Structural basis for misfolding in myocilin-associated glaucoma

Rebecca K. Donegan, Shannon E. Hill, Dana M. Freeman, Elaine Nguyen, Susan D. Orwig, Katherine C. Turnage and Raquel L. Lieberman*

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

*To whom correspondence should be addressed. Tel: +1 404 385 3663; Fax: +1 404 894 2295; Email: raquel.lieberman@chemistry.gatech.edu

Abstract

Olfactomedin (OLF) domain-containing proteins play roles in fundamental cellular processes and have been implicated in disorders ranging from glaucoma, cancers and inflammatory bowel disorder, to attention deficit disorder and childhood obesity. We solved crystal structures of the OLF domain of myocilin (myoc-OLF), the best studied such domain to date. Mutations in myoc-OLF are causative in the autosomal dominant inherited form of the prevalent ocular disorder glaucoma. The structures reveal a new addition to the small family of five-bladed β-propellers. Propellers are most well known for their ability to act as hubs for protein–protein interactions, a function that seems most likely for myoc-OLF, but they can also act as enzymes. A calcium ion, sodium ion and glycerol molecule were identified within a central hydrophilic cavity that is accessible via movements of surface loop residues. By mapping familial glaucoma-associated lesions onto the myoc-OLF structure, three regions sensitive to aggregation have been identified, with direct applicability to differentiating between neutral and disease-causing non-synonymous mutations documented in the human population worldwide. Evolutionary analysis mapped onto the myoc-OLF structure reveals conserved and divergent regions for possible overlapping and distinctive functional protein–protein or protein–ligand interactions across the broader OLF domain family. While deciphering the specific normal biological functions, ligands and binding partners for OLF domains will likely continue to be a challenging long-term experimental pursuit, atomic detail structural knowledge of myoc-OLF is a valuable guide for understanding the implications of glaucoma-associated mutations and will help focus future studies of this biomedically important domain family.

Introduction

The olfactomedin (OLF) domain, first identified as part of an eponymous multidomain protein involved in olfactory chemoreception in bullfrogs >20 years ago (1,2), is now recognized as a large domain family (PFAM: PF02191) comprising seven phylogenetic branches (3) (Supplementary Material, Table S1). OLFs characterized to date are found largely as part of extracellular proteins (2,4–7) in multicellular organisms. They are implicated in an expansive and growing number of human disorders including a host of cancers (8), inflammatory bowel disorder (9), defense against infection (10), attention deficit hyperactivity disorder (11) and childhood obesity (12), among others.

Myocilin, the first to be associated with disease and the best studied OLF domain-containing protein, is one of the few veritable disease-modifying drug targets for glaucoma (13), the second leading cause of blindness worldwide (14). Non-synonymous lesions localized almost exclusively to its OLF domain (myoc-OLF) lead to an early-onset, autosomal-dominant inherited form of the most prevalent subtype, primary open angle glaucoma (15), and account for ~3–10% of such patients (13). Nearly 100 myoc-OLF variants have been documented in individuals, through a combination of familial and population glaucoma case-control genetic studies (16). The leading proposed pathogenic mechanism involves a gain of toxic function, namely...
aggregation, in the region of the anterior eye called the trabecular meshwork (TM). In cell culture, disease-causing myocilin missense variants are all similarly prone to aggregation in the endoplasmic reticulum (ER) (17), leading to ER stress (18-23) and cell death (17,20). This cascade is presumed to lead to reduced TM-associated fluid outflow and accelerated onset of ocular hypertension, the major risk factor for vision loss and glaucoma. Although the ER chaperone machinery should be capable of efficient degradation of such mutant proteins, there appears to be an aberrant interaction between mutant myocilin and chaperones that subverts proper retrotranslocation for proteasomal degradation (19). Studies of purified recombinant myoc-OLF variants in vitro demonstrate that while wild-type myoc-OLF is a well-folded monomer, single glaucomatous point mutations, such as P370L, can be sufficient to impair folding to the extent that only insoluble aggregates are isolated (24). More generally, thermal stabilities of folded, disease-associated variants, measured by the melting temperature ($T_m$), are statistically correlated with the age the associated patients were diagnosed with glaucoma (24). Myoc-OLF aggregates exhibit hallmarks of amyloid (25,26) and appear either as long straight fibrils or an atypical large-diameter circular species. These morphologies are similar to aggregates formed by two well-conserved peptide stretches within myoc-OLF, P1 ($G_{202}AVVYSGSLYFQ$) and P3 ($V_{337}ANAFIICGTLYTVSSY$), respectively (25), suggesting that a propensity to misfold might be an inherent property of OLF domains. Finally, high levels of wild-type myocilin are associated with steroid-induced glaucoma, a secondary disease subtype (27). Although not directly linked to pathogenesis, myocilin expression is stimulated by glucocorticoids in TM cells (27), which, as seen with disease variants, may lead to aggregation and involve ER stress (28,29).

Like other disease-associated proteins that exhibit a toxic gain-of-function misfolding phenotype, studies of overexpressed wild-type or disease-associated missense variants of myocilin do not have an obvious connection to normal biological function(s), which remains poorly defined. Whereas mice expressing mutant myocilin display symptoms of glaucoma (30,31), myocilin knock out mice do not (32). Recent studies reveal that although there is no major systemic phenotype, myocilin null mice do exhibit structural changes in the retina (33), optic nerve (34), as well as in bone (35), and in the peripheral nervous system (36), suggesting a loss of function. Predominantly, but not exclusively, extracellular locations have been proposed for myocilin (8), and evidence suggests distinct functions in each region, including a regulatory role in signaling (37), cell proliferation and survival (33,38). Interacting partners identified to date have localized primarily to the myocilin N-terminal coiled-coil region (13). Individual studies have identified flotillin-1 (39), a membrane-bound protein involved in vesicular trafficking, and optineurin (40), an extracellular protein, as myoc-OLF-specific interacting partners. The implications of these interactions for myocilin function, however, remain unclear (39).

Here we report three crystal structures of myoc-OLF, revealing a new addition to the five-bladed β-propeller structural protein class. Propellers are known for their involvement in protein–protein interactions, and analysis of the structure reveals potential binding sites for protein–protein and/or ligand–protein interactions. Evolutionary considerations indicate that OLF domains share a common scaffold but likely interact with distinct surface binding partners. The myoc-OLF structure provides a three-dimensional framework from which to understand the molecular basis of glaucoma-associated misfolding, and differentiate between neutral and disease-causing non-synonymous mutations found in the human population. Our study offers new insight into and lays the groundwork for understanding pathogenesis and biology of this enigmatic yet important domain family.

### Results

#### Overall architecture

The OLF crystal structure was solved by anomalous phasing using seleno-methionine (SeMet)-substituted myoc-OLF(E396D) (myoc-OLF(E396D)I222), a single nucleotide polymorphism (SNP) variant that grew superior quality crystals to those of wild type. Molecular replacement was then used to obtain the high resolution structure at 1.9 Å resolution of native myoc-OLF(E396D) (myoc-OLF(E396D)I222) and the 2.1 Å resolution structure of wild-type myoc-OLF (see Materials and Methods and Table 1). The final models include residues 244–502 (numbering scheme for full-length myocilin). The N-terminal 16 residues of myoc-OLF, which complete the C-terminal fragment of myocilin reported upon proteolytic cleavage in cell culture (41), are not visible in the structures, likely because they are part of a linker between the OLF domain and N-terminal regions of full-length myocilin.

The OLF domain is an ~40 Å diameter by ~30 Å high β-propeller with five blades, each composed of four antiparallel β-strands (Fig. 1), arranged radially around a central water-filled cavity ~11 Å in diameter (Supplementary Material, Fig. S1). Compared with other known propellers of varying blade number (42), the blades in OLF are notably asymmetric (see also comparison of five-bladed propellers below). Nearly half of the toroïd-shaped molecule is occupied by Blades D and E. Discontinuous Strands D-18/D-18b and E-21/E-21b in the myoc-OLF(E396D) structures (Fig. 1a) appear as single continuous strands in the wild-type myoc-OLF structure (not shown). The electron density in this region is well defined in all structures, but for clarity, in our subsequent descriptions, we refer to the blades as labeled in Figure 1 for our highest resolution structure, that of myoc-OLF(E396D)I222. Other structural features include several helical turns and a short α-helix. The α-helix resides between myoc-OLF Strands A-5 and A-6, and packs against the E-A blade interface (Fig. 1a). Despite the lack of sequence similarity in this region (Supplementary Material, Fig. S2), the α-helix is predicted by secondary structure prediction algorithms (not shown) and is likely a general feature of the OLF domain.

Two features of OLF stabilize the propeller in a closed circular conformation. First, extensive interactions from the sequentially discontinuous two outer strands of Blade E, Strands E-1 to E-2 and E-21 (Fig. 1c), form the propeller feature colloquially termed ‘Velcro’ or ‘molecular clasping’ (43). Second, a single disulfide bond is found at the bottom face of the propeller (Fig. 1b), between Cys 245 at the N-terminus of the visible structural domain prior to the start of E-1, and Cys 433, located within Loop D-15/D-16 between the internal two strands of Blade D. A similar disulfide bond is seen in four-bladed propellers (43,44) but not five-bladed propeller structures. The disulfide bond is likely present in most OLF domains, as the C-terminal Cys is highly conserved, and in sequence alignments (Supplementary Material, Fig. S2), there is typically an N-terminal cysteine near the equivalent position of myoc-OLF Cys 245.

