FEBS Letters 443 (1999) 370–374 FEBS 21488

Three receptor genes for plasminogen related growth factors in the genome of the puffer fish *Fugu rubripes*

Amanda Cottage[a], Melody Clark[a], Kelvin Hawker[b], Yagnesh Umrania[a], Denis Wheller[c], Martin Bishop[a], Greg Elgar[a,*]

[a] *MRC HGMP Resource Centre, Hinxton, Cambridge CB10 1SB, UK*
[b] *MRC Growth Factors Group, MRC Centre, Hills Road, Cambridge CB2 2QH, UK*
[c] *Faculty of Built and Environmental Science, Anglia Polytechnic University, East Road, Cambridge CB1 1PT, UK*

Received 27 October 1998; received in revised form 15 December 1998

Abstract Plasminogen related growth factors (PRGFs) and their receptors play major roles in embryogenesis, tissue regeneration and neoplasia. In order to investigate the complexity and evolution of the PRGF receptor family we have cloned and sequenced three receptors for PRGFs in the teleost fish *Fugu rubripes*, a model vertebrate with a compact genome. One of the receptor genes isolated encodes the orthologue of mammalian MET, whilst the other two may represent *Fugu rubripes* orthologues of RON and SEA. This is the first time three PRGF receptors have been identified in a single species.

Key words: Hepatocyte growth factor/receptor; Plasminogen; Oncogene; Genomic sequence; *Fugu rubripes*

1. Introduction

Plasminogen related growth factors (PRGFs) are a family of high molecular weight proteins with a mechanism of proteolytic activation that resembles that of the serine proteinase precursor plasminogen [1]. They play important roles in embryogenesis [2–4], tissue regeneration [5,6] and neoplasia [7]. In mammals, two plasminogen related growth factors, HGF/SF [8] (hepatocyte growth factor/scatter factor) and HGFl/MSP [9] (hepatocyte growth factor-like/macrophage stimulating protein), have been identified that act on target cells through receptor tyrosine kinases known as MET [10] and RON [11] respectively. Recently a receptor of the MET/RON family was identified in chicken as the cellular homologue of the v-SEA oncogene [12]. SEA does not appear to be the avian orthologue of human RON because it is not activated by either human or chicken HGFl/MSP. These data suggest the presence of a third ligand receptor system within the PRGF and PRGF receptor (PRGFR) families although attempts at cloning the avian orthologue of RON or the mammalian orthologue of SEA have been unsuccessful. In order to investigate the complexity and evolution of the PRGF receptor family we have cloned and sequenced three receptors for PRGFs in the teleost fish *Fugu rubripes*, a model vertebrate with a compact genome [13].

*Corresponding author. Fax: (44) (1223) 494512.
E-mail: gelgar@hgmp.mrc.ac.uk

Accession numbers: EMBL: AJ010316 (*CAV* and *CAV2*), EMBL: AJ010317 (Cosmid 165K09, *PRGFR3*, *TRIP*, *Sand*, *GRM7*), EMBL: AJ009961 (*PRGFR1*) and EMBL: AJ010348 (Cosmid 168J21, *PRGFR2*, *UBE1*, *calmodulin binding protein* gene).

2. Materials and methods

2.1. Cloning of PRGF receptors in Fugu rubripes

From an alignment of PRGF receptor sequences (available in EMBL [14]) from a number of vertebrates, degenerate primers were designed within the tyrosine kinase domain to amplify coding and intron sequence from *Fugu rubripes* genomic DNA: A, forward primer: 5′ ctg gcn gcn mgg aay tgy atg (translation LAARNCM); B, reverse primer: 5′ gac can acr tcn gay ttn gtn gtn gtr aa 3′ (translation FTTKSDVW).

