Characterization of a mammalian cDNA encoding a protein with high sequence similarity to the *Drosophila* regulatory protein Rhomboid

John C. Pascall*, Kenneth D. Brown

The Babraham Institute, Babraham Hall, Cambridge CB2 4AT, UK

Received 23 April 1998; revised version received 18 May 1998

Abstract The *Drosophila* regulatory protein Rhomboid has been demonstrated genetically to facilitate signalling within the Spitz/epidermal growth factor receptor/mitogen-activated protein kinase pathway. Using a polymerase chain reaction (PCR)-based strategy, we have cloned a human cDNA which encodes a protein that has high sequence similarity to Rhomboid. The encoded protein, termed rhomboid-related protein (RRP), is predicted to contain seven transmembrane domains. Northern analysis indicates that RRP mRNA is expressed at highest levels in brain and kidney.

Key words: Rhomboid; Epidermal growth factor receptor signalling; Protein processing; Spitz; *Drosophila melanogaster*

1. Introduction

The *Drosophila* epidermal growth factor (EGF) receptor, DER, serves many functions during development (see [1]). The major agonist for DER during embryonic and larval development is Spitz (see [2]). Spitz is a member of the EGF/transforming growth factor-α family [3] and, in common with most other EGF-related proteins characterized to date, is synthesised as a transmembrane precursor protein [3]. Processing of the Spitz precursor to release it from the plasma membrane is obligatory for it to function as a DER agonist [4]. The product of a second gene, *rhomboid*, has been genetically linked to Spitz processing [4], although this has yet to be confirmed biochemically, and other functions have been proposed (see below). *Rhomboid* was first cloned in 1990 [5] and codes for a 355 amino-acid protein with six or seven putative transmembrane domains. There are no other identifiable domains that give clues to its function.

DNA sequences related to *rhomboid* have been identified as part of the *Caenorhabditis elegans* genome sequencing project (see [2]), suggesting that Rhomboid function may be evolutionarily conserved. However, to our knowledge, there are no reports of a mammalian *rhomboid* gene. Since the key components of the EGF receptor signalling pathway are conserved between *Drosophila* and mammals (see [6]), it seems probable that such a homologue will exist. Here we report the cloning and analysis of the distribution of expression of a potential mammalian Rhomboid homologue, which we have named rhomboid-related protein (RRP). A detailed characterisation of RRP, in association with the studies of Rhomboid function in *Drosophila*, may provide important insights into novel mechanisms for the regulation of EGF receptor signalling.

2. Materials and methods

2.1. Materials

Marathon cDNA and a human multiple tissue Northern blot (both Clontech products) were purchased from Cambridge Bioscience (Cambridge, UK). *Taq* DNA polymerase was purchased from Boehringer Mannheim (Lewes, East Sussex, UK). Cloned *Pfu* DNA polymerase, *TaqPlus Precision* PCR system, and pCR-Script Amp cloning kit were from Stratagene Ltd. (Cambridge, UK). The pGEM-T Easy vector system was purchased from Promega (Southampton, UK). All oligonucleotides were synthesised by the Microchemical Facility at the Babraham Institute.

2.2. Cloning of mammalian cDNAs encoding Rhomboid-related proteins

Comparison of the *Drosophila rhomboid* cDNA sequence with that of a randomly sequenced human genomic DNA fragment (accession number Z92544) allowed us to predict oligonucleotide sequences that were used to derive a series of overlapping cDNA clones encoding human RRP mRNA. The clones were derived by a PCR-based strategy using double-stranded cDNA derived from RNA isolated from the HL-60 cell line (Marathon cDNA; Clontech) as starting material. We have also isolated a cDNA clone encoding a fragment of a rat RRP cDNA by reverse transcription-PCR starting from RNA isolated from a rat epithelial cell line. Full details of the oligonucleotides and the PCR and cloning conditions used to derive the human and rat clones can be obtained from the authors. All clones were sequenced on both strands using an ABI377 sequencer and the sequences analysed using programs from the University of Wisconsin Genetic Computing Group [7].

2.3. RNA isolation and Northern blot analysis

RNA was isolated as described [8]. A human multiple tissue Northern blot (Clontech) was probed with a ^{32}P-labelled cDNA spanning the human RRP coding region. Hybridization and washing conditions were as recommended by the manufacturer.

3. Results and discussion

A search of the EMBL and GenBank databases with the *Drosophila* Rhomboid protein sequence, using the program TFASTA, identified strong similarity with a translation of DNA sequences within a human genomic clone (accession number Z92544) derived from chromosome 16. Using oligonucleotides predicted from this genomic sequence, a series of overlapping cDNA fragments were cloned by a PCR-based strategy, using cDNA (Marathon; Clontech) derived from RNA isolated from the human leukaemic HL-60 cell line. The consensus cDNA sequence and deduced protein sequence of human RRP are shown in Fig. 1. The cDNA is approximately 1.6 kb in length, similar to the size of the correspond-

*Corresponding author. Fax: (44) (1223) 836481.
E-mail: john.pascall@bbsrc.ac.uk

Abbreviations: EGF, epidermal growth factor; MAP kinase, mitogen-activated protein kinase; PCR, polymerase chain reaction; RRP, rhomboid-related protein

The nucleotide sequences described in this paper will appear in the DDBJ, EMBL and GenBank databases with accession numbers Y17108 (human) and Y17258 (rat).