Nearly half of the residues are found within well-ordered loops above and below the propeller (Figs 1a and 2a), with minor differences among the solved structures (Fig. 2a). The longest loop is composed of residues 360–379 (Fig. 1d), connects Strands B-10 and C-11, and caps the top entrance to the central
cavity (Supplementary Material, Fig. S1). Although no global changes are observed among our structures, Loop B-10/C-11 appears to serve as a gate for access to the central cavity via the postioning of the side chains of myoc-OLF Trp 373 and Tyr 442, which resides on Loop D-16/D-17. In total, three states were captured that are not attributable to crystal contacts (Fig. 2a and b). The side chain of myoc-OLF residue Trp 373 is visible in two conformations in our structures, one pointed into the central cavity toward the bottom face, a second major pocket electron density peak was identified (Fig. 3a). In our structures, based on previously by metal analysis (46). The ion was modeled into a 25Å peak in the Fc – Fo difference electron density map present after initial model building. The heptacoordinate Ca2+ is ligated by the side chains of Asp 380, Asn 428 and Asp 478, all in a monodentate fashion, as well as the carbonyl backbones of Ala 429 and Ile 477, and two water molecules (Fig. 3a). The structure confirms Asp 380, the only previously identified Ca2+ ligand (46). The ligand environment in myoc-OLF is typical for Ca2+ (47), but not in a documented Ca2+ binding motif. Consistent with the previous finding that the apo disease variant D380A is a moderately stable protein (46), the internal position of the myoc-OLF Ca2+ ion appears as an ionic tether for Blades C, D and E (Fig. 1a) to confer stability. Adjacent to the Ca2+ site, at a distance of ~3.4 Å deeper into the cavity toward the bottom face, a second major Fc – Fo difference electron density peak was identified (Fig. 3a). In our structures, based on final refined metal–ligand distances (48), this density is modeled as a Na+ ion coordinated by Asp 380 and Asp 478, as well as the carbonyl backbone of Leu 381 and a water molecule (Fig. 3a). Lastly, within the hydrophilic cavity in both closed and semi-open states of myoc-OLF(E396D), a glycerol molecule is bound, in the region between Strands A-3 and B-7 above the Ca2+ site (Fig. 3b). Glycerol was not introduced during crystallization of wild-type myoc-OLF and is thus not observed in that structure.

**Comparison to known five-bladed β-propellers**

Myoc-OLF is distinct from the four different proteins with known five-bladed propeller folds (Supplementary Material, Fig. S3), with root mean squared deviations > 3 Å, consistent with sequence analyses that confirm OLF domains are not phylogenetically related to these propellers (<5–9% identity scattered previously by metal analysis (46). The ion was modeled into a 25Å peak in the Fc – Fo difference electron density map present after initial model building. The heptacoordinate Ca2+ is ligated by the side chains of Asp 380, Asn 428 and Asp 478, all in a monodentate fashion, as well as the carbonyl backbones of Ala 429 and Ile 477, and two water molecules (Fig. 3a). The structure confirms Asp 380, the only previously identified Ca2+ ligand (46). The ligand environment in myoc-OLF is typical for Ca2+ (47), but not in a documented Ca2+ binding motif. Consistent with the previous finding that the apo disease variant D380A is a moderately stable protein (46), the internal position of the myoc-OLF Ca2+ ion appears as an ionic tether for Blades C, D and E (Fig. 1a) to confer stability. Adjacent to the Ca2+ site, at a distance of ~3.4 Å deeper into the cavity toward the bottom face, a second major Fc – Fo difference electron density peak was identified (Fig. 3a). In our structures, based on final refined metal–ligand distances (48), this density is modeled as a Na+ ion coordinated by Asp 380 and Asp 478, as well as the carbonyl backbone of Leu 381 and a water molecule (Fig. 3a). Lastly, within the hydrophilic cavity in both closed and semi-open states of myoc-OLF(E396D), a glycerol molecule is bound, in the region between Strands A-3 and B-7 above the Ca2+ site (Fig. 3b). Glycerol was not introduced during crystallization of wild-type myoc-OLF and is thus not observed in that structure.
Figure 1. Structural features of the OLF domain. (A) Top and side view of representative myoc-OLF propeller. Blades A–E with corresponding numbered strands, listed according to myoc-OLF(E396D)_122 structure. (B) Internal calcium-binding site and disulfide bond at bottom face. (C) Molecular clasp region. (D) Loop B-10/C-11, comprising residues 360–379. Dashed lines represent relevant stabilizing interactions. Coloring: N (blue) to C (red).

Figure 2. OLF molecular surface. (A) Superposition of wild-type myoc-OLF and two myoc-OLF(E396D) structures. Yellow, β-strand; red, α-helix or turn; green, loop. Zoomed regions highlight where loop structures differ. (B) Top face surface representation and open–closed Loop B-10/C-11 snapshots observed crystallographically. Cartoon representations of boxed region appear below each surface representation as well as observed ligands in the corresponding structure (PEG or glycerol, red) with final 2Fo – Fc electron density contoured at 1σ.
throughout the sequence, not shown). Five-bladed propellers are primarily associated with glycosyl hydrolase or cyclotransferase activity; other five-bladed propellers include a human apyrase (49) and the highly symmetric founding member, tachylectin-2, which binds carbohydrates (42). The five tachylectin-2 binding sites, located between connecting segments of one sheet and the outer two strands of a second adjacent sheet (50), are absent in OLF (Supplementary Material, Fig. S3a). Superposition of myoc-OLF with arabinase (51) or glutaminyl cyclase (52) reveals that the only potential catalytic residue is the unlikely myoc-OLF Ca²⁺ ligand Asp 380 (Supplementary Material, Fig. S3b,c). While glutaminyl cyclase harbors a Ca²⁺ binding site, this ion is not part of the glutaminyl cyclase active site (52); arabinase has a His-ligated Cl⁻ ion ∼3 Å away with no obvious relation to the myoc-OLF Ca²⁺ site. The surface substrate-binding cleft of arabinase is likewise not observed in OLF and clashes with Loop B-10/C-11 (Supplementary Material, Fig. S3b). Similar differences with apyrase are observed (Supplementary Material, Fig. S3d). In sum, the OLF domain expands the diversity of sequentially unrelated β-propellers (42) and is a new representative template for five-bladed propellers. However, as such, the ligand binding and active sites of these five-bladed propellers do not directly shed light on the specific function for the OLF domain (see Discussion).

Surface electrostatics

Inspection of the electrostatic surface potential (Fig. 4a) correlates with the predicted pI ∼5. The top surface of myoc-OLF contains several charged regions. There is a cluster of five positively charged surface Lys and Arg residues, in the loops connecting segments of one sheet and the outer two strands of a second adjacent sheet (50), are absent in OLF (Supplementary Material, Fig. S3a). Superposition of myoc-OLF with arabinase (51) or glutaminyl cyclase (52) reveals that the only potential catalytic residue is the unlikely myoc-OLF Ca²⁺ ligand Asp 380 (Supplementary Material, Fig. S3b,c). While glutaminyl cyclase harbors a Ca²⁺ binding site, this ion is not part of the glutaminyl cyclase active site (52); arabinase has a His-ligated Cl⁻ ion ∼3 Å away with no obvious relation to the myoc-OLF Ca²⁺ site. The surface substrate-binding cleft of arabinase is likewise not observed in OLF and clashes with Loop B-10/C-11 (Supplementary Material, Fig. S3b). Similar differences with apyrase are observed (Supplementary Material, Fig. S3d). In sum, the OLF domain expands the diversity of sequentially unrelated β-propellers (42) and is a new representative template for five-bladed propellers. However, as such, the ligand binding and active sites of these five-bladed propellers do not directly shed light on the specific function for the OLF domain (see Discussion).

Evolutionarily conserved and divergent features of the OLF domain

To gain better insight into the implications of the observed structural features of myoc-OLF on the OLF domain family as a whole,
the evolutionary trace (54) of 1000 OLF domain sequences was mapped onto the OLF structure (Fig. 4b), revealing regions of conserved and divergent features of OLF domains. Spatially clustered, highly conserved residues define regions with a high likelihood of functional significance (54). Interestingly, the interior of OLF contains a high density of conserved residues,
including Blade C and its loops, the interior strands of Blade D, the central cavity metal ion and ligand-binding sites, as well as near the site of a bound PEG at the bottom face (not shown). A major conserved feature revealed by this analysis is a cation–π interaction involving Tyr 371 within Loop B-10/C-11 and Lys 423 on Loop C-14/D-15 (Fig. 4c). Cation–π interactions are important for molecular recognition and other biochemical processes (55) and thus may prove to be a common functional design feature of OLFs. On the opposite end of the spectrum, the OLF surface clearly exhibits the highest degree of sequence divergence across the entire domain family (Fig. 4b; see Discussion).

Locations of familial glaucoma variants and identification of misfolding trigger zones

The myoc-OLF structure provides the first opportunity to understand the molecular basis of glaucoma-associated misfolding at the atomic level. Analysis of the locations of 21 myoc-OLF variants with Mendelian inheritance patterns in affected families or in selected groups of juvenile/early onset glaucoma patients, whose pathogenicity has been correlated quantitatively with thermal stability and aggregation propensity in cells and in vitro (17, 24, 56–58), and three non-disease-associated SNPs, reveals three distinct destabilizing regions—(i) core hydrophobic β-sheet belt including the molecular clasp, (ii) Loop B-10/C-11 and cation–π interaction, and (iii) Ca2+ environs (Table 2).

The largest number of variants is found within the core β-sheet belt of the propeller (Table 2), particularly those variants with the lowest $T_m$, i.e., the most destabilized, such as W286R and I477N/S (24, 56), where side chain alterations would disrupt hydrophobic packing and be least tolerated. Part of the core region, where substitutions overall appear better tolerated, is the molecular clasp (Fig. 1c). Such variants, namely moderate variants G246R, G252R, R272G, N480K, I499F and S502P, affect

### Table 2. Glaucoma-associated lesions correlated with stability and myoc-OLF structure

<table>
<thead>
<tr>
<th>Mutation</th>
<th>$T_m$ (°C)</th>
<th>Structural location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Familial mutations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Hydrophobic β-sheet belt (40%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G246R</td>
<td>42.5 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>G252R</td>
<td>43.0 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>R272G</td>
<td>41.0 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>W286R</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>C433R</td>
<td>40.4 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>Y437H</td>
<td>40.3 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>I477N</td>
<td>37.7 ± 0.8</td>
<td></td>
</tr>
<tr>
<td>I477S</td>
<td>39.7 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>N480K</td>
<td>42.4 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>I499F</td>
<td>42.8 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>S502P</td>
<td>41.0 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>B. Loop B-10/C-11 and cation–π (33%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E323K</td>
<td>44.0 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>G364V</td>
<td>45.0 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>G367R</td>
<td>42.7 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>P370L</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>T377M</td>
<td>44.3 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>K423E</td>
<td>34.2 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>V426F</td>
<td>41.5 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>C. Ca2+ site environs (17%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D380A</td>
<td>46.6 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>A427T</td>
<td>48.3 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>P481L</td>
<td>45.5 ± 0.4</td>
<td></td>
</tr>
<tr>
<td><strong>II. Single nucleotide polymorphisms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E352Q</td>
<td>54.8 ± 0.5</td>
<td>Loop B-9/B-10</td>
</tr>
<tr>
<td>E396D</td>
<td>53.1 ± 0.7</td>
<td>Loop C-12/C-13</td>
</tr>
<tr>
<td>K398R</td>
<td>53.8 ± 0.2</td>
<td>Loop C-12/C-13</td>
</tr>
</tbody>
</table>

*Percentages in each category of (I) reflect those as listed in (16).

$I477N/S and N480K also belong to C. $T_m$ values as reported in (46) except when denoted, which are reported in (24). N/A indicates that no soluble recombinant protein could be isolated for $T_m$ measurement. $T_m$ value for wild-type myoc-OLF = 52.2°C (24). D380A also belongs to B.
the tethering of the strands and/or alignment of the disulfide bond in this region. The C433R mutation abrogates disulfide bond formation. Interestingly, however, the effect of C433R on stability is not more drastic than that of Y437H (24), the variant used in myocilin glaucoma mouse models (21,30); both are highly conserved residues within OLF domains (Supplementary Material, Fig. S2). In the case of Tyr 437, a water-mediated side chain interaction stabilizes the region near the disulfide bond (Supplementary Material, Fig. S4).

Variants involving Loop B-10/C-11, and adjacent cation–π interaction (Table 2; Figs 1d and 4c), include those that are both moderately stable and severely destabilized. The substitution K423E, which would abolish the highly conserved cation–π interaction, is one of the most thermally unstable mutations measured to date (56). Similarly, although not assessed for stability in vitro due to the fact that it was reported just once, Y371D is associated with a teenage onset, severe glaucoma (59). Variants of moderate stability Y426F and T377M affect the cation–π interaction: Val 426 holds Lys 423 in position via main chain interaction, and the side chain of Lys 423 interacts with the main chain carbonyl of Thr 377 (Fig. 4c). The side chain of Thr 377 also forms a hydrogen bonding interaction with the main chain of Tyr 371, which would be lost upon substitution to Met (Fig. 4c). P370L, a severe mutant (58,60), is a site that appears to organize nearby residues on Loop B-10/C-11. Curiously, Pro 370 itself does not appear to participate directly in regional stabilization (Fig. 1d and 4c), nor is it highly conserved among non-myocilin OLF domains (Supplementary Material, Fig. S2). Other moderately stable variants include charge inversion mutant E323K, as well as G364V and G367R, which may be better tolerated, because these residues are surface exposed and form no native side chain interactions.

Several additional substitutions cluster near the Ca2+ coordination sphere (Table 2; Fig. 1b and 3a). These include I477N/S, whose main chain interaction with Ca2+ may be compromised somewhat by the disruption of the β-sheet core but not enough to abolish Ca2+ binding, and the moderately stable apo variant D380A (46), which also serves to stabilize the side chain of cation–π residue Tyr 371 (Fig. 4c). Also in this region is A427T, identified within a small group of family members of variable glaucoma diagnosis (61), and a mild variant from a protein stability/aggregation standpoint (24). The substitution likely weakens Ca2+ coordination to Asn 428 and Ala 429; thus, our structure supports the initial pathogenic assignment.

Notably, neutral polymorphisms (http://www.ncbi.nlm.nih.gov/SNP), E352Q (rs61745146), K398R (rs56314834, (62,63)) and E396D (rs61730975), are located on surface loops, remote from the identified misfolding regions (Table 2). Similarly, A445V, identified in glaucoma patients (61–65), has a reported age of onset of 63 years, beyond the cutoff considered early onset (61,63). Like the SNPs (Table 2), Ala 445 is located on a remote surface and exhibits wild-type-like stability (24) (Table 3). Thus, our analysis would predict that the charge neutral A445V mutation would not be prone to misfolding; additional work will be required to strengthen the argument for glaucoma causality in associated individuals, for example, via one of the alternative mechanisms proposed for myocilin-associated glaucoma including altered endocytosis (66), mitochondrial membrane disruption (23) or apoptosis due to susceptibility toward reactive oxygen species (67).

### Table 3. Evaluated variants based on location in myoc-OLF structure

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Tm (°C)</th>
<th>Location</th>
<th>Original assignment</th>
<th>Proposed new assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>A445V</td>
<td>54.2 ± 0.2</td>
<td>Surface loop</td>
<td>Disease</td>
<td>SNP</td>
</tr>
<tr>
<td>V329M</td>
<td>49.6 ± 0.4</td>
<td>β-Sheet belt</td>
<td>SNP</td>
<td>Disease</td>
</tr>
<tr>
<td>S425P</td>
<td>40.8 ± 0.3</td>
<td>β-Sheet belt</td>
<td>SNP</td>
<td>Disease</td>
</tr>
<tr>
<td>R422C</td>
<td>52.1 ± 0.1</td>
<td>Surface loop</td>
<td>SNP</td>
<td>No change</td>
</tr>
<tr>
<td>Y473C</td>
<td>54.2 ± 0.1</td>
<td>Surface loop</td>
<td>SNP</td>
<td>No change</td>
</tr>
<tr>
<td>T293K</td>
<td>52.9 ± 0.6</td>
<td>Surface loop</td>
<td>Disease</td>
<td>SNP</td>
</tr>
<tr>
<td>T353I</td>
<td>53.1 ± 0.3</td>
<td>Surface loop</td>
<td>Ambiguous</td>
<td>SNP</td>
</tr>
</tbody>
</table>

*As listed in (16).

### Assessment of population-based myoc-OLF variants

For variants identified from large population-based genetic studies, designation as pathogenic or neutral polymorphism is complicated by the heterogeneous nature of glaucoma and age onset of disease, and is largely dependent on whether the variant is identified in control subjects or relatives of an affected person (62). In addition, few of the many tens of variants identified through these studies have been biochemically characterized for stability and aggregation propensity. Similar to the treatment of A445V above, we applied the structural inferences from familial mutations and SNPs to intuit the likelihood of pathogenicity of an additional ~60 disease-causing genetic lesions identified in such population studies but lacking confirmed protein behavior (16). Approximately 90% of the annotated disease or SNP variants in myoc-OLF fall within the categories described above. Six outliers were identified with possible alternative assignments, however. These proteins were recombinantly expressed and purified, and their Tm’s measured (Table 3) as a direct experimental determinant of structural deformation, and thus aggregation propensity and disease liability. First, V329M has been identified several times in glaucoma patients (63,68,69) but is labeled as a likely polymorphism, perhaps because Met is the residue found in the zebrafish myocilin homolog. Due to its location within a β-sheet, we predicted V329M to be a pathogenic variant. The V329M myoc-OLF variant is indeed somewhat thermally destabilized (Table 3), suggesting a mild pathogenic label, similar to the familial A427T variant described above, is likely more appropriate. S425P is also assigned as a neutral polymorphism variant (62), but we posited S425P would be prone to misfolding and aggregation, and thus instead be better characterized as pathogenic. Specifically, the Pro would compromise the position of Lys 423, which is involved in the stabilizing cation–π interaction (Fig. 4c). Indeed, the S425P myoc-OLF variant exhibits a low Tm (Table 3) and mediocre yield of folded protein previously associated with the least stable familial variants, even lower than that of Y437H (24). Variants harboring R422C and Y473C (62,63), both labeled as neutral polymorphisms, are further away from the aforementioned misfolding regions, but it was not clear whether a lone Cys would interfere with proper disulfide bond formation upon folding. Both the R422C and Y473C variants are thermally indistinguishable from wild type and are likely non-pathogenic, as originally suggested. Finally, labeled pathogenic mutation T293K (62–65) and T353I (63,68,70,71) whose pathogenicity is labeled as uncertain (16) are remote surface exposed residues.
unlikely to promote misfolding, in agreement with our structure-based prediction, these myoc-OLF variants are well folded, have wild-type-like stability and are more likely SNPs than causative via a misfolding phenotype in glaucoma.

Discussion

OLF domains are broadly involved in development and implicated in a host of human ailments, but molecular details of their biological activities and corresponding dysfunction in a variety of contexts remain active areas of investigation. Clarification of structure-(dys)function relationships continues to grow in importance as the biological reach of OLF domain-containing proteins expands in the context of the interactome (72–75). For myocilin and its relation to glaucoma, our results demonstrate how the myoc-OLF structure can be used to strengthen or dispute disease causality in identified mutants, and understand physicochemical features associated with a variant prone to pathogenic misfolding. Propellers do not fall into the traditional paradigm for globular proteins, namely propellers do not possess a hydrophobic core. Thus, they are both more difficult than others to accurately predict by sequence (76), and folding pathways are not well understood.

We identified three structural trigger zones for misfolding—the hydrophobic inner blades, Loop B-10/C-11, and the cavity region with Ca2+—where substitutions correlate directly with destabilizing variants. Mutations within the categorized regions yield a folded, but non-native structure that enhances myoc-OLF fibrillation propensity (25). The previously identified amyloidogenic peptides P1 and P3 (Supplementary Material, Figs S2 and S5) are found within the innermost two strands of opposing Blades B (Strands B–7 and B–8) and D (Strands D–15 and D–16). Their internal position explains why fibrillation does not proceed except upon accessing a partially folded state, facilitated by non-native amino acid substitutions. In addition, we were able to shed new light onto a variant of ambiguous pathogenicity, suggest that selected mutations thought previously to be neutral polymorphisms are likely pathogenic, and identify probable neutral polymorphisms on the surface of myoc-OLF for further interrogation. As continued genetic testing of glaucoma patients throughout the world reveals new myocilin variants, it will now be possible to integrate such structural knowledge to help evaluate whether a patient is diagnosed with glaucoma because of or unrelated to the myoc-OLF amino acid substitution. Moreover, the structure affords the opportunity to identify myocilin-directed molecules to stabilize pathogenic mutants against aggregation using rational drug discovery as a new therapeutic direction. No such targeted treatment for any glaucoma subtype is currently available; indeed, a new glaucoma drug has not been approved in over a decade (77).

The discovery that the OLF domain possesses a propeller fold falls in line with the existing and complex functional picture. Propellers are generally known for their ability to orchestrate multiple signaling events across diverse pathways and have a broad portfolio of interacting partners [see, for example RACK1 (78)], both intracellularly and extracellularly (79). Evolutionary trace reveals convergent and divergent features of OLFs that could not be inferred simply from sequence alignments and homology considerations. Available evidence from our structures suggests that the central cavity of the OLF domain possesses a common ligand-binding site. Access to the central cavity appears gated in myoc-OLF at the top face where exogenous small molecules appear in our structures. In support of the solvent accessibility of the internal hydrophilic core of myoc-OLF, glycerol, only introduced during cryoprotection of myoc-OLF(E396D) crystals, is seen within the central cavity even when the structure reveals the closed Loop B-10/C-11 conformation. Entrance to the internal cavity could come from the observed movement of Trp 373 or via other loop motions on the top or bottom face that have not yet been detected. Biologically relevant metabolites that, like glycerol, could occupy the identified internal pocket may include those that regulate or signal extracellular matrix remodeling, development, or other changes in activity in neurons or aqueous humor. For example, lactate is thought to be an energy source for neurons (80,81), which, like urea and ascorbate, has been identified in aqueous humor (82). Ascorbate is also related to collagen (83) and elastin (84) synthesis, and may play a role in extracellular matrix remodeling by matrix metalloproteinases (85). In support of these suggestions, lactate stabilizes myoc-OLF (86).

Furthermore, though prior metal analysis of myoc-OLF clearly identified a single stable ion within a chelator-resistant site (46), we cannot rule out the possibility that the crystallographic Na+ site is a labile site for Ca2+. A secondary Ca2+ site could enable OLFs to be responsive to calcium release or other extracellular function (87), or, in appropriate orthologs, neurological triggers. Internal sensing could propagate relevant changes to other regions of the OLF domain to facilitate higher order protein–protein interactions or complex formation. Additional structures of OLF domains from different subfamilies will be valuable in expanding our appreciation of the structural and surface divergence of the OLF domain family.

Analysis of the structures further indicates the possibility of enzymatic function for the OLF domain is remote. Known five-bladed β-propeller enzymes have their active sites at the top face of the propeller near the central cavity. While there is noted overlap of the highly conserved myoc-OLF Asp 380 residue with that of a catalytic aspartate of arabinase, it is an unlikely nuclease. Asp 380 is bound to the Ca2+ and Na+ ions as well as provides stabilization to and proper positioning of Tyr 371 of the cation–π interaction. Removal of Asp 380 as a Ca2+ ligand by its replacement with disease-associated Ala yields a destabilized apo protein (24,46) that readily aggregates in physiological buffers at 37°C (25). The remaining active site residues associated with hydrolysis in other five-bladed propellers are Thr residues in myoc-OLF, which are not commonly associated with catalysis (88). Major conformational changes in surface loops would be further required at the top face to create a binding pocket of sufficient size and depth to accommodate a small molecule substrate. Though it is possible to envision such motion, myoc-OLF variants that are expected to modulate the strength of the cation–π interaction are destabilized, suggesting that significant stabilization energy would need to be compensated upon such hypothetical substrate binding. The remaining loops other than B-10/C-11 are rather short and likewise may not be reconfigurable. Without candidate catalytic residues or a clear substrate-binding pocket, there is no obvious chemical reaction to propose and test.

In sum, the availability of OLF structures substantially advances our ability to interpret current data and design further experiments on function and dysfunction of the five-bladed β-propeller OLF domain. Knowledge gleaned through continued structure-(dys)function studies of myocilin-associated glaucoma can serve as a model to understand the molecular significance of identified mutations in other OLF-associated disorders (11,12,89) and adds to our comprehension of propeller protein folding more generally. Functionally, the OLF domain appears poised for interactions with other biological small molecules and/or binding partners. In the hydrophilic cavity, the binding entity is likely similar among OLFs. At the OLF surface, the lack of sequence conservation among subfamilies suggest dissimilar interactions and
thus function. While the specific functions, ligands and binding partners for OLF domains will likely continue to be a challenging long-term experimental pursuit, atomic-detail structural knowledge will be a valuable guide to focus future studies of this family of biomedically important proteins.

Materials and Methods

Protein expression, purification and characterization

Wild-type myoc-OLF and myoc-OLF(E396D) were expressed and purified as described previously (24,56). Myoc-OLF variants T353I, R422C, S425P, Y473C, V329M and T293K were generated by site-directed mutagenesis (QuikChange Lightening kit, Stratagene), verified by DNA sequencing (MWG Operon), expressed, purified and Tm calculated as the midpoint of unfolding using differential scanning fluorimetry, as described previously (24,56). Primers are listed in Supplementary Material, Table S2. SeMet-substituted myoc-OLF(E396D) was produced by expression of the corresponding MBF-OLF(E396D) fusion protein (24) in the auxotrophic T7 Express Crystal E.coli cell line (New England Biolabs). Individual colonies were inoculated into 10 ml starter cultures, grown in Superior Broth (US Biological) overnight at 37°C, further diluted into 1 l of Superior Broth and shaken at 200 rpm at 37°C until an OD600 = 2.0 was reached. Cells were then pelleted by centrifugation and resuspended in pre-warmed M9 minimal media supplemented with 0.4% glucose, 1 mM MgSO4, 0.1 mM CaCl2 and 0.0002% ferric ammonium citrate, but lacking methionine or seleno-methionine. Cells were allowed to shake at 37°C for 2.5 h to deplete any remaining methionine. Fifty micrograms per milliliter of Tm isopropyl β-d-thiogalactopyranoside (IPTG), and cells were allowed to grow overnight. Cells were pelleted, lysed and purified as above. Overall, yield was ~0.6 mg purified SeMet-substituted myoc-OLF(E396D)/L Superior Broth cell culture. Incorporation of SeMet was confirmed by using a Bruker S2 Picofox instrument.

Crystallization, data collection and structure determination

Purified SeMet-substituted myoc-OLF(E396D)221 was concentrated to 10 mg/ml, and crystals grown by the hanging drop method by equilibration against a solution containing 19% PEG 3350, 0.1M Bis-Tris pH 6 and 0.1M magnesium formate. Wild-type myoc-OLF (10 mg/ml, 50 mM HEPES, pH 7.5, supplemented with 0.36 mg/ml subtilisin A) grew crystals by equilibration against a reservoir solution containing 5% PEG 300, 39% PEG 200, 100 mM MES, pH 6.0, and crystals of myoc-OLF(E396D)222 (10 mg/ml in 10 mM HEPES, pH 7.5) were grown from 32% PEG 400, 5% PEG 3000 and 50 mM CAPSO, pH 9.5. Crystals of myoc-OLF(E396D) and myoc-OLF(E396D)221 were cryo-cooled in a solution containing the respective reservoir solution supplemented with 5% glycerol and 20% PEG 200. The wild-type myoc-OLF crystals were cooled from mother liquor directly. Diffraction data were collected at the Advanced Photon Source, Argonne National Labs beamlines, The National Institute of General Medical Sciences and National Cancer Institute (GM/CA-CAT) 23-ID (wild-type myoc-OLF) and Southeast Regional Collaborative Access Team (SER-CAT) 22-ID (myoc-OLF(E396D)). Data were processed using XDS/XSCALE (90). The initial structure was solved by single wavelength Se anomalous phasing using AutoSol (91) (myoc-OLF(E396D)221, mean anomalous difference in units of estimated standard deviation > 1 across all resolution bins in processed dataset, phasing figure of merit = 0.5) and the other two (wild-type myoc-OLF, high-resolution myoc-OLF(E396D)222) by molecular replacement using Phaser (92), using a model generated from experimental phasing. The models were iteratively built and refined using Coot (93) and Phenix.refine (92). PEG 200 was modified from the PEG 400 entry from HicUP database (94). For wild-type myoc-OLF and myoc-OLF(E396D)222, the pseudomerohedral twinning operator identified using Xtriage (95) was incorporated in the refinement algorithm. Structures have been deposited to the protein databank with PDB codes 4WXQ, 4WXS and 4WXU.

Structure analysis

Comparison to known structures in the PDB was enabled using DALI (96), crystal contact analysis using PDBiPISA (97), metal coordination analysis conducted using FINDSITE (98) and Metals(3) (99), and structural alignments using SSM (100). Sequence alignment for Supplementary Material, Figure S2 was prepared using PROMALS3D (101) and rendered in ESPript (102). Electrostatic surface was calculated using PDB2PQR (103) APBS (104) and figures generated in PyMOL (www.pymol.org), using the default secondary structure assignment settings. Evolutionary trace was conducted using default parameters except for increasing the number of OLF sequences to 1000, results of which were mapped on to the structure for rendering in PyMOL (53) (alignment in Supplementary Material, Table S3).

Supplementary Material

Supplementary Material is available at HMG online.

Conflict of interest statement. None declared.

Funding

This work was supported by grants from the American Health Assistance Foundation, Glaucosa Research Foundation, National Institutes of Health (R01EY021205), and Pew Scholar in Biomedical Sciences program to R.L.L., Georgia Tech Molecular Biophysics Senior Award to R.K.D., US Department of Education Graduate Assistance in Areas of National Need (P200A060188) to S.D.O., and a Georgia Tech Presidential Undergraduate Research Award to D.M.F. GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). Institutions supporting SER-CAT may be found at www.ser-cat.org/members.html. The Advanced Photon Source is a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contracts No. DE-AC02-06CH11357 and W-31-109-Eng-38.

References


sequence variants in German normal tension glaucoma patients. Mol. Vis., 11, 284–287.
100. Krysinski, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure...