The PCR products generated were sub-cloned into *Eco*RV cut pBluescript (Stratagene) and sequenced using KS primer and Amersham Thermo Sequenase dye terminator reactions. The sequences generated were then used in BLAST v2.0 [15] searches of the SwissProt, TREMBL [16] and EMBL databases. Specific primers were then generated to positive clones: 1, forward tk83F: 5′ gtc acg gac atg tgg gga gc 3′; reverse tk83R: 5′ gac tcg tcc agc ctg aac cc 3′; 2, forward tk102F: 5′ gtt cat cta cgt gag ttc ggt g 3′; reverse tk102R: 5′ gct ctc atc cag cct gtc gg 3′; 3, forward tk332F: 5′ gtg aca ccc tga gac gacg 3′; reverse tk332R: 5′ gta cgt ttc atc cac cct ggg 3′.

2.2. Isolation and sequencing of Fugu rubripes genomic clones containing PRGF receptors

The specific primers were used to generate three PCR products from *Fugu rubripes* genomic DNA, these were used to probe a *Fugu rubripes* unamplified phage lambda library and an unamplified cosmid high density gridded library (MRC HGMP Resource Centre UK). Positive clones were Southern blotted and overlapping clones determined. Selected clones were then sequence scanned (see *Fugu rubripes* landmark mapping web page for details http://fugu.hgmp.mrc.ac.uk).

2.3. Characterisation of PRGF receptors and neighbouring genes

Sequences were transferred to a UNIX environment and clipped using a modified Pregap [17] script. Sequences were screened against sequencing (pBluescript and M13) and cosmid (Lawrist4) vectors and matching regions were masked. Passed sequences were then assembled using Gap4 [18] and Pint (P. Weston, unpublished data, http://menu.hgmp.mrc.ac.uk/cgi-bin/pint.pl). Areas of discrepancy, those not covered by a clone and single stranded regions were verified by PCR and sequenced or verified by primer walking. Analyses of the finished sequences were carried out using Nix (G. Williams, P. Woollard and P. Hingamp, unpublished data, http://menu.hgmp.mrc.ac.uk/menu-bin/Nix.pl) and ClustalW [19], for the PRGF receptors this was verified by sequencing RT-PCR products. Phylogenetic analysis was undertaken using SEAVIEW and PHYLO_WIN [20].

3. Results and discussion

PRGF receptor fragments have been isolated by degenerate PCR and the full genomic sequences elucidated from cosmid clones selected by hybridisation with the PCR products. We have designated the three PRGF receptor genes *PRGFR1*, *PRGFR2* and *PRGFR3*. The *PRGFR1* gene spans at least 36 790 nucleotides on two overlapping cosmids (Fig. 1). It is flanked by *Fugu rubripes* homologues of mammalian *caveolin 1* and *2* genes which have recently been mapped to chromo-

PII: S0014-5793(99)00011-3

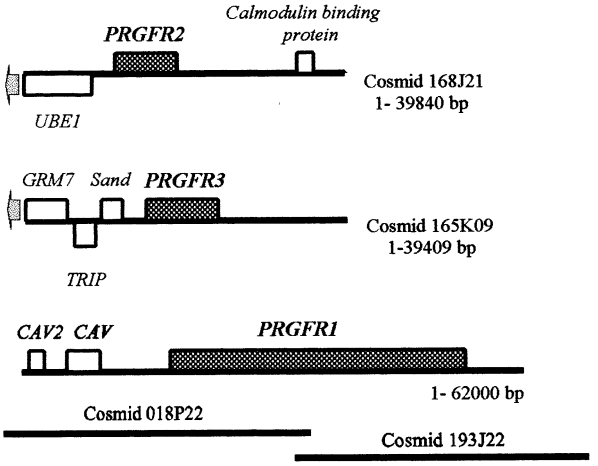

Fig. 1. Schematic organisation of *Fugu rubripes PRGFR1*, *PRGFR2* and *PRGFR3* loci. Arrangement of genes is shown to scale along the genomic locus as a series of open or shaded boxes. Open boxes represent genes neighbouring *PRGFRs*, filled boxes represent the *PRGFRs*, those below the locus bar are on the complementary strand. For brevity, only the canonical tiling path for genomic cosmid clones is shown below the locus bar. Each genomic region is represented with multiple independent genomic clones (not shown). *CAV* and *CAV2* represent the *Fugu rubripes* homologues of human caveolin (EMBL accession number: Z18951) and caveolin-2 (EMBL accession number: AF035752). *UBE1* represents the *Fugu rubripes* homologue of human ubiquitin activating enzyme E1 or E1-like (EMBL accession number: X56976); the 3′ end of the *UBE1* gene is not present on cosmid 168J21 as indicated by the arrow. The *calmodulin binding protein* gene represents the *Fugu rubripes* homologue of a *Rattus norvegicus* calmodulin binding protein (EMBL accession number: L22557). *GRM7* represents the *Fugu rubripes* homologue of human metabotropic glutamate receptor 7 (EMBL accession number: X94552); the 5′ end of the *GRM7* gene is not present on cosmid 165K09 as indicated by the arrow. *TRIP* represents the *Fugu rubripes* orthologue of human *TRIP* (TRAF (tumor necrosis factor receptor associated factor) interacting protein) (EMBL accession number: U77845). *Sand* is a novel gene with homology to predicted proteins in *Saccharomyces cerevisiae* (EMBL accession number: X94106), *Schizosaccharomyces pombe* (EMBL accession number: Z69239) and *Caenorhabditis elegans* (EMBL accession number: U61954).

somes 6-A2 in mouse and 7q31 in man [21], the same chromosomal location as the *MET* gene. The *PRGFR2* gene spans at least 7589 nucleotides on cosmid 168J21 (Fig. 1). It is flanked at the 3′ end by a gene homologous to a rat calmodulin binding protein and at the 5′ end on the complementary strand, by a gene homologous to ubiquitin activating enzyme E1 (UBE1). The latter gene maps to chromosome Xp11 in man but its paralogue maps to 3p21 [22], the same chromosomal location as the *RON* gene. Cosmid 122L06 (sequence available from http://fugu.hgmp.mrc.ac.uk) overlaps 168J21 at the 5′ end and contains sequences with homology to the human cosmid LUCA15 (EMBL accession number: U73166) which maps to 3p21 and to the G protein GNAI2 [23] which also maps to 3p21 in man. The *PRGFR3* gene spans at least 9541 nucleotides in cosmid 165K09 and is flanked at the 5′ end by three genes. One has similarity only to *Caenorhabditis elegans* (EMBL accession number: U61954) and yeast genes (EMBL accession numbers: X94106 and Z69239). A second gene shows highest homology to the metabotropic glutamate receptor 7 (GRM7) [24] which we have mapped to chromosome 3 in man (unpublished data). The third gene, present on the complementary strand, appears to be the *Fugu rubripes* orthologue of the TRAF interacting protein (TRIP) [25]. Human TRIP has recently been mapped to 3p21 (Fig. 1). Although the *Fugu rubripes* loci for *PRGFR2* and *PRGFR3* both contain genes with human homologues mapping to 3p21, we do not know whether they are linked in the *Fugu rubripes* genome.

The intron-exon structure of the *PRGFR1*, *PRGFR2* and *PRGFR3* genes is described in Table 1. Despite the greater length of the *PRGFR1* gene, all three genes have a similar structure with 20 coding exons and 19 introns. This reflects the organisation of human *MET* [26] and mouse *ron* [27] which are the only members of this family to have been characterised so far at the genomic level. The intron phases of the corresponding exons are identical in the five genes, although intron sizes are significantly larger in *PRGFR1* and human *MET*; *PRGFR1* and human *MET* span ∼37 kb and ∼130 kb respectively compared with 7.6 kb for *PRGFR2*, 9.5 kb for *PRGFR3* and 13.2 kb for mouse *ron* (Table 1). These data

Table 1
Comparison of exon/intron sizes between *Fugu rubripes PRGFR1*, *PRGFR2* and *PRGFR3* genes

Coding exon	PRGFR2 length (bp)	PRGFR3 length (bp)	PRGFR1 length (bp)	IVS	PRGFR2 length (bp)	PRGFR3 length (bp)	PRGFR1 length (bp)	Type ALL
1	1200	1206	1221	A	1316	1405	15215	0
2	177	177	204	B	319	74	1887	0
3	120	123	129	C	125	482	85	0
4	171	171	174	D	76	529	415	0
5	164	161	158	E	112	365	673	II
6	133	136	115	F	75	73	95	0
7	137	137	137	G	104	269	100	II
8	180	180	165	H	88	81	211	II
9	100	100	100	I	76	250	4518	0
10	228	219	210	J	89	87	2433	0
11	147	147	150	K	75	150	1616	0
12	163	154	154	L	68	154	373	I
13	90	90	135	M	154	95	817	I
14	210	216	261	N	87	223	1393	I
15	81	81	81	O	83	91	1073	I
16	182	182	182	P	72	134	153	0
17	110	110	110	Q	332	83	102	II
18	166	166	166	R	73	117	118	0
19	137	137	137	S	141	757	1234	II
20	229	229	286	–	–	–	–	–

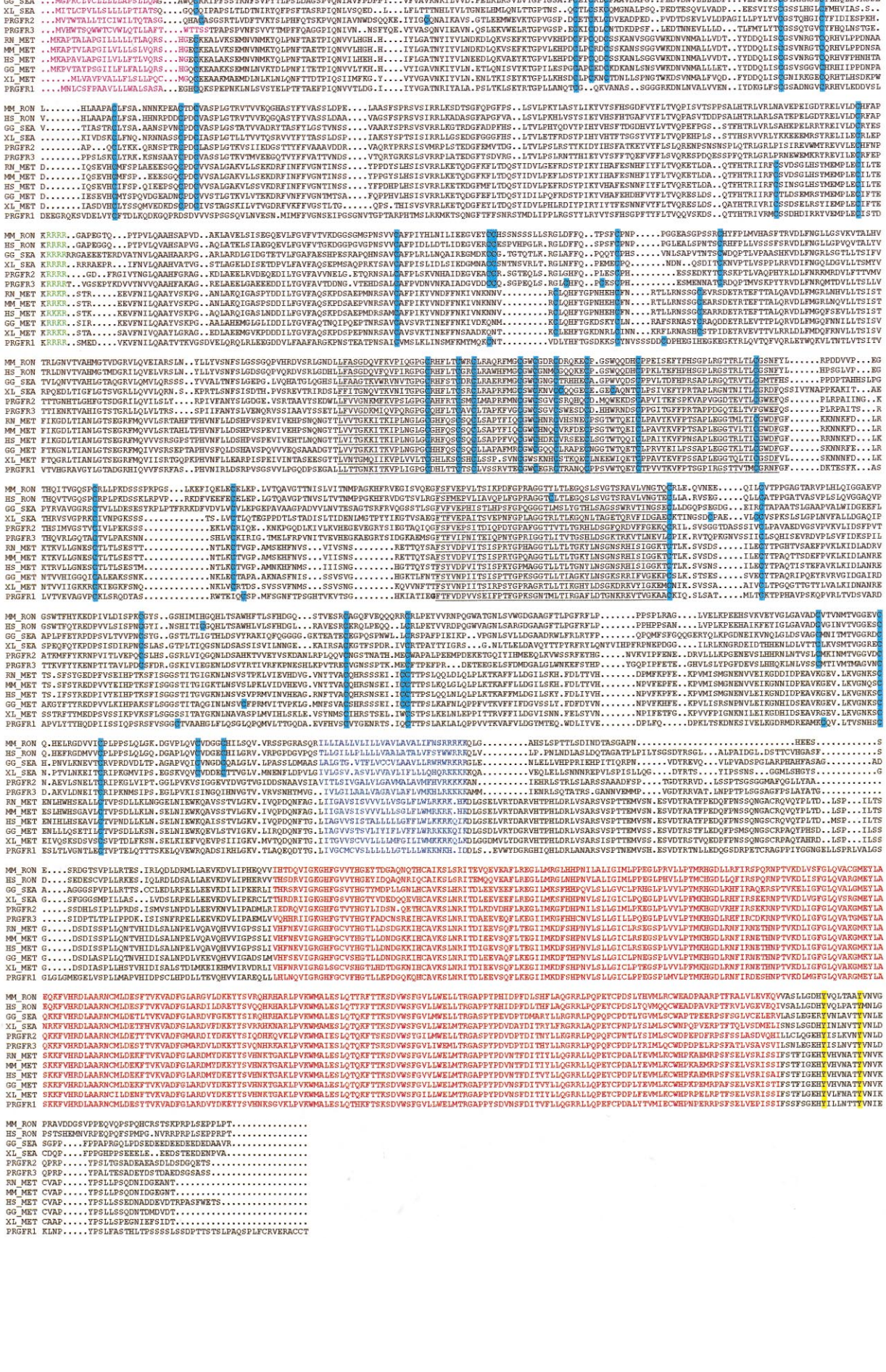

Fig. 2. Protein alignment and domain boundaries of the plasminogen related growth factor receptors. The sequences used: Hs_Met (SwissProt accession number: P08581), Hs-Ron (SwissProt accession number: Q04912), Mm-Met (SwissProt accession number: Q62125), Mm_Ron (SwissProt accession number: G2231241)*, Rn_Met (TREMBL accession number: P97579), X1_Sea (SwissProt accession number: G1679600)**, Xl_Met (OWL accession number: JC5148), Gg_Met (TREMBL accession number: Q90975), Gg_Sea (TREMBL accession number: Q08757), PRGFR1, PRGFR2 and PRGFR3. The alignment was produced using CUSTALW [19]. Cysteines in the extra-membrane domain are shown on a cyan background. The signal peptide is shown in magenta. The furin cleavage site is shown in green. The single trans-membrane domain is shown in blue. The typical cytoplasmic kinase domain is shown in red. Shown on yellow background are two tyrosine residues in the C-terminal peptide essential for downstream signalling. These features are present in all known PRGF receptors. Shown underlined are two conserved sequences present in the extra-membrane domain of the β-chain: the first is 81 residues in length and is cysteine-rich; the second is 45 residues in length and is rich in glycine, proline and hydrophobic residues. Both sequences may be important for the three-dimensional structure of the extra-membrane domain of the β-chain. *X1_Sea: designated *RON* in EMBL and *SEA* in SwissProt. **Mm_Ron: an exon is missing after exon 12 from the submitted protein sequence.
←

further strengthen the conclusion that *PRGFR1* encodes *Fugu rubripes* MET and emphasise that the PRGF receptors of the RON/SEA subfamily have a compact gene structure compared to that of MET.

The protein sequences encoded by the *PRGFR1*, *PRGFR2* and *PRGFR3* genes are shown in Fig. 2 and aligned with the other known sequences of PRGF receptors. The alignment highlights the conserved features of this family of receptor tyrosine kinases, namely an open reading frame in the region of 1400 amino acids and a furin cleavage site (Fig. 2) located on average 300 residues from the translation start site. Processing of the furin site during receptor secretion and translocation results in the typical heterodimeric (α/β) structure of the PRGF receptors, with an α-chain of 275–291 residues located outside the membrane disulphide-linked to the larger β-chain. The β-chain comprises a large extracellular domain of 627–665 residues, a single trans-membrane domain of 27–29 residues (Fig. 2), a linker of 69–123 residues which joins the trans-membrane and the kinase domains, a typical cytoplasmic kinase domain of 259–260 residues (Fig. 2) and a C-terminal peptide containing the two tyrosine residues (Fig. 2) essential for downstream signalling of both the MET and RON receptors. These features are present in all known PRGF receptors [28,29] (Fig. 2). Notable are two conserved sequences in all PRGFRs (Fig. 3), in the extra-membrane domain of the β-chain (Fig. 2): the first is 81 residues in length and is cysteine-rich; the second is 45 residues in length and is rich in glycine, proline and hydrophobic residues. Both sequences may be important for the three-dimensional structure of the extra-membrane domain of the β chain.