PII S0014-5793(98)00622-X

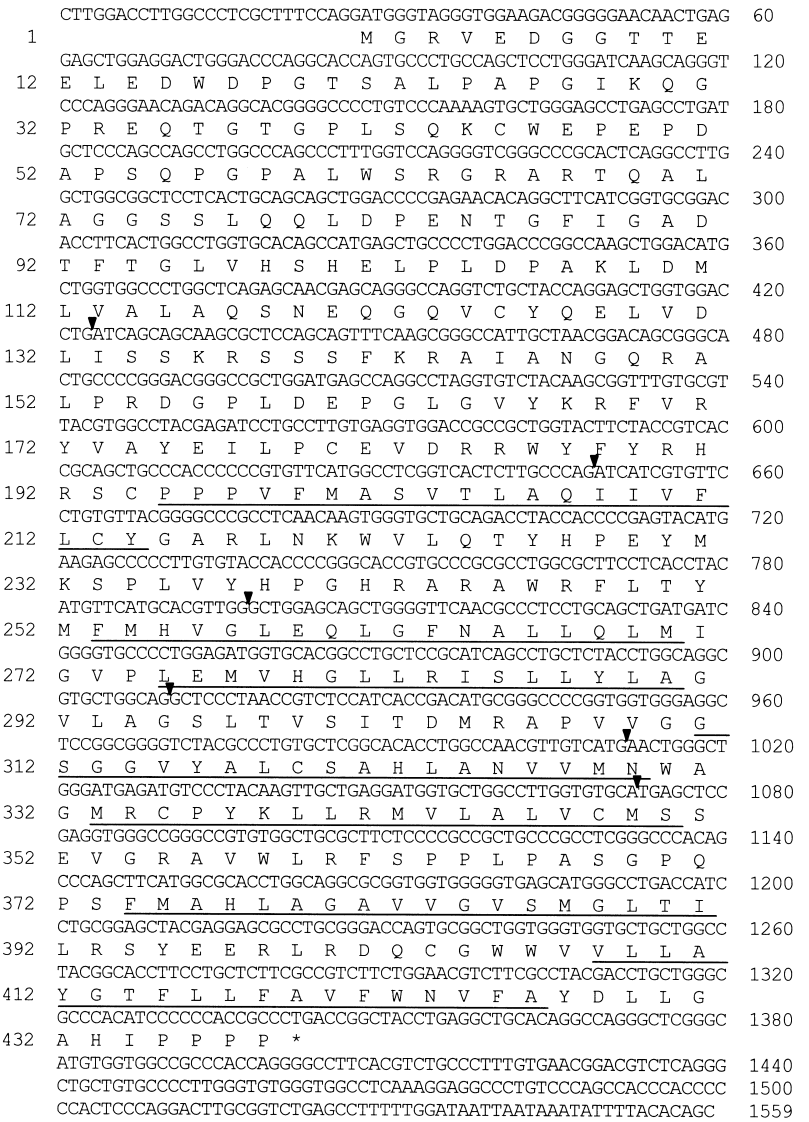

Fig. 1. The cDNA and predicted protein sequences of human rhomboid-related protein. Potential transmembrane domains identified by the program PhDTopology are indicated by underlining, and the exon junctions by arrowheads.

ing mRNA in human brain and kidney (see below), and encodes a protein of 438 amino acids. Although the cDNA sequence does not contain an in-frame stop codon upstream of the predicted initiator methionine, we believe this residue is likely to be the N-terminus of the protein as, firstly, the methionine is located in a strong consensus sequence for protein synthesis initiation [9] and secondly, additional 5′-RACE reactions performed on Marathon cDNA derived from either the HL-60 cell line or foetal human brain, using oligonucleotides located within 200 bases of the predicted 5′ end of the mRNA, failed to identify any larger species. Comparison of the cDNA and genomic sequences indicates that the RRP gene contains seven exons that span approximately 2.2 kb pairs. The location of intron-exon boundaries (see Fig. 1) bears no obvious relationship to the structure of the protein. They do not, for example, correspond to the predicted transmembrane domains of the protein. In contrast with the mammalian gene, the *Drosophila rhomboid* gene comprises only two exons within the protein coding region [5] and the single exonic junction does not correspond to any of those in the mammalian gene.

Computer-aided protein sequence analysis suggests that, like *Drosophila* Rhomboid [5], human RRP is a multispan transmembrane protein. Transmembrane domains predicted by the program PhdTopology [10] are indicated by underlining in Fig. 1. Both this program, and a second, TMpred [11], suggest that human RRP has seven transmembrane spans with an intracellular N-terminus. However, detailed experimental analysis will be needed to confirm the protein's cellular location and topology. A comparison of the protein sequences of *Drosophila* Rhomboid and human RRP using the program GAP showed 30% identity (56% similarity). An alignment of human RRP with *Drosophila* Rhomboid and two related *C. elegans* proteins is shown in Fig. 2. All four sequences show strong similarity, which is particularly apparent within the predicted transmembrane regions (amino acids 195–426 of human RRP).

