

SHORT REPORT

Three novel human VMD2-like genes are members of the evolutionary highly conserved RFP-TM family

Heidi Stöhr¹, Andreas Marquardt¹, Indrajit Nanda¹, Michael Schmid¹ and Bernhard HF Weber^{*,1}

¹Institut für Humangenetik, Universität Würzburg, D-97074 Würzburg, Germany

The RFP-TM protein family was first described in *Caenorhabditis elegans* as hypothetical transmembrane proteins containing a conserved 350–400 amino acid domain including the invariant peptide motif RFP. The VMD2 gene underlying Best disease was shown to represent the first human member of the RFP-TM protein family. More than 97% of the disease-causing mutations are located in the N-terminal RFP-TM domain implying important functional properties. Here, we have identified three novel VMD2-related human genes (VMD2L1, VMD2L2 and VMD2L3) demonstrating a high degree of conservation in their respective RFP-TM domains. Each of the VMD2-like proteins has a unique C-terminus that lack similarity to other proteins or motifs. By FISH analysis, VMD2L1 was localised to chromosome 19p13.2-p13.12, VMD2L2 to 1p32.3-p33 and VMD2L3 to 12q14.2-q15. RT-PCR analyses revealed tissue-restricted expression of the three genes with both VMD2L1 and VMD2L2 abundantly transcribed in colon. VMD2L1 is present in the retinal pigment epithelium while VMD2L3 shows predominant expression in skeletal muscle.

European Journal of Human Genetics (2002) 10, 281–284. DOI: 10.1038/sj/ejhg/5200796

Keywords: RFP-TM protein family; VMD2; Best vitelliform macular dystrophy

Introduction

To functionally classify protein sequences predicted from the *Caenorhabditis elegans* genome, more than 7000 proteins have been clustered into domain families on the basis of multiple sequence alignments.¹ One of these clusters is described as worm family 8 and assembles 26 paralogous transmembrane (TM) spanning proteins that share a homology region of 350–400 amino acids (aa) with a high content of aromatic residues including an invariant arginine (R), phenylalanine (F) and proline (P) motif.

Previously, we and others have cloned the gene underlying the autosomal dominant vitelliform macular dystrophy (VMD2; Best disease).^{2,3} The VMD2 gene on chromosome 11q13 consists of 11 exons including a 5' untranslated exon and encodes a 585 aa protein, named bestrophin, which is

predominantly expressed in the retinal pigment epithelium (RPE).^{2,3} Sequence analysis of bestrophin revealed significant homology across the entire RFP-TM domain.^{2,3} Recent biochemical and immunohistochemical data provide confirmatory evidence for the integration of bestrophin in the plasma membrane of the RPE.⁴

With only one possible exception of a frameshift mutation in exon 10, all other known variants causing Best disease are affecting the RFP-TM domain of bestrophin suggesting a critical role of this domain in the etiology of the disorder.^{2–3,5} In addition, the majority of alterations (>97%) represent missense mutations which modify residues highly conserved between VMD2 and related proteins from worm and fly.

In this study we report the identification of three novel putative human proteins closely related to bestrophin, thus constituting a new subfamily of RFP-TM proteins.

Materials and methods

Bioinformatics

Genomic sequences were examined with the NIX application (<http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/>).

*Correspondence: Bernhard H. F. Weber, Institut für Humangenetik, Biozentrum, Am Hubland, D-97074 Würzburg, Germany.

Tel: 49 931 888 4062; Fax: 49 931 888 4069;

E-mail: bweb@biozentrum.uni-wuerzburg.de

Received 30 November 2001; revised 11 February 2002; accepted 14 February 2002

BLAST programs at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) were used for homology searches. Pattern and profile searches were performed with the SMART tool at EMBL (<http://smart.embl-heidelberg.de/>). Multiple sequence alignments were done with the ClustalW1.8 program, (<http://www.ebi.ac.uk/clustalw/>) and shading of the aligned sequences was achieved with BOXSHADE 3.21 (http://www.isrec.isb-sib.ch:8080/software/BOX_form.html).

Chromosomal localisation and expression analysis

Fluorescent in situ hybridisation (FISH) was performed as described earlier.⁶ Total RNA from human retina and RPE tissue was isolated and reverse transcribed as described elsewhere.⁷ Total RNA from the remaining tissues were purchased from BD Biosciences Clontech (Palo Alto, USA). The synthesised cDNA served as template for subsequent PCR assays using intron-spanning primer pairs specific for each gene (sequence information on primer pairs upon request).

Results and Discussion

BLASTX searches in the human draft sequence with the 1758 bp coding cDNA of VMD2 revealed homologous regions that correspond to exons 2 to 9 of the VMD2 gene on three distinct chromosomes including #1, #12 and #19 (GenBank Acc. nos AL592166, AC025263 and AC018761, respectively). The sequences encode three entire RFP-TM domains starting with putative translation initiation codons (ATG) that are found within a context that conform well to the Kozak consensus sequence. No homologies to exons 10 and 11 of the VMD2 gene were found.

Based on the VMD2 homology and multiple-algorithm exon predictions, eight putative exons were defined for each of the three novel VMD2-like genes, named VMD2L1, VMD2L2, VMD2L3. To complete the respective 3' ends of the coding sequences of the VMD2-like genes we applied an EST assembly strategy.

In dbEST, VMD2L1 is represented by two overlapping ESTs (AA573517, AA621745). The corresponding cDNA clones were completely sequenced confirming the sequence of exons 3 to 8. In addition, a colon cDNA clone (AK000139) extended the VMD2L1 transcript by a novel exon 9. In total, 1908 bp of VMD2L1 cDNA sequence were assembled consisting of 1530 bp of coding sequence and 378 bp of 3' untranslated region (UTR). The ORF encodes a putative 509 aa protein with a calculated molecular weight (MW) of 57.1 kDa.

The eight RFP-TM-related exons of VMD2L2 are partially covered by human ESTs. One EST (AW117683) contains the first two exons, another EST (BI762068) encompasses part of exon 8. Complete sequencing of clone wz24f11 confirmed the sequences corresponding to exon 6 to 8 and extended the VMD2L2 transcript by an additional 3' exon. Overall, the assembled VMD2L2 cDNA sequence is 2045 bp long and contains 623 bp of 3' UTR. The ORF of 1422 bp encodes a putative protein of 473 aa with a calculated MW of 53.5 kDa.

BLASTN searches with the available sequences of VMD2L3 identified ESTs which represent six of the eight predicted exons (exons 3 to 6, exons 8 and 9). Similar to the other VMD2-like genes, one EST (BE793366) was found to extend the transcript by a novel downstream exon. In addition, a 5' untranslated exon was detected in various ESTs. The assembled 1506 bp transcript consists of a 5' UTR of 172 bp in size, a 1197 bp coding region and a 3' UTR of 137 bp. The putative 398 aa protein has a calculated MW of 46.6 kDa. Alignment of the ESTs to genomic sequence revealed putative splice variants of VMD2L3. For example, all five ESTs containing the 5' untranslated exon skip the first (AL598355) or the first and second coding exon (e.g. BC006440, AV757060) and are spliced to either of two different sequences located in intron 4. Translation of these putative mRNAs leads to truncated 95 or 72 aa VMD2L3 protein isoforms with an alternative start codon in exon 4 and altered C-termini. The isoforms lack putative TM helices. Another group of ESTs was found to skip exon 7 (e.g. BE793366) causing an internal deletion of 30 aa. The functional relevance of these transcripts remains to be determined.

VMD2 and its novel human relatives VMD2L1, VMD2L2 and VMD2L3 share a conserved gene structure exemplified by almost identical sizes of the eight RFP-TM domain encoding exons and the highly conserved positions of their corresponding exon-intron boundaries (Figure 1). The elongation of exon 6 in VMD2L2 is caused by an internal insertion of 45 bp, thus retaining the conserved splice sites. Each of the four paralogous genes contains a unique 3' end of variable length. The length of the respective introns show greater variability giving rise to genomic loci with distinct sizes for VMD2L1 (5.9 kb), VMD2L2 (4.1 kb), VMD2L3 (54.4 kb) and VMD2 (11.5 kb).

Pairwise protein sequence comparison between bestrophin and the three putative VMD2-like proteins display an overall 60–67% identity and 73–81% similarity for the respective RFP-TM domains (Figure 1). Seventy-seven of 175 invariant residues are also identical in 70% of the 34 non-mammalian RFP-TM proteins. This remarkable conservation is not only restricted to the compositionally-biased TM regions within the RFP-TM segment. One of the most highly conserved regions is found C-terminal to the last putative TM helix and includes a region enriched in charged residues defining a novel motif KVAE-x-L-[IL]-NP-[FLM]-GEDDDDFE-[TFLVC]-N-x(2)-[IVL]-DRN (Figure 1). Interestingly, 22 of the known 82 distinct disease causing mutations in the VMD2 gene affect one of these amino acids further supporting a functional significance of this sequence. None of the C-termini of the VMD2-like proteins show significant homology to known proteins or motifs.

