10–12], but residues 450–601 appear to be a major immunogenic determinant for inhibitory monoclonal antibodies as well as patients' autoimmune anti-IR antibodies. Therefore, this region was called 'major immunogenic region' (MIR) [13]. A blunted tyrosine autophosphorylation response in deletion studies of this region [11] suggests that it may be involved in the transmission of the signal, after insulin interaction, to the tyrosine kinase.

Here we suggest that the connecting domain is another FnIII repeat and that the IR family, therefore, has three consecutive FnIII domains. We call this novel FnIII domain FnIII0, because the next two domains have already been designated FnIII1 and FnIII2.

2. Materials and methods

The PSI-Blast [14] search was performed on the NCBI WWW server. The secondary structure prediction with PHD [15–17] and Jpred [18] was carried out on the EMBL and EBI WWW servers respectively. Other secondary structure prediction programs Spetor [19,20] and Predator [21,22] were run locally.

The list of FnIII domains with known structures was taken from SCOP [23] and the structure-based alignment was obtained using COMPARER [24,25] followed by manual adjustment. The sequences of the connecting domains of the IR family were added by CLUSTALX [26]. The final alignment was manually adjusted.

Two known structures of FnIII domains were chosen as the template for the modelling of IR and IGF1R connecting domain. These were domain two of the growth hormone receptor (PDB code 3hhr) and domain eight of fibronectin (PDB code 1fnf). The model was built using MODELLER [27]. Evaluation of the model was carried out with Verify 3D [28] and Procheck [29]. The cycle of realignment, modelling and structure evaluation was repeated until no further improvements on the structure were observed. Similarly, the model of the three-domain structure (FnIII0, FnIII1 and FnIII2) of IGF1R was built with MODELLER using domains seven, eight and nine of fibronectin (PDB code 1fnf), domain two of neuroglian (PDB code 1cfb) and domain two of growth hormone receptor (PDB code 3hhr).

3. Results and discussion

The high degree of sequence similarity between IR, IGF1R

*Corresponding author. Fax: (44) (1223) 766082.
E-mail: tom@cryst.bioc.cam.ac.uk

PII: S0014-5793(98)01509-9

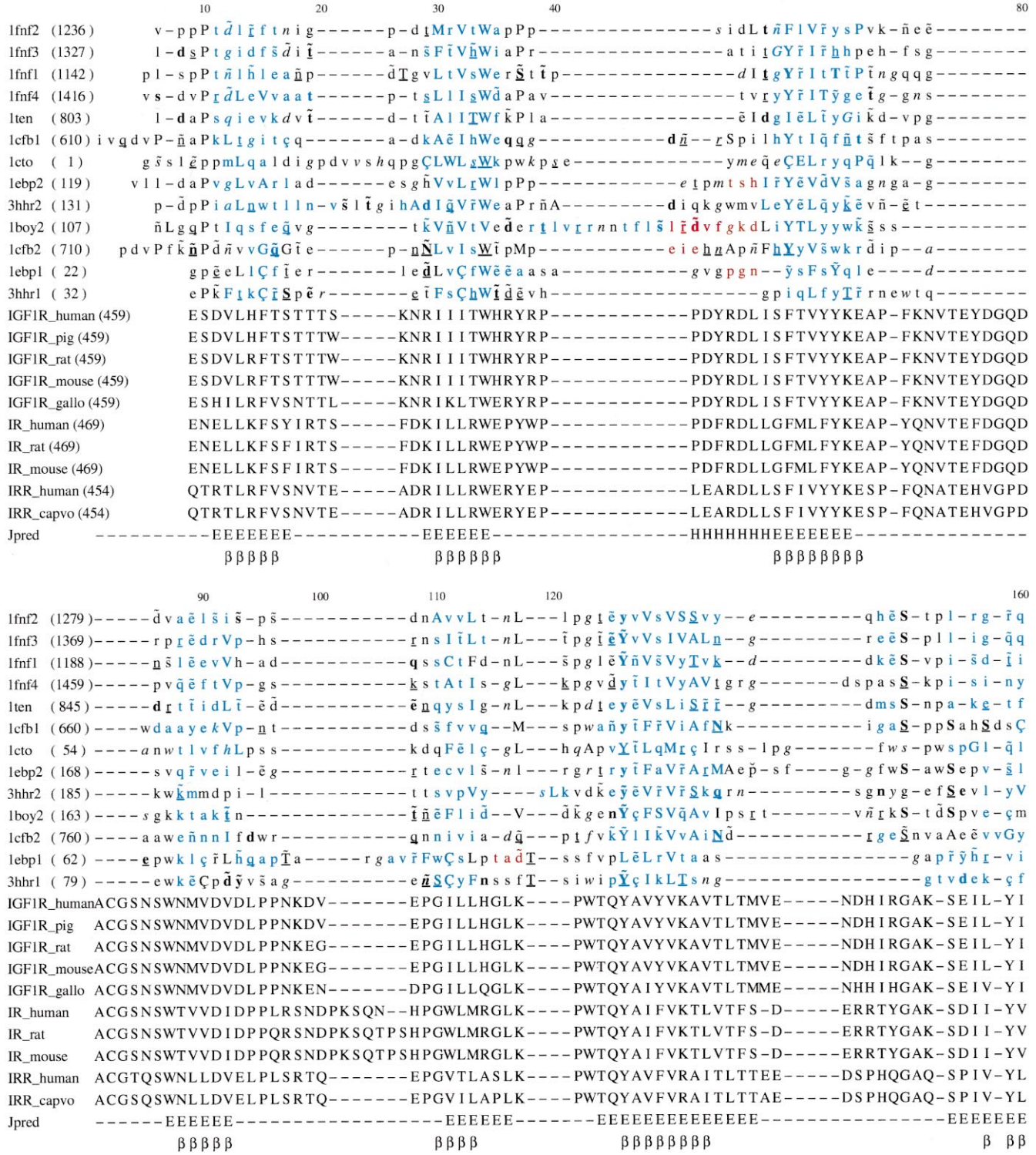

Fig. 1. Sequence alignment of FnIII domains with the sequences of the connecting domains of the IR family. Structure-based alignment of FnIII domains were generated with COMPARER [24,25] and formatted by JOY [40]. Aligned proteins are: 1fnf1, 1fnf2, 1fnf3 and 1fnf4: fibronectin, domains seven, eight, nine and ten (PDB code 1fnf); 1ten: tenascin (PDB code 1ten); 1cfb1 and 1cfb2: neuroglian, domains one and two (PDB code 1cfb); 1cto: granulocyte colony stimulating factor receptor (g-csf-r) (PDB code 1cto); 1ebp1 and 1ebp2: erythropoietin receptor (epo-r) domains one and two (PDB code 1ebp); 3hhr1 and 3hhr2: growth hormone receptor (gh receptor) domains one and two (PDB code 3hhr); IGF1R_human: human IGF1R; (Swiss-Prot accession P08069); IGF1R_pig: pig IGF1R (Q29000; Q28951); IGF1R_rat: rat IGF1R (P24062); IGF1R_mouse: mouse IGF1R (Q60751; Q62123); IGF1R_gallo: *Gallus gallus* IGF1R (TrEMBL P79773); IR_human: human IR (P06213); IR_rat: rat IR (P15127; P9768); IR_mouse: mouse IR (P15208); IRR_human: human IRR (P14616); IRR_capvo: *Cavia porcellus* IRR (P14617). Residue numbers are shown in parentheses. The bottom line of the alignment (Jpred) shows the secondary structure prediction result by Jpred. E: β-strand; H: α-helix.

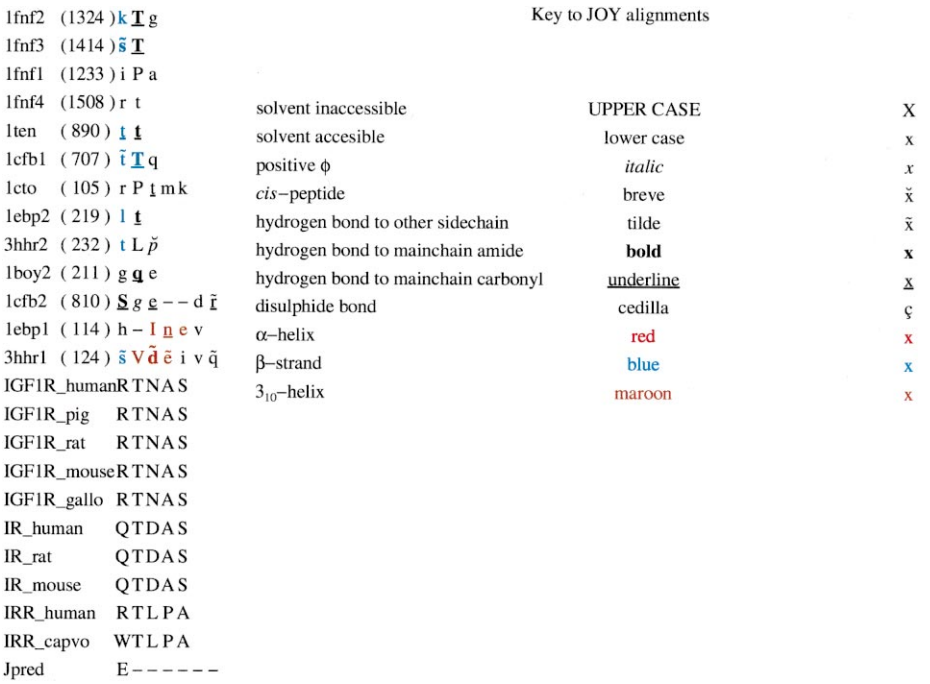

Fig. 1 (*continued*).

and IRR indicated that all three will have similar structures. In the following we will mainly discuss IR, but similar arguments apply to the other members.

A PSI-Blast search [14] for the connecting domain of the three proteins did not give any significant hit. We then ran several secondary structure prediction programs, including PHD [15–17], Spetor [19,20], Predator [21,22] and Jpred [18], all of which strongly indicated that there are seven β-strands in the domain. The two highest scoring proteins predicted by Threader [30] were also both β-proteins; these were immunoglobulin (Ig) and fibronectin. After a careful examination of the structure-based alignments, the sequence of the connecting domain showed significantly higher compatibility with the fibronectin domain than with the Ig fold (see below).

FnIII domains all adopt a conserved β-sandwich with seven strands (see Fig. 2C), despite the low sequence identity (between 8 and 20%) among the members of the family. Thus, structure-based alignment is essential to characterise the sequence conservation pattern within this family. We used COMPARER [24,25], in conjunction with some manual adjustments, to obtain a structure-based alignment of FnIII domains (Fig. 1).

Residues highly conserved in the FnIII domain are: Pro in the A-strand, Trp in the B-strand, Tyr in the C-strand, Tyr in the F-strand and Leu six residues N-terminal of this Tyr, which is located in the E-F loop [31]. Most of these residues are buried in the structure and form a hydrophobic core. There are other conserved buried hydrophobic residues in the FnIII fold, suggesting that they might have an important role in stabilising the hydrophobic core of the domain. These residues are located two and four residues N-terminal of the conserved Trp in the B-strand, two residues C-terminal of the conserved Tyr in the C-strand and two and four residues C-terminal of the conserved Tyr in the F-strand (Fig. 1).

We added the sequences of the connecting domain to the structure-based alignment of FnIII. It can be seen that most of the key residues are also conserved in the connecting domain of the IR family sequences (Figs. 1 and 2A), except for Pro in the A-strand. Trp (residue 489) in the B-strand is conserved. Phe-503, instead of Tyr, is in the C-strand. This is a conservative substitution and Phe is also present in some of the FnIII domains with known structure. Tyr-562 in the F-strand is conserved, as well as Leu-556 in the loop between strands F and E. Other buried hydrophobic residues important in the FnIII fold are present in the sequence of the connecting domain, e.g. Leu-487 in the B-strand, Leu-505 in the C-strand, Ile-564 and Val-566 in the F-strand.

The alignment is also in good agreement with the secondary structure prediction. As an example, the secondary structure prediction by Jpred [18] is shown in Fig. 1. Note that our input to the secondary structure prediction programs was either a single sequence or a multiple alignment of the connecting domains of the IR family. No program detected any other sequence similarity and therefore, information that would have been gained from the multiple alignment of sequences with those of FnIII domains was never used in these secondary structure predictions.

Although both FnIII and Ig domains adopt a seven-stranded β-sandwich, key residues are not conserved between them and they are regarded as different superfamilies [23]. A careful examination of alignments of the connecting domain with several Ig domains was carried out. Alignments of the sequences of the connecting domain with C1 set -constant immunoglobulin-, C1 set -constant non immunoglobulin-, V set -variable non immunoglobulin-, I set and the C2 set, were made and analysed using the structure-based alignments in the HOMSTRAD database [32]. The conserved residues of Ig, as well as the sequence separation of the main residues in the characteristic 'pin' structure of Ig-like molecules [33], were not present in the connecting domain.

We built a model of the IR and IGF1R connecting domain. Two known structures that show the highest sequence identity

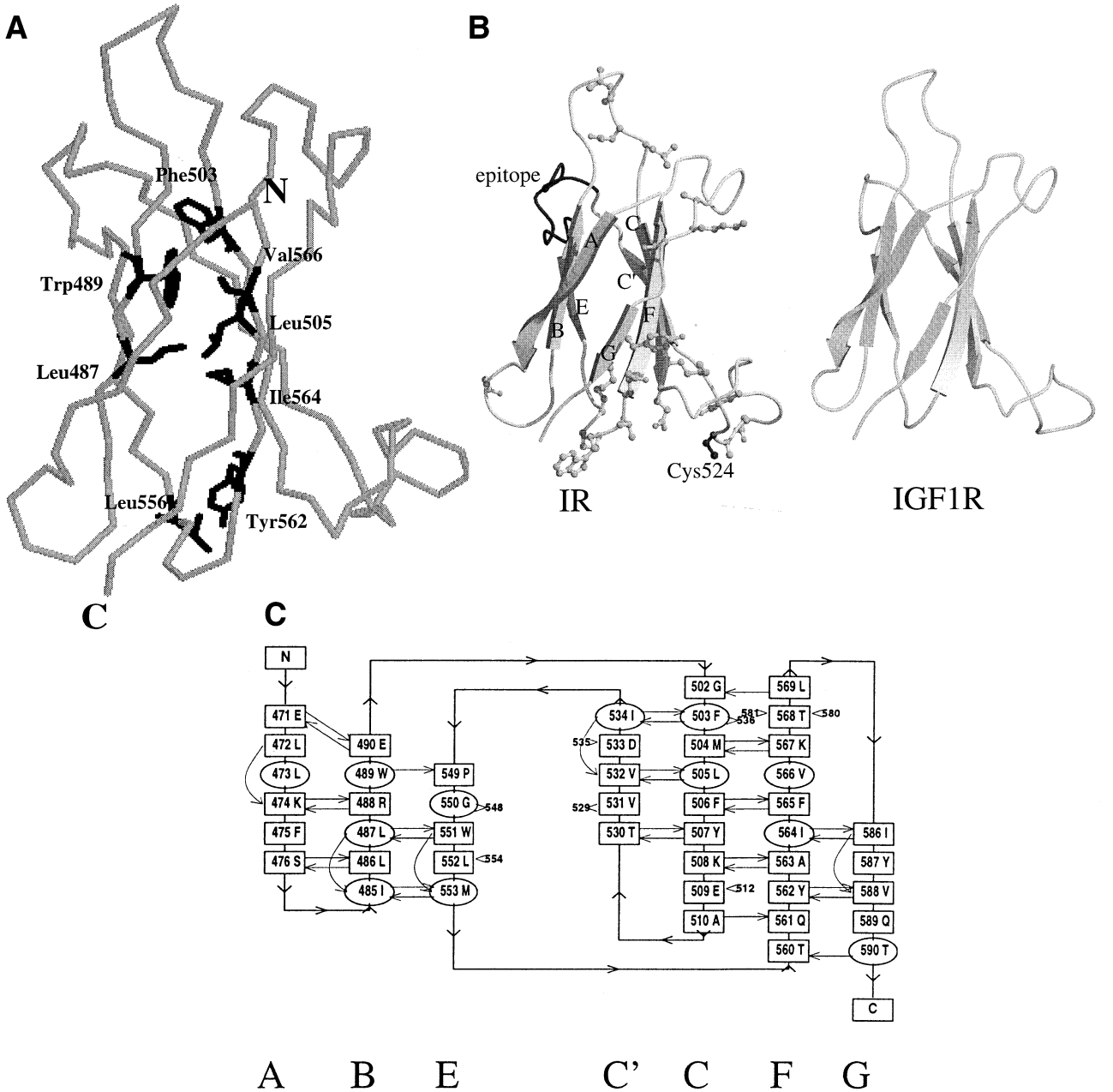

Fig. 2. A: Backbone representation of the model of the FnIII0 domain of IR. Shown in black are the side chains of the conserved residues in the FnIII fold. Figure generated by RASMOL [41]. B: Cartoon representation of the models of the FnIII0 domain of IR and IGF1R. In the model of IR, Cys-524 forming the α-α disulphide bond is labelled. Residues in the principal immunogenic loop are shown in black. Residues, conserved in the IR family and not in the FnIII family, are shown in ball-and-stick representation. Figure generated with MOLSCRIPT [42] and RASTER3D [43,44]. C: Schematic representation of the hydrogen bond patterns in the model of the FnIII0 domain of IR. Only residues located in a β-strand are shown in a circle (buried) or in a square box (exposed). Figure generated by HERA [45].

with the target sequences were chosen as the template for modelling. These structures were domain two of the growth hormone receptor (Protein Data Bank (PDB [34,35]) code, 3hhr) and domain eight of fibronectin (PDB code 1fnf). Their sequence identities with the target were 25 and 20% respectively.

The model is shown in Fig. 2. In the model, all the conserved residues point inwards, making a tightly packed hydrophobic core. Evaluation with Verify 3D [28] produced all positive scores. Evaluation of the stereo-chemical quality of the model with Procheck [29] showed that only a few residues (8.4% for IR and 1.9% for IGF1R) are in disallowed regions in the Ramachandran plot.

Eight monoclonal antibodies that inhibit insulin binding to its receptor were found to bind to an epitope in residues 450–601, within FnIII0 of the human IR [13]. The finding that none recognises the mouse IR helped narrow the epitopes recognised by these antibodies. In this region all the differences in amino acid sequence between mouse and human receptors (except for a single substitution of Phe by Tyr in residue 476) are located between residues 535 and 548. These 13 amino acids are therefore likely to constitute an epitope

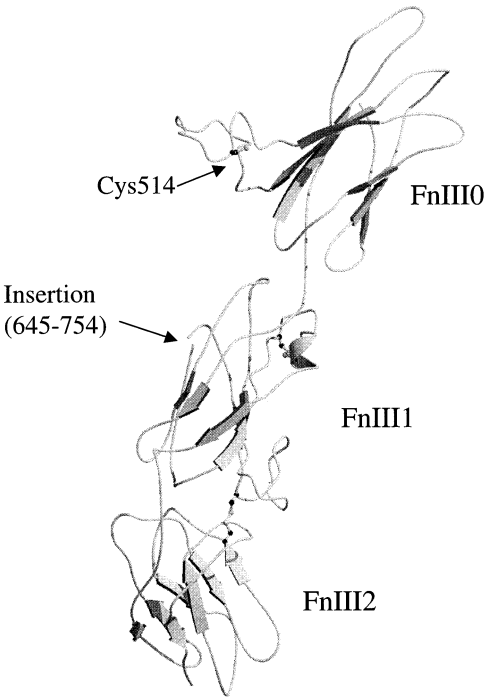

Fig. 3. Model of the three domains FnIII0, FnIII1 and FnIII2 of IGF1R. Figure generated with MOLSCRIPT and RASTER3D.

recognised by the eight antibodies [13]. This is consistent with our model, where these residues are located in the loop between strands C′ and E (Fig. 2B).

Residues conserved among the members of the IR family should play structural and/or functional roles. Structural key residues should be also conserved in the FnIII family, assuming that they adopt a similar fold. Thus, the ones conserved only within the IR family, but not the FnIII family, are likely to be involved in the activity. These residues, displayed as ball-and-stick in Fig. 2B, are indeed mainly polar and located on the surface. Furthermore, many of these residues cluster in two regions. First, the loops C-C′ and E-F contain highly conserved residues (Glu-517 to Gly-525 and Lys-557 to Ala-563 respectively), which cluster on the same face of the β-sandwich. The loop C-C′ includes Cys-524 that forms the disulphide bridge between the protomers and therefore this cluster is likely to be involved in protomer-protomer interactions. This is consistent with the putative location of the epitope on the opposite side of the fold (Fig. 2B). The identification of the protomer-protomer interaction region can help in the design of mutations, e.g. introducing another Cys to make the dimer more stable. The second cluster is found on the loop B-C (Arg-498 to Leu-500) and this might be involved in the interactions with the first three domains of the extracellular region.

There are some functional similarities between the connecting domain and other FnIII domains. FnIII domains often exist proximal to the membrane [36,37], suggesting that they may play a role in dimerisation, e.g. binding to similar FnIII modules [38]. That is also the case of the connecting domain in the IR family where the α-α disulphide bridge (residue 524 in IR) plays an important role in dimerisation.

Our finding that the connecting domain belongs to the FnIII superfamily, indicates that the extracellular region of the IR family contains three FnIII domains in the extracellular juxtamembrane region. A model of the three FnIII domains of IGF1R is shown in Fig. 3. Although there is not much information about the relative orientation of the three domains, a single disulphide bond between FnIII1 and FnIII2 appears to be a strong constraint. Fig. 3 shows that the three domains can be arranged linearly, an arrangement of FnIII domains that has been previously observed in the crystal structure of fibronectin [39].

4. Conclusions

Several lines of evidence strongly suggest that the connecting domain in the IR family is a FnIII domain.

Very little is known about the last three domains of the extracellular region and a better knowledge might be useful in understanding the mechanism of how IGFI, IGFII and insulin transmit signals through the receptor to produce their metabolic and mitogenic effect. Identification of particular residues and knowledge of their possible position in the 3D structure may help design further mutagenesis studies, such as making an additional disulphide bond to stabilise the α2β2 dimer. Our model of the 3D structure provides detailed hypothesis about domain linkers and this may assist crystallisation studies. The model may also help engineer a peptide to use in enzyme-linked immunoassays for screening for anti-receptor antibodies.

Acknowledgements: Cristina Marino-Buslje is a European Community fellow. Kenji Mizuguchi thanks HFSP and SmithKline Beecham for financial support. Ken Siddle thanks the British Diabetic Association for project grant support RD97/0001536.

References

[1] Sparrow, L.G., McKern, N.M., Gorman, J.J., Strike, P.M., Robinson, C.P., Bentley, J.D. and Ward, C.W. (1997) J. Biol. Chem. 272, 29460–29467.
[2] McKern, N.M., Meizhen, L., Frenkel, M.J., Verkuylen, A., Bentley, J.D., Lovrecz, G., Ivanic, N., Elleman, T.C., Garret, T.P.J., Cosgrove, L.J. and Ward, C.W. (1997) Protein Sci. 6, 2663–2666.
[3] Garret, T.P.J., McKern, N.M., Meizhen, L., Frenke, M.J., Bentley, J.D., Lovrecz, G.O., Elleman, T.C., Cosgrove, L.J. and Ward, C.W. (1998) Nature 394, 395–399.
[4] Hubbard, S.R., Wei, L., Ellis, L. and Hendrickson, W.A. (1994) Nature 372, 746–754.
[5] O'Brian, J.P., Frye, R.A., Cogswell, P.C., Neubauer, A., Kitch, B., Prokop, C., Espinosa III, R., Le Beau, M.M., Earp, H.S. and Liu, E.T. (1991) Mol. Cell. Biol. 11, 5016–5031.
[6] Schafer, E.M., Erickson, H.P., Federwisch, M., Wollmer, A. and Ellis, L. (1992) J. Biol. Chem. 267, 23393–23402.
[7] Schumacher, R., Soos, M.A., Schlessinger, J., Brandenburg, D., Siddle, K. and Ullrich, A. (1993) J. Biol. Chem. 268, 1087–1094.
[8] Williams, P.F., Mynarcik, D.C., Yu, G.Q. and Whittaker, J. (1995) J. Biol. Chem. 270, 3012–3016.
[9] Mynarcik, D.C., Williams, P.F., Schaffer, L., Qin Yu, G. and Whittaker, J. (1997) J. Biol. Chem. 272, 2077–2081.
[10] Kristensen, C., Wiberg, F.C., Schaffer, L. and Andersen, A.A. (1998) J. Biol. Chem. 273, 17780–17786.
[11] Sung, C.K., Wong, K.Y., Yip, C.C., Hawley, D.M. and Goldfine, I.D. (1994) Mol. Endocrinol. 8, 315–324.
[12] Kadowaki, H., Kadowaki, T., Cama, A., Markus-Samuels, B., Rovira, A., Bevins, C.L. and Taylor, S.I. (1990) J. Biol. Chem. 265, 21285–21296.
[13] Zhang, B. and Roth, R.A. (1991) Biochemistry 88, 9858–9862.
[14] Altschul, S.F. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.

