Structure and misfolding of the flexible tripartite coiled coil domain of glaucoma-associated myocilin

Shannon E. Hill[1], Elaine Nguyen,[1] Rebecca K. Donegan[1], Anthony Hazel[2], James C. Gumbart[2], Raquel L. Lieberman[1*]

Affiliations

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta GA 30332

[2] School of Physics, Georgia Institute of Technology, Atlanta GA 30332

*Correspondence and Lead Contact: Raquel Lieberman,

Raquel.lieberman@chemistry.gatech.edu, phone: 404-385-3663

1

Summary

Glaucoma-associated myocilin is a member of the olfactomedins, a protein family broadly involved in neuronal development and human disease. Molecular studies of the myocilin N-terminal coiled coil demonstrate a unique tripartite architecture: a disulfide-linked, parallel dimer-of-dimers Y-shaped molecule, with distinct tetramer and dimer regions. The structure of the C-terminal 7-heptad dimer elucidates an unexpected repeat pattern involving electrostatic inter-strand stabilization. Molecular dynamics simulations reveal an alternate conformation in which the terminal inter-strand disulfide bond limits the extent of unfolding and results in a kinked configuration. Taken together, full-length myocilin is also branched, with two pairs of C-terminal olfactomedin domains. Selected variants within the N-terminal region alter the apparent quaternary structure of myocilin but do so without compromising stability or causing aggregation. In addition to increasing our structural knowledge of extracellular coiled coils for protein design and biomedically important olfactomedins, this work broadens the scope of protein misfolding in the pathogenesis of myocilin-associated glaucoma.

Keywords

Glaucoma, small angle X-ray scattering, extracellular matrix, coiled-coil, X-ray crystallography, molecular dynamics simulations, protein design, protein misfolding

Highlights

- Glaucoma-causing extracellular protein associated with amyloid forming propensity

- Structural studies confirm tripartite parallel dimer-of-dimers coiled coil

- Leucine zipper exhibits non-canonical heptad repeat pattern with disulfide cap

- Glaucoma-associated variants in tetramer alter structure but not stability

eTOC blurb

Hill et al describe the structure of the coiled-coil region of myocilin, the extracellular olfactomedin family member closely associated with the ocular disorder glaucoma. Myocilin's coiled coil adopts a unique Y-shaped parallel dimer-of-timers employing an unusual heptad repeat pattern. Selected disease variants alter quaternary structure.

Introduction

Olfactomedins (PFAM: PF02191), a family of extracellular, modular, multidomain proteins each with a 5-bladed β-propeller olfactomedin (OLF) (Donegan et al., 2015) domain, are broadly involved in eukaryotic development (Tomarev and Nakaya, 2009; Zeng et al., 2005). The majority of olfactomedin subfamilies also contain a high probability signal sequence for cellular secretion followed by homo-oligomerizing coiled-coil (CC) domains of varying lengths (Figure 1). Despite their prevalence in multicellular organisms and implication in a variety and increasing number of human diseases (Anholt, 2014), however, the molecular architecture and specific functions of olfactomedins remain largely unknown.
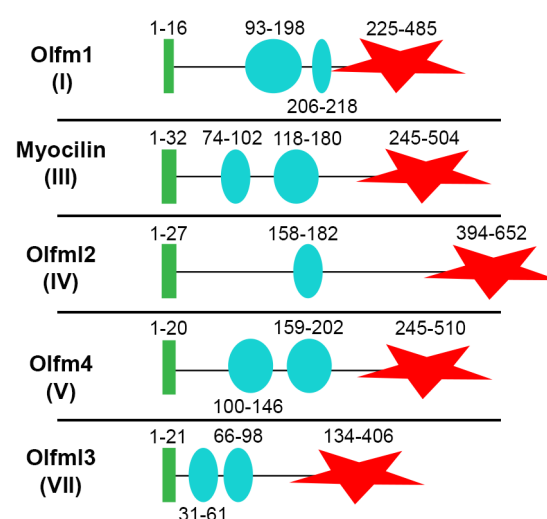

Figure 1. Distinctive domain structures among CC-containing olfactomedin subfamilies. Green bar, predicted signal sequence for cellular secretion; cyan circles, predicted CC regions; red star, olfactomedin domain. Subfamily II is membrane-anchored; subfamily VI has a collagen domain.

4

The structures of CCs, one of the most common structural domains in nature (Wolf et al., 1997) and of considerable interest for protein design and biomaterials (Lupas and Bassler, 2017; Woolfson, 2005), can be challenging to predict *in silico*, particularly beyond dimers that are common among extracellular CCs (Kammerer, 1997). The CCs encoded within olfactomedins tend to have non-canonical heptad repeat (*abcdefg*)$_n$ patterns, diverging from the regular period of hydrophobic residues at positions '*a*' and '*d*' and electrostatic residues at positions '*e*' and '*g*' (Mason and Arndt, 2004). Thus, bioinformatics programs detect CCs in olfactomedins, but vary in predictions for register, stoichiometry, and orientation. Olfactomedin-1 (subfamily 1) is the only olfactomedin protein for which a supramolecular organization has been experimentally characterized: an apparent V-shaped molecule seen by transmission electron microscopic imaging and small angle X-ray scattering (SAXS) envelope of the full-length protein(Pronker et al., 2015). However, because olfactomedin family CC domains share very low sequence similarity (< 15% identity overall, Supplementary Figure S1), results from olfactomedin-1 are not predictive of other olfactomedin family members.

Here we characterized the unique architecture of full-length myocilin (olfactomedin subfamily 3), one of the best-studied olfactomedins, in molecular detail. Myocilin is expressed at high levels in the trabecular meshwork (TM), an extracellular matrix in the anterior segment of the eye that forms part of the anatomical ultrafiltration system for aqueous humor (Abu-Hassan et al., 2014), and is a critical tissue for maintaining intraocular pressure. The myocilin CC-containing region has been reported to interact with matrix components (Filla et al., 2002; Ueda et al., 2002; Wentz-Hunter et al., 2004) and confer its adhesion properties (Goldwich et al., 2009; Wentz-Hunter et al., 2004), but there is much still to learn about the structure and biological function of myocilin. Myocilin is associated with two subtypes of glaucoma (Polansky et al.,

5

1997; Rozsa et al., 1998; Stone et al., 1997), a leading cause of blindness worldwide typically associated with elevated intraocular pressure (Quigley and Broman, 2006). Mutations in myocilin are causative for Mendelian-inherited form of open angle glaucoma (~4% of ~70 million), the most common subtype (Hewitt et al., 2006). OLF-resident myocilin variants are destabilized (Burns et al., 2010; Burns et al., 2011) and prone to amyloid-like aggregation (Hill et al., 2014; Orwig et al., 2012), which is believed to cause a toxic gain of function, leading to early-onset glaucoma (Stothert et al., 2016).

We present biophysical and biochemical experiments demonstrating that the myocilin CC domain is a distinct Y-shaped molecular entity composed of two parallel dimer-of-dimers that are stabilized on the three ends by disulfide bonds. The 1.9 Å resolution X-ray structure of the C-terminal 7 heptads of the CC further reveals an unexpected repeat pattern involving inter-strand stabilization between oppositely charged residues at heptad position '*a*' and '*g*'. Molecular dynamics simulations demonstrate that the C-terminal capping disulfide bond limits unfolding of the CC and enables an alternative kinked structure to form. Selected glaucoma-associated mutations within the N-terminal coiled coil region adopt non-native structures but are not destabilized or aggregation prone. Our study increases structural knowledge of extracellular coiled coils, biomedically important olfactomedins, and broadens our appreciation of protein misfolding in the pathogenesis of myocilin-associated glaucoma.

Results and Discussion

The myocilin coiled-coil imparts a unique tri-branched supramolecular architecture. After unsuccessful attempts to express soluble full-length myocilin in *E. coli* or purify sufficient quantities from spent media of primary human TM cells (not shown), we generated well-behaved constructs of myocilin outside of the OLF domain whose structure we solved previously (Donegan et al., 2015). Biophysical characterization of NTD_{33-227}, which includes linker regions before and after the predicted CC domain (Figure 1), reveals the typical circular dichroism (CD) signatures for α-helices. Thermal unfolding is reversible (Figure 2a, Table 1, melting temperature ((T_m) ~ 67 °C), a value considerably higher than the ~52 °C measured for the OLF domain (Orwig and Lieberman, 2011). To investigate the contribution of Cys residues to the stability and oligomeric state of NTD_{33-227}, Cys-to-Ser mutants were prepared. The Cys residues thermally stabilize NTD_{33-227} nearly 20 °C, with Cys185 contributing the biggest thermal increase (Table 1), but do not change the elution profile from size exclusion chromatography (SEC) and thus the apparent quaternary structure (Figure 2b). In each case, the predominant species present upon disulfide bond formation is a dimer (Figure 2c).
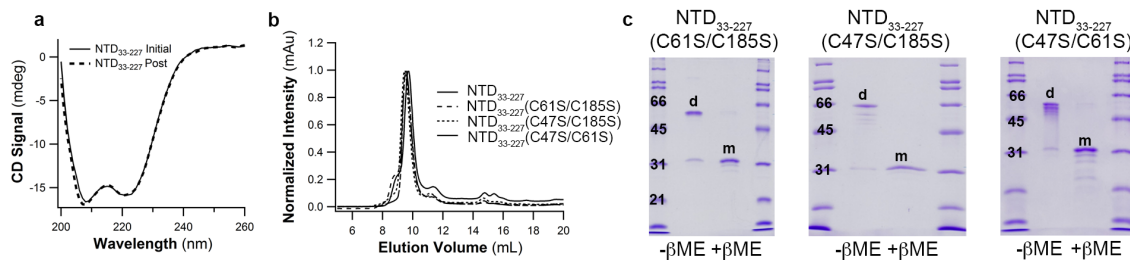
Figure 2. Biophysical characterization of NTD$_{33-227}$. (a) Initial NTD$_{33-227}$ circular dichroism spectrum reveals α-helical signature. Overlay with post-melt NTD$_{33-227}$ spectrum confirms thermal unfolding is reversible. (b) SEC elution profiles demonstrate unchanged quaternary structure for wild type and Cys-to-Ser NTD$_{33-227}$ variants. (c) SDS-PAGE analysis of NTD$_{33-227}$ Cys-to-Ser variants under non-reducing conditions show predominantly dimer (d) species. Calculated molecular weight for NTD$_{33-227}$ monomer (m) = 27 kDa.

Table 1. Tm values for constructs in this study.

Construct	Tm (°C)	Rationale
NTD$_{33-227}$	66.7	All except OLF domain
NTD$_{33-227}$ (C61S/C185S)	56.8	Cys$_{47}$ only
NTD$_{33-227}$ (C47S/C185S)	56.3	Cys$_{61}$ only
NTD$_{33-227}$ C47S/C61S	64.4	Cys$_{185}$ only
NTD$_{33-227}$ (C47S/C61S/C185S)	44.2	Cys-less construct
CC$_{69-184}$	44.2	
CC$_{69-185}$	64.5	+Cys$_{185}$
NTD$_{33-227}$ V53A	67.0	GCM[*], within linker
NTD$_{33-227}$ R82C	68.0	GCM, within tetramer
NTD$_{33-227}$ L95P	66.1	Neutral, within tetramer
NTD$_{33-227}$ R126W	66.2	GCM, within dimer
NTD$_{33-227}$ R128W	68.1	Neutral, within dimer
NTD$_{33-227}$ L215Q	67.3	GCM, within linker

[*]GCM (glaucoma causing mutation) or Neutral as annotated in the database myocilin.com

8

To clarify further the oligomeric state and shape of the structural coiled coil of myocilin, we next prepared a CC construct bracketed by the two Cys residues closest to the predicted CC-(CC_{60-185}). The SEC elution profile of this species is consistent with a 4- or 5-mer (Supplemental Figure S2a). The molecular envelope determined by SEC-SAXS analysis (Supplemental Figure S2) reveals a branched, tripartite structure that looks like the letter 'Y' (Figure 3a, Table 2). There is a short stem (3-4 nm wide x 4 nm long) and two arms (each 3-4 nm wide and 9-11 nm long) ~130° apart. The molecule appears nearly rotationally symmetric, but symmetry is broken due to apparent flexibility of the arms (Figure 3a).

To the best of our knowledge, the Y-shape topology is unprecedented among both naturally occurring and designed CCs. Modular proteins containing repeat structures are common within the extracellular matrix proteome (Hynes and Naba, 2012), and the addition of the myocilin CC diversifies the modular arrangement. The myocilin CC further expands our appreciation for the versatility of deceptively straightforward CCs and should motivate the design of bio-inspired materials with novel supramolecular arrangements, a highly active research area of protein science (Lupas and Bassler, 2017).
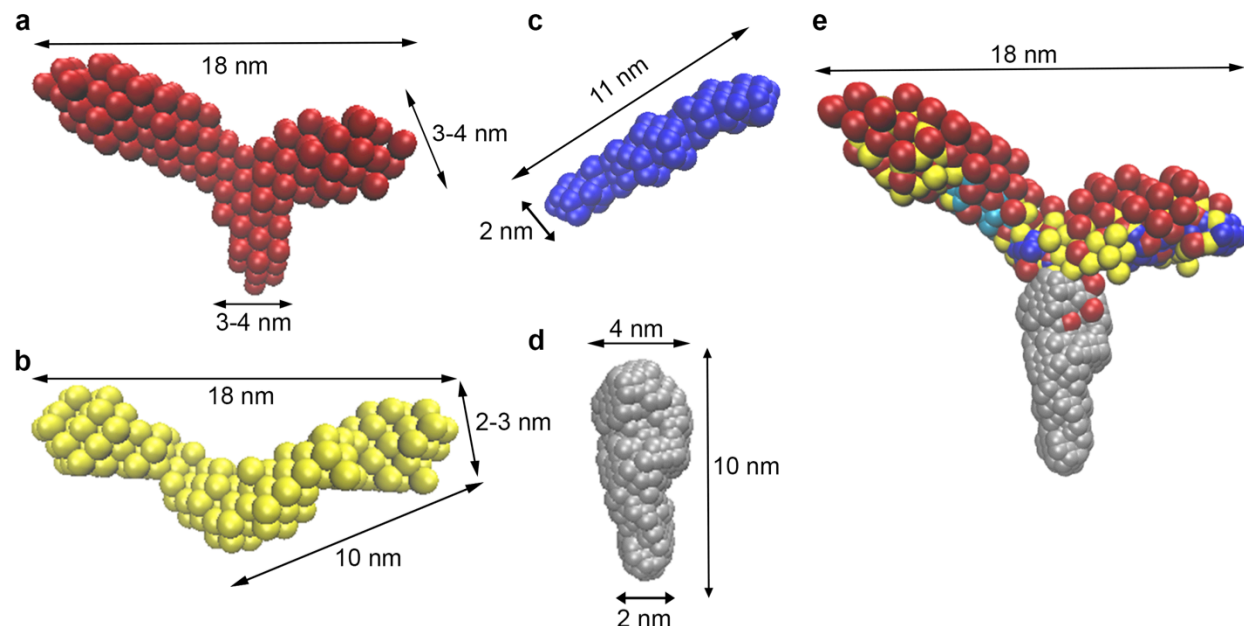
Figure 3. *Ab initio* models from SEC-SAXS reveal tripartite structure for myocilin CC. (a) CC_{60-185} is Y-shaped with arms spanning 18 nm and 130° apart from a stem ~3-4 nm in width. (b) CC_{69-185} is a P2-symmetric V-shaped molecule with similar dimensions to arms of CC_{60-185}. (c) $CC_{112-185}$ is a P2-symmetric, ~11 nm long rod (d) CC_{33-111} is an oblong molecule with varying widths of 2-4 nm. (e) Overlay of SAXS molecular envelopes in a-d. See also Table 2.

Table 2. SEC-SAXS Statistics

Construct	R_g (Å)	Monomer MW	SAXS Mol (Oligomer)	Chi^2	NSD
CC_{60-185}	60.2	14.6 kDa	72 kDa (4.9)	0.9692	0.900
CC_{69-185}	56.5	13.7 kDa	49 kDa (3.6)	1.104	0.695
				1.181 (P2)	1.532 (P2)
CC_{33-111}	29.4	8.8 kDa	25 kDa (2.8)	1.006	0.606
$CC_{112-185}$	32.9	8.9 kDa	21 kDa (2.4)	1.036	0.526
				1.118 (P2)	1.801 (P2)
mLZ_{93-171}	32.2	9.1 kDa	20 kDa (2.2)	1.072	0.586
				1.078 (P2)	1.268 (P2)

Chi^2 and normalized spatial discrepancy (NSD) values are from *ab initio* modeling. P2 indicates values obtained from *ab initio* modeling when constrained to P2 symmetry.

10

Identification of parallel tetramer and dimer CC subdomains within Y-shaped molecule. To probe the molecular principles governing the adoption of the Y-shaped structure, we next tested the hypothesis that the Y-shaped molecule apparent from the SEC-SAXS envelope of CC_{60-185} is composed of distinct tetramer and dimer regions corresponding to the two CC sub-regions predicted by Marcoil (Delorenzi and Speed, 2002) and Logicoil (Vincent et al., 2013), residues ~74-102 and ~118-180 (Figure 1). The former stretch escapes detection by other CC prediction programs but even when predicted, assignment of its oligomeric state is of low confidence (not shown). The latter stretch is well predicted as it has regularly spaced Leu residues indicative of the leucine zipper CC subtype.

Biophysical characterization of the Cys-free CC_{69-184} exhibits moderate stability (Table 1), high helical content, and elutes from SEC at a volume consistent with a molecular mass consistent of a tetramer (Supplemental Fig. S3a,b). The addition of Cys_{185} to yield CC_{69-185} increases expression yields and increases thermal stability by 20 °C (Table 1) but secondary and quaternary structures do not change (Supplemental Fig. S3b,c). Non-reducing gel analysis of disulfide bond contributions to the oligomeric state CC_{69-185} reveal only dimers, not higher ordered species (Supplemental Fig. S3d). The SEC-SAXS envelope of CC_{69-185} reveals a P2 symmetric, V-shaped molecule with dimensions 18 nm x 10 nm (Figure 3b, Table 2, Supplemental Fig. S4, Supplemental Fig S5a). The envelope of CC_{69-185} overlays well with the larger CC_{60-185} construct (Supplemental Fig S5b), and appears to be missing the third 'stem'. We suggest that the stem is not resolved in this construct because, without the disulfide bond from Cys_{61}, the protein is susceptible to proteolysis.

To clarify the orientation of CC_{69-185}, relative to the Y-shaped CC_{60-185}, we characterized two additional constructs, CC_{33-111} and $CC_{112-185}$, biochemically and by SEC-SAXS

11

(Supplemental Fig. S6, S7, Table 2). Both proteins are, as expected, helical (Supplemental Fig. S6a,d) but the SEC-SAXS envelopes reveal distinct rods for each constructs (Figure 3c, d). By analysis of its SEC profile, CC_{33-111} is a tetramer (Supplemental Fig. S6b). The corresponding rod is not symmetric, with half of the rod length ~4 nm wide and half ~2 nm wide, perhaps reflecting structural differences in the predicted unstructured N-terminal region and the predicted CC region starting near Gln 75. The SEC profile of $CC_{112-185}$ indicates a dimer (Supplemental Fig. S6e), which is disulfide bonded (Supplemental Fig. S6f,g). $CC_{112-185}$ best matches the length of the arms of the V-shaped CC_{69-185} (Figure 3c, Supplemental Fig. S5c,d). There is a bulge observed approximately in the middle of the $CC_{112-185}$ rod, a feature seen in the pairwise distribution (P(r)) plot of the SEC-SAXS data (Supplemental Fig. S5d, Supplemental Fig. S7h, and see below). These data indicate that CC_{60-185} is composed of two sequentially identical C-terminally disulfide-bonded dimers, which converge at the N-terminal region to form a tetrameric stem (Figure 3e).

Crystal structure of the 7 C-terminal heptad repeats reveals atypical stabilization for a leucine zipper. Next, we solved a 1.9 Å resolution crystal structure of the C-terminal 7 heptad repeats of the mouse myocilin leucine zipper region, $mLZ_{122-171}$, (residues 136-185 in human myocilin, 80% sequence identity, Figure 4a, Supplemental Table S1). This crystal grew after *in situ* proteolysis of mLZ_{93-171}, a construct nearly identical to $CC_{112-185}$ in sequence and structure (Figure 3c, Supplemental Fig. S8). The susceptibility of mLZ_{93-171} to further proteolysis is not immediately clear, but the presence of a so-called skip residue at Thr89/Thr103 (Figure 4a) suggests the possibility of extended unraveling as seen in myosin (Taylor et al., 2015). The structure of $mLZ_{122-171}$, which was solved by combining *ab initio* modeling with molecular replacement, reveals the C-terminal disulfide bond (Figure 4b) anticipated from biochemical

experiments (Table 1, Figure 2c, Supplemental Fig. S3d, Fig. S6f, g), capping a parallel dimer with a left-handed supercoiled twist (Figure 4c). The two coiled coils in the asymmetric unit are nearly identical to each other, as are the two helices within each coiled coil, with root mean squared deviations (r.m.s.d.) ~ 1.5 Å (not shown).
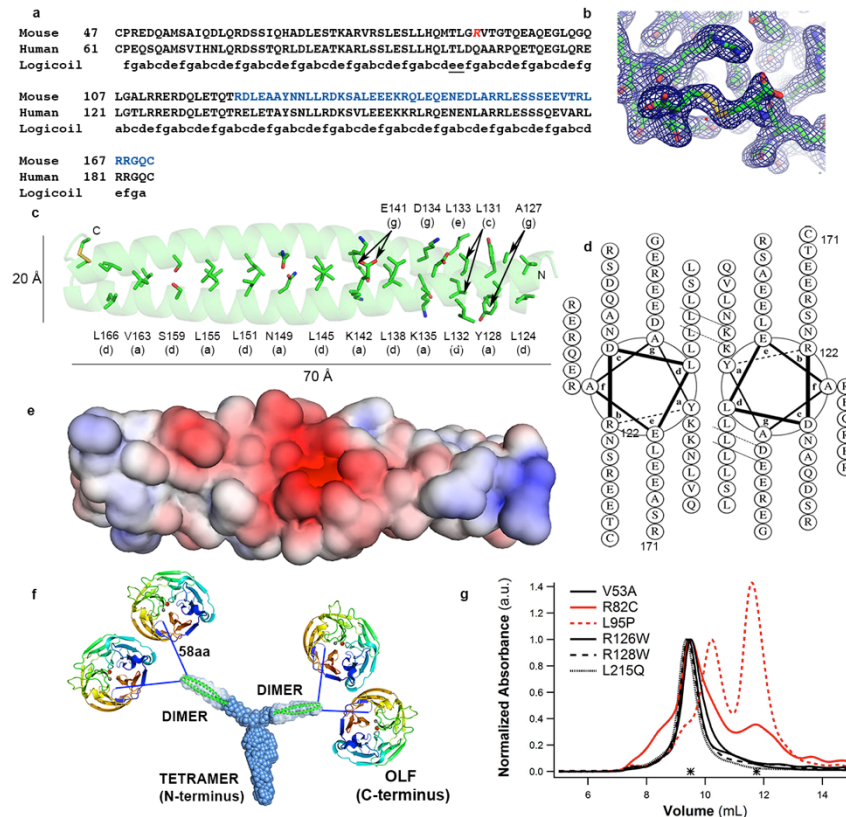
Figure 4. Crystal structure of mLZ$_{122\text{-}171}$, overall myocilin model, and effects of selected glaucoma-relevant variants. (a) Partial sequence alignment of human and mouse myocilin with predicted register of heptad repeats. Red Arg, start of the construct used for crystallization (mLZ$_{93\text{-}171}$). Blue stretch, crystallized fragment mLZ$_{122\text{-}171}$. (b) Two mLZ$_{122\text{-}171}$ helices are covalently tethered by a C-terminal disulfide bond. Final 2Fo-Fc (blue) electron density map is contoured at 1.0 σ and Fo-Fc (red/green) is contoured at ±3.5 σ. (c) Ribbon diagram of mLZ$_{122\text{-}171}$ parallel dimer with labeled heptad positions (a) and (d) and disulfide represented as sticks. (d) mLZ$_{122\text{-}171}$ in helical wheel representation. (e) Electrostatic surface (negative (red, -5 kT/e-) to positive (blue, + 5 kT/e-)) of mLZ$_{122\text{-}171}$ dimer. (f) Overall model of full-length myocilin: Y-shaped dimer-of-dimers composed of an N-terminal parallel tetrameric domain leading to two parallel CC dimers, followed by 58 amino acid long linkers, and C-terminal OLF domains. (g) SEC profile comparison of six NTD$_{33\text{-}227}$ glaucoma-relevant variants reveals altered elution and thus quaternary structure for R82C and L95P (red traces).

14

Analysis of the $mLZ_{122-171}$ structure reveals key deviations from expected general patterns found in the stereotypical "knobs-in-holes" CC (Parry et al., 2008) (Figure 4c, d). Overall, $mLZ_{122-171}$ follows the pattern of a hydrophobic residue, Leu, in position '*d*' except for the penultimate heptad in which it is a polar Ser159. By contrast, fewer than half of the residues in the '*a*' register are hydrophobic. The four remaining residues at '*a*' positions, which are highly conserved among myocilin from human, mouse, rat, fish and cow (Supplementary Figure S9a), are positively charged or polar. Two Lys residues are in positions poised to stabilize negatively charged aspartate and glutamates found one residue N-terminal in the adjacent helix at the '*g*' position, though we note these distances are somewhat long for electrostatic interactions (~3-5 Å). Consensus from prior studies of CCs has suggested that '*a*' positions can tolerate polar residues because they are less buried. Polar residues like Asn are specifically associated with parallel dimers (Mason and Arndt, 2004); in our structure, Asn149 is in position '*a*'. Charged residues in '*a*', particularly the frequency of their occurrence is unexpected. One protein design study identified this atypical pairing as important for the specificity of CC interactions among closely related sequences (Grigoryan et al., 2009), although the need for such specificity is not obvious for myocilin. In support of the importance of the *a-g* pairing for specificity, and not the traditional '*e*' and '*g*' positions, there are three negatively charged residues in position '*e*' in $mLZ_{122-171}$ and four in '*g*', with just one positively charged residue in '*e*' and '*g*' (Figure 4d). Two of the acidic residues in '*g*', which do not participate in salt bridges with 'core' Lys residues in '*a*' positions, are proximal to Asn149 and impart the overall negative electrostatic surface charge of $mLZ_{122-171}$ (Figure 4e). Remarkably, despite these non-canonical repeat features, $mLZ_{122-171}$ overlays well with the closest structural, but distant sequential, homologs, the prototypical leucine zipper GCN4 and geminin (24% and 19% sequence homology across 40

15

residues, and r.m.s.d.= 1.05 Å and 1.54 Å, respectively, not shown).

Molecular dynamics (MD) simulations suggest the potential for mLZ$_{122-171}$ to adopt an alternative structure. After 50 ns of equilibration at 310 K, the helical structures of mLZ$_{122-171}$ and the disulfide-deletion variant mLZ$_{122-170}$ remain essentially unchanged (Supplemental Fig. S10a). However, after 100 ns at 400 K, a bend emerges in the structure of mLZ$_{122-171}$ near the C-terminus (Ser159-Gln170), concomitant with a shift in paired interactions between oppositely charged residues on either helix (Supplemental Fig. S10b,c). This region of the leucine zipper includes paired Ser159 residues located in the heptad 'd' position as well as other polar/charged residues. Without the disulfide and consistent with stability measurements (Table 1), the mLZ$_{122-170}$ dimer is considerably more susceptible to unfolding. For example, in the 400 K simulation, the distinct kinked structure was not formed. In steered molecular dynamics simulations at 310 K in which the centers-of-mass of each coil were pulled apart at a constant velocity of 1 Å/ns for 10 ns, the C-terminal ends separated completely (Supplemental Fig. S10d). In line with its more hydrophobic character, the N-terminal region of mLZ$_{122-171}$ and mLZ$_{122-170}$ dimers remained unaltered in all simulations. In sum, the disulfide bond at the C-terminus of the myocilin leucine zipper appears to limit the extent to which each protomer can unfold and the prevalence of complementary charged residues allows the coil to adopt an alternative structure. Such resistance to unfolding is likely a feature that helps myocilin cope with the regular mechanical distortions of the TM ocular tissue (Acott et al., 2014).

Model of full-length myocilin. With the molecular structures of the myocilin CC and OLF domains (Donegan et al., 2015) in hand, we propose the following model of full-length myocilin (Figure 4f). Myocilin adopts a distinct oligomeric state defined by the dimer-stabilizing trio of Cys residues 47, 61, and 185, and the CCs, which confer a unique Y-shaped architecture

16

of a branched dimer-of-dimers tetramer. We found no evidence for higher ordered states for myocilin seen in previous studies (Fautsch et al., 2004; Russell et al., 2001). Higher ordered states might form *in vivo*, perhaps mediated by a yet-unidentified binding partner not present in our purified samples. It is also possible that high molecular weight myocilin was observed because myocilin in prior studies was misfolded, either due to incomplete posttranslational processing of the signal sequence or due to handling of antioxidant aqueous humor fluid (Richer and Rose, 1998) *ex vivo*, both of which could lead to aberrant inter-molecular disulfide bond formation. Two parallel dimers emerge at an obtuse angle from the N-terminal tetrameric stalk, imparting multivalency to the otherwise monomeric C-terminal OLF domain, likely to increase avidity and space OLF pairs apart from one another. Based on our constructs the transition from tetramer to the two dimers begins between Ser69 and Glu112, and most likely includes the lower-CC probability region containing the skip residue, Thr103 (Figure 4a, Supplemental Figure S9a). Aside from structural regions, myocilin also contains lengthy, linker regions on both N- and C-terminal ends of the structural Y, the functional/structural significance of which is not currently known.

Implications for other olfactomedin family members. The high level of sequence conservation across myocilin-like subfamily 3 members (Supplemental Fig. S9b) indicates the Y-shape architecture is shared across eukaryotes. The topology of myocilins is distinct from the dimer-of-dimers V of olfactomedin-1 (subfamily 1), however, likely a result of the different coil prediction pattern (Figure 1) and low overall sequence similarity. For example, angle of the V of olfactomedin-1 appears acute, not obtuse like measured for CC_{69-185}. In addition, the C-terminal region of the olfactomedin-1 CC, which harbors a disulfide-forming Cys found within the CC (Pronker et al., 2015), is not aligned with the CC-capping residue Cys 185 in myocilin (Figure

17

4b, Supplemental Fig. S1b). In olfactomedin-1, the end of the disulfide-harboring CC is immediately adjacent to the start of the OLF domain (Pronker et al., 2015) whereas in myocilin Cys185 is 60 residues away from the start of the structural OLF at Cys245. Whether the remaining three CC-containing olfactomedin subfamilies adopt a dimer-of-dimers and to what extent they are similar to myocilin or olfactomedin-1 remains to be seen. Other open questions across the olfactomedin family surround the functional consequences of the unusual olfactomedin topologies. A naive interpretation would suggest the CCs are tailored "molecular rulers," to space pairs of OLF domains, but mechanotransduction or as molecular scaffolds for complex protein-protein interactions are intriguing alternative possibilities.

Structural defects imparted by single nucleotide polymorphism and glaucoma-associated myocilin variants. In contrast to mutations found within the myocilin OLF domain, many of which are associated with early-onset and familial glaucoma (Hewitt et al., 2006), the vast majority of mutations in the N-terminal module have been identified via case-control genome sequencing studies. As such, their annotation as pathogenic is predicted statistically. The extent to which the toxic gain-of-function hypothesis for myocilin-associated glaucoma (misfolding/non-secretion/aggregation) holds for mutations found beyond OLF is unclear because such variants have not been systematically studied. Just one study investigated cellular trafficking and adhesion properties of two annotated N-terminal disease variants (R82C and R126W, a rare late-onset familial mutation (Faucher et al., 2002)) and one neutral (L95P) mutant (Gobeil et al., 2006). Compared to most OLF-resident mutations, which were retained intracellularly (Liu and Vollrath, 2004; Vollrath and Liu, 2006; Zhou and Vollrath, 1999), all three N-terminal variants tested (Gobeil et al., 2006) were secreted, but for R82C and L95P variants, partial intracellular sequestration was observed in a cell-based assay. R82C- and L95P-

18

mutant myocilins also lost adhesion properties in a cell assay (Gobeil et al., 2006), but R126W did not.

On the basis of the examined wild-type CC constructs, which exhibit reversible thermal unfolding (Figure 2a, Supplementary Fig. S3a), it is apparent that the CC domain is not intrinsically prone to aggregation like OLF, but disease-associated N-terminal mutants might still misfold by exhibiting structural anomalies without aggregation. To test our hypothesis, we biophysically characterized six NTD$_{33-227}$ variants (Table 1, Figure 4g, Supplemental Fig. S11). Linker-resident, conservative substitutions V53A and L215Q are annotated as glaucomatous (Hewitt et al., 2007). Biophysically, however, they are indistinguishable from wild-type in stability, thermal unfolding reversibility, and SEC elution profile, indicating no major structural changes have occurred. In the dimer region of the CC, the two changes R126W and R128W, the latter annotated as neutral, are expected to be poorly tolerated. This is especially true for R128W, in which the Arg appears to fall in the electrostatically positive '*a*' position for pairing with a negatively charged Glu '*g*' in an adjacent helix. Surprisingly, both variants behaved as wild-type. In the tetramer region, which has thus far eluded crystallization, similarly non-conservative substitutions R82C and L95P previously mentioned, were also not expected to be well tolerated. Both of these variants exhibited wild-type like stability and reversible thermal unfolding, with no evidence for aggregation (Supplemental Fig. S11). The elution profiles from SEC differ from wild-type, however, in that they are both mixtures of wild-type and smaller, non-native species (Figure 4g).

In sum, our characterization of myocilin reveals a new topology for CCs, an unprecedented branched Y-shaped protein structure. The tetramer region is sensitive to mutation, as changes in quaternary structure were observed for R82C and L95P, with a consequence of

19

non-native cellular secretion and poor cellular adhesion observed previously. Further studies will be required to assess whether both variants are glaucoma-causing, and to what extent gain-of-function (misfolding, intracellular sequestration) and/or loss-of-function (adhesion, mechanical properties, other) mechanisms are at play. By contrast, the dimer region appears relatively resilient to mutation and unfolding, and may intentionally access multiple distinct conformations. Documented linker-resident mutations, which are not expected to affect the quaternary structure nor located at positions with high likelihood for posttranslational modification, may not be causative for glaucoma. Beyond expanding the scope of naturally occurring protein structures, defects associated with new variants identified in the population are synergistic with efforts to identify at-risk patients to manage glaucoma prior to onset of vision loss (Souzeau et al., 2016) and to develop personalized medicine for glaucoma.

Materials and Methods

Molecular Cloning, Protein Expression, and Purification

Constructs of N-terminal portions of human myocilin were obtained by subcloning from an *E. coli* codon-optimized construct of full-length human myocilin (DNA 2.0) as described in Supplemental Methods. Proteins were expressed in *E. coli* BL21(DE3)plysS cells using the same protocol. A single transformed colony was inoculated into a selective 150-250 mL Luria-Bertani culture (LB, Fisher, supplemented with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol) and agitated at 225 r.p.m., 37 °C, overnight (16-24 hours). A starter culture (20 mL) was used to inoculate 1L of Superior Broth media (U.S. Biological), which was then agitated at 225 r.p.m., 37 °C until an optical density at 600 nm of 1.0-2.0 was reached. Protein expression was induced with 0.5 mM ispropyl β-D-thiogalactopyranoside (IPTG) at 18 °C and induced cells were grown overnight (16-20 hours) by agitation at 225 r.p.m., 18 °C. Cells were harvested by centrifugation and flash frozen in liquid nitrogen for storage at -80 °C. Cells were lysed by French Press, protein constructs purified by Ni^{2+}-affinity and size exclusion chromatography, and further evaluated as detailed in Supplemental Methods.

Circular Dichroism

Far-UV circular dichroism (CD) spectra and thermal melts were acquired on a Jasco J-815 spectropolarimeter equipped with a Jasco PTC-4245/15 temperature control system. Protein samples at a concentration range of 5-30 μM were measured in 50 mM Tris pH 7.5, 200 mM NaCl buffer at 4 °C. Scans were measured from 200 nm to 300 nm at a rate of 200 nm/min and a data pitch of 1 nm using a 0.1 cm cuvette. Each measurement was blank-subtracted with buffer and is an average of 10 scans. Far-UV thermal melts were performed from 4 to 80 °C at a rate of 1 °C min^{-1} increase in temperature and a data pitch of 2 °C. Ten scans from 300 to 200 nm at a

21

200-nm min^{-1} scan rate were averaged for each temperature. The T_m was determined by plotting the ratio of values recorded at 208 nm and 222 nm versus temperature and analyzed by Boltzmann Sigmoid analysis using Igor Pro.

SEC-SAXS

SEC–SAXS experiments were performed at BioCAT (beamline 18ID, Advanced Photon Source at Argonne National Labs); details of the experimental set up and data analysis are presented in Supplemental Methods.

Crystallization, data collection, structure determination

Crystals of mouse LZ_{93-171} (later determined to be mouse $LZ_{122-171}$) at 20 mg/mL in 10 mM Hepes pH 7.5 No NaCl grew after 90 days in 0.1 M ammonium acetate, 0.1 M BisTris pH 5.5, and 17% PEG 10,000 from a Hampton Index, HR2-134, sparse matrix sitting drop tray. Crystals were cryoprotected in 0.1 M ammonium acetate, 0.1 M BisTris pH 5.5, 25% PEG 10,000, and 25% glycerol. Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline and processed using HKL-2000 (Otwinowski and Minor, 1997). An initial molecular replacement solution was found using AMPLE (Bibby et al., 2012), which consisted of three ~3 kDa helical poly-alanine chains with reasonable electron density. Using one of these ~3 kDa chains another round of molecular replacement was performed with Phaser (McCoy et al., 2007) but with a search criteria of eight molecules in the asymmetric unit as suggested by analysis of Matthews Coefficient (Kantardjieff and Rupp, 2003). Visualization of neighboring unit cells in Coot (Emsley et al., 2010) suggested a probable connection of pairs of chains, leading to the final solution of four ~6 kDa poly-alanine chains in the asymmetric unit. The poly-alanine chains were iteratively built and refined using Coot (Emsley et al., 2010) and Phenix.refine (Afonine et al., 2012) by slowly replacing alanines with respective amino acids to

fit the electron density. The final model consists of mouse residues 122-171. The structure has been deposited to the PDB with accession number 5VR2.

Molecular dynamics (MD) simulations

Two structures were prepared for MD simulations: the dimer of $mLZ_{122-171}$ both with (WT) and $mLZ_{122-171}$ without (ΔC171) the C-terminal cysteines. Each structure was solvated in a water box of dimensions $105 \times 105 \times 105 \text{ Å}^3$ using VMD (Humphrey et al., 1996). The system was neutralized with 0.15 M NaCl, resulting in a total of ~110,000 atoms for each system. Molecular dynamics simulations were carried out using both the NAMD 2.12 (Phillips et al., 2005) and Amber16 (Case et al., 2016) programs with the CHARMM36 all-atom force field (Best et al., 2012a; Best et al., 2012b). The temperature was fixed using Langevin dynamics; the pressure was kept constant at 1 atm using the Langevin piston method (Feller et al., 1995). The equations of motion were integrated using the RESPA multiple time-step algorithm with a time step of 2 fs used for all bonded interactions, 2 fs for short-range non-bonded interactions, and 4 fs for long-range electrostatic interactions. Long-range electrostatic interactions were calculated using the particle-mesh Ewald method (Darden et al., 1993). Bonds involving hydrogen atoms were constrained to their equilibrium length employing the Rattle algorithm (Andersen, 1983).

Bioinformatics

Bioinformatics programs used to analyze sequences and structures are listed in Supplemental Methods.

References

Abu-Hassan, D.W., Acott, T.S., and Kelley, M.J. (2014). The Trabecular Meshwork: A Basic Review of Form and Function. J Ocul Biol *2*, 9-18.

Acott, T.S., Kelley, M.J., Keller, K.E., Vranka, J.A., Abu-Hassan, D.W., Li, X., Aga, M., and Bradley, J.M. (2014). Intraocular pressure homeostasis: maintaining balance in a high-pressure environment. J Ocul Pharmacol Ther *30*, 94-101.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr *68*, 352-367.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Andersen, H.C. (1983). Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Calculations. J Comput Phys *52*, 24-34.

Anholt, R.R. (2014). Olfactomedin proteins: central players in development and disease. Front Cell Dev Biol *2*, 6.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U S A *98*, 10037-10041.

Best, R.B., Mittal, J., Feig, M., and MacKerell, A.D., Jr. (2012a). Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. Biophys J *103*, 1045-1051.

Best, R.B., Zhu, X., Shim, J., Lopes, P.E., Mittal, J., Feig, M., and Mackerell, A.D., Jr. (2012b). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. J Chem Theory Comput *8*, 3257-3273.

Bibby, J., Keegan, R.M., Mayans, O., Winn, M.D., and Rigden, D.J. (2012). AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. Acta Crystallogr D Biol Crystallogr *68*, 1622-1631.

Burns, J.N., Orwig, S.D., Harris, J.L., Watkins, J.D., Vollrath, D., and Lieberman, R.L. (2010). Rescue of glaucoma-causing mutant myocilin thermal stability by chemical chaperones. ACS Chem. Biol. *5*, 477-487.

Burns, J.N., Turnage, K.C., Walker, C.A., and Lieberman, R.L. (2011). The stability of myocilin olfactomedin domain variants provides new insight into glaucoma as a protein misfolding disorder. Biochemistry *50*, 5824-5833.

Case, D.A., Betz, R.M., Botello-Smith, W., Cerutti, D.S., T.E. Cheatham, I., Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W.*, et al.* (2016). Amber 16.

Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J Chem Phys *98*, 10089-10092.

Delorenzi, M., and Speed, T. (2002). An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics *18*, 617-625.

Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., and Baker, N.A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res *35*, W522-W525.

Donegan, R.K., Hill, S.E., Freeman, D.M., Orwig, S.D., Turnage, K.C., and Lieberman, R.L. (2015). Structural basis for misfolding in glaucoma-associated myocilin. Hum. Mol. Genet. *24*, 2111-2124.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. *D66*, 486-501.

Faucher, M., Anctil, J.L., Rodrigue, M.A., Duchesne, A., Bergeron, D., Blondeau, P., Cote, G., Dubois, S., Bergeron, J., Arseneault, R.*, et al.* (2002). Founder TIGR/myocilin mutations for glaucoma in the Quebec population. Hum. Mol. Genet. *11*, 2077-2090.

Fautsch, M.P., Vrabel, A.M., Peterson, S.L., and Johnson, D.H. (2004). In vitro and in vivo characterization of disulfide bond use in myocilin complex formation. Mol Vis *10*, 417-425.

Feller, S.E., Zhang, Y.H., Pastor, R.W., and Brooks, B.R. (1995). Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. J Chem Phys *103*, 4613-4621.

Filla, M.S., Liu, X., Nguyen, T.D., Polansky, J.R., Brandt, C.R., Kaufman, P.L., and Peters, D.M. (2002). In vitro localization of TIGR/MYOC in trabecular meshwork extracellular matrix and binding to fibronectin. Invest Ophthalmol Vis Sci *43*, 151-161.

Fischer, H., Neto, M.D., Napolitano, H.B., Polikarpov, I., and Craievich, A.F. (2010). Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J Appl Cryst *43*, 101-109.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Totowa, NJ: Humana Press), pp. 571-607.

Gobeil, S., Letartre, L., and Raymond, V. (2006). Functional analysis of the glaucoma-causing TIGR/myocilin protein: integrity of amino-terminal coiled-coil regions and olfactomedin homology domain is essential for extracellular adhesion and secretion. Exp. Eye Res. *82*, 1017-1029.

Goldwich, A., Scholz, M., and Tamm, E.R. (2009). Myocilin promotes substrate adhesion, spreading and formation of focal contacts in podocytes and mesangial cells. Histochem Cell Biol *131*, 167-180.

Grigoryan, G., and Keating, A.E. (2008). Structural specificity in coiled-coil interactions. Curr Opin Struct Biol *18*, 477-483.

Grigoryan, G., Reinke, A.W., and Keating, A.E. (2009). Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature *458*, 859-864.

Hewitt, A.W., Craig, J.E., and Mackey, D.A. (2006). Complex genetics of complex traits: the case of primary open-angle glaucoma. Clin Exp Ophthal *34*, 472-484.

Hewitt, A.W., Mackey, D.A., and Craig, J.E. (2007). Myocilin allele-specific glaucoma phenotype database. Hum. Mutat.

Hill, S.E., Donegan, R.K., and Lieberman, R.L. (2014). The glaucoma-associated olfactomedin domain of myocilin forms polymorphic fibrils that are constrained by partial unfolding and peptide sequence. J. Mol. Biol. *426*, 921-935.

Hill, S.E., Donegan, R.K., Nguyen, E., Desai, T.M., and Lieberman, R.L. (2015). Molecular details of olfactomedin domains provide pathway to structure-function studies. PloS one *10*, e0130888.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38, 27-38.

Hynes, R.O., and Naba, A. (2012). Overview of the matrisome--an inventory of extracellular matrix constituents and functions. Cold Spring Harb Perspect Biol *4*, a004903.

Kammerer, R.A. (1997). Alpha-helical coiled-coil oligomerization domains in extracellular proteins. Matrix Biol *15*, 555-565; discussion 567-558.

Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. *12*, 1865-1871.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography *36*, 1277-1282.

Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography *34*, 33-41.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. *D60*, 2256-2268.

Krissinel, E., and Henrick, K. (2005). Multiple alignment of protein structures in three dimensions. Lect Notes Comput Sc *3695*, 67-78.

Liu, Y., and Vollrath, D. (2004). Reversal of mutant myocilin non-secretion and cell killing: implications for glaucoma. Hum. Mol. Genet. *13*, 1193-1204.

Lupas, A.N., and Bassler, J. (2017). Coiled Coils - A Model System for the 21st Century. Trends Biochem Sci *42*, 130-140.

Mason, J.M., and Arndt, K.M. (2004). Coiled coil domains: stability, specificity, and biological implications. Chembiochem *5*, 170-176.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658-674.

Orwig, S.D., and Lieberman, R.L. (2011). Biophysical characterization of the olfactomedin domain of myocilin, an extracellular matrix protein implicated in inherited forms of glaucoma. PloS one *6*, e16347.

Orwig, S.D., Perry, C.W., Kim, L.Y., Turnage, K.C., Zhang, R., Vollrath, D., Schmidt-Krey, I., and Lieberman, R.L. (2012). Amyloid fibril formation by the glaucoma-associated olfactomedin domain of myocilin. J. Mol. Biol. *421*, 242-255.

26

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol *276*, 307-326.

Parry, D.A., Fraser, R.D., and Squire, J.M. (2008). Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. J Struct Biol *163*, 258-269.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr *45*, 342-350.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J Comput Chem *26*, 1781-1802.

Polansky, J.R., Fauss, D.J., Chen, P., Chen, H., Lutjen-Drecoll, E., Johnson, D., Kurtz, R.M., Ma, Z.D., Bloom, E., and Nguyen, T.D. (1997). Cellular pharmacology and molecular biology of the trabecular meshwork inducible glucocorticoid response gene product. Ophthalmologica *211*, 126-139.

Pronker, M.F., Bos, T.G., Sharp, T.H., Thies-Weesie, D.M., and Janssen, B.J. (2015). Olfactomedin-1 Has a V-shaped Disulfide-linked Tetrameric Structure. J Biol Chem *290*, 15092-15101.

Quigley, H.A., and Broman, A.T. (2006). The number of people with glaucoma worldwide in 2010 and 2020. Br J Ophthalmol *90*, 262-267.

Richer, S.P., and Rose, R.C. (1998). Water soluble antioxidants in mammalian aqueous humor: interaction with UV B and hydrogen peroxide. Vision Res *38*, 2881-2888.

Rozsa, F.W., Shimizu, S., Lichter, P.R., Johnson, A.T., Othman, M.I., Scott, K., Downs, C.A., Nguyen, T.D., Polansky, J., and Richards, J.E. (1998). GLC1A mutations point to regions of potential functional importance on the TIGR/MYOC protein. Mol Vis *4*, 20.

Russell, P., Tamm, E.R., Grehn, F.J., Picht, G., and Johnson, M. (2001). The presence and properties of myocilin in the aqueous humor. Invest Ophthalmol Vis Sci *42*, 983-986.

Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol *1079*, 105-116.

Souzeau, E., Tram, K.H., Witney, M., Ruddle, J.B., Graham, S.L., Healey, P.R., Goldberg, I., Mackey, D.A., Hewitt, A.W., Burdon, K.P.*, et al.* (2017). Myocilin Predictive Genetic Testing for Primary Open-Angle Glaucoma Leads to Early Identification of At-Risk Individuals. Ophthalmology *124*, 303-309.

Stone, E.M., Fingert, J.H., Alward, W.L., Nguyen, T.D., Polansky, J.R., Sunden, S.L., Nishimura, D., Clark, A.F., Nystuen, A., Nichols, B.E.*, et al.* (1997). Identification of a gene that causes primary open angle glaucoma. Science *275*, 668-670.

Stothert, A.R., Fontaine, S.N., Sabbagh, J.J., and Dickey, C.A. (2016). Targeting the ER-autophagy system in the trabecular meshwork to treat glaucoma. Exp Eye Res *144*, 38-45.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879-2886.

Taylor, K.C., Buvoli, M., Korkmaz, E.N., Buvoli, A., Zheng, Y., Heinze, N.T., Cui, Q., Leinwand, L.A., and Rayment, I. (2015). Skip residues modulate the structural properties of the myosin rod and guide thick filament assembly. Proc Natl Acad Sci U S A *112*, E3806-3815.

Testa, O.D., Moutevelis, E., and Woolfson, D.N. (2009). CC+: a relational database of coiled-coil structures. Nucleic Acids Res *37*, D315-322.

Tomarev, S.I., and Nakaya, N. (2009). Olfactomedin domain-containing proteins: possible mechanisms of action and functions in normal development and pathology. Mol. Neurobiol. *40*, 122-138.

Ueda, J., Wentz-Hunter, K., and Yue, B.Y. (2002). Distribution of myocilin and extracellular matrix components in the juxtacanalicular tissue of human eyes. Invest Ophthalmol Vis Sci *43*, 1068-1076.

Vincent, T.L., Green, P.J., and Woolfson, D.N. (2013). LOGICOIL--multi-state prediction of coiled-coil oligomeric state. Bioinformatics *29*, 69-76.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography *36*, 860-864.

Vollrath, D., and Liu, Y. (2006). Temperature sensitive secretion of mutant myocilins. Exp. Eye Res. *82*, 1030-1036.

Wentz-Hunter, K., Kubota, R., Shen, X., and Yue, B.Y. (2004). Extracellular myocilin affects activity of human trabecular meshwork cells. J Cell Physiol *200*, 45-52.

Wheeler, T.J., Clements, J., and Finn, R.D. (2014). Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics *15*, 7.

Wolf, E., Kim, P.S., and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci *6*, 1179-1189.

Woolfson, D.N. (2005). The design of coiled-coil structures and assemblies. Adv Protein Chem *70*, 79-112.

Zeng, L.C., Han, Z.G., and Ma, W.J. (2005). Elucidation of subfamily segregation and intramolecular coevolution of the olfactomedin-like proteins by comprehensive phylogenetic analysis and gene expression pattern assessment. FEBS Lett. *579*, 5443-5453.

Zhou, Z., and Vollrath, D. (1999). A cellular assay distinguishes normal and mutant TIGR/myocilin protein. Hum. Mol. Genet. *8*, 2221-2228.

28

Author Contributions

S.E.H., R.L.L., J.C.G. designed experiments, interpreted results, wrote, and edited the manuscript. S.E.H., E.N., R.K.D., A.H. conducted experiments.

Supplemental Materials and Methods

Molecular Cloning and Protein Expression

Constructs of N-terminal portions of human myocilin were obtained by subcloning from an *E. coli* codon-optimized construct of full-length human myocilin (DNA 2.0). Primers for NTD_{33-227}, CC_{60-185}, CC_{69-184}, CC_{33-111}, and $CC_{112-184}$, were designed to incorporate a C-terminal TAA stop codon as listed in Supplemental Table S2. Amplified DNA was purified via a Promega PCR purification procedure, treated with T4 DNA polymerase, and annealed into the pET-30 Xa/LIC vector following the manufacturer's directions. Cys-to-Ser variants of NTD_{33-227} ($C_{61}S/C_{185}S$, $C_{47}S/C_{185}S$, $C_{47}S/C_{61}S$, $C_{47}S/C_{61}S/C_{185}S$), CC_{69-185}, $CC_{112-185}$, and NTD_{33-227} disease-causing variants were generated by site directed mutagenesis (QuikChange Stratagene mutagenesis kit, primers listed in Supplemental Table S2). Plasmids were propagated by *E. coli* NovaBlue GigaSingles and the correct in-frame DNA sequence was confirmed (MWG Operon). The plasmid for Mouse LZ_{55-171}, designed to be identical to the pET-30 Xa/LIC vector with a Factor Xa protease cleavage site was purchased from Genscript. After proteins were expressed in *E. coli* BL21(DE3)plysS cells, they were purified as detailed below.

Protein Purification

NTD_{33-227} (WT)

Cell pellets (20 g) were resuspended using a serological pipette into 40 mL Ni^{2+}-affinity purification wash buffer (50 mM Hepes pH 7.5, 200 mM NaCl, 40 mM imidazole, 10% glycerol) supplemented with 2 EDTA-free Roche Protease Inhibitor Cocktail tablets. Cells were lysed by 2-3 passages through a French press (12,000 psi). To increase protein yield, 1% Trion X-100 was added and lysed cells were rocked gently in a cold room for 1 hour. Cellular debris was removed by ultracentrifugation at $130,000 \times$ g, 4 °C for 1 hour and the supernatant loaded onto a 1 mL Ni^{2+}-affinity purification column (GE Healthcare) using an AKTA purification system (GE Healthcare). After 120 column-volumes of wash buffer was loaded to remove contaminating proteins, NTD_{33-227} was eluted with a gradient from 40-500mM imidazole using a mixture of wash buffer and elution buffer composed of 50mM Hepes pH 7.5, 200mM NaCl, 500 mM imidazole, 10% glycerol. Fractions containing NTD_{33-227} were concentrated using Amicon 100K MWCO filters (EMD Millipore) and loaded onto a Superdex 200 10/300 GL (GE Healthcare) size exclusion column (SEC) equilibrated with gel filtration buffer (50 mM Hepes pH 7.5, 200 mM NaCl, 10% glycerol). The fractions containing NTD_{33-227} were identified by SDS-PAGE analysis (15% polyacrylamide) with Coomassie staining and found to have a contaminating band of DnaK, a result confirmed by trypsin digest mass spectrometry (Supplemental Fig. S12a, Systems Mass Spectrometry Core Facility, Georgia Tech). Taking advantage of the finding that NTD_{33-227} can unfold reversibly (Figure 2a), an on-column unfolding refolding protocol (see below) was used to remove the contaminating DnaK for selected preparations (e.g. Figure 2 and NTD_{33-227} Cys-to-Ser variants below). Removal of DnaK did not alter the elution profile of NTD_{33-227} from Superdex 200 (Supplemental Fig. S12b,c). Protein concentration was estimated by absorbance at 280 nm using a molecular extinction coefficient ($8,605$ $M^{-1}cm^{-1}$) and a molecular weight (27,014.8 g/mol) calculated by ExPASy ProtParam (Gasteiger et al., 2005)

NTD_{33-227} Cys-to-Ser Variants (C47S/C61S, C61S/C185S, C47S/C185S, C47S/C61S/C185S)

Cell pellets (20 g) were lysed using the Triton X-100 step and purified by Ni^{2+}-affinity column chromatography as outlined above. Fractions containing NTD_{33-227} were concentrated to 1 mL using Amicon 100K MWCO filters (EMD Millipore) followed by dilution into 50mL urea unfolding buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 8 M Urea). Unfolded NTD_{33-227} was

30

reloaded onto the Ni^{2+}-affinity column, washed with 20 column volumes of urea unfolding buffer to remove contaminating DnaK, and then refolded on the column by a gradient into refolding buffer (50mM Hepes pH 7.5, 200mM NaCl, 10% glycerol). Folded NTD_{33-227} was eluted by a gradient from 0-500 mM imidazole also containing 50mM Hepes pH 7.5, 200mM NaCl, 10% glycerol, fractions containing NTD_{33-227} were concentrated using Amicon 30K MWCO filters (EMD Millipore), and samples were then purified by Superdex 200 10/300 GL SEC as described above. Purified proteins were assessed and quantified as for WT NTD_{33-227}.

Annotated NTD_{33-227} Variants (V53A, R82C, L95P, R126W, R128W, L215Q)

Cell pellets (10 g) were resuspended using a serological pipette into 20 mL Ni^{2+}-affinity purification wash buffer supplemented with 1 EDTA-free Roche Protease Inhibitor Cocktail tablet. Cells were lysed, debris removed, and protein purified as described for NTD_{33-227}. Fractions containing NTD_{33-227} were assessed by SDS-PAGE analysis, concentrated with a 100K MWCO filter, and re-loaded onto a Superdex 200 10/300 GL equilibrated with gel filtration buffer to obtain the final presented chromatogram.

CC_{60-185}

Cell pellets (20 g) were resuspended, lysed, and purified by Ni^{2+}-affinity and SEC as for NTD_{33-227}, except that the initial cell resuspension buffer was further supplemented with 1 mL of 10 mg/mL DNase, 200 µL of 1 M $CaCl_2$, 200 µL of 1 M $MgCl_2$ and the Triton X-100 step was not used. To remove tags, purified CC_{60-185} was incubated with Factor Xa (Roche, ~50:1 mass ratio) plus ~5-10 mM $CaCl_2$ for 96 hours at 4 °C. Digested CC_{60-185} was diluted ~20 fold into heparin wash buffer (20 mM Hepes, pH 7.5, 10% glycerol), loaded onto a 5 mL HiTrap Heparin HP column (GE Healthcare), and eluted by a gradient from 0-1 M NaCl with buffers also containing 20 mM Hepes pH 7.5, 10% glycerol, which also removed Factor Xa and DnaK contaminants. Heparin elution fractions were assessed by SDS-PAGE analysis as above, and those containing CC_{60-185} were concentrated with Amicon 10K MWCO filters and loaded onto a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare) equilibrated with gel filtration buffer. Fractions corresponding to a molecular weight of ~60 kDa based on a standard calibration curve of the SEC column, containing pure CC_{60-185} as assessed by SDS-PAGE analysis with Coomassie staining, were concentrated in an Amicon 10K MWCO filter. Protein concentration was measured by absorbance at 280 nm using a molecular extinction coefficient $(1,490 M^{-1}cm^{-1})$ and a molecular weight (14,643 g/mol) calculated by Expasy ProtParam.

CC_{69-185}

Cell pellets (20 g) were resuspended and lysed as for CC_{60-185} except that the 1% Triton X-100 procedure was performed as for WT NTD_{33-227}. Purification by Ni^{2+} affinity and SEC proceeded as described for WT NTD_{33-227}. After assessment by SDS-PAGE analysis, fractions containing CC_{69-185} were digested and purified by heparin affinity and Superdex 75 following procedures for CC_{60-185} Fractions corresponding to a molecular weight of ~50 kDa based on an SEC standard calibration curve containing pure CC_{69-185} as assessed by SDS-PAGE analysis with Coomassie staining were concentrated in an Amicon 10K MWCO filter. Protein concentration was measured by absorbance at 280 nm using a molecular extinction coefficient $(1,490 M^{-1}cm^{-1})$ and a molecular weight (13,681 g/mol) calculated by Expasy ProtParam.

CC_{33-111}

Cell pellets (20 g) were resuspended, lysed, and purified as described for CC_{60-185} except Amicon 10K MWCO filters were used to concentrate the protein. Following assessment by SDS-PAGE analysis, fractions containing pure CC_{33-111} were incubated with Factor Xa (Roche, ~50:1 mass ratio) plus ~5-10 mM $CaCl_2$ for 96 hours at 4 °C. Digested CC_{33-111} was concentrated with

two Amicon 10K MWCO filters and loaded onto a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare) equilibrated with gel filtration buffer. Fractions corresponding to a molecular weight of 40 kDa based on a standard calibration curve containing pure CC_{33-111} as assessed by SDS-PAGE analysis with Coomassie staining were concentrated in an Amicon 10K MWCO filter. Protein concentration was measured by absorbance at 280 nm using a molecular extinction coefficient ($1,490$ $M^{-1}cm^{-1}$) and a molecular weight ($8,801$ g/mol) calculated by ExPasy ProtParam.

$CC_{112-185}$

Cell pellets (20 g) were resuspended, lysed, and purified as described for CC_{33-111}. Purified $CC_{112-185}$ was digested and further purified using the procedure described for CC_{60-185}. Fractions corresponding to a molecular weight of 23 kDa based on a calibration curve for SEC containing pure $CC_{112-185}$ as assessed by SDS-PAGE analysis were concentrated in an Amicon 10K MWCO filter. Protein concentration was measured by absorbance at 280 nm using a molecular extinction coefficient ($1,490$ $M^{-1}cm^{-1}$) and a molecular weight ($8,858$ g/mol) calculated by Expasy ProtParam.

Mouse LZ_{55-171} and LZ_{93-171}

Cell pellets (20 g) were resuspended, lysed, and purified as described for CC_{60-85} except 100K MWCO filtration devices were used to concentrate the protein. After assessment by SDS-PAGE, fractions containing pure mouse LZ_{55-171} were incubated with Factor Xa (Roche, ~50:1 mass ratio) plus ~5-10 mM $CaCl_2$ for 72 hours at 4 °C. Digested mouse LZ_{55-171} was diluted ~30 fold into heparin wash buffer (20 mM Hepes, pH 7.5, 10% glycerol), loaded onto a 5 mL HiTrap Heparin HP column (GE Healthcare). Unlike the human constructs, mouse LZ_{55-171} did not bind the heparin column, so flow through fractions were concentrated with Amicon 10K MWCO filters and loaded onto a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare) equilibrated with gel filtration buffer. The majority of protein was found in fractions corresponding to a molecular weight of ~30 kDa based on a calibration curve; however, after assessment by SDS-PAGE the dominant band on the gel was below the predicted $13,500$ g/mol molecular weight for cleaved mouse LZ_{55-171}. Mass spectrometry (Supplemental Fig. S12d) confirmed mouse LZ_{55-171} was cleaved by Factor Xa after a second Gly-Arg amino acid within the amino acid sequence, yielding mouse LZ_{93-171} used in subsequent experiments. Fractions containing mouse LZ_{93-171} were concentrated in an Amicon 10K MWCO filter. Protein concentration was measured by absorbance at 280 nm using a molecular extinction coefficient ($1,490$ $M^{-1}cm^{-1}$) and a molecular weight ($9,144$ g/mol) calculated by ExPASy ProtParam for mouse LZ_{93-117}.

SEC-SAXS

SEC–SAXS experiments were performed at BioCAT (beamline 18ID, Advanced Photon Source at Argonne National Labs). The camera included a focused 12 KeV (1.03 Å) X-ray beam, a 1.5 mm quartz capillary sample cell, a sample to detector distance of ~3.5 m, and a Pilatus 3 1M detector (Dectris). The Q-range sampled was ~0.0042–0.4 Å−1. A size-exclusion chromatography setup in line with the SAXS camera was used to separate the sample from potential aggregates, breakdown products and other contaminants using a Superdex 75 10/300 GL (GE Healthcare Lifesciences) column. The elution trajectory after the UV monitor was redirected to the SAXS sample flow-cell. Exposures (1 s) were collected every 2 s during the gel-filtration chromatography run. Data reduction to generate the I(q) vs. q curves for all constructs except CC_{60-185} was performed with the BioCAT beam line specific pipeline, which uses the ATSAS program suite(Petoukhov et al., 2012). For CC_{60-185} multiple I(q) vs. q curves

32

were manually generated and inspected to optimize frame choice, eliminating frames for which signal was being derived from aggregated protein. Exposures flanking the elution peak were averaged and used as the buffer curve for each run, and the exposures corresponding to the UV peak on the chromatogram were treated as protein plus buffer curves. Data were corrected for background scattering by subtracting the buffer curve from protein plus buffer curves. The radius of gyration (R_g) was calculated using Guinier approximation and a P(r) curve, both of which were done using PRIMUS (Konarev et al., 2003). D_{max} was also calculated using the P(r) curve and the output was used to calculate molecular envelopes using DAMMIN (Svergun, 1999) The envelopes were then averaged using DAMAVER and DAMFILT (Volkov and Svergun, 2003). For $CC_{69-185,}$ $CC_{112-185}$, and mLZ_{93-171} results of imposing two-fold symmetry (P2) was compared to results with no imposed symmetry (P1). For all constructs and symmetry constraints presented, at least 9 out of the 10 envelopes were accepted by DAMAVER on the basis of the basis of the Normalized Spatial Discrepancy (NSD). SUPCOMB(Kozin and Svergun, 2001) was used to overlay envelopes of differing symmetry, compare CC_{60-185} with CC_{69-185}, and compare $CC_{112-185}$ compared to mLZ_{93-171} (Supplemental Figure S5 and Supplemental Figure S8e,f). Overlay of $CC_{112-185}$ and CC_{33-111} (and Figure 3e) was performed manually within the PyMOL Molecular Graphic System v. 1.7.6.0. Final *ab initio* models were rendered using Visual Molecular Dynamics (Humphrey et al., 1996). Molecular masses were calculated using SAXS MolW2 (Fischer et al., 2010).

Bioinformatics

Multiple sequence alignments were generated in CLUSTAL Omega (Sievers and Higgins, 2014). Myocilin sequences from multiple organisms were identified using the human myocilin OLF domain sequence as a query within BLAST (Altschul et al., 1990). The list was manually culled to remove partial sequences and non-myocilin proteins, which were identified either by annotation or by non-myocilin sequence stretches recognized by us through other related structures of OLF domain proteins (Hill et al., 2015). Then the OLF domains were deleted beginning with the first structural residue, Cys (Donegan et al., 2015) leaving 75 N-terminal sequences intact. After multiple sequence alignments, the consensus sequence was visualized using Skylign (Wheeler et al., 2014). Coiled coil prediction servers Marcoil (Delorenzi and Speed, 2002) and Logicoil (Vincent et al., 2013) were utilized to generate models for the coiled-coil domain of myocilin; a consensus cutoff of 80% probability was used to label the limits of the domains presented in Figure 1. Logicoil was used to obtain the coil register for Figure 4a and DrawCoil 1.0 (Grigoryan and Keating, 2008) was used to generate the image presented in Figure 4d. A structure similarity search was conducted with PDBeFOLD (Krissinel and Henrick, 2005) and sequence similarity searches limited to CCs were conducted using CC+ Builder (Testa et al., 2009). Structures were aligned using the secondary structure matching algorithm (Krissinel and Henrick, 2004) to obtain r.m.s.d. values in Coot. Electrostatic surfaces were generated using default options within the PDB2PQR web server (Dolinsky et al., 2007) and the APBS plugin (Baker et al., 2001) within PyMOL.

Supplemental Figures

(a) Overall sequence alignment

```
olfm4      MRPGLSFLLALLFFLGQAAGDLGDVGPPIPSPGFSSFPGVDSSS----SFSS---SSRSG
olfm1      ----------------------------MSVPLLKIGVVLSTMAMITNWMSQTLPSLVG
olfml2     ---------------------MAAAALPPRPLLLLLPLV-------LLLS------G
myoc       ----------MRFFCAR-CC---SFGPEMPAVQLLLLACL-------VWDV-------G
olfml3     -----------------------MGPSTPLL---ILFLL--------SWSG-------P
                                                      :   :

olfm4      SSSS-RSLGSGGSV-----------SQLFSNFTGSVDDRGTCQCSVSLPDTTFPVDR-VE
olfm1      LNTTKLSAAGGGTLDRSTGVLPTNPEESWQVYSSAQDSEGRCICTVVAPQQTMCSRDART
olfml2     RPTRADS-----------KVF----GDLDQVRMTSEGSDCRCKCIMRPLSKDACSRVRSG
myoc       ART-AQL-----------------------RK-ANDQSGRCQYTFSVAS----------
olfml3     LQG-QQH-----------------------HL-VEYME----RRLAALE----------
                                                         .    .

olfm4      RLE---F-TAHVL---------------------------SQKFEKELSKVREYVQLIS
olfm1      KQ-----------------------------------L---RQLLEKVQNMSQSIE
olfml2     RARVEDFYTVETVSSGTDCRCSCTAPPSSLNPCENEWKMEKLKKQAPELLKLQSMVDLLE
myoc       ----------------PNESSCPEQSQAMSVIHNLQRD--SSTQRLDLEATKARLSSLE
olfml3     ----------------ERLAQCQDQSSR--------------HAAELRDFKNK-----
                                                          *    :

olfm4      VYEK--KL--------------------------LNLTVRIDIMEKDTISYTELDFELI
olfm1      VL--------------DRRTQRDLQYVEKMENQMKGLESKFKQVEESHKQHLARQFKAI
olfml2     GTLYSMDLMKVH------------AYV-------HKVASQMNTLEESIKANLSRENEVV
myoc       SLLHQLTLDQAARPQETQEGLQRELGTLRRER---DQLETQTRELET-AYSNLLRDKSVL
olfml3     ----------------------------------------------MLPLLEVA
                                                              .

olfm4      KVEVKEMEKLVIQLKESFGGSSEIVDQLEVEIRNMTLLVEKLETLDKNNVLAIRREIVAL
olfm1      KAKMDELRPLIPVLE-EYKADAKLVLQ------------------------FKEEVQNL
olfml2     KDSV---RHLSEQLR-HYENHSAIMLG------------------------IKKELSRL
myoc       EEEK---KRLRQE----NENLARRLES------------------------SSQEVARL
olfml3     EKER---EALRTE----ADTISGRVDR------------------------LEREVDYL
           : .    . *              :  :                          .*:  *

olfm4      KTKLKECEASKDQNTPV-----VHPPPTPGS------------------------------
olfm1      TSVLNE---------------------LQEEIGAYDYDEL--QSRVSNLEERLRACM
olfml2     GLQLLQKDAAAAPATPATGTGSKAQDTARGKGKDISKYGSVQKSFADRG--LPKPPKEKL
myoc       -----RRG----QCPQTRDTARAVPPGSRE----VSTWNLDTLAFQELKSELTEVPASRI
olfml3     -----ETQNPALPCVEFDEKV-TGGPGTKGKGRRNEKYDMVTD----------------
                    .

olfm4      ------------------------------------------------------------
olfm1      ---QKLA------------------------------------------------------
olfml2     LQVEKLRKESGKGSFLQPTAKPRALAQQQAVIRGFTYYKAGKQEVTEAVADNTLQGTSWL
myoc       L------KESPSG--------------------YLRSGEGDTG--------------
olfml3     ------------------------------------------------------------
```

34

(b) Myocilin/olfactomedin-1 (22% identity)

```
myocilin       -------------MRFFCARCCSFGPEMPAVQLL--------------------LLAC
olfm1          MSVPLLKIGVVLSTMAMITNWM---SQTLPSLVGLNTTKLSAAGGGTLDRSTGVLPTNPE
                            * ::        .  .:*::  *

myocilin       LVWDVGARTAQLRKANDQSGRCQYTFSVASPNESSCPEQSQAMSVIHNL---QRDSST-Q
olfm1          ESWQV------YSSAQDSEGRCICTV--VAPQQTMCSRDARTKQLRQLLEKVQNMSQSIE
                *:*          .*:*..***   *.  .:*::: * .:::: .: : *    *. *.: :

myocilin       RLD---------LEATKARLSSLESLLHQLTLDQAARPQETQEGLQRELGTLRRERDQLE
olfm1          VLDRRTQRDLQYVEKMENQMKGLESKFKQVEES-------HKQHLARQFKAIKAKMDELR
                **           :*   : ::..*** ::*:   .        :: * *:: ::: : *:*.

myocilin       TQTRELETAYSNLLRDKSVLEEEKKRLRQENENLARRLESSSQEVARLRRGQCPQTRDTA
olfm1          PLIPVLEE-Y---KADAKLV----LQFKEEVQNLTSVLNEL-------------------
                    **   *      * .::     :::::* :**:  *:.

myocilin       RAVPPGSREVSTWNLDTLAFQELKSELTEVPASRILKESPSGYLRSGEGDTG
olfm1          ------QEEIGAYDY-----DELQSRVSNLEER--LR----ACMQK----LA
                    ..*:.:::       :**:*.:::::    *:     . ::.     .
```

(c) Myocilin/olfactomedin-like-2 (22% identity)

```
myocilin       MRFFCARCCSFGPEMPAVQLLLLLACLVWDVGARTA----------QLRKA-NDQSGRCQY
olfml2         ---------MAAAALPPRPLLLLPLVLLLSGRPTRADSKVFGDLDQVRMTSEGSDCRCKC
                      .  :*    ****  ::    *  *        *:* : :... **:

myocilin       TFS-----------------------VASPNE--SSCPEQSQAMSVIH------NLQR
olfml2         IMRPLSKDACSRVRSGRARVEDFYTVETVSSGTDCRCSCTAPPSSLNPCENEWKMEKLKK
                :                        *:* .: .**    .::.  .       :*::

myocilin       DSSTQRLDLEAT----KAR-----LSSLESLLHQLTLDQAARPQETQEGLQRELGTLRRE
olfml2         Q-APELLKLQSMVDLLEGTLYSMDLMKVHAYVHKVASQMNTLEESIKANLSRENEVVKDS
                : : : *.*::       :.      * .:.: :*::: :  :  :. : .*.**  .:: .

myocilin       RDQLETQTRELETAYSNLLRDKSVLEEEKKRLRQENENLARRLESSSQEVARLRRGQCPQ
olfml2         VRHLSEQLRHYENHSAIMLG-------IKKELSRLGLQLLQKDAA--AAPATPATGTGSK
                 :*. * *. *.  : :*       **.* : . :* ::  :     *     *    :

myocilin       TRDTARAVPPGSREVSTWNLDTLAFQELKSELTEVPASRIL------------------
olfml2         AQDTARGK---GKDISKYGSVQKSFAD--RGLPKPPKEKLLQVEKLRKESGKGSFLQPTA
                ::****.    .:::*.:.   :* :    * : * .::*

myocilin       ----------------------------------------------------------
olfml2         KPRALAQQQAVIRGFTYYKAGKQEVTEAVADNTLQGTSWLEQLPPKVEGRSNSAEPNSAE

myocilin       ----------------------------------------KESPSGYLRSGEGDTG-
olfml2         QDEAEPRSSERVDLASGTPTSIPATTTTATTTPTPTTSLLPTEPPSGPEVSSQGREAS
                                                        .* ***   *.:*   .
```

35

(d) Myocilin/olfactomedin-4 (22% identity)

```
myocilin      ----------MRFFCAR----CCSFGPEMPAVQLLLLACLV----WD-------------
olfm4         MRPGLSFLLALLFFLGQAAGDLGDVGPPIPSPGFSSFPGVDSSSSFSSSSRSGSSSSRSL
                 : ** .:        ..** :*:  :  :  :      :.

myocilin      -VGAR----TAQLRKANDQSGRCQYTFSVASPNESSCPEQSQAMSVI-HN----LQRDS-
olfm4         GSGGSVSQLFSNFTGSVDDRGTCQCSVSLPDTTFP--VDRVERLEFTAHVLSQKFEKELS
                *.        :::   : *: * ** :.*: . .       :: : :.. *     ::::

myocilin      -------------------------------STQRLDLEATKARLSSLESLLHQLTLDQA
olfm4         KVREYVQLISVYEKKLLNLTVRIDIMEKDTISYTELDFELIKVEVKEMEKLVIQ------
                                               *  .**:*  *..:..:*.*: *

myocilin      ARPQETQEGLQRELGTLRRERDQLETQTRELETAYSNLLRDKSVLEEEKKRLRQENENLA
olfm4         ---------LKESFGGSSEIVDQLEVEIRNM----------TLLVEKLETLDKN-NVLA
                       *:..:*    .  ****.: *::          ::* *: : * :: : **

myocilin      RRLESSSQEVARLRRGQCPQTRDTA--RAVPPGSREVSTWNLDTLAFQELKSELTEVPAS
olfm4         IRREIV---ALKTKLKECEASKDQNTPVVHPPPTPGS----------------------
                * *      . : :  :*  ::*      . ** :

myocilin      RILKESPSGYLRSGEGDTG
olfm4         -------------------
```

(e) Myociin/olfactomedin-like-3 (21% identity)

```
myocilin      MRFFCARCCSFGPEMPAVQLLLLACLVWDVGARTAQLRKANDQSGRCQYTFSVASPNESS
olfml3        ----------MGPSTPLL---ILFLLSWSGPLQGQQHHLVEYME----RRLAALEERLAQ
                  :**. * :   :*  * * *.   :  * : .:  .      ::. . . :.

myocilin      CPEQSQAMSVIHNLQRDSSTQRLDLEATKARLSSLESLLHQLTLDQAARPQETQEGLQRE
olfml3        CQDQSSR------------HAAELRDFKNK----------------------------
              * :**.               :  :*.  * :

myocilin      LGTLRRERDQLETQTRELETAYSNLLRDKSVLEEEKKRLRQENENLARRLESSSQEVARL
olfml3        -----------------------MLPLLEVAEKEREALRTEADTISGRVDRLEREVDYL
                                     :*    .* *:*:: ** * :.:: *::  .:** *

myocilin      RRG----QCPQTRDTARAVPPGSRE----VSTWNLDTLAFQELKSELTEVPASRILKESP
olfml3        ETQNPALPCVEFDEKV-TGGPGTKGKGRRNEKYDMVTD--------------------
              .        * : :.. : **:: :      ..::: *

myocilin      SGYLRSGEGDTG
olfml3        ------------
```

Supplemental Figure S1. Sequence alignments of CC-containing regions of olfactomedin-containing proteins. Related to Fig. 1. (a) Overall sequence alignment of CC-containing regions of human proteins found within different olfactomedin subfamilies. (b) Pairwise sequence alignment of myocilin and olfactomedin-1. (c) Pairwise sequence alignment of myocilin and olfactomedin-like-2. (d) Pairwise sequence alignment of myocilin and olfactomedin-4. (e) Pairwise sequence alignment of myocilin and olfactomedin-like-3.
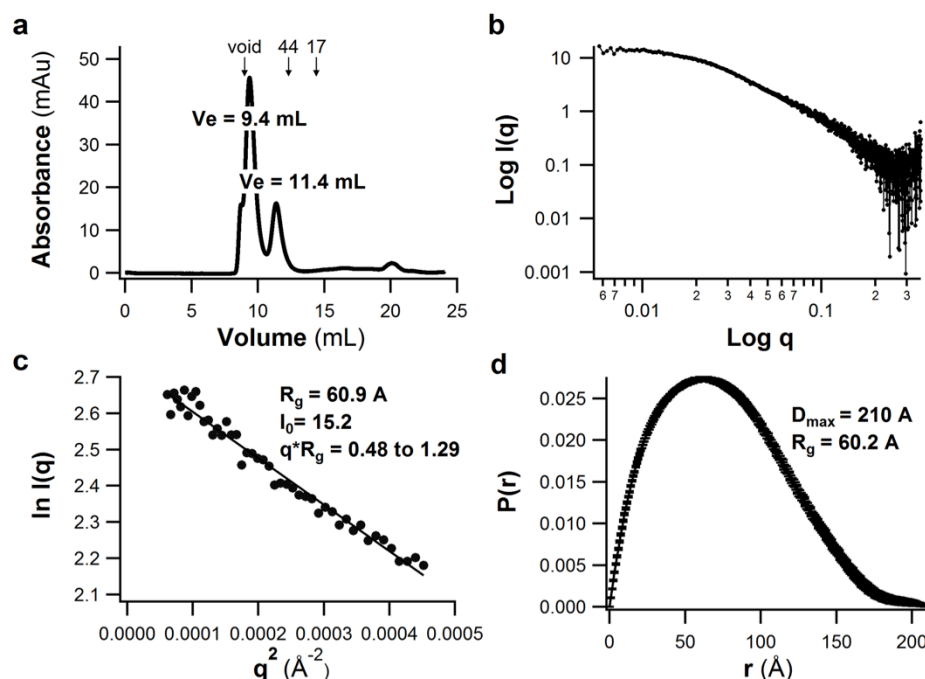
Supplemental Figure S2. SEC-SAXS analysis of CC_{60-185}. Related to Fig. 3. (a) Superdex-75 GL trace of CC_{60-185} obtained at the beamline reveals two species, one in the void volume and another with a molecular weight of 70 kDa based on a standard calibration curve. Given the calculated mass of 14.6 kDa/monomer, the second species is a 4 or 5-mer. Further SAXS analysis focused exclusively on this second species. (b) Scattering intensity profile. I(q): scattered intensity, q:scattering vector. (c) Guinier plot with calculated radius of gyration (Rg). (d) Pair wise distribution plot with calculated maximum particle size (Dmax) and radius of gyration (Rg).
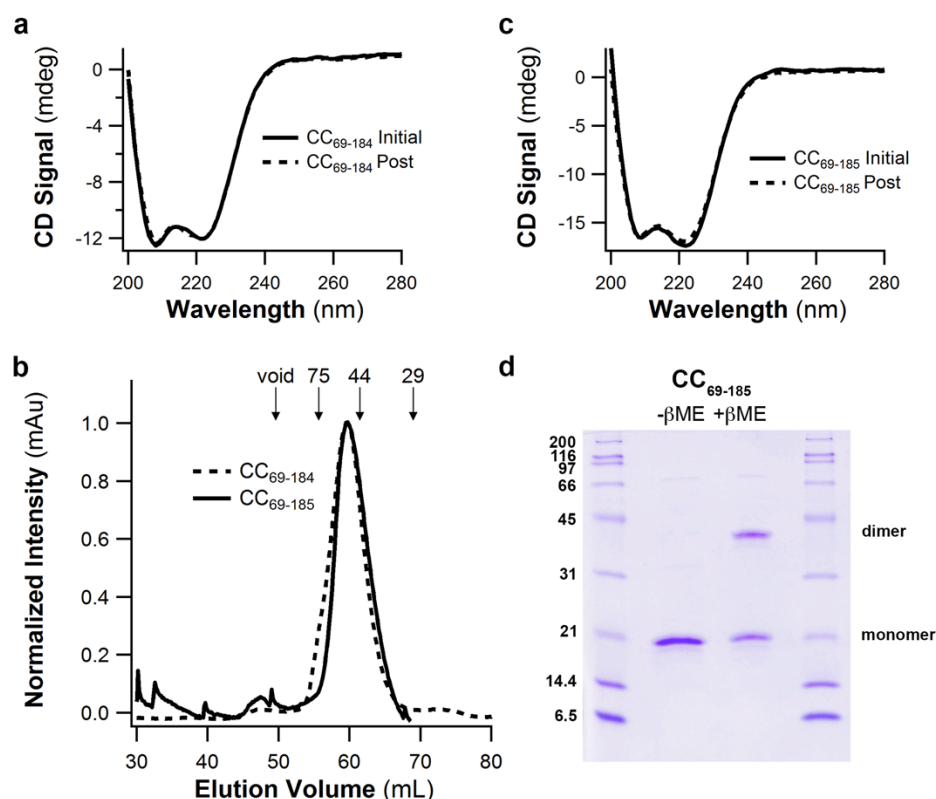
Supplemental Figure S3. Biochemical characterization of CC$_{69-184}$ and CC$_{69-185}$. Related to Fig. 3. (a) Circular dichroism spectra of CC$_{69-184}$ reveal alpha-helical signatures and reversible thermal unfolding (b) Superdex75 preparative SEC traces of CC$_{69-184}$ and CC$_{69-185}$ reveal a total molecular mass of 53 kDa based on a standard calibration curve. Given the calculated mass of 13.7 kDa/monomer the species is a tetramer. (b) Circular dichroism spectra of CC$_{69-185}$ reveal an alpha-helical signature and reversible thermal unfolding. (d) SDS-PAGE analysis of CC$_{60-185}$ under non-reducing conditions demonstrate predominantly disulfide-dependent dimer species.

38
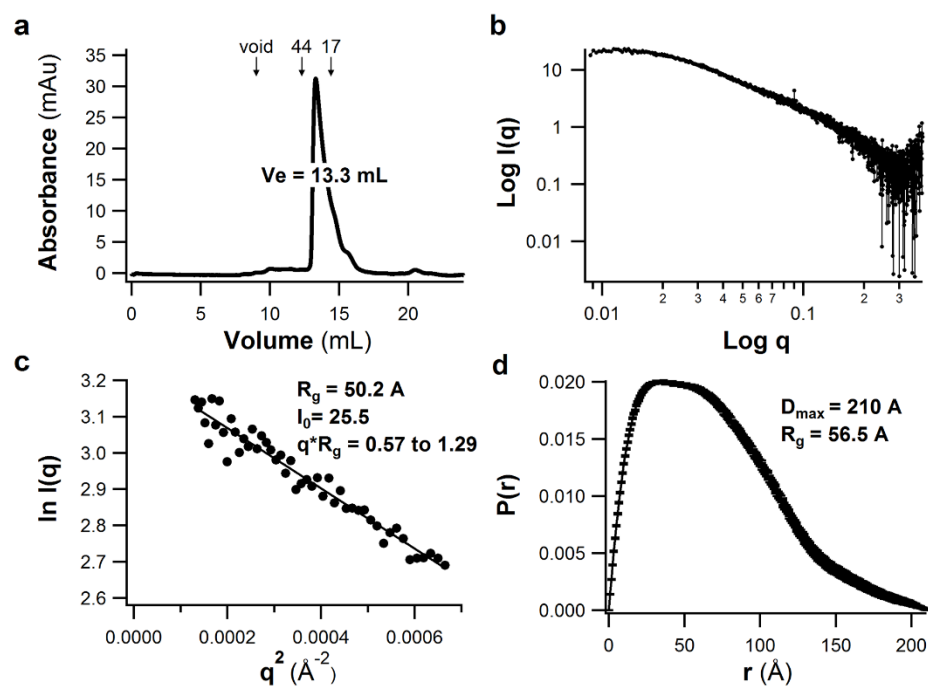
Supplemental Figure S4. CC$_{69\text{-}185}$ SEC-SAXS analysis. Related to Fig. 3. (a) Superdex-75 GL trace of CC$_{69\text{-}185}$ obtained at the beamline reveals a monodisperse sample. (b) Scattering intensity profile. I(q): scattered intensity, q:scattering vector. (c) Guinier plot with calculated radius of gyration (Rg). (d) Pair wise distribution plot with calculated maximum particle size (Dmax) and radius of gyration (Rg).
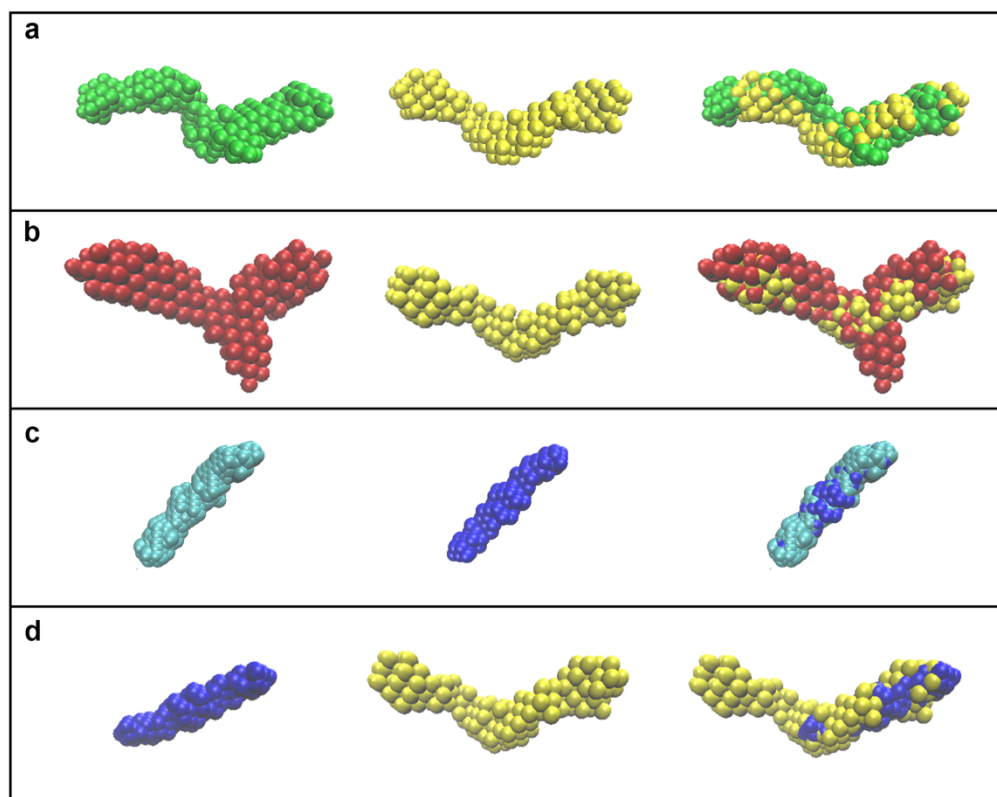
Supplemental Figure S5. Pairwise comparisons of *ab initio* SEC-SAXS models. Related to Fig. 3 and Table 2. Molecular envelopes of (a) CC_{69-185} (b) CC_{60-185} (c) $CC_{112-185}$ with no (P1) symmetry (left), P2 symmetry (middle) and superpositions with SUPCOMB (right). (d) Molecular envelopes of $CC_{112-185}$ with P2 symmetry (left), CC_{69-185} with P2 symmetry (middle) and superposition by manual manipulation in PyMOL (right).
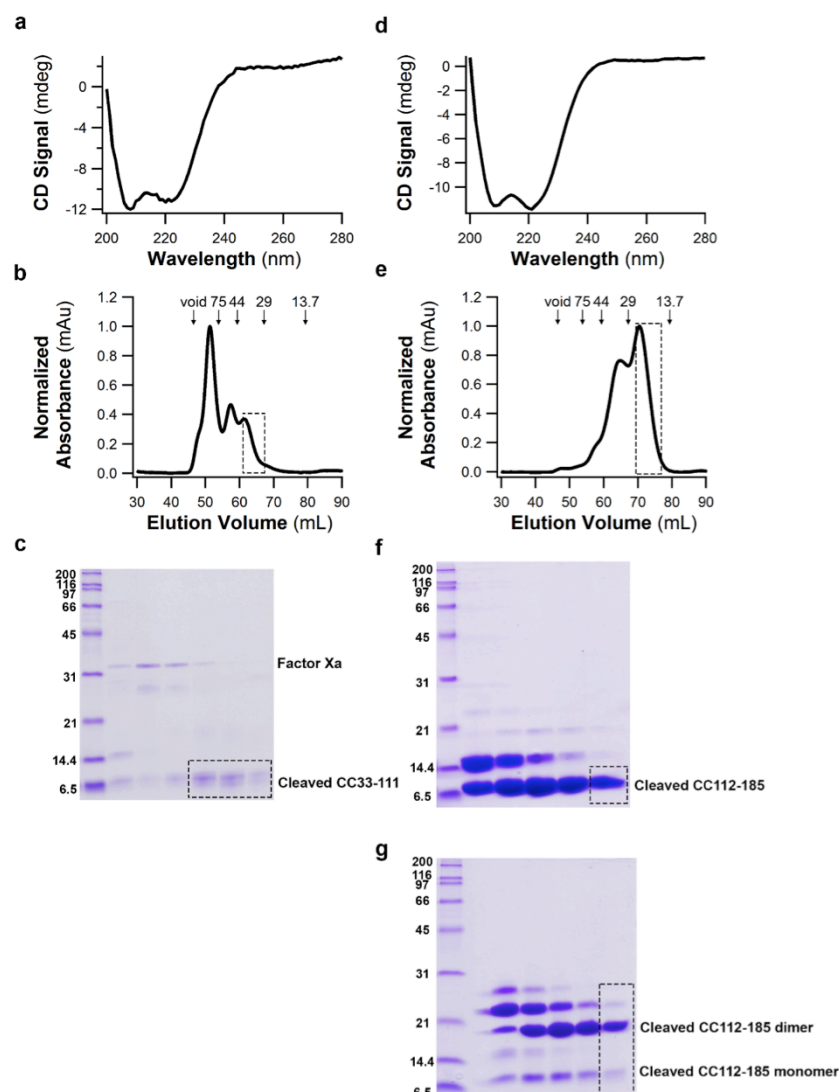
Supplemental Figure S6. Biochemical characterization of CC$_{33\text{-}111}$ and CC$_{112\text{-}185}$. Related to Fig. 3. (a) Circular dichroism spectrum of CC$_{33\text{-}111}$ reveals an alpha-helical signature. (b) Superdex-75 size exclusion chromatogram of CC$_{33\text{-}111}$. The first peak contains uncleaved and cleaved CC$_{33\text{-}111}$, the second peak contains mostly Factor Xa protease, and the third peak (indicated by a dashed box) contains cleaved CC$_{33\text{-}111}$ used in subsequent experiments. The molecular weight of the peak within the dashed box is 40 kDa based on a standard calibration curve, which consistent with a 4 or 5-mer of an ~8.8 kDa CC$_{33\text{-}111}$ monomer. (c) SDS-PAGE analysis of CC$_{33\text{-}111}$ fractions from Superdex- 75. Dashed box indicates fractions concentrated for SEC-SAXS. (d) Circular dichroism spectrum of CC$_{112\text{-}184}$ reveals an alpha-helical signature. (e) Superdex-75 size exclusion chromatogram of CC$_{112\text{-}185}$. The first peak contains uncleaved and cleaved CC$_{112\text{-}185}$, and the second peak (indicated by a dashed box) contains mostly cleaved CC$_{112\text{-}185}$ used in subsequent experiments. The molecular mass of the species is 23 kDa based on a standard calibration curve and thus consistent with a 2- or 3-mer of an ~9 kDa CC$_{112\text{-}185}$ species. (f) SDS-PAGE analysis of CC$_{112\text{-}185}$-containing fractions from Superdex 75 showin in (e). Dashed box indicates fraction selected for SEC-SAXS.

41
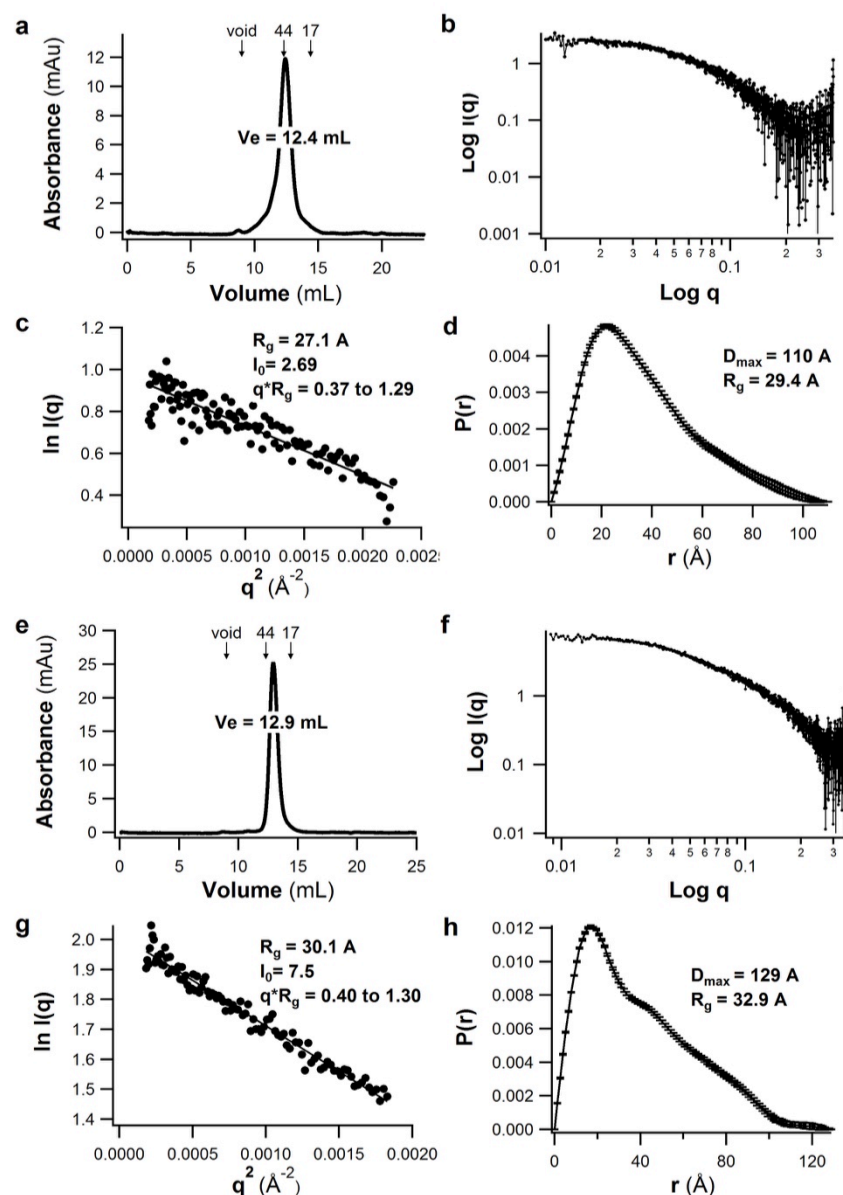
Supplemental Figure S7. CC$_{33-111}$ and CC$_{112-185}$ SEC-SAXS analysis. Related to Fig. 3. (a) Superdex-75 GL trace of CC$_{33-111}$ obtained at the beamline reveals monodisperse sample with a molecular weight of 44 kDa. Given the monomeric molecular mass of ~8.8 kDa the species is an apparent 5-mer. (b) Scattering intensity profile. I(q): scattered intensity, q:scattering vector. (c) Guinier plot with calculated radius of gyration (Rg). (d) Pair wise distribution plot with calculated maximum particle size (Dmax) and radius of gyration (Rg). (e) Superdex-75 GL trace of CC$_{112-185}$ obtained at the beamline reveals monodisperse sample (f) Scattering intensity profile. I(q): scattered intensity, q:scattering vector. (g) Guinier plot with calculated radius of gyration (Rg). (h) Pair wise distribution plot with calculated maximum particle size (Dmax) and radius of gyration (Rg).
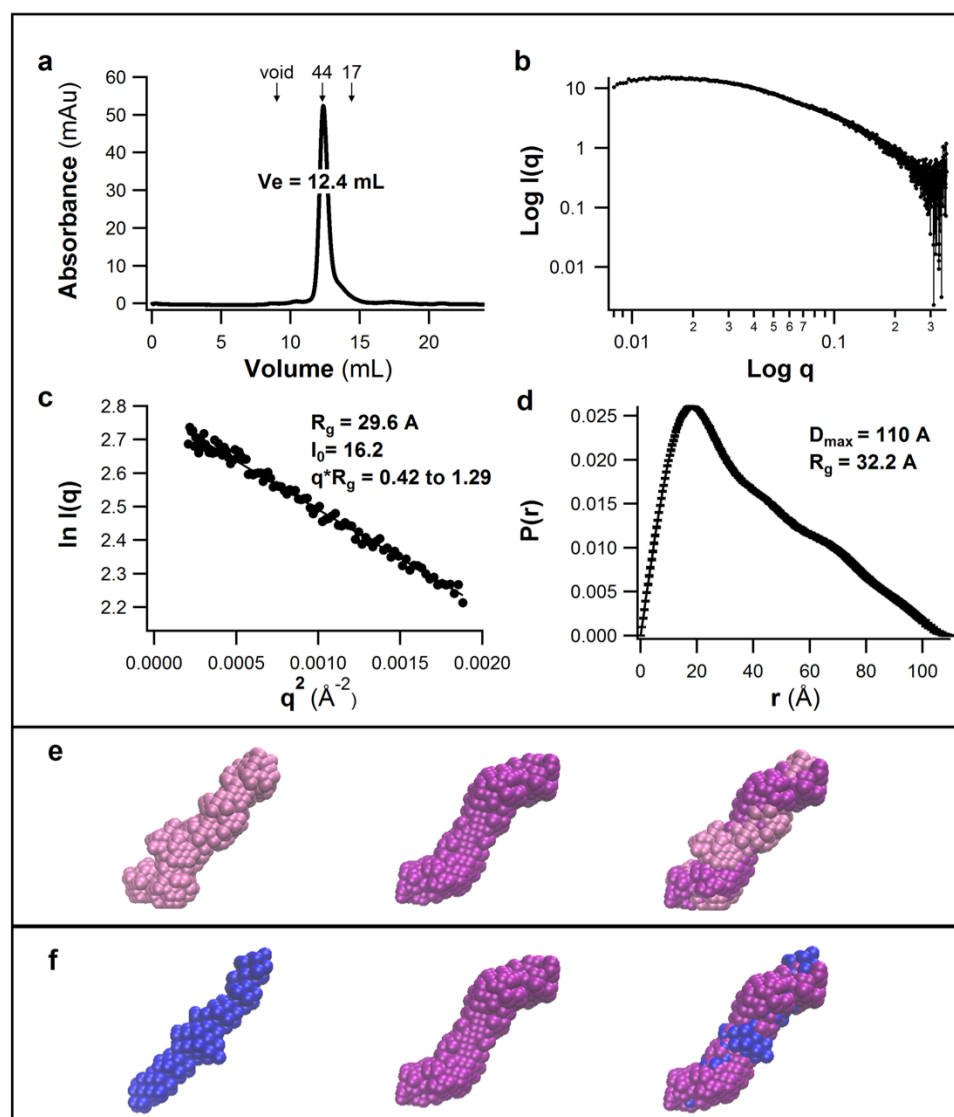
Supplemental Figure S8. mLZ$_{93-171}$ SEC-SAXS analysis. Related to Fig. 3. (a) Superdex-75 GL trace of mCC$_{93-171}$ reveals monodisperse sample (b) Scattering intensity profile. I(q): scattered intensity, q:scattering vector. (c) Guinier plot with calculated radius of gyration (Rg). (d) Pair wise distribution plot with calculated maximum particle size (Dmax) and radius of gyration (Rg). (e) Molecular envelopes of mCC$_{93-171}$ with no (P1) symmetry constraints (left), P2 symmetry constraints (middle), and the two envelopes superimposed with SUPCOMB (right). (f) Molecular envelopes of CC$_{112-185}$ with P2 symmetry constraint (left), mCC$_{93-171}$ with P2 symmetry constraint (middle) and superposition with SUPCOMB (right).

a

```
                                              x
gi_3065674_h._sapiens      1  MRFFCARCCSFGPEMPAVQLLLLACLVWDVGARTAQLRKANDQSGRCQYTFSVASPNESS   60
gi_15077142_m._musculus    1  --------------MPALHLLFLACLVWGMGARTAQFRKANDRSGRCQYTFTVASPNESS   46
gi_74356501_b._taurus      1  --------------MPAVQLLLLACLLGGVGARTAQFQKANDRSGRCQYTFSVASPSESS   46
gi_3845607_r._norvegicus   1  MP-SCARCCSCGPKMPALQLLFLACLVWGMGARTAQFRKANDRSGRCQYTFTVASPSESS   59
gi_62632725_d._rerio       1  MW-FL------------AVLWISSLLMGSQVQSSANLRRANAGNGRCQYTFMVDSPTEAS   47
                              ***  *: :::* :  ::::*::::**: :******* *:** *:*


                                        x             x
gi_3065674_h._sapiens     61  CPEQSQAMSVIHNLQRDSSTQRLDLEATKARLSSLESLLHQLTLDQA--ARPQETQEGLQ  118
gi_15077142_m._musculus   47  CPREDQAMSAIQDLQRDSSIQHADLESTKARVRSLESLLHQMTLGRV--TGTQEAQEGLQ  104
gi_74356501_b._taurus     47  CPEQGQAMLAIQELQRDSSEQRATLESTKARLSSLEALLHRLTSGQP--AGPLETHQGