Published in final edited form as:

Invest Ophthalmol Vis Sci. 2015 October ; 56(11): 6255–6264. doi:10.1167/iov.15-17726.

Evolutionary Characterization of the Retinitis Pigmentosa GTPase Regulator Gene

Rakesh Kotapati Raghupathy¹, Philippe Gautier², Dinesh C. Soares², Alan F. Wright², and Xinhua Shu¹

¹Department of Life Sciences, Glasgow Caledonian University, Glasgow, United Kingdom

²MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Purpose—The evolutionary conservation of the retinitis pigmentosa GTPase regulator (*RPGR*) gene was examined across vertebrate and invertebrate lineages to elucidate its function.

Methods—Orthologous *RPGR* sequences from vertebrates and invertebrates were selected. Multiple sequence alignments, phylogenetic analyses, synteny, and gene structure comparisons were carried out. Expression of the alternatively spliced constitutive (*RPGR*^{const} or *RPGR*^{ex1-19}) and *RPGR*^{ORF15} isoforms was examined in developing and adult zebrafish.

Results—Phylogenetic analyses and syntenic relationships were consistent with the selected sequences being true orthologues, although whole genome duplications in teleost fish resulted in a more complex picture. The splice form *RPGR*^{const} was present in all vertebrate and invertebrate species but the defining carboxyl (C)-terminal exon of *RPGR*^{ORF15} was absent from all invertebrates. The regulator of chromosome condensation (RCC1)-like domain adopts a seven-bladed β -propeller structure, which was present in both major splice forms and strongly conserved across evolution. The repetitive acidic region of *RPGR*^{ORF15} showed a high rate of in-frame deletions/insertions across nine primate species, compared with flanking sequences, consistent with an unstable and rapidly evolving region. In zebrafish, *RPGR*^{const} transcripts were most strongly expressed in early development, while the *RPGR*^{ORF15} isoform showed highest expression in adult eye.

Conclusions—The regulator of chromosome condensation 1-like domain of *RPGR* was conserved in vertebrates and invertebrates, but *RPGR*^{ORF15} was unique to vertebrates, consistent with a proposed role in the ciliary-based transport of cargoes such as rhodopsin, which is ~10 times more abundant in vertebrate than invertebrate photoreceptors. The repetitive acidic region of *RPGR*^{ORF15} shows a rapid rate of evolution, consistent with a mutation “hot spot.”

Correspondence: Xinhua Shu, Department of Life Sciences, Glasgow Caledonian University, Glasgow G4 0BA, UK; Xinhua.Shu@gnu.ac.uk. Alan F. Wright, MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK; Alan.Wright@igmm.ed.ac.uk.
RKR and PG are joint first authors.

Disclosure: R.K. Raghupathy, None; P. Gautier, None; D.C. Soares, None; A.F. Wright, None; X. Shu, None

Keywords

retinitis pigmentosa; RPGR; gene structure; evolution; cilia; photoreceptors

Retinitis pigmentosa (RP) is a heterogeneous group of retinal dystrophies characterized by impaired dark adaption and premature death of rod and cone photoreceptors, leading to early onset of night blindness and progressing to severe visual loss.^{1,2} Retinitis pigmentosa affects approximately 1 in 3500 individuals and can be inherited as an autosomal dominant, autosomal recessive, X-linked, oligogenic, or mitochondrial trait.^{3,4} One of the most severe forms of human retinal degeneration, X-linked RP (XLRP), accounts for 10% to 20% of all RP cases.⁵ It is also genetically heterogeneous, with at least six identified loci (<http://www.sph.uth.tmc.edu/RetNet/>, in the public domain). Mutations in the retinitis pigmentosa GTPase regulator (*RPGR*) gene are responsible for the most common (RP3) type of XLRP, accounting for 70% to 80% of cases.^{6–9}

The human *RPGR* gene exhibits a complex expression pattern with multiple alternatively spliced isoforms.^{7,10} *RPGR*^{const} (also called *RPGR*^{ex1–19}) and *RPGR*^{ORF15} represent the two main protein-coding splice variants, which share an amino (N) terminal domain homologous to the RCC1 (regulator of chromosome condensation 1).^{6,11,12} In humans, the “constitutive” isoform, *RPGR*^{const}, contains 19 exons and is widely expressed, whereas *RPGR*^{ORF15} has an alternatively spliced C-terminal exon (ORF15) which shows a more restricted expression, including the retina in species examined to date.^{7,13–17} Mutations in *RPGR*^{ORF15} give rise to XLRP (95% of cases), cone and cone–rod dystrophies, atrophic macular degeneration or syndromal RP with respiratory infection, hearing loss and primary ciliary dyskinesia.^{12,18} Naturally occurring mutations in *RPGR*^{ORF15} have been reported in the mouse and in dogs with progressive retinal atrophy.^{13,19}

Retinitis pigmentosa GTPase regulator is widely expressed in vertebrate tissues (e.g., brain, eye, kidney, lung, and testis).^{10,13,15–17} In the eye, it is predominantly localized to photoreceptor connecting cilia¹⁵ and, in proliferating cells, to basal bodies/centrosomes.²⁰ Less consistently, it has been reported in ciliary rootlets, nuclei and photoreceptor outer segments of some species.^{14,16,21} Retinitis pigmentosa GTPase regulator has been reported to interact with a wide range of proteins including scaffold proteins (*RPGRIP1*, whirlin); ciliary proteins (*NPHP5/IQCB1*, *NPHP6/CEP290*, *RPGRIP1*, *RPGRIP1L*); nuclear or nucleocytoplasmic shuttling proteins (*SMC1*, *SMC3*, nucleophosmin); microtubule-associated proteins (*Rab8*, *IFT88*, dynein, kinesin II, *KIF3A*); and the lipid trafficking protein *PDEδ* (*PDED*) (see Ref. 22 for review).

In the face of this array of proposed interactions, no clear picture of *RPGR* function has yet emerged. We therefore examined its evolutionary conservation across 37 vertebrate and 18 invertebrate species by assembling orthologous *RPGR* sequences, using GenBank as well our own sequencing data, and performed multiple sequence alignments, phylogenetic analyses, and we examined syntenic relationships to further elucidate its function.

Methods

Cloning and Sequencing of RPGR Exon ORF15 and RPGR cDNA Orthologues

We examined *RPGR*^{ORF15} orthologues from 37 vertebrate species. These were identified from public sequence database but also included our own sequence data from eight primates (chimpanzee, cynomolgus monkey, gibbon, marmoset, owl monkey, African green monkey, rhesus, and gorilla), and six other vertebrates (cat, pig, sheep, rabbit, rat, and hamster). These were cloned by PCR using individual species' genomic DNA as template with degenerate primers. The genomic DNA samples were obtained from European Collection of Cell Culture (ECACC) through Sigma-Aldrich (Dorset, UK). We ligated the PCR products into a commercial cloning system (pGEM-T Easy vector; Promega Corp., Southampton, UK) and sequenced with M13 forward and reverse primers together with other primers (primer details provided on request). The remaining vertebrate species were obtained from GenBank and are listed in Supplementary Table S1. Similarly for 18 putative invertebrate RPGR orthologue sequences, 15 were obtained from GenBank (listed in Supplementary Table S1) and three were from our own sequence data (amphioxus, fruit fly, and sea anemone).

We isolated total RNAs from chicken (provided by Roslin Institute); fruit fly eyes (provided by MRC Human Genetics Unit); sea anemone (gifted by Professor Ulrich Technau, University of Vienna); and amphioxus (gifted by Sebastian Shimeld, PhD, Oxford University) using an RNA purification kit (Absolutely RNA Miniprep kit; Agilent Technologies, Stockport, UK) according to the manufacturer's protocol. Complementary DNAs (cDNAs) were synthesized using a cDNA synthesis kit (Cat. 04379012001; Roche Diagnostics, Ltd., Burgess Hill, UK). The full-length RPGR cDNA sequences from the above species were obtained by PCR using primers complementary to the predicted exonic sequences derived from these species' genomes. The PCR products were sequenced after cloning into a commercial cloning system (Promega Corp., Southampton, UK).

Comparative Sequence Analysis

To identify orthologues to human RPGR in other species, the human RPGR amino acid sequences were compared with the translated protein sequences from genomic sequence data of different species using BLAST.23 Orthology was tested with a reciprocal BLAST search of the candidate protein against the human proteome, and reciprocal best-hits were considered to be orthologous sequences. Gene predictions were made through GeneWise24 (European Bioinformatics Institute, Cambridgeshire, UK) using the published human RPGR protein sequences as templates. These predictions were then manually curated and refined using other sources of information available via the Ensembl website (<http://www.ensembl.org/>, in the public domain), including Ensembl's own gene prediction models, de novo GENSCAN25 predictions, expressed sequence tag, cDNA and protein homologies.

Multiple sequence alignments of the proteins were constructed using MUSCLE26 and CLUSTALW.27 Sequences were manipulated using programs from the EMBOSS28 package. Sequence alignments were displayed using the GeneDoc29 software. The phylogenetic tree was constructed by MEGA version 6.30 using default options and a

maximum-likelihood method, based on the proportion of amino acid sites at which the compared sequences were different, omitting alignment gaps as necessary using the pairwise-deletion option. The Poisson substitution model was used for distance calculation. The reliability of each branch was assessed using 1000 bootstrap replicates.

We divided the ORF15-encoded region into five repetitive glutamic acid-rich bins (aa729-767, aa768-849, aa850-940, aa941-981, aa982-1085) and three nonrepetitive flanking regions (aa586-681, aa682-728, aa1086-1152). The number of different insertions or deletion (indel) events was counted in each bin and normalized to the amino acid sequence length. The number of indels per amino acid was compared between repetitive and nonrepetitive regions using a 2-sample unequal variance *t*-test (2-tailed).