The PAC/BAC clones containing the VMD2-like genes have been assigned to chromosome 1 (VMD2L2), chromosome 12 (VMD2L3) and chromosome 19 (VMD2L1). FISH to human metaphase spreads confirmed and refined the localisation of

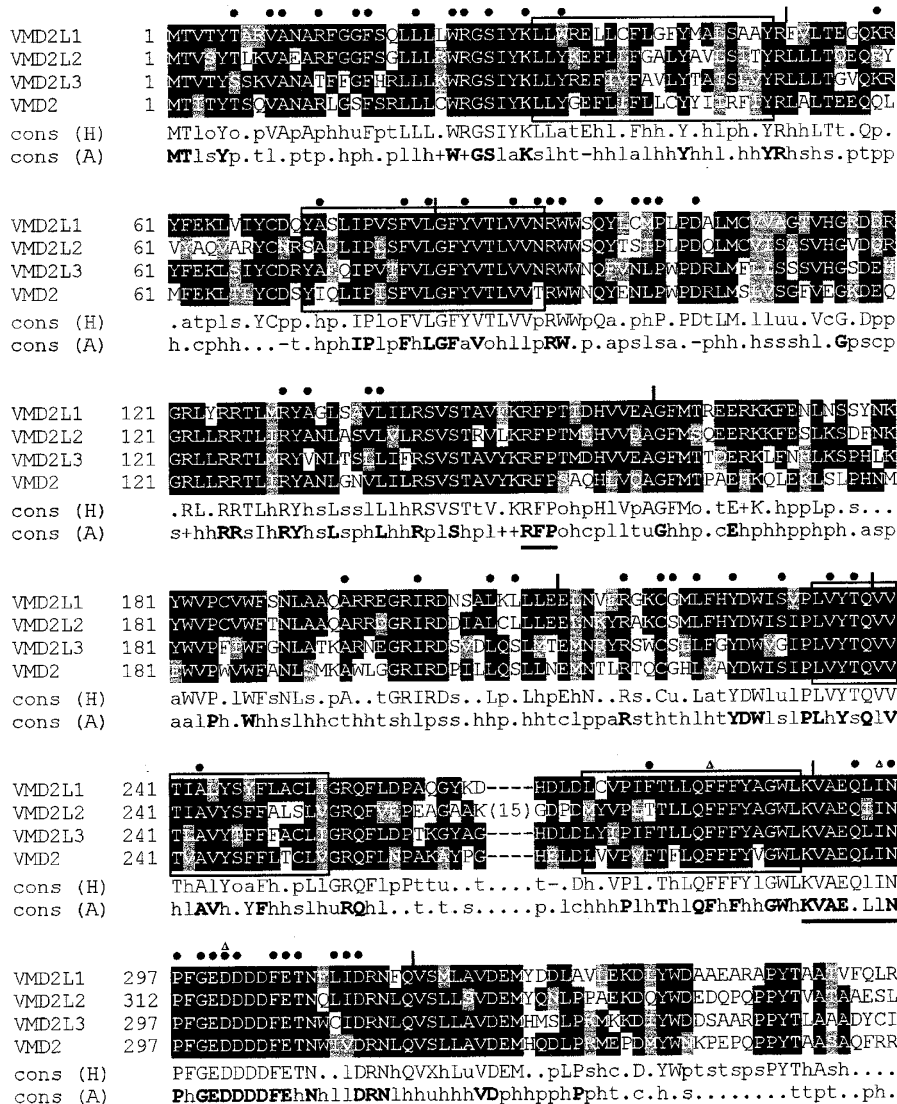


Figure 1 Multiple sequence alignment of the RFP-TM domains of human VMD2L1, VMD2L2, VMD2L3 and VMD2. Conserved exon/intron boundaries are shown by vertical bars. Predicted TM regions are framed. Residues affected by mutations in Best disease patients are indicated by black dots (missense mutations) or open triangles (deletions). Note that several distinct mutations may affect a single codon. The consensus sequence of the four human proteins (100% identity) [cons (H)] as well as of all 34 known RFP-TM proteins (>70% identity) [cons (A)] is given below the alignment. Invariant residues are shown in bold. Sidegroups of other conserved residues are indicated (o, alcohol; l, aliphatic; a, aromatic; c, charged; h, hydrophobic; -, negatively charged, +, positively charged, p, polar, s, small, u, tiny, t, turnlike). The invariant RFP motif and the highly conserved KVAE-x-L-[IL]-NP-[FLM]-GEDDDDFE-[TFLVC]-N-x(2)-[IVL]-DRN sequence are underlined.

VMD2L1 to chromosome 19p13.2-p13.12, VMD2L2 to 1p32.3-p33 and VMD2L3 to 12q14.2-q15 (Figure 2a).

RT-PCR analysis demonstrated tissue-restriction of the VMD2-like transcripts (Figure 2b). The transcription of VMD2L1 was mainly confined to the RPE and colon and the VMD2L2 expression was predominantly observed in colon and weakly in fetal brain, spinal cord, retina, lung, trachea, testis and placenta. VMD2L3 was strongly present in skeletal muscle RNA and weaker in brain, spinal cord, bone marrow and retina as well as thymus and testis (Figure 2b).