To investigate RRP expression, polyadenylated RNA sam-

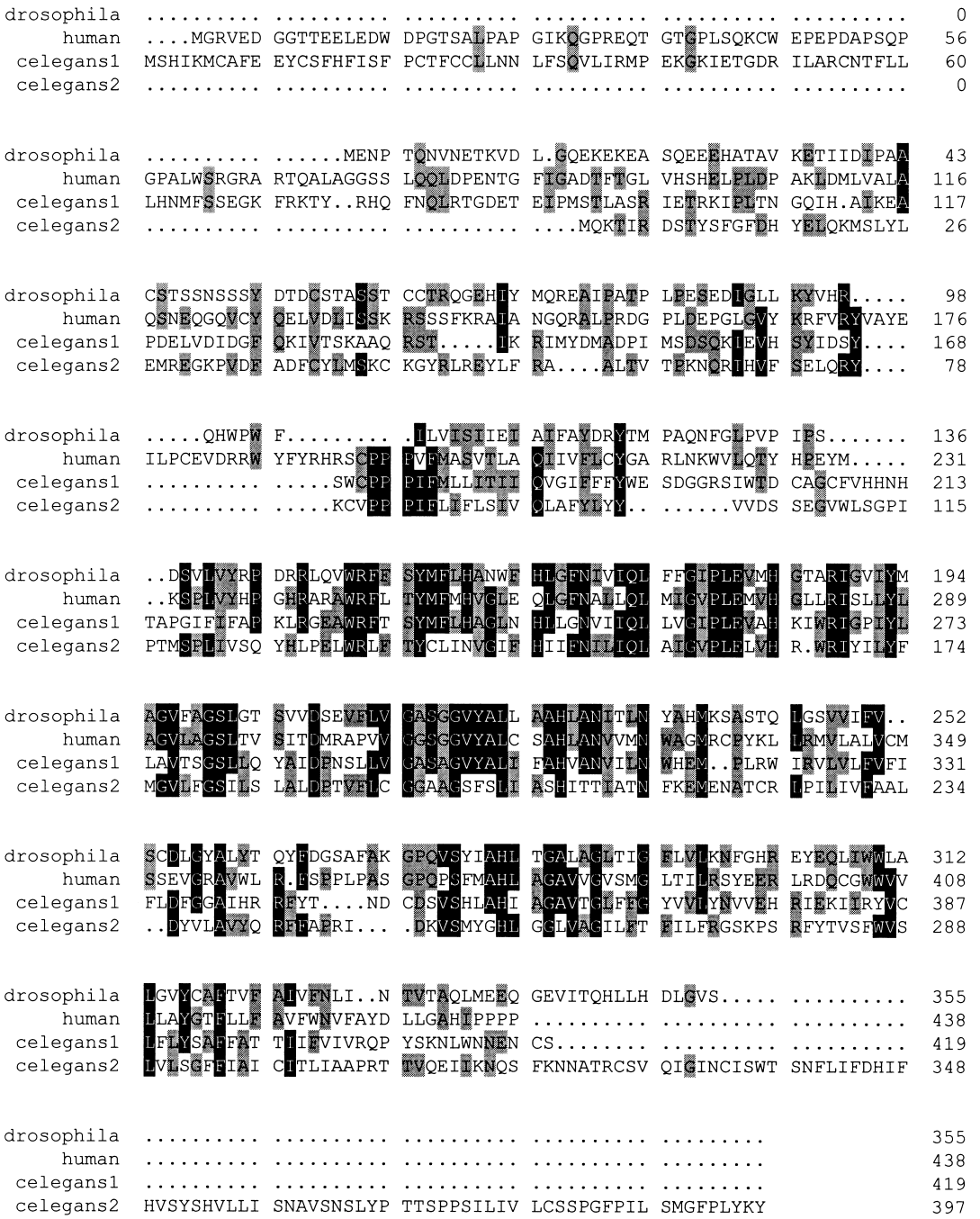

Fig. 2. Protein sequence comparison between human rhomboid-related protein, *Drosophila* Rhomboid and two related proteins in *Caenorhabditis elegans*. Sequences were aligned using the program PILEUP from the University of Wisconsin Genetic Computing Group Package. Sites at which two of the amino acids are identical are highlighted in grey, and at which three or four are identical in black. The protein sequences for celegans1 and celegans2 can be found at TREMBL:q19821 and SWISSPROT:p34356, respectively.

ples isolated from human tissues were analysed by Northern blotting using a cDNA spanning the human RRP coding region as a probe (Fig. 3). The probe hybridized strongly to a transcript of approximately 1.7 kb in RNA isolated from brain (lane 2) and kidney (lane 7). Interestingly, larger hybridizing species were detected in RNA from heart (lane 1), brain (lane 2) and skeletal muscle (lane 6), raising the possibility either of tissue-specific alternative splicing of the RRP precursor RNA or of cross-hybridizing mRNA species derived from related genes. In support of this latter suggestion, database searches with the RRP sequence identified a human expressed sequence tag (accession number AA351652) coding for a protein fragment with high sequence similarity to the C-terminal 50 amino acids of human RRP, suggesting that at least two rhomboid-related genes are expressed in human tissues.