The Expression of *ZFRPGR2^{const}* and *ZFRPGR2^{ORF15}* in Zebrafish

Total RNAs were isolated from adult zebrafish tissues (kidney, muscle, ovary, liver, intestine, skin, eye, heart, brain, testis) and from various developmental stages (immature oocyte, oocyte, 32-cell, 256-cell, sphere, gastrula, bud, eye appearance, 24 hours, 2 days, 3 days) using an RNA purification kit (Agilent Technologies). The resultant total RNAs were reverse-transcribed using a cDNA synthesis kit (Roche Diagnostics, Ltd.). Polymerase chain reaction was carried out with the resultant cDNAs as templates using primers based on the published zebrafish *RPGR* sequences.¹⁷ The 621-bp fragment of *ZFRPGR2^{const}* (*ZFRPGR2^{ex1-17}*) transcript was amplified using the forward primer (5' CACATAATGAAGATGGACCCTAGC3') and reverse primer (5' TCAAAGCAGACTGCAAGTGGCTGAG3'). The 672-bp fragment of *ZFRPGR2^{ORF15}* transcript was amplified using forward primer (5' AAAAAGTCTGGATCTGAAAAGAC3') and reverse primer (5' ACTTTTATCCCTTCTGTTCCCGAC3'). A 517-bp fragment of β -actin was amplified as a control for PCR using forward primer (5' TGCCATGTATGTGGCCATCCA3') and reverse primer (5' ACCTCCAGACAGCACTGTGT3'). Each cycle of PCR started with 30 seconds of denaturation, 1 minute of primer annealing at 60°C, 1 minute of extension at 72°C with 35 cycles, and a final extension at 72°C for 10 minutes. Gel electrophoresis was carried out on 1% agarose gel.

To quantify the expression of *ZFRPGR2^{const}* and *ZFRPGR2^{ORF15}*, quantitative real-time PCR analysis was carried using a commercial kit (platinum SYBR Green QPCR SuperMix-UDG; Invitrogen, Paisley, UK). The two specific primers for *ZFRPGR2^{ORF15}* in exon 14 (5' GCAAAAAGTCTGGATCTGAAAAG3' and 5' TAAAGAAGTCACCTTGGTGAG3') and two specific primers for *ZFRPGR2^{const}* on exon 15 (5' GTCCCTGATGGACCAATGAAA3') and exon 17 (5' AAGCAGACTGCAAGTGGCTGA3') at the C-terminal region of the transcripts, amplified 258-bp and 315-bp fragments, respectively. Expression of 18S RNA was used for normalization; the primers used for 18S were 5' CCACTCCCGAGATCAACTA3' and 5' CAAATTACCCATTCCCGACA3'. The relative expression levels of the target (*ZFRPGR2^{ex1-17}* and *ZFRPGR2^{ORF15}*) and control housekeeping gene (18S rRNA) were determined using the 2^{-Ct} method. Statistical analysis was carried out using graphing and statistics software (Graph-Pad Prism; GraphPad Software, Inc., La Jolla, CA, USA). An

unpaired *t*-test was carried out to find out the expression difference between *ZRPGR2^{const}* and *ZRPGR2^{ORF15}*.

Results

RPGR Gene Structure in Vertebrates and Invertebrates

To determine the gene structure of *RPGR* in vertebrates and invertebrates, we analyzed genomic sequences from a variety of vertebrate and invertebrate species using human *RPGR* as a reference. Our analysis showed that all the vertebrate species examined have exons capable of encoding both the *RPGR^{const}* and *RPGR^{ORF15}* alternative splice forms described previously in human and mouse tissues.^{6–8,10,15} In contrast, the invertebrate species examined only had exons capable of encoding the *RPGR^{const}* splice form (Fig. 1). Structural analysis of all *RPGR* exons and introns showed that the *RPGR^{const}* splice form had between 15 exons in sea anemone and 19 exons in the human gene. The *Drosophila melanogaster* *dmRPGR* gene was the most divergent, with only five exons. The orthologous *RPGR^{const}* exon sizes were similar in vertebrate and invertebrate species, except for *Drosophila*, which had a large C-terminal exon. As expected, intron sizes in vertebrate species were generally larger than those in invertebrates (Fig. 1). A search for the unique C-terminal exon of the *RPGR^{ORF15}* splice form showed that all vertebrates but no invertebrates contained this large (~1 kb) repetitive glutamic acid-rich region and short basic C-terminal domain (Fig. 1B).

Conservation of *RPGR* Protein in Vertebrates and Invertebrates

The retinitis pigmentosa GTPase regulator exon ORF15 has a purine-rich domain, which contains most of the reported frameshift indel mutations found in XLRP, and encodes a peptide of 567 amino acids in humans (Fig. 2).^{7,22} The evolutionary conservation of this region was analyzed by aligning the amino acid sequences encoded by exon ORF15 in humans and eight other primate species using CLUSTALW (Fig. 2). The number of different sequence gaps (reflecting the minimum number of independent indel events) was counted within five repetitive sequence “bins” compared with three flanking nonrepetitive region bins (see Methods) across the nine primate species (Fig. 2). The results were normalized for bin sequence length. The result showed that there were 22.6 ± 10.0 gaps ($\pm$ SEM; per residue $\times 10^{-3}$; $n = 3$ bins) in the nonrepetitive regions compared with 102.1 ± 16.6 gaps ($\pm$ SEM; per residue $\times 10^{-3}$; $n = 5$ bins) in the repetitive region. This difference was statistically significant ($P < 0.007$) although provided only a rough estimate of the true evolutionary rate because: (1) alternative alignments were possible, (2) it ignored the possibility that apparently similar gap events occurred independently; and (3) it ignored amino acid substitutions (although these contribute virtually no XLRP mutations in the repeat region). However, it provided an underestimate of the rate of sequence evolution in these regions and is consistent with the proposal that ORF15 is rapidly evolving due to its inherent instability and is indeed a “hot spot” for mutation.⁷

