Molecular Vision 2002; 8:271-9 <http://www.molvis.org/molvis/v8/a34>
Received 31 May 2002 | Accepted 19 July 2002 | Published 22 July 2002

Did myocilin evolve from two different primordial proteins?

Arijit Mukhopadhyay, Arnab Gupta, Saibal Mukherjee, Keya Chaudhuri, Kunal Ray

Indian Institute of Chemical Biology, Human Genetics & Genomics Division, Jadavpur, Kolkata - 700 032, India

Purpose: Myocilin, a 57 kDa glycoprotein, has been of much interest because of its association with primary open-angle glaucoma, lack of understanding of its biological function, and sequence homology of its N- and C-termini with two distinctly different proteins, myosin and olfactomedin, respectively. In that context, the molecular evolution of myocilin was investigated.

Methods: The human myocilin protein was used as query in sequence alignment program and the similar protein sequences were searched in the protein databases for different species. The secondary structure analysis of human myocilin and the prediction of disulfide bonded cysteine residues in the protein were done using PSIPRED and CYSPRED software, respectively. Presence of putative motifs in the protein sequences was determined using the ScanProsite tool with the option of including patterns with the high probability of occurrence. The phylogenetic analyses of human, mouse, rat, and bovine myocilin were done at the DNA Data Bank of Japan (DDBJ) server.

Results: It was observed that while two different protein sequences from *Drosophila melanogaster* contained significant homology with either C-terminal or N-terminal of myocilin, a single protein from *Xenopus laevis* showed homology covering entire C-terminal and most of the N-terminal region of myocilin. These observations are noteworthy in the context of the previously reported homology of N-terminal domain of myocilin with a single protein (non-muscle myosin) in *Dictyostelium discoideum*, and C-terminal domain with a single protein (olfactomedin-like) in *Caenorhabditis elegans*, both representing lower organisms. Further, specific amino acids and putative functional motifs were observed to be conserved between the two sets of proteins having homology to two ends (N- and C-termini) of myocilin. At the secondary structure level, myocilin shows two distinctly different domains: (a) the N-terminal region is primarily α-helical type, and (b) the C-terminal region consists mostly of β-sheet and turn.

Conclusions: These observations led to the hypothesis that during evolution myocilin might have resulted from fusion of genes for at least two different proteins with functional implications relevant to mammals. It is noteworthy that mutations in the myocilin gene, causal to Primary Open Angle Glaucoma, have only been detected in the first exon (corresponding to myosin like region) and the last exon (corresponding to olfactomedin domain) but not the region (exon 2) between the two domains. Phylogenetic analysis of mammalian myocilin revealed that rat and mouse myocilins demonstrate a closer relationship compared to its human or bovine homologues.

Myocilin was originally described as myosin-like acidic protein (isoelectric point 5.2) expressed predominantly in the photoreceptor cells of retina and localized particularly in the rootlet and basal body of connecting cilium. Hence the protein was named as myocilin [1]. Polansky et al. [2] identified the same protein while studying the effects of steroids on the trabecular meshwork cells in culture. In the eye, the trabecular meshwork cells help regulate eye pressure by controlling the drainage of fluid from the eye as new fluid is produced. The cultured cells, when treated with steroids, secreted the same protein, which Nguyen et al. [3] called TIGR (for trabecular meshwork inducible glucocorticoid response protein). Mutations in the MYOC/TIGR gene have been associated with most if not all familial cases of juvenile onset primary open-angle glaucoma (POAG) as well as about 3% of all POAG in different populations tested [4,5]. The gene encodes for a protein of unknown function and has a broad range of expression not limited to the eye. The protein has been reported to have high level of sequence similarity to non-muscle myosin of *Dictyostelium discoideum* in the N-terminal region and to olfactomedin of the bullfrog in the C-terminal region [1]. Recently, it has been reported that myocilin interacts with the regulatory light chain (RLC) of myosin, a component of the myosin motor protein complex, independent of its olfactomedin domain, which implies a role for myocilin in the actomyosin system [6].

Olfactomedin is a secreted polymeric glycoprotein of unknown function, originally discovered at the mucociliary surface of the amphibian olfactory neuroepithelim and later identified throughout the mammalian brain [7]. The molecular evolution of olfactomedin has also been studied by examining its phylogenetic history to identify conserved structural motifs. The study based on comparison of protein sequences revealed that the evolution of the N-terminal half of the molecule involved extensive insertions and deletions while the C-terminal region evolved mostly through point mutations, suggesting evolutionary constrains in the C-terminal region for a predictably important functional role. The sequence similarity of myocilin at its C-terminal segment is restricted to the C-terminal segment of olfactomedin [7]. Also, Kulkarni et al. [8] reported widespread occurrence of olfactomedin and related

Correspondence to: Dr. Kunal Ray, Indian Institute of Chemical Biology, 4 Raja S. C. Mullick Road, Jadavpur, Kolkata - 700 032, India; Phone: 91-33-473-3491 (or 0492 or 6793); FAX: 91-33-473-5197 (or 0284); email: kunalray@vsnl.com

proteins among vertebrates and invertebrates, and studied the phylogenetic relation of 18 olfactomedin-related proteins, including human myocilin.

Kubota et al. [1] had originally proposed that myocilin contained a leucine zipper domain in the N-terminal region which was not detected in bullfrog olfactomedin, but was present in non-muscle myosin of *D. discoideum*, suggesting the retention of a very primitive structure in human myocilin. Later it was demonstrated by site directed mutagenesis of selective leucine residues (amino acids 117-166) of myocilin with glycine that myocilin forms dimers and multimers with itself through its leucine-zipper domain [9]. Though pathogenesis in POAG caused by a myocilin gene defect is not yet clear, mutations in myocilin gene causing glaucoma have been identified, primarily in the C-terminal (olfactomedin like) region and to a lesser extent in N-terminal region, implicating functional importance of both the regions of the protein. A recent study showed that secretion of human myocilin protein almost ceased when it lacked the olfactomedin-like region [10] and another study [11] reported that a few mutations in the olfactomedin region of myocilin decreased its secretion. Both these studies support the secretory role of myocilin. It is possible that the N-terminal region helps to form dimer and/or oligomers and forms the structural basis of trabecular meshwork.