The precise role and mode of action of Rhomboid during *Drosophila* development has yet to be determined. Although DER and its ligand, Spitz, are widely expressed in *Drosophila*, Rhomboid expression appears to be restricted to those tissues where active DER signalling occurs [5,12,13]. Whilst the

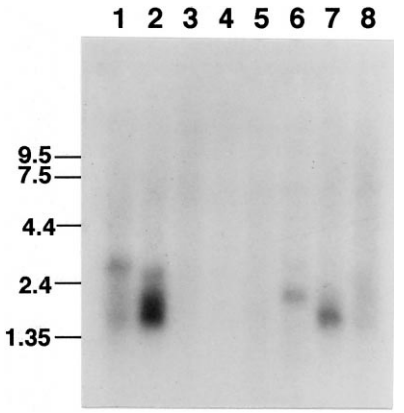

Fig. 3. Expression of rhomboid-related protein mRNA in human tissues. Polyadenylated RNA samples (2 μg) from heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7) and pancreas (lane 8) were analysed by Northern blotting. The blot was hybridized with a ^{32}P-labelled cDNA encoding human rhomboid-related protein, washed and then exposed to X-ray film at −70°C with intensifier screens. The positions of co-electrophoresed RNA size markers are indicated.

weight of evidence favours a role for Rhomboid in Spitz processing, alternative functions have been suggested (see [2]). Firstly, Rhomboid may regulate DER/mitogen-activated protein (MAP) kinase activity directly, either by facilitating ligand-receptor interaction or by an involvement at some point in the downstream signalling pathway from the receptor. The evidence against this proposal is that in vitro studies have failed to find any difference in EGF receptor signalling between *Drosophila* cell lines expressing and lacking Rhomboid, at least to the level of MAP kinase activation [4]. A second suggestion is that Rhomboid may function in association with proteins at intercellular junctions to facilitate cell-cell interaction and hence to allow ligand-receptor interaction. In support of this idea is the demonstration that in some embryonic cells Rhomboid co-localises with the *Drosophila* β-catenin-like protein Armadillo at adherens junctions [14].

In conclusion, we have cloned a cDNA encoding a mammalian protein which shows high sequence similarity with the *Drosophila* protein Rhomboid, which is strongly implicated in the modulation of EGF receptor signalling in the fly. The demonstration of a putative human homologue of Rhomboid suggests that the protein's function may be evolutionarily conserved. If so, RRP may represent a novel component in the mammalian EGF receptor signalling pathway. Our identification of RRP should stimulate analysis of the protein's function and facilitate a search for additional, related proteins.

References

[1] Schweitzer, R. and Shilo, B.-Z. (1997) Trends Genet. 13, 191–196.
[2] Wasserman, J.D. and Freeman, M. (1997) Trends Cell Biol. 7, 431–436.
[3] Rutledge, B.J., Zhang, K., Bier, E., Jan, Y.N. and Perrimon, N. (1992) Genes Dev. 6, 1503–1517.
[4] Schweitzer, R., Shaharabany, M., Seger, R. and Shilo, B.-Z. (1995) Genes Dev. 9, 1518–1529.
[5] Bier, E., Jan, L.Y. and Jan, Y.N. (1990) Genes Dev. 4, 190–203.
[6] Duffy, J.B. and Perrimon, N. (1996) Curr. Opin. Cell Biol. 8, 231–238.
[7] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–394.
[8] Pascall, J.C. and Brown, K.D. (1997) Biochem. J. 324, 869–875.
[9] Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8148.
[10] Rost, B., Fariselli, P. and Casadio, R. (1996) Protein Sci. 5, 1704–1718.
[11] Hofmann, K. and Stoffel, W. (1993) Biol. Chem. Hoppe-Seyler 374, 166–170.
[12] Ruohola-Baker, H., Grell, E., Chou, T.-B., Baker, D., Yah, L.Y. and Jan, Y.H. (1993) Cell 73, 953–965.
[13] Sturtevant, M.A., Roark, M. and Bier, E. (1993) Genes Dev. 7, 961–973.
[14] Sturtevant, M.A., Roark, M., O'Neill, J.W., Biehs, B., Colley, N. and Bier, E. (1996) Dev. Biol. 174, 298–309.