To date, one can only speculate regarding the functions of the RFP-TM proteins. Our expression study demonstrated that three of the four human proteins are abundantly transcribed in RPE and/or colon. These tissues consist of polarised epithelial cells whose plasma membranes are divided into distinct apical and basolateral surfaces. The basolateral membrane communicates with neighbouring cells as well as the extracellular matrix via membrane proteins. Interestingly, bestrophin has been demonstrated to be specifically localised to the basolateral membrane of the

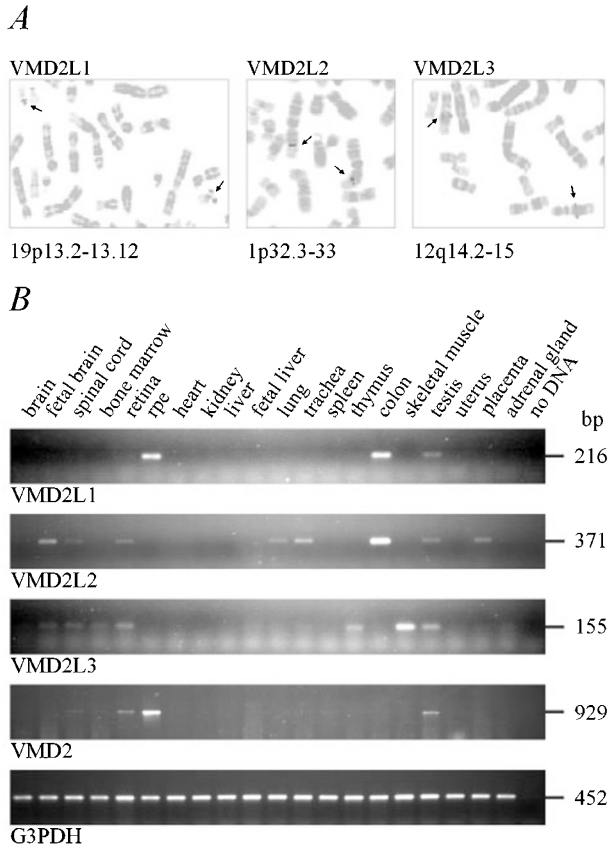


Figure 2 Chromosomal location and expression profile of the human VMD2-like genes. **(A)** FISH and 4'-6'-diamino-2-phenylindole counterstaining on human metaphase spreads. **(B)** RT-PCR analysis in 20 human tissues. Amplification of G3PDH served as a control to assess RNA integrity.

RPE.⁴ It is therefore conceivable, that RFP-TM proteins may exert similar functions in cells with a polarised phenotype, possibly as receptors involved in a vectorially oriented transport of molecules.

Acknowledgments

This work was supported in part by a grant from the Bundesministerium für Bildung und Forschung (BMBF) under 01KW9921 and the Deutsche Forschungsgemeinschaft (DFG) (WE 1259/13-1). The sequence data reported in this paper have been submitted to GenBank and have been assigned the accession numbers AF440756 (cDNA sequence for VMD2L1), AF440757 (cDNA sequence for VMD2L2) and AF440758 (cDNA sequence for VMD2L3).

References

- 1 Sonnhammer EL, Durbin R: Analysis of protein domain families in *Caenorhabditis elegans*. *Genomics* 1997; **46**: 200–216.
- 2 Marquardt A, Stöhr H, Passmore LA *et al*: Mutations in a novel gene, VMD2, encoding a protein of unknown properties cause juvenile-onset vitelliform macular dystrophy (Best's disease). *Hum Mol Gen* 1998; **7**: 1517–1525.
- 3 Petrukhin K, Koisti MJ, Bakall B *et al*: Identification of the gene responsible for Best macular dystrophy. *Nat Genet* 1998; **19**: 241–247.
- 4 Marmorstein AD, Marmorstein LY, Rayborn M *et al*: Bestrophin, the product of the Best vitelliform macular dystrophy gene (VMD2), localises to the basolateral plasma membrane of the retinal pigment epithelium. *Proc Natl Acad Sci USA* 2000; **97**: 12758–12763.
- 5 VMD2 Mutation Database at <http://www.uni-wuerzburg.de/humangenetics/vmd2.html>
- 6 Kohler MR, Vogt PH: Interstitial deletions of repetitive DNA blocks in dicentric human Y chromosomes. *Chromosoma* 1994; **103**: 324–330.
- 7 Stöhr H, Mah N, Schulz H *et al*: EST mining of the UniGene dataset to identify retina-specific genes. *Cytogenet Cell Genet* 2000; **91**: 267–277.