The alternative splice form *RPGR^{ORF15}*, as defined by the presence of a C-terminal exon encoding a glutamic acid-rich region 300 to 400 amino acids long,^{7,22} was identified in all of the vertebrate but none of the invertebrate species examined. At the extreme C-terminus of *RPGR^{ORF15}* is a ~60 amino acid segment enriched for basic amino acids (aa1086–1152 in

the human sequence, Fig. 2), which binds two interacting proteins, nucleophosmin, and whirlin.^{20–22} This basic domain was well conserved in the 23 vertebrate species examined. This included nine primates (human, chimpanzee, cynomolgus monkey, gorilla, gibbon, African green monkey, rhesus monkey, marmoset, owl monkey), which showed very little sequence divergence in this region (Fig. 2). Strong conservation of the basic C-terminal region was also found in 14 other vertebrates, including carnivores (cat, dog), artiodactyls (pig, sheep, cow), lagomorpha (rabbit), rodents (mouse, rat, hamster), birds (chicken), amphibians (*Xenopus laevis*), and fish (zebrafish, medaka, stickleback) (data not shown).²⁰

In contrast, sequences encoding the RPGR^{const} isoform were present in all vertebrate and invertebrate species examined (Fig. 3). The only well-defined functional domain in the N-terminal region of both RPGR isoforms is the RCC1-like domain. This corresponds to a 7-bladed β -propeller (B1–B7) made up from seven RCC1-like sequence repeats: six of these blades (B2–B7) are made up from continuous RCC1-like sequence repeats, whereas blade 1 (B1) is made up from disparate sequence regions where the C and D β -strands are provided by the N-terminus (aa8–30), while the A and B β -strands are contributed by the C-terminal region (aa349–370) of the domain.³¹ This sequence and structure arrangement are highly similar to the RCC1 protein.¹¹ Previous published data indicated that the RPGR RCC1-like domain is conserved in vertebrates.¹⁶ In order to understand whether this RPGR domain varied within a larger group of vertebrates and invertebrates, we aligned the RCC1-like domain protein sequences from 18 representative vertebrate species and 16 invertebrate species. The results from all-against-all pairwise percentage sequence identity calculation (Supplementary Table S2) showed that conservation of the RCC1-like domain is high across all examined species. To illustrate this point, using the human RCC1-like domain sequence as a comparator, its sequence is 98% identical (highest) to gibbon and 53% identical (lowest) with one of the medaka sequences, among vertebrates; it is 53% identical to amphioxus (highest) and 40% identical to the butterfly (lowest) RPGR sequence, among invertebrates. A comparative assessment of the seven individual repeats indicates that the human sequence exhibits highest identity within repeat 2 (98% gibbon to 46% polychaete/louse) and highest variability within repeat 6 (98% gibbon to 25% butterfly); a large insertion is present within some invertebrates between B6-D and B7-A (Fig. 3). Within vertebrates, the individual gibbon repeat sequences are consistently the most identical (94%–100%) to human sequence, and one of the medaka or zebrafish sequences, the most variable (38%–65%). Surprisingly, among invertebrates, the limpet sequence displays the lowest degree of sequence identity to the human sequence within repeat 4 (29%), but the highest within repeat 6 (62%) (Supplementary Table S2).

A total of 32 disease-associated RPGR missense mutations have been reported in the RCC1-like domain,³² the majority of which are conserved or conservatively substituted among all the examined vertebrate and invertebrate species in the alignment. However, variable positions are also seen between vertebrates and invertebrates. For example, F130C is strictly conserved among all vertebrates, while it is replaced almost exclusively by a Proline residue among invertebrates (Fig. 3).⁸ Likewise, known disease-mutation R127G,⁹ is occupied by a positively charged or polar residue among vertebrates, but is highly variable among the invertebrates. These examples might reflect a divergence in function between vertebrate and invertebrate species (Fig. 3).

Finally, the presumptive $RPGR^{const}$ isoform in vertebrate and invertebrate species has a predicted isoprenylation signal motif at the C-terminus, which was conserved in all species examined (data not shown).^{6,8}

Phylogenetic Analysis

To understand the evolutionary history of the *RPGR* gene, we constructed a phylogenetic tree using *RPGR* RCC1-like domain protein sequences from 25 representative vertebrate and 16 invertebrate species (Fig. 4). The human HERC4 ubiquitin ligase contains an RCC1-like domain, which was used as an outgroup to root the tree, and was found to be only distantly related to the *RPGR* RCC1-like domains. In contrast, homologues of *RPGR* fall into well-supported clades, consistent with the known classification in vertebrates (primates, carnivores, artiodactyls, rodentia, birds, amphibians, and teleost fish). The invertebrate sequences were all well separated from the vertebrate *RPGR*s with the ancestral chordate amphioxus at the boundary between vertebrate and invertebrates. The results support the choice of putative *RPGR* RCC1-like domains as true orthologues of the human gene.

RPGR Gene Synteny

Shared synteny is a well-established criterion for identifying orthologous chromosomal regions in different species. We compared the genomic organization of our selected *RPGR* orthologues and found that the chromosome regions containing the syntenic gene arrangement (*OTC*, *RPGR*, *SRPX*) was similar in primates, carnivores, artiodactyls, rodentia, birds, amphibians and teleost fish (Fig. 5 and data not shown). Due to whole-genome duplication events in teleosts, there were two copies of the *RPGR* gene in zebrafish, medaka fish, and stickleback. One of the two *RPGR* orthologues in zebrafish, medaka fish, and stickleback shared syntenic genes with mammals, but not the other. The *Fugu* genomic sequence (scaffold 42; Fig. 5) shows loss of the putative *RPGR* orthologue in the region flanked by the usual syntenic genes (*OTC*-*RPGR*-*SRPX*), while the other *RPGR*-like *Fugu* gene (scaffold 57) shows the same syntenic relationships as the nonorthologous *RPGR*-like genes found in zebrafish (chr11), stickleback (grl), and Medaka fish (chr2). The true *RPGR* orthologue in *Fugu* therefore appears to have been lost during evolution. Most invertebrate species did not display shared synteny with vertebrate *RPGR* sequences except for the sea anemone *RPGR* which is flanked by the *OTC* gene, similar to most vertebrates (Fig. 5).

Analyses of *ZFRPGR2^{const}* and *ZFRPGR2^{ORF15}* Expression in Zebrafish Development and Adult Tissues

To further understand the relative expression levels and functional relationships of the *RPGR^{const}* and *RPGR^{ORF15}* transcripts, we assessed their relative expression levels in zebrafish. Both *ZFRPGR2^{const}* (*ZFRPGR2^{ex1-17}*) and *ZFRPGR2^{ORF15}* expression was detected during zebrafish embryogenesis and early development (Fig. 6A). High expression levels of *ZFRPGR2^{const}* were detected in immature oocyte, oocyte and 32-cell stage embryos, declining thereafter, perhaps reflecting the importance of cilia at these development stages (Fig. 6B). *ZFRPGR2^{ORF15}* expression was more similar across developmental stages, although declined to expression levels similar to *ZFRPGR2^{const}* after the sphere stage (Fig. 6B). Expression of *ZFRPGR2^{const}* and *ZFRPGR2^{ORF15}* in adult tissues were examined in total RNA isolated from kidney, muscle, ovary, liver, intestine, skin, eye,

heart, brain, and testis. Expression of *ZRPGR2^{ORF15}* was readily detected in all the examined tissues. Expression of *ZFRPGR2^{const}* was readily detected in muscle, ovary, intestine, skin, brain, and testis, but was only just detectable in kidney, liver, eye, and heart (Fig. 6C). Expression levels of *ZFRPGR2^{ORF15}* were significantly higher than those of *ZFRPGR2^{const}* in the eye ($P < 0.0001$), brain, and testis (Fig. 6D). Eye showed the highest level of *ZFRPGR2^{ORF15}*, followed by brain and testis. In adults, the expression level of *ZFRPGR2^{const}* was similar across tissues.

Discussion

The major finding in this study is that the alternatively spliced RPGR^{ORF15} isoform is confined to vertebrates, since the defining C-terminal exon ORF15 is absent from all invertebrates examined, suggesting a role that is unique to the ciliary-derived photoreceptors of “simple” vertebrate eyes, as opposed to the rhabdomeric photoreceptors that are typical of “compound” invertebrate eyes.³³ The common ancestor of vertebrates and invertebrates may have used both types of photoreceptor, as in some urbilateria such as the marine ragworm, but with few exceptions the association of cilia with vertebrate photoreceptors holds true.^{34,35}

There are features in both the RPGR^{const} and RPGR^{ORF15} isoforms that are consistent with a shared function, especially the strong conservation of the RCC1-like domain. This domain encodes a seven-bladed β -propeller structure (Fig. 3),³¹ which is well conserved throughout vertebrate and invertebrate evolution. The function of this domain is not yet clear but it interacts with at least two proteins: the first is RPGRIP1, which appears to be a ciliary scaffold,^{36,37} and the second is PDE6 (PDE6D), a prenyl binding protein^{31,38} which solubilizes prenylated protein cargos from membranes and shuttles them to the photoreceptor connecting cilium (see Ref. 22 for review). Together, the two proteins have been proposed to regulate protein cargoes unloading and/or transport at the base of the connecting cilium.^{31,37} Both proteins have been proposed to affect photoreceptor ciliary function, but the RCC1-like domain is not unique to the RPGR^{ORF15} isoform, which appears to be of particular importance in the retina, both because of its high expression in this tissue and because retinal disease-causing mutations are all present in this transcript.^{7–9,22}

The photoreceptor cilium is unusual in the high volume cargoes that are transported through the cilium en route to the outer segment, especially rhodopsin, which occupies 50% of rod disk space at a density of 30,000 to 50,000 molecules/ μm^2 .³⁹ There is evidence that regulation of rhodopsin transport is one of the primary functions of RPGR (see Ref. 22 for review). In vertebrates, visual sensitivity is directly proportional to the amount of outer segment rhodopsin, whereas in invertebrates it is much less dependent on rhodopsin concentration.⁴⁰ In vertebrate rods, rhodopsin is present in flattened membranous disks rather than in microvilli, as in invertebrates. Disks can hold more protein than microvilli and the density of rhodopsin in rods is approximately ten times higher than in the microvillar membranes of *Drosophila*, which increases the efficiency of light capture accordingly.⁴¹ Rhodopsin also has a unique structural role in vertebrates since rhodopsin knockout mice fail to develop rod outer segments.⁴²

Vertebrate photoreceptors may therefore have required a series of adaptations to the ancestral primary cilium, one of which was perhaps the evolution of the RPGR^{ORF15} isoform. The expression pattern of the zebrafish RPGR^{const} and RPGR^{ORF15} isoforms is consistent with this, since RPGR^{ORF15} was most strongly expressed in the adult eye, whereas RPGR^{const} isoform was most strongly expressed in the oocyte and early cleavage stages of the embryo where motile cilia, which contain RPGR^{const},¹⁵ are important (Fig. 6). In contrast to the RCC1-like domain, the basic C-terminal domain, which is unique to RPGR^{ORF15}, interacts with at least two proteins: the chaperone nucleophosmin (NPM)²⁰ and the PDZ-containing scaffold protein whirlin.²¹ Neither of these proteins is uniquely expressed in vertebrate photoreceptors, but NPM is present in metaphase centrosomes while whirlin is enriched in ciliary structures within the eye and ear.

The unusual structure of the repetitive glutamate/glycinerich region within C-terminal exon ORF15 of RPGR^{ORF15} deserves comment. This is present but poorly conserved in vertebrates and, in humans, it includes 27 imperfect direct AG repeats with 97.5% purines within a ~1-kb region encoding the consensus peptide motif EEEGEGEGE.⁷ Despite difficulties in aligning the repetitive segment, its rate of indels within primate lineages is surprisingly high relative to flanking unique sequences (Fig. 2). This suggests that this region is rapidly evolving and, despite being a coding region, that it can tolerate a high rate of indels, provided that the reading frame and acidic charge are maintained.

The acidic repeat region was previously shown to be a mutation “hotspot” for X-linked retinal dystrophies.^{7,22} This type of homo(purine/pyrimidine) (R/Y) tract forms stable triplexes that can block replication and transcription and so tends to promote genetic rearrangements, perhaps promoting recombination or genome diversity in evolution,⁴³ explaining the high rate of evolutionary change seen even within primate lineages (Fig. 2). The function of this repeat-region is unknown but it is typical of unstructured or intrinsically disordered regions⁴⁴ which tend to show a strong net charge and commonly interact with protein or nucleic acid partners, often with significant avidity via multiple short repetitive motifs or posttranslationally modified sites, for example, FG repeats in nucleoporins.⁴⁵ The acidic domain of ORF15 contains typical 3 to 10 amino acid motif repeats, which often facilitate the formation of new and diverse multi-protein complexes, serving as docking platforms, as in BRCA1.⁴⁶ The glutamic acid-rich domain has no known binding partners as yet. Although this domain is not easily aligned across orthologues and is poorly conserved at the individual amino acid level (Fig. 2), the overall charge and repeat structure length are conserved in vertebrates,⁷ suggesting that binding partners or complexes involving this domain remain to be identified.

In summary, RPGR is a multidomain protein whose functions remain to be fully elucidated but the RPGR^{ORF15} isoform appears to be unique to vertebrates, perhaps because it is necessary for the docking or transport of high-volume cargoes such as rhodopsin through the connecting cilium to the outer segment, a feature that is unique to the ciliary photoreceptors of vertebrates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Royal Society of London, TENOVUS Scotland, the Medical Research Council, National Eye Research Centre, Visual Research Trust, Fight for Sight, RP Fighting Blindness, the W.H. Ross Foundation, the Rosetrees Trust, and the Yorkhill Children's Charity for supporting this work. We also thank Chris Ponting, PhD (University of Oxford, Oxford, UK), for advice and suggestions; Ulrich Technau, PhD, for gifting sea anemone samples (University of Vienna, Vienna, Austria); and Sebastian Shimeld, PhD (University of Oxford), for gifting amphioxus sample. The Open Access was supported by the Rosetrees Trust.

References

1. Wright AF, Chakarova CF, Abd El-Aziz M, Bhattacharya SS. Photoreceptor degeneration - genetic and mechanistic dissection of a complex trait. *Nat Rev Genet.* 2010; 11:273–284. [PubMed: 20212494]
2. Bramall AN, Wright AF, Jacobson SG, McInnes RR. The genomic, biochemical and cellular responses of the retina in inherited photoreceptor degenerations, and prospects for the treatment of these disorders. *Annu Rev Neurosci.* 2010; 33:441–472. [PubMed: 20572772]
3. Weleber RG. Inherited and orphan retinal diseases: phenotypes, genotypes, and probable treatment groups. *Retina.* 2005; 25:S4–S7. [PubMed: 16374328]
4. Bunker CH, Berson EL, Bromley WC, Hayes RP, Roderick TH. Prevalence of retinitis pigmentosa in Maine. *Am J Ophthalmol.* 1984; 97:357–65. [PubMed: 6702974]
5. Breuer DK, Yashar BM, Filippova E, et al. A comprehensive mutation analysis of RP2 and RPGR in a North American cohort of families with X-linked retinitis pigmentosa. *Am J Hum Genet.* 2002; 70:1545–1554. [PubMed: 11992260]
6. Meindl A, Dry K, Herrmann K, et al. A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nat Genet.* 1996; 13:35–42. [PubMed: 8673101]
7. Vervoort R, Lennon A, Bird AC, et al. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat Genet.* 2000; 25:462–466. [PubMed: 10932196]
8. Roepman R, van Duijnhoven G, Rosenberg T, et al. Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. *Hum Mol Genet.* 1996; 5:1035–1041. [PubMed: 8817343]
9. Sharon D, Sandberg MA, Rabe VW, Stillberger M, Dryja TP, Berson EL. RP2 and RPGR mutations and clinical correlations in patients with X-linked retinitis pigmentosa. *Am J Hum Genet.* 2003; 73:1131–1146. [PubMed: 14564670]
10. Kirschner R, Rosenberg T, Schultz-Heienbrok R, et al. RPGR transcription studies in mouse and human tissues reveal a retina-specific isoform that is disrupted in a patient with X-linked retinitis pigmentosa. *Hum Mol Genet.* 1999; 8:1571–1578. [PubMed: 10401007]
11. Renault L, Nassar N, Vetter I, et al. The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller. *Nature.* 1998; 392:97–101. [PubMed: 9510255]
12. Shu X, Black GC, Rice JM, et al. RPGR mutation analysis and disease: an update. *Hum Mutat.* 2007; 28:322–328. [PubMed: 17195164]
13. Zhang Q, Acland GM, Wu WX, et al. Different RPGR exon ORF15 mutations in Canids provide insights into photoreceptor cell degeneration. *Hum Mol Genet.* 2002; 11:993–1003. [PubMed: 11978759]
14. Mavlyutov TA, Zhao H, Ferreira PA. Species-specific subcellular localization of RPGR and RPGRIP isoforms: implications for the phenotypic variability of congenital retinopathies among species. *Hum Mol Genet.* 2002; 11:1899–1907. [PubMed: 12140192]

15. Hong DH, Pawlyk B, Sokolov M, et al. RPGR isoforms in photoreceptor connecting cilia and the transitional zone of motile cilia. *Invest Ophthalmol Vis Sci*. 2003; 44:2413–2421. [PubMed: 12766038]
16. Shu X, Zeng Z, Eckmiller MS, et al. Developmental and tissue expression of *Xenopus laevis* RPGR. *Invest Ophthalmol Vis Sci*. 2006; 47:348–356. [PubMed: 16384984]
17. Shu X, Zeng Z, Gautier P, et al. Zebrafish *Rpgr* is required for normal retinal development and plays a role in dynein-based retrograde transport processes. *Hum Mol Genet*. 2010; 19:657–670. [PubMed: 19955120]
18. Shu X, McDowall E, Brown AF, Wright AF. The human retinitis pigmentosa GTPase regulator gene variant database. *Hum Mutat*. 2008; 29:605–608. [PubMed: 18361418]
19. Thompson DA, Khan NW, Othman MI, et al. Rd9 is a naturally occurring mouse model of a common form of retinitis pigmentosa caused by mutations in RPGR-ORF15. *PLoS One*. 2012; 7:e35865. [PubMed: 22563472]
20. Shu X, Fry AM, Tulloch B, et al. RPGR ORF15 isoform colocalizes with RPGRIP1 at centrioles and basal bodies and interacts with nucleophosmin. *Hum Mol Genet*. 2005; 14:1183–1197. [PubMed: 15772089]
21. Wright RN, Hong DH, Perkins B. *Rpgr*ORF15 connects to the usher protein network through direct interactions with multiple whirlin isoforms. *Invest Ophthalmol Vis Sci*. 2012; 53:1519–1529. [PubMed: 22323458]
22. Megaw RD, Soares DC, Wright AF. RPGR: its role in photoreceptor physiology, human disease, and future therapies. *Exp Eye Res*. 2015; 138:32–41. [PubMed: 26093275]
23. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25:3389–3402. [PubMed: 9254694]
24. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res*. 2004; 14:988–995. [PubMed: 15123596]
25. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 1997; 268:78–94. [PubMed: 9149143]
26. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32:1792–1797. [PubMed: 15034147]
27. Higgins DG, Thompson JD, Gibson TJ. Using CLUSTALW for multiple sequence alignments. *Methods Enzymol*. 1996; 266:383–402. [PubMed: 8743695]
28. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000; 16:276–277. [PubMed: 10827456]
29. Nicholas KB, Nicholas HB Jr, Deerfield DW II. GeneDoc: analysis and visualization of genetic variation. *EMBnet News*. 1997; 4:14–17.
30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013; 30:2725–2729. [PubMed: 24132122]
31. Wätzlich D, Vetter I, Gotthardt K, et al. The interplay between RPGR, PDE6 and Arl2/3 regulate the ciliary targeting of farnesylated cargo. *EMBO Rep*. 2013; 14:465–472. [PubMed: 23559067]
32. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014; 133:1–9. [PubMed: 24077912]
33. Land MF. The optical structures of animal eyes. *Curr Biol*. 2005; 15:R319–R323. [PubMed: 15893273]
34. Arendt D, Tessmar-Raible K, Snyman H, Dorresteijn AW, Wittbrodt J. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science*. 2004; 306:869–871. [PubMed: 15514158]
35. Kozmik Z, Ruzickova J, Jonasova K, et al. Assembly of the cnidarian camera-type eye from vertebrate-like components. *Proc Natl Acad Sci U S A*. 2008; 105:8989–8993. [PubMed: 18577593]
36. Patil H, Guruju MR, Cho KI, et al. Structural and functional plasticity of subcellular tethering, targeting and processing of RPGRIP1 by RPGR isoforms. *Biol Open*. 2012; 1:140–160. [PubMed: 23213406]