[15] Rost, B. and Sander, C. (1993) Proc. Natl. Acad. Sci. USA 90, 7558–7562.
[16] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[17] Rost, B. and Sander, C. (1994) Proteins 19, 55–72.
[18] http://barton.ebi.ac.uk/servers/jpred.html
[19] Blundell, T.L. and Zhu, Z.Y. (1995) Biophys. Chem. 55, 167–184.
[20] Zhu, Z.Y. and Blundell, T.L. (1996) J. Mol. Biol. 260, 261–276.
[21] Frishman, D. and Argos, P. (1996) Protein Eng. 9, 133–142.
[22] Frishman, D. and Argos, P. (1997) Proteins 27, 329–335.
[23] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[24] Sali, A. and Blundell, T. (1990) J. Mol. Biol. 212, 403–428.
[25] Zhu, Z.Y., Sali, A. and Blundell, T.L. (1992) Protein Eng. 5, 43–51.
[26] Higgins, D.G., Thompsom, J.D. and Gibson, T.J. (1996) Methods Enzymol. 266, 383–402.
[27] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[28] Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Nature 356, 83–85.
[29] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) J. Appl. Crystallogr. 26, 283–291.
[30] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) Nature 358, 86–89.
[31] Leahy, D.J., Hendrickson, W.A., Aukhill, I. and Erickson, H.P. (1992) Science 258, 987–991.
[32] Mizuguchi, K., Deanem, C.M., Blundell, T.L. and Overington, J.P. (1998) Protein Sci. (in press).
[33] Smith, D.K. and Hong, X. (1997) J. Mol. Biol. 274, 530–545.
[34] Abola, E.E., Sussman, J.L., Prilusky, J. and Manning, N.O. (1997) Methods Enzymol. 277, 556–571.
[35] Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J. (1987) in: Crystallographic Databases – Information Content, Software Systems, Scientific Applications (Allen, F.H., Bergerhoff, G. and Sievers, R., Eds.), pp. 107–132, Data Commission of the International Union of Crystallography, Bonn.
[36] de Vos, A.M., Ultsch, M. and Kossiakoff, A.A. (1992) Science 257, 306–312.
[37] Somers, W., Ultsch, M., de Vos, A.M. and Kossiakoff, A.A. (1994) Nature 372, 478–481.
[38] Bork, P., Downing, K., Kieffer, B. and Campbell, I.D. (1996) Q. Rev. Biophys. 29, 119–167.
[39] Leahy, D.J., Aukhil, I. and Erickson, H.P. (1996) Cell 84, 155–164.
[40] Mizuguchi, K., Deane, C.M., Blundell, T.L., Johnson, M.S. and Overington, J.P. (1998) Bioinformatics 14, 617–623.
[41] Sayle, R.A. and Milner-White, E.J. (1995) Trends Biochem. Sci. 20, 374–376.
[42] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[43] Bacon, D.J. and Anderson, W.F. (1988) J. Mol. Graph. 6, 219–220.
[44] Merritt, E.A. and Murphy, M. (1994) Acta Crystallogr. D 50, 869–873.
[45] Hutchinson, E.G. and Thornton, J.M. (1990) Proteins 8, 203–212.