Parallel to enrichment of the databases with sequencing of more genomes, it is becoming increasingly possible to assign functions to newly discovered proteins by searching for their homologues in the database, together with other strategies. Evolutionary analysis is becoming very powerful to discover conserved motifs and thereby predicting the functions. In that context we attempted to trace the evolutionary path of human myocilin, a protein yet to be defined in terms of its structure and function.

METHODS

Identification of non-mammalian protein sequences that are homologous to human myocilin: The proteins sharing significant sequence similarity to human myocilin (NP_000252) was initially searched in the NCBI server using the BLASTp program [12] in primarily two different ways: (i) The query sequence (human myocilin) was searched against the entire non-redundant (nr) dataset; and (ii) The same query sequence was also searched against the protein database specific for individual species. Using the first method, the majority of hits were olfactomedin-related proteins and noelin-like proteins of mammals and other higher eucaryotes showing homology to myocilin primarily in the C-terminal portion of the protein. To make sure that we do not fail to identify less homologous proteins from lower organisms which might be important to trace the evolution of myocilin, the second method was used, and a number of proteins were identified showing significant sequence similarity to both the N- and C-termini of human myocilin.

The search was done against available databases for different species which include bacterial kingdom, slime mold (*D. discoideum*), nematode (*C. elegans*), arthropod (*D. melanogaster*), amphibia (*X. laevis*) and aves (*G. gallus*) in an attempt to trace the evolution of mammalian myocilin. The default specifications of the BLAST analysis were used (low complexity filter, word size; 3, expect value; 10, BLOSUM 62 matrix and the gap penalties; 11 for opening; and 1 for extension).

Identification of mammalian myocilin protein sequences: To identify myocilin protein sequences in mammals we searched in the protein database at the NCBI server and found 24 entries. After consolidating duplicate entries six distinct protein sequences were identified which include two from human (NP_000252 and BAA24532), two from rat (AAK83081 and Q9R1J4), and one each from mouse (AAK83081) and bovine (BAA82152).

Phylogenetic analyses of mammalian myocilin: Phylogenetic analysis of the amino acid sequences of human, mouse, rat, and bovine myocilin was done at the DDBJ server. The following myocilin protein entries were chosen for phylogenetic analysis for the reason described in the Results section: Human (NP_000252; 504 amino acids), Bovine (BAA82152; 490 amino acids), Rat (AAK83081; 489 amino acids), and Mouse (AAK83082; 490 amino acids). The sequences were aligned using the multiple alignment tool ClustalW [13]. Unrooted phylogenetic tree was estimated using the BLOSUM matrix and the neighbor-joining tree-building algorithm [14]. Bootstrap values based upon 1000 iterations provide estimates of statistical support for the tree.

Secondary structure analysis of human myocilin protein: The secondary structure analysis of human myocilin was done using the PSIPRED software [15].

Cysteine disulfide bond prediction in mammalian myocilin: Prediction of disulfide bonded cysteine residues in the entire myocilin protein was done using CYSPRED software [16].

Prediction of putative motifs in the protein sequences: The presence of different motifs in the protein sequences was predicted using the ScanProsite tool with the option of including patterns with the high probability of occurrence [17].

RESULTS

The reported high homology of myocilin to nonmuscle myosin of *Dictyostelium discoideum* in the N-terminal region and to olfactomedin of the bullfrog in the C-terminal region [1] prompted us to examine the secondary structure of myocilin for signature of myosin-like coiled-coil proteins and olfactomedin. Myosin is mainly an α-helical protein, which forms coiled-coils to bind to itself and to other proteins like actin through its leucine-zipper domain. On the other hand the C-terminal region of olfactomedin consists mainly of β-sheet and turn. In this context we observed that the N-terminal region of myocilin consisted of primarily α-helix while the C-terminal contained mostly β-sheet (Figure 1). Based on this observation we hypothesized that myocilin might have evolved by fusion of genes for two different proteins.

To test this hypothesis we searched for peptide sequences in the available protein database for different species to identify peptides showing statistically significant sequence simi-

larity to myocilin. For each species only one protein was selected which gave the best score for homology with a specific region of myocilin. For those species where more than one protein was identified with a similar level of high score, it was observed that those were functionally related. For example, in *X. laevis* two noelins are found which really are isoforms of each other. Hence, only one of the related proteins was used for comparison with myocilin. Thus, the search led to identification of a single protein (M3 protein) in *Streptococcus pyogenes* in addition to a single protein (myosin heavy chain) in *D. discoideum* reported previously [1] to have homology to the N-terminal region of human myocilin. A similar search in the protein database for *C. elegans* identified, as reported earlier [7], a single protein having homology to the olfactomedin domain of myocilin located at its C-terminal region. Interestingly, on searching the protein database for *D. melanogaster*, two proteins (myosin heavy chain like protein, and CG6867 gene product) were detected to have significant homology to the N- and C-terminal region of human myocilin (Figure 2). Similarly, a search of *Gallus gallus* led to identification of two proteins, Dynactin 1 and Neuronal olfactomedin-related ER localized protein. However, it is noteworthy that the database for *G. gallus*, which evolved much later in speciation, did not contain any myocilin-like protein. The failure to detect the myocilin-like protein in the *G. gallus* database suggests lack of entry of the putative protein as yet or absence of the protein in the bird. A search of the *X. laevis* protein database resulted in identification of an olfactomedin-like protein (noelin) which has significant sequence similarity to entire C-terminal and most of the N-terminal regions of myocilin (Figure 2). The failure to identify a single protein in any species except *X. laevis*, prior to evolution of mammals, having homology covering entire myocilin, and detection of two distinct proteins in *D. melanogaster* having homology to two terminal regions of myocilin strongly suggested that molecular evolution of this highly conserved mammalian protein might have occurred by fusion of genes for two proteins prior to evolution of the mammalian lineage. The source of the homologous proteins, their names, regions of the proteins bearing homology to human myocilin and the level of identity are presented in Table 1.

Next, to identify the signature of myocilin (if any) in the predicted precursor proteins in other species that evolved prior to mammals, conservation of motifs in these proteins was examined. In the analysis we also included proteins from *G. gallus* (a bird) that evolved parallel to mammals from reptiles. As shown in Figure 2, among the proteins that have ho-

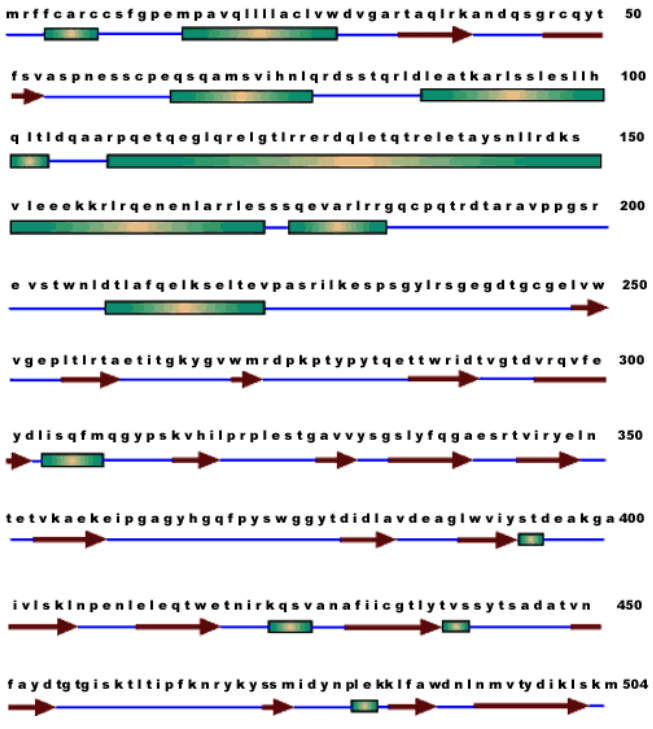

Figure 1. Predicted secondary structure of human myocilin. Open rectangle, solid arrows, and straight lines represent α-helix, β-sheet and turn, and coiled-coils, respectively. The N-terminal half is primarily α-helix; the C-terminal half contains mostly β-sheet and turn, and coiled-coils.

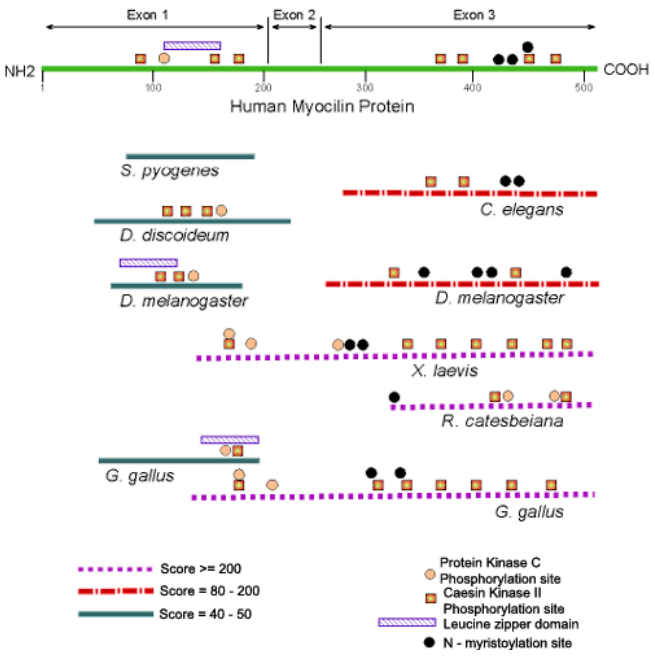

Figure 2. Protein sequences which share significant sequence homology with human myocilin. Human myocilin sequence was used as query and homology search was done using BLASTp program at the NCBI server as described in the Methods section. Regions of the proteins (identified in Table 1) bearing homology to human myocilin is shown by bar diagrams, depending on the level of homology, as indicated by a score (in arbitrary units). Relative locations of sequence motifs with putative functions are marked. The *S. pyogenes* protein, which contains sequences that match the motif for protein kinase C (PKC) phosphorylation sites, have not been shown because bacteria do not have PKC. The genomic structure of myocilin gene (i.e. exon boundaries) is shown with respect to the encoded protein. Exon 2 approximately corresponds to the region that does not have homology to non-myocilin proteins in lower organisms, and have not yet been associated to any among 34 characterized mutations to date causal to primary open angle glaucoma.

mology to the N-terminal region of myocilin, (a) *D. discoideum* protein has PKC and casein kinase 2 (CK2) phosphorylation sites, (b) *D. melanogaster* protein contains PKC and CK2 phosphorylation sites and a leucine zipper domain, (c) *X. laevis* contains PKC and CK2 phosphorylation sites and (d) *G. gallus* contains PKC, CK2, and tyrosine phosphorylation sites as well as leucine zipper domain. All three phosphorylation sites and leucine zipper domain are located in the region of human myocilin sharing homology to proteins present in slime mold, arthropod, and aves mentioned above. The *D. discoideum* protein also contains a leucine zipper domain but located in the N-terminal region, beyond the area of similarity with myocilin. A similar search for motifs in the proteins showing homology to the C-terminal region of myocilin revealed the presence of CK2 phosphorylation and N-myristoylation sites in proteins identified in all five species, i.e. *C. elegans*, *D. melanogaster*, *X. laevis*, *R. catesbeina*, and *G. gallus* representing nematods, arthropods, two amphibians and aves, respectively. It is noteworthy that only the set of proteins having similarity to the C-terminal region of myocilin contains N-myristoylation sites. Lack of any database of proteins for reptiles precludes further in silico analysis of evolution of myocilin at the present time.

We also examined the level of amino acid identity between two sets of proteins identified in different species having homology to N- and C-termini of human myocilin (Table 2). Since birds are not the predecessors of mammals and they independently evolved from a separate stalk of reptiles, the proteins identified in *G. gallus* mentioned above have not been included in this analysis. It was observed that in the N-terminal domain of human myocilin a stretch of peptide sequence (corresponding to the residues 116 to 176) aligned with overlapping sequences from 3 species (*S. pyogenes*, *D. discoideum*, and *D. melanogaster*) without any gap. In this region, 21.3% amino acids (13/61) were identical to human myocilin in at least two species (Table 2A) and homology increased to 24.5% if the locations were included which also contained conservative changes (Table 2A). Likewise, in the C-terminal domain of human myocilin a stretch of peptide sequence (corresponding to the residues 372 to 501) aligned with overlapping sequences from *C. elegans*, *D. melanogaster*, and *R. catesbeiana* with 2 gaps introduced for *C. elegans* and one gap for *D. melanogaster* to maximize amino acid identities. In this region 24.4% amino acids (32/131) were identical to human myocilin in at least two species (Table 2B) and homology increased to 41.2% if the locations were included which also contained conservative changes (Table 2B). As the homology of *X. laevis* noelin protein encompasses both N- and C-termini of human myocilin, the sequence alignment is shown separately (Table 2C). Since two isoforms of noelin (noelin-1 and noelin-2) are identical in the region having homology to myocilin and differ only in the length of the protein in the N-terminal region having no homology to myocilin, noelin 1 was arbitrarily selected for our analysis (Table 1 and Table 2). Thus, noelin 1 was found to have highest homology with myocilin among all other non-mammalian proteins (34% identity which increases to 51.9% if the conserved changes are included). These results provided additional support towards the possible evolution of myocilin from genes for two different proteins.

Table 1. Proteins in non-mammalian species with homology to human myocilin

Name of protein	Protein accession number	Source	E value	Length of protein	Region of myocilin bearing homology	Identity
Proteins with homology to N-terminal region of myocilin						
M3 Protein	AAA96960	S. pyogenes	5e-05	539 aa	84-196	38/129 (29%)
Myosin heavy chain	AAA33227	D. discoideum	9e-05	2116 aa	58-222	36/168 (21%)
Myosin-heavy-chain-like	AAF55271	D. melanogaster	8e-04	1706 aa	74-176	27/103 (26%)
Dynactin 1	P35458	G. gallus	5e-05	1224 aa	50-181	40/141 (28%)
Proteins with homology to C-terminal region of myocilin						
Olfactomedin-like	CAB04088	C. elegans	1e-23	598 aa	260-503	71/262 (27%)
CG6867	AAF48788	D. melanogaster	2e-28	935 aa	251-503	74/254 (29%)
Olfactomedin	AAA49527	R. catesbeiana	2e-20	464 aa	323-501	56/179 (31%)
Proteins with homology to both N- & C-terminal regions of myocilin						
Noelin-1	AAL66227	X. laevis	3e-57	485 aa	128-501	138/399 (34%)
Neuronal olfactomedin-related ER localized protein	AAF40413	G. gallus	5e-60	485 aa	110-501	134/407 (32%)

Regions of human myocilin (NP_000252) bearing homology to the protein in other species are indicated by amino acid numbers of myocilin. Number (and percentage) of amino acids that are identical to human myocilin within the homologous segment is given. The homology search between proteins by BLASTp provides the E value ("expect" value) which is an estimation of the probability of finding the homology by chance. Hence, the lower the E value, the higher the significance of homology.

TABLE 2. SEQUENCE ALIGNMENTS OF PROTEINS WITH HOMOLOGY TO HUMAN MYOCILIN

A

```
    Species          Protein                      Segments with homology to human myocilin
---------------   --------------   ------------------------------------------------------------------------

S. pyogenes       M3 Protein        383 GLRRDLDASREAKKQVEKALEEANSKLAALEKLNKELEESKKLTEKEKAELQAKLEAEAKA  443
D. discoideum     Myosin HC         965 GQSDTISRLEKIKDELQKEVEELTESFSESKDKGVLEKTRVRLQSELDDLTVRLDSETKD 1025
D. melanogaster   Myosin HC-like   1233 GTEEEFAQLRRSKNETERRAKEQEEELDEMAGQIQLLEQAKLRLEMTLETMRKEARRESQQ 1293
H. sapiens        Myocilin          116 GLQRELGTLRRERDQLETQTRELETAYSNLLRDKSVLEEKKRLRQENENLARRLESSQE  176
                                          ^    * ++   *   +    ^             ^^  + ++  +   +   +
```