37. Remans K, Bürger M, Vetter IR, Wittinghofer A. C2 domains as protein-protein interaction modules in the ciliary transition zone. *Cell Rep.* 2014; 8:1–9. [PubMed: 24981858]
38. Zhang H, Li S, Doan T, et al. Deletion of PrBP/delta impedes transport of GRK1 and PDE6 catalytic subunits to photoreceptor outer segments. *Proc Natl Acad Sci U S A.* 2007; 104:8857–8862. [PubMed: 17496142]
39. Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K. Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature.* 2003; 421:127–128. [PubMed: 12520290]
40. Goldsmith TH. Photoreceptor processes: some problems and perspectives. *J Exp Zool.* 1975; 194:89–101. [PubMed: 453]
41. Fain GL, Hardie R, Laughlin SB. Phototransduction and the evolution of photoreceptors. *Curr Biol.* 2010; 20:R114–R124. [PubMed: 20144772]
42. Humphries MM, Rancourt D, Farrar GJ, et al. Retinopathy induced in mice by targeted disruption of the rhodopsin gene. *Nat Genet.* 1997; 15:216–219. [PubMed: 9020854]
43. Bacolla A, Collins JR, Gold B, et al. Long homopurine*homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Res.* 2006; 34:2663–2675. [PubMed: 16714445]
44. Latysheva NS, Flock T, Weatheritt RJ, Chavali S, Babu MM. How do disordered regions achieve comparable functions to structured domains? *Protein Sci.* 2015; 24:909–922. [PubMed: 25752799]
45. Labokha AA, Gradmann S, Frey S, et al. Systematic analysis of barrier-forming FG hydrogels from *Xenopus* nuclear pore complexes. *EMBO J.* 2013; 32:204–218. [PubMed: 23202855]
46. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys.* 2008; 37:215–246. [PubMed: 18573080]

Figure 1.

Exon-intron structure of the *RPGR* gene orthologues in vertebrates and invertebrates. (A) $RPGR^{const}$ ($RPGR^{ex1-19}$) isoform and (B) $RPGR^{ORF15}$ isoform. Sizes (in nucleotides) of exons (drawn to scale) and introns are indicated above and below each box respectively. The genomic sequence for each species was obtained through NCBI and Ensembl. *Black* indicates the coding RCC1-like domain and *white* indicates other coding regions.

Figure 2.

Alignment of RAGRORF15 amino acid sequences from nine primate species showing high levels of insertion/deletion events within the central repetitive region compared with the nonrepetitive regions flanking it. Glutamic acid (*green shading*) and glycine (*red shading*) amino acids are highlighted. The central repetitive region was variable in length, consistent with a high rate of indels during primate evolution over the past 65 to 70 million years.

Figure 3.

Alignment of RPGR RCC1-like domain amino acid sequences from vertebrates and invertebrates. Vertebrate species are shown within the *brown box* and invertebrates within the *green box*. Conserved residues are shown with a *red background*. The seven blades that form the RCC1-like domain β -propeller is shown and numbered based upon structure. The first blade is made up from two different regions of sequence. The corresponding secondary structure assignment from PDB ID: 4JHN32 is also shown above the alignment (*arrows*: β -strands, *squiggles*: α -helices; *TT*: hydrogen-bonded turns) along with solvent accessibility

of each position ('acc') shown below the alignment (*white*: buried residues; *cyan*: intermediate residues; *blue*: accessible residues). Positions of known disease-causing missense mutations³² are indicated by *red stars*.

Figure 4.

Phylogenetic analysis of RPGR RCC1-like domain amino acid sequences from 25 vertebrate (*red*) and 16 invertebrate (*green*) species carried out using MEGA software and maximum likelihood methods. This shows that the chosen orthologues are consistent with accepted phylogenetic relationships.

Figure 5.

Syntenic analysis of chromosomal loci bearing the *RPGR* gene in vertebrates and invertebrates. Genes are color-coded and their orientations shown by the *arrows*.

Figure 6.

(A) Temporal expression of *ZFRPGR2^{ORF15}* and *ZFRPGR^{const}* (*ZFRPGR2^{ex1-17}*) in zebrafish development detected by RT-PCR. (B) Expression levels of *ZFRPGR2^{const}* and *ZFRPGR2^{ORF15}* in oocytes and at different developmental stages. **** $P < 0.0001$. (C) Expression of *ZFRPGR2^{ORF15}* and *ZFRPGR2^{const}* among different tissues of zebrafish detected by RT-PCR. (D) Relative expression levels of *ZFRPGR2^{ORF15}* and *ZFRPGR2^{const}* in different zebrafish tissues. **** $P < 0.0001$; * $P < 0.001$.