The predicted coding sequences of the *PRGFR1*, *PRGFR2* and *PRGFR3* genes indicate that all three encode potentially functional proteins. The three genes were investigated by RT-PCR and sequence data from the PCR products confirmed expression. *PRGFR1* is expressed predominantly in the liver, gonads, kidney, gill and ovaries, *PRGFR2* is expressed in all tissues examined and *PRGFR3* is expressed in the gonads, kidney and gill.

The amino acid sequences of the *PRGFR1*, *PRGFR2* and *PRGFR3* genes are significantly diverged as indicated by sequence similarities of only 36% between *PRGFR1* and *PRGFR2*, 32% between *PRGFR1* and *PRGFR3* and 58% between *PRGFR2* and *PRGFR3*. Whilst the *PRGFR1* gene shows higher levels of sequence identity with the MET receptors (50%) than with the RON/SEA receptors (35%), the opposite is true for the *PRGFR2* and *PRGFR3* genes which show levels of sequence identity with the MET receptors (39%) and levels of sequence identity with the RON/SEA (44%). The *PRGFR2* and *PRGFR3* genes show similar levels of sequence identity with the RON receptors (40%) but the *PRGFR3* gene shows the greatest level of sequence identity with the SEA receptors (46%). On this basis and incorporating linkage data (Fig. 1), we can identify the *PRGFR1* gene as the orthologue of the *MET* gene. It is difficult to predict whether *PRGFR2* and *PRGFR3* are the orthologues of *RON* and *SEA* respectively as linkage data place both at 3p21 in man, the locus of the RON receptor gene. There are no mapping data for SEA in any vertebrate. Phylogenetic analysis (Fig. 3) is complicated by the fact that comparisons to *PRGFR2* are being made with mammalian genes and comparisons to *PRGFR3* are being made with chicken and *Xenopus* genes. In fact it appears that *PRGFR2* and *PRGFR3* are most similar to each other and may have arisen by a dupli-

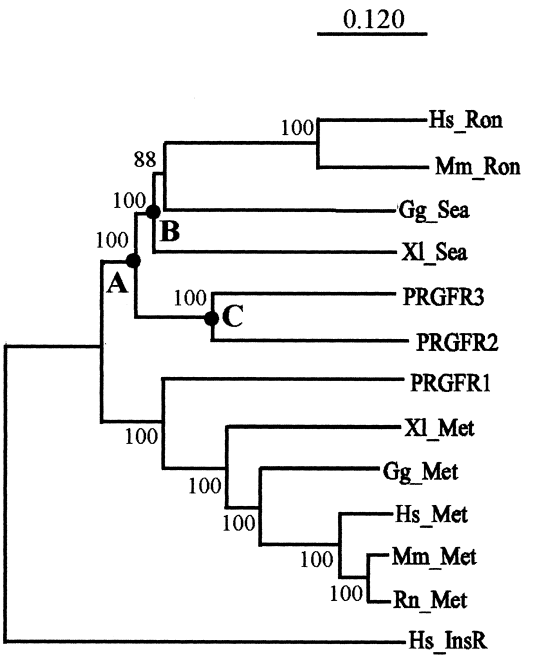

Fig. 3. Phylogeny of the plasminogen related growth factor receptor amino acid sequences. The sequences used were as in Fig. 2, but include the human insulin growth factor receptor Hs_InsR (SwissProt accesion number: P06213). The tree was constructed using PHYLO_WIN [20] selecting the neighbour joining method and using Hs_InsR as an outgroup. As is clearly indicated the PRGF receptors can be clustered into two subfamilies: the MET subfamily and the RON/SEA subfamily. Only *Fugu rubripes* has two members of the RON/SEA subfamily characterised to date, thus in this analysis the relationship between these subfamily members is unclear and may be more representative of relationships between vertebrates. A: Divergence of common ancestral gene for *SEA*, *RON*, *PRGFR2* and *PRGFR3*. B: Divergence of SEA and RON. C: Divergence of PRGFR2 and PRGFR3. Numbers on the nodes indicate percentage recovery of these nodes per 500 bootstrap replications.

cation event that post dates the divergence of the *SEA* and *RON* genes (Fig. 3, node C). However, the common ancestor of the *PRGFR2* and *PRGFR3* genes (Fig. 3, node A) predates the divergence of the *RON* and *SEA* genes (Fig. 3, node B). It would appear that this common ancestor has been duplicated in at least one other vertebrate as neither human nor chicken HGFl/MSP activates chicken SEA, therefore a third PRGF/receptor system which is of the Sea/Ron lineage would seem to be operating in chicken as well as *Fugu rubripes*.

The data presented here clearly indicate that the PRGF receptors can be clustered into two subfamilies: the MET subfamily including PRGFR1 and the RON/SEA subfamily which includes both PRGFR2 and PRGFR3. The two subfamilies can be recognised through pair-wise sequence comparisons, the length of the extra-membrane domain of the β-chain (on average 629 residues in the MET subfamily and 649 residues in the RON/SEA subfamily) and the length of the linker between the trans-membrane and the kinase domains (on average 116 residues in the MET subfamily and 89 residues in the RON/SEA subfamily) (Fig. 2). Several sequence signatures in the α-chain, in the extra-cellular domain of the β-chain and in the linker between the trans-membrane and the kinase domains further discriminate PRGF receptors of the MET and RON/SEA subfamilies (Fig. 2). Phylogenetic analysis of the PRGF receptors, using the sequence of the human insulin receptor as an outgroup, confirms that the two subfamilies form distinct clusters (Fig. 3).

We have utilised sequence, domain and phylogenetic analysis to investigate three *Fugu rubripes PRGFR* genes. We have shown conservation of the intron-exon structure of the PRGF receptor genes across vertebrates and have provided the basis for classifying the PRGF receptor genes into two subfamilies. Significantly we have shown, for the first time, direct evidence that a single species expresses three PRGF receptors, supporting receptor/ligand data from chicken and man implying that other species, including mammals, may express a third receptor and corresponding ligand. Receptor/ligand binding studies of the *Fugu rubripes* genes should confirm their function and will clearly be required in order to reconstruct the roles of PRGFs and their receptors in vertebrate species.

Acknowledgements: This work has been supported by the Medical Research Council. We thank Steven Meek, Sarah Warner, Sarah Smith, Yvonne Edwards, Ermanno Gherardi, Samuel Aparicio and Paul Weston for their support and advice.

References

[1] Donate, L.E., Gherardi, E., Srinivasan, N., Sowdhamini, R., Aparicio, S. and Blundell, T.L. (1994) Protein Sci. 3, 2378–2394.
[2] Schmidt, C., Bladt, F., Goedecke, S., Brinkmann, V., Zschiesche, W., Sharpe, M., Gherardi, E. and Birchmeier, C. (1995) Nature 373, 699–702.
[3] Uehara, Y., Minowa, O., Mori, C., Shiota, K., Kuno, J., Noda, T. and Kitamura, N. (1995) Nature 373, 702–705.
[4] Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A. and Birchmeier, C. (1995) Nature 376, 768–771.
[5] Nagaike, M., Hirao, S., Tajima, H., Noji, S., Taniguchi, S., Matsumoto, K. and Nakamura, T. (1991) J. Biol. Chem. 266, 22781–22784.
[6] Yanagita, K., Matsumoto, K., Sekiguchi, K., Ishibashi, H., Niho, Y. and Nakamura, T. (1993) J. Biol. Chem. 268, 21212–21217.
[7] Jeffers, M., Rong, S. and Woude, G.F. (1996) J. Mol. Med. 74, 505–513.
[8] Nakamura, T., Nawa, K., Ichihara, A., Kaise, N. and Nishino, T. (1989) Nature 342, 440–443.
[9] Yoshimura, T., Yuhki, N., Wang, M.H., Skeel, A. and Leonard, E.J. (1993) J. Biol. Chem. 268, 15461–15468.
[10] Bottaro, D.P., Rubin, J.S., Faletto, D.L., Chan, A.M., Kmiecik, T.E., Vande Woude, G.F. and Aaronson, S.A. (1991) Science 251, 802–804.
[11] Wang, M.H., Ronsin, C., Gesnel, M.C., Coupey, L., Skeel, A., Leonard, E.J. and Breathnach, R. (1994) Science 266, 117–119.
[12] Huff, J.L., Jelinek, M.A., Borgman, C.A., Lansing, T.J. and Parsons, J.T. (1993) Proc. Natl. Acad. Sci. USA 90, 6140–6144.
[13] Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. and Aparicio, S. (1993) Nature 366, 265–268.
[14] Stoesser, G., Sterk, P., Tuli, M.A., Stoehr, P.J. and Cameron, G.N. (1997) Nucleic Acids Res. 25, 7–14.
[15] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[16] Bairoch, A. and Apweiler, R. (1998) Nucleic Acids Res. 26, 38–42.
[17] Bonfield, J.K. and Staden, R. (1996) DNA Seq. 6, 109–117.
[18] Bonfield, J.K., Smith, K.F. and Staden, R. (1995) Nucleic Acids Res. 23, 4992–4999.
[19] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[20] Galtier, N., Gouy, M. and Gautier, C. (1996) Comput. Appl. Biosci. 12, 543–548.
[21] Engelman, J.A., Zhang, X.L., Galbiati, F. and Lisanti, M.P. (1998) FEBS Lett. 429, 330–336.
[22] Kok, K., Hofstra, R., Pilz, A., van den Berg, A., Terpstra, P., Buys, C.H. and Carritt, B. (1993) Proc. Natl. Acad. Sci. USA 90, 6071–6075.
[23] Magovcevic, I., Ang, S.L., Seidman, J.G., Tolman, C.J., Neer, E.J. and Morton, C.C. (1992) Genomics 12, 125–129.
[24] Okamoto, N., Hori, S., Akazawa, C., Hayashi, Y., Shigemoto, R., Mizuno, N. and Nakanishi, S. (1994) J. Biol. Chem. 269, 1231–1236.
[25] Lee, S.Y., Lee, S.Y. and Choi, Y. (1997) J. Exp. Med. 185, 1275–1285.
[26] Lin, J.C., Naujokas, M., Zhu, H., Nolet, S. and Park, M. (1998) Oncogene 16, 833–842.
[27] Waltz, S.E., Toms, C.L., McDowell, S.A., Clay, L.A., Muraoka, R.S., Air, E.L., Sun, W.Y., Thomas, M.B. and Degen, S.J. (1998) Oncogene 16, 27–42.
[28] Bardelli, A., Ponzetto, C. and Comoglio, P.M. (1994) J. Biotechnol. 37, 109–122.
[29] Ponzetto, C., Bardelli, A., Zhen, Z., Maina, F., dalla Zonca, P., Giordano, S., Graziani, A., Panayotou, G. and Comoglio, P.M. (1994) Cell 77, 261–271.