B

```
    Species            Protein                      Segments with homology to human myocilin
---------------   -----------------   ------------------------------------------------------------------------

C. elegans        Olfactomedin-like   460 LYDRPHNYVDFAVDENGLWAIYAG-ADSETMRMAKIEP-SLFVVNIWNVEVNTTEIADSFIMCGV  522
D. melanogaster   CG6867              796 LYTTDYNYMDFNVDEVGLWVIYSTYNSN-NTLVAKLDAETLKMQYNFNITLDHHQFGEMFIVCGN  859
R. catesbeiana    Olfactomedin        332 YAGTMFQDMDFSSDEKGLWVIFTTEKSAGKIVVGKVNVATFTVDNIWITTQNKSDASNAFMICGV  396
H. sapiens        Myocilin            372 YSWGGYTDIDLAVDEAGLWVIYSTDEAKGAIVLSKLNPENLELEQTWETNIRKQSVANAFIICGT  435
                                              * *^* +^^ ^^^+^+ +        * ^* * *+ * * +        *   ^+*^^

C. elegans        Olfactomedin-like   523 WYGLKSANNLQTQITHAYDLFRNDTIPGQVEWYNPYQGLTMLGYNPLDARLYFFDNSSLLSVNVRI 588
D. melanogaster   CG6867              860 LYAIDSGTDKNTQIRYVVDLYKGKLLNTNLPFSNPFSHTTTVGYNPLTVELYSWDKGNALTYPIRY 925
R. catesbeiana    Olfactomedin        397 LYVTRSLGPKMEEVFYMFDTKTGKEGHLSIMMEKMAEKVHSLSYNSNDRKLYMFSEGYLLHYDIAL 462
H. sapiens        Myocilin            436 LYTVSSYTSADATVNFAYDTGTGISKTLTIPFKNRYKYSSMIDYNPLEKKLFAWDNLNMVTYDIKL 501
                                          +^   ^ *     *  * ^   +     ** * +       * ^^++   ^  *+  *** + +
```

C

```
 Species     Protein                    Segments with homology to human myocilin
----------   --------   ------------------------------------------------------------------------

X. laevis    Noelin-1    87 RDARTKQLRQLLEKVQNMSQSIEVLDRRTQRDLQYVERMENQMKGLESKFKQVEETHRQHQARQFK 152
H. sapiens   Myocilin   128 RDQLETQTRELETAYSNLLRDKSVLEEKKRLRQ---ENENLARRLESSQEVARL-RRGQCPQTR 189
                             ^^    ^ ^ ^*   *^*     ^^*    ^  ^     ^^ ** ^^^    ^    ^  ^  ^ *

X. laevis    Noelin-1   153 AIKAKMEELRPLIPVLEEYKADAKLVLQFKEEIQNLTSVLNE-LQEEIGAYDYEELQSRVSNLEER 217
H. sapiens   Myocilin   190 ------DTARAVPPGSREVSTWNLDTLAFQELKSELTEVPASRILKESPSGY-------------- 248
                                 * *^****^   ^        ^ ^ ^  * ^^ ^*   *  ^   *

X. laevis    Noelin-1   218 LRACMQKLACGKLTGISEPVTIKTSGS---RFGSWMTD--PLAPEGDNRVWYMDGYHNN-RFVREY 277
H. sapiens   Myocilin   249 LRSGEGDTGCGELVWVGEPLTLRTAETITGKYGVWMRDPKPTYPYTQETTWRIDTVGTDVRQVFEY 301
                             ^^ * *   ^^ ^  **^^*^**^  *   * ^ ^^ ^  ^  ^ *    ^ *^*  *  ^ ^ ^^

X. laevis    Noelin-1   278 KSMEDFMNTDNFTSHRLPHPWSGTGQVVYNGSIYFNKFQSHIIIRFDLKSETILKTRSLDSAGYTN 343
H. sapiens   Myocilin   302 DLISQFMQGYPSKVHILPRPLESTGAVVYSGSLYFQGAESRTVIRYELNTETVKAEKEIPGAGYHG 367
                               *  ^^**     ^ ^^*^* *^^ ^^^*^^*^^ * *^* *^^ *^ *^^*   * * *^^^ *

X. laevis    Noelin-1   344 VYHYAWGGQSDIDLMVDENGLWVVYATNQNAGNIVISKLDPNTLQILKTWNTGYPKRSAGEAFMIC 409
H. sapiens   Myocilin   368 QFPYSWGGYTDIDLAVDEAGLWVIYSTDEAKGAIVLSKLNPENLELEQTWETNIRKQSVANAFIIC 433
                                ^ ^^^**^^^^*^^^ ^^^^*^ ^    ^ ^^*^^^ ^ *^ *  ^^ ^*  ^ ^*  ^^*^^

X. laevis    Noelin-1   410 GTLYVTNGY-SGGTKVHYAYQTNTSNYEYIDIPFHNLYSHISMLDYNPKDRALYAWNNGHQILYNV 474
H. sapiens   Myocilin   434 GTLYTVSSYTSADATVNFAYDTGTGISKTLTIPFKNRYKYSSMIDYNPLEKKLFAWDNLNMVTYDI 499
                             ^^^^  **^ ^    ^  ^^ ^*^* * ** ^^^*^ ^   ^^*^^^^ ** ^ ^^ ^   * ^ *

X. laevis    Noelin-1   475 TL 476
H. sapiens   Myocilin   500 KL 501
                              ^
```

Segments of N-terminal and C-terminal regions of human myocilin are shown with overlapping regions of other proteins in **A** and **B**, respectively. In **C**, the homologous segments of human myocilin and *Xenopus* noelin are shown. Co-ordinates (amino acid numbers) of the region of each protein are indicated on both sides of the amino acid sequence. The sequences were aligned by eye to maximize amino acid identities. Human myocilin was the reference protein used as query. Locations where amino acids are identical are marked by a caret (^). Locations where any change is restricted to conservative amino acids are marked by asterisk (*), and locations where the amino acids are conserved in all but one species are marked by a plus sign (+). Putative sequence motifs are colored: green for PKC, blue for CK2, and red for N-myristoylation. Overlap between two putative sites have been shown by a purple marker. For example, in **B**TGKE residues represent an overlapping PKC (TGK) and a CK2 (TGKE) site. Among all the proteins mentioned in Figure 2, only the leucine zipper motifs for myocilin are present in the homologous region presented in this table (from amino acids 116-166).

Retention of a higher level of homology in the C-terminal domain might be due to functional importance of this region of protein. Kulkarni et al. [8] have shown that the C-terminal domain of olfactomedin and olfactomedin-like proteins are more conserved through evolution. The putative functional motifs marked in Figure 2 and shared between homologous regions of these proteins are shown in Table 2.

A search of the NCBI database for myocilin resulted in identification of the molecule only from mammalian species. Among two putative human myocilin protein sequences, one is of same length (490 aa; 55 kDa) as its homologues in mouse and bovine and the other is 14 amino acids longer at its N-terminal end (504 amino acids; 57 kDa). It has been reported [18] that western blots of human myocilin detected a 57 kDa band (predicted size for larger variant) and a 59 kDa band predicted to have resulted from post-translation modification of 57 kDa protein. These observations strongly suggest that the larger myocilin variant is either expressed exclusively or at a higher level relative to the shorter variant. As shown in Table 3, the sequence neighboring the first AUG in the human myocilin mRNA fits better with the Kozak consensus sequence [19]. Two rat myocilin protein sequences in the database are predicted to contain 489 and 502 amino acids. In the case of rat mRNAs, the down stream AUG would be favored as a translation start site according to the Kozak consensus sequence (Table 3). Hence, for comparison, we selected the larger human myocilin (NP_000252) and shorter rat homologue (AAK83081) along with the single myocilin proteins known to be present in mouse and bovine. It is an intriguing though that none of the putative translational start sites in human and rat myocilin mRNA contained the conserved G down stream to the ATG triplet codon, as would be expected according to the Kozak consensus sequence. Four mammalian myocilins submitted to the protein database were found to be highly conserved (73-89% identity) with a higher level of identity between mouse and rat myocilin. To assess the evolutionary interrelationship between mammalian myocilins, a phylogenetic analysis was carried out. Since the level of sequence homology between these proteins is very high, a neighbor-joining tree constructed using the entire sequence of the proteins suggested that the rat and mouse homologues of the protein would be closely related among the myocilin homologues characterized so far (Figure 3). Consistency between the detected and expected evolutionary relationships suggests that all known mammalian myocilin proteins are orthologs.

In the human myocilin protein (NP_000252), there are a total nine cysteine residues at positions 5, 8, 9, 25, 47, 61, 185, 245, and 433. In other mammalian myocilins, the first 3 cysteine residues (positions 5, 8, and 9) of the human homologue are not present, and the other six are conserved in all four mammalian myocilins. Prediction of disulfide bonded cysteine residues in the entire myocilin protein, using CYSPRED software, suggested that the probability of such bondage is maximum for residues 33 and 47 for rat, mouse and bovine myocilin. The predicted disulfide bond would be expected to be stable in the myocilin molecules that are secreted. It has been reported that in addition to the intracellular form of myocilin, the protein is also secreted into culture media of primary trabecular meshwork cells [10]. Human myocilin (NP_000252), which is 14 amino acids longer at the N-terminal region compared to its homologues in other mammals and contains a signal peptide that spans the first 32 amino acids, is not predicted to harbor a disulfide bond. We observed that the

Table 3. Putative site for initiation of translation in myocilin mRNA

Kozak consensus sequence	G	C	C	G A	C	C	AUG	G	Predicted protein
	-	-	-	-	-	-	---	-	---------
Human myocilin									
Neighboring sequence to first AUG	U	C	U	G	C	A	AUG	A	504 aa
Neighboring sequence to second AUG	C	C	U	G	A	G	AUG	C	490 aa
Rat myocilin									
Neighboring sequence to first AUG	A	U	A	C	C	G	AUG	C	502 aa
Neighboring sequence to second AUG	C	C	C	A	A	G	AUG	C	489 aa

The known mammalian myocilin human and rat homologues contain more than one AUG in the 5'-region of the mRNA with a potential of being the site for initiation of translation. The putative start sites for both the human and rat myocilins were determined following Kozak's consensus sequence for initiation of translation for eukaryotic mRNA. The bases marked in red matches the Kozak consensus sequence [19].

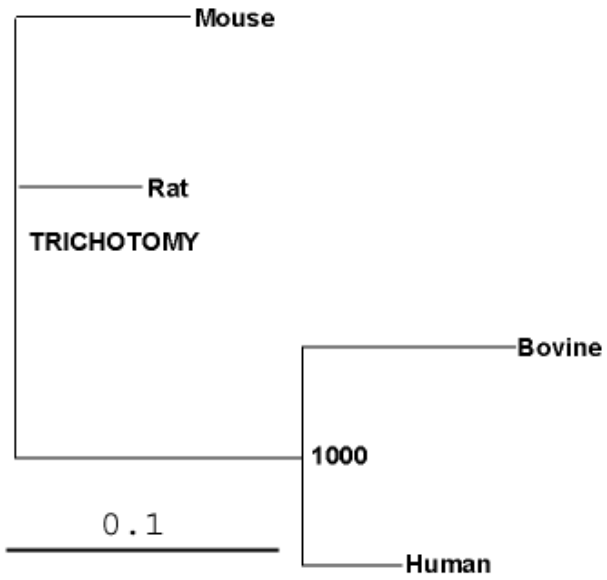

Figure 3. Phylogram of four different mammalian myocilin proteins. To assess the evolutionary interrelationship between the mammalian myocilins, a phylogenetic analysis was carried out. Since the level of sequence homology between these proteins is very high, a neighbor-joining tree was constructed using the entire sequence of the proteins. The scale denotes the branch length corresponding to substitutions per site. Trichotomy means that the tree is divided into three branches and the number 1000 denote the bootstrap value.

previously reported [1] conservation of the human myocilin cysteine at residue 433 in *C. elegans* olfactomedin-like protein is also retained in *D. melanogaster* as well as *R. catesbeiana* proteins (Table 2). A comparison of human olfactomedin-like proteins [8] also revealed that 4 of 5 such proteins (Olf A, Olf B, Olf C, and myocilin, but not Olf D) retain this cysteine residue. Thus, conservation of this cysteine residue in olfactomedin and olfactomedin-like proteins through evolution strongly suggests its important functional role.

DISCUSSION

Our search of the entire database for myocilin led to an observation that this protein is only found in mammals. This suggests that it might represent a protein that evolved later in evolution with functional implications not yet well appreciated. The initial observation [1] that two termini of myocilin have significant homology to two other proteins in lower organisms provided us a direction to follow and prompted us to investigate the evolution of myocilin. Recently, in an attempt to detect protein function and protein-protein interactions from genome sequences, Marcotte et al. [20] identified some pairs of interacting proteins which have homologues in another organism fused into a single protein chain. One, among a few other examples, is fusion of two *E. coli* proteins (γ-glutamyl phosphate reductase and glutamate 5-kinase) to a single human protein (δ-1-pyrroline-5-carboxylate synthatase). This observation led us to speculate that myocilin, which has been reported to be present in only mammals, might have evolved from more than one protein in lower organisms with which it has significant homology, such as myosin-like coiled-coil protein and olfactomedin. It is noteworthy that myocilin was originally described as myosin-like acidic protein [1], and later the same protein has been described as an olfactomedin-like protein, based on the evolutionary studies of olfactomedin [8]. More recently, it has been reported that myocilin interacts with regulatory light chain (RLC) of myosin, a component of the myosin motor protein complex, independent of its olfactomedin domain, which implies a role for myocilin in the actomyosin system [6]. Though the function of myosin is well defined, there is a lack of clear understanding about the biological functions of olfactomedin and myocilin. Despite a lack of knowledge of possible interaction between myosin and olfactomedin, or between such related proteins, we hypothesized that myocilin might have evolved from these two proteins. The hypothesis gained substantial support from the observed presence of similar putative functional motifs and conservation of amino acid sequences in the two set of proteins in lower organisms that have significant homology to N- and C-termini of human myocilin. One could argue that the observed similarity of the N-terminal region of myocilin with other proteins in lower organisms may be due to the structural constraints present in coiled-coils and not due to evolutionary relatedness. However, it is to be noted that all coiled-coil proteins do not necessarily share all these suggested similarities. For example human clusterin protein (AAH25381), despite being a coiled-coil protein, does not have the homology with human myocilin at any level identified for other proteins described in this manuscript, nor does it contain a putative leucine zipper domain. A clearer picture might emerge in the future from both experimental and in silico analyses. The chicken database does not contain any entry for myocilin but lists two proteins that, like *D. melanogaster* proteins, have significant homology to human myocilin. Since mammals and birds evolved separately from their common root at the level of reptiles, it is reasonable to examine the proteins expressed in reptiles and amphibians to understand the evolution of myocilin, and determine that whether it is present in these organisms. As mentioned in the Results section, there is no database for reptiles, which limits the scope of further investigation at this time.

Myocilin, which is expressed in optic nerve-head [21], retina [1], aqueous humor [22], and in lower amounts in skeletal muscles [1], still needs to be associated with a biological function. Compelling genetic evidence established its causal association with juvenile open angle glaucoma. However, the observations that homozygous nonsense mutations in the myocilin gene did not affect normal vision in human [23] and disruption of the gene in mouse did not interfere with its vision [24] strongly argued against a specific biological function of myocilin related to vision. It is argued that the mutations in the gene actually cause glaucoma due to a gain of function of the protein [24]. This argument is consistent with the observed transmission of juvenile open angle glaucoma as a dominant trait. It is noteworthy that *X. laevis* noelin, which has the highest level of homology with myocilin compared to all other proteins in other species that predates evolution of mammals and could potentially be a precursor of myocilin, has been reported to be involved in promotion of neurogenesis [25]. Since myocilin is an olfactomedin-like protein, it is provocative to speculate that noelin, as a close 'cousin' of myocilin, among other olfactomedin-like proteins, might be involved in some eye disorders including glaucoma. Recently, another novel olfactomedin-related gene optimedin, located on chromosome 1p21 in humans, has been described [26]. In the human eye, optimedin is expressed in the retina and the trabecular meshwork. Interaction between optemedin and myocilin have been demonstrated by in vitro experiments and it has been proposed that mutant myocilin interferes with secretion of optimedin in transfected cells. Incidentally, no information regarding the genetic basis of glaucoma has been described in animal models [27]. Glaucoma has not been described in amphibians. No specific gene defect is described in hereditary glaucoma of birds, nor is it known if olfactomedin-like protein is involved in light induced glaucoma of the chicken [28].

Myocilin has a leucine-zipper domain in its N-terminal region, which is present in the non-muscle myosin of *D. discoideum* and *D. melanogaster* but not present in the proteins containing olfactomedin like domain and having similarity to C-terminal region of myocilin. Recently, it was reported that myocilin forms a complex oligomer when present in human aqueous humor, through its leucine-zipper domain [9]. On the other hand, the cysteine at residue 433 of human

myocilin, which is predicted to participate in intermolecular disulfide bonding [29], is conserved among all the proteins in other organisms that have been identified to have significant similarity to myocilin. This cysteine residue is conserved in olfactomedin and olfactomedin-like proteins through evolution [8]. All of 34 mutations characterized so far in the myocilin gene in POAG patients are located in either exons 1 and 3 which corresponds to the myosin-like and olfactomedin-like domains respectively located at two ends of the gene, and none in exon 2 (corresponding to amino acid 202 to 244). Surprisingly, the gap in the alignment of *Drosophila* proteins corresponds to this apparently non-functional region of myocilin. The N-terminal homologue ends at amino acid 176 and the C-terminal homologue starts at amino acid 251. Further investigation would be required to test the possibility that the region corresponding to exon 2 of the myocilin gene is a linker region between the myosin-like and the olfactomedin-like domains in myocilin that appears to have been acquired by fusion of two proteins.

ACKNOWLEDGEMENTS

The financial help from Council of Scientific and Industrial Research, India for the study as well as pre-doctoral fellowship to AM, AG and SM is gratefully acknowledged. The study has been partially supported by grants to KR from Department of Science & Technology and Indian Council of Medical Research, Government of India.

REFERENCES

1. Kubota R, Noda S, Wang Y, Minoshima S, Asakawa S, Kudoh J, Mashima Y, Oguchi Y, Shimizu N. A novel myosin-like protein (myocilin) expressed in the connecting cilium of the photoreceptor: molecular cloning, tissue expression, and chromosomal mapping. Genomics 1997; 41:360-9.
2. Polansky JR, Fauss DJ, Chen P, Chen H, Lutjen-Drecoll E, Johnson D, Kurtz RM, Ma ZD, Bloom E, Nguyen TD. Cellular pharmacology and molecular biology of the trabecular meshwork inducible glucocorticoid response gene product. Ophthalmologica 1997; 211:126-39.
3. Nguyen TD, Chen P, Huang WD, Chen H, Johnson D, Polansky JR. Gene structure and properties of TIGR, an olfactomedin-related glycoprotein cloned from glucocorticoid-induced trabecular meshwork cells. J Biol Chem 1998; 273:6341-50.
4. Alward WL, Fingert JH, Coote MA, Johnson AT, Lerner SF, Junqua D, Durcan FJ, McCartney PJ, Mackey DA, Sheffield VC, Stone EM. Clinical features associated with mutations in the chromosome 1 open-angle glaucoma gene (GLC1A). N Engl J Med 1998; 338:1022-7.
5. Stone EM, Fingert JH, Alward WL, Nguyen TD, Polansky JR, Sunden SL, Nishimura D, Clark AF, Nystuen A, Nichols BE, Mackey DA, Ritch R, Kalenak JW, Craven ER, Sheffield VC. Identification of a gene that causes primary open angle glaucoma. Science 1997; 275:668-70.
6. Wentz-Hunter K, Ueda J, Yue BY. Protein interactions with myocilin. Invest Ophthalmol Vis Sci 2002; 43:176-82.
7. Karavanich CA, Anholt RR. Molecular evolution of olfactomedin. Mol Biol Evol 1998; 15:718-26.
8. Kulkarni NH, Karavanich CA, Atchley WR, Anholt RR. Characterization and differential expression of a human gene family of olfactomedin-related proteins. Genet Res 2000; 76:41-50.
9. Fautsch MP, Johnson DH. Characterization of myocilin-myocilin interactions. Invest Ophthalmol Vis Sci 2001; 42:2324-31.
10. Caballero M, Rowlette LL, Borras T. Altered secretion of a TIGR/MYOC mutant lacking the olfactomedin domain. Biochim Biophys Acta 2000; 1502:447-60.
11. Jacobson N, Andrews M, Shepard AR, Nishimura D, Searby C, Fingert JH, Hageman G, Mullins R, Davidson BL, Kwon YH, Alward WL, Stone EM, Clark AF, Sheffield VC. Non-secretion of mutant proteins of the glaucoma gene myocilin in cultured trabecular meshwork cells and in aqueous humor. Hum Mol Genet 2001; 10:117-25.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-10.
13. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22:4673-80.
14. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4:406-25.
15. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000; 16:404-5.
16. Fariselli P, Riccobelli P, Casadio R. Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. Proteins 1999; 36:340-6.
17. Bairoch A. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res 1991; 19[Suppl]:2241-5.
18. Fingert JH, Ying L, Swiderski RE, Nystuen AM, Arbour NC, Alward WL, Sheffield VC, Stone EM. Characterization and comparison of the human and mouse GLC1A glaucoma genes. Genome Res 1998; 8:377-84.
19. Kozak M. How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell 1978; 15:1109-23.
20. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. Science 1999; 285:751-3.
21. Clark AF, Kawase K, English-Wright S, Lane D, Steely HT, Yamamoto T, Kitazawa Y, Kwon YH, Fingert JH, Swiderski RE, Mullins RF, Hageman GS, Alward WL, Sheffield VC, Stone EM. Expression of the glaucoma gene myocilin (MYOC) in the human optic nerve head. FASEB J 2001; 15:1251-3.
22. Russell P, Tamm ER, Grehn FJ, Picht G, Johnson M. The presence and properties of myocilin in the aqueous humor. Invest Ophthalmol Vis Sci 2001; 42:983-6.
23. Lam DS, Leung YF, Chua JK, Baum L, Fan DS, Choy KW, Pang CP. Truncations in the TIGR gene in individuals with and without primary open-angle glaucoma. Invest Ophthalmol Vis Sci 2000; 41:1386-91.
24. Kim BS, Savinova OV, Reedy MV, Martin J, Lun Y, Gan L, Smith RS, Tomarev SI, John SW, Johnson RL. Targeted Disruption of the Myocilin Gene (Myoc) Suggests that Human Glaucoma-Causing Mutations Are Gain of Function. Mol Cell Biol 2001; 21:7707-13.
25. Moreno TA, Bronner-Fraser M. The secreted glycoprotein Noelin-1 promotes neurogenesis in Xenopus. Dev Biol 2001; 240:340-60.
26. Torrado M, Trivedi R, Zinovieva R, Karavanova I, Tomarev SI. Optimedin: a novel olfactomedin-related protein that interacts with myocilin. Hum Mol Genet 2002; 11:1291-301.
27. Gelatt KN. Animal models for glaucoma. Invest Ophthalmol Vis Sci 1977; 16:592-6.

28. Salter DW, Payne WS, Ramsey DT, Blair M, Render JA. A new inherited ocular anomaly in pigmented white Leghorn chickens. J Vet Diagn Invest 1997; 9:407-9.

29. Yokoe H, Anholt RR. Molecular cloning of olfactomedin, an extracellular matrix protein specific to olfactory neuroepithelium. Proc Natl Acad Sci U S A 1993; 90:4655-9.

The print version of this article was created on 23 Jul 2002. This reflects all typographical corrections and errata to the article through that date. Details of any changes may be found in the online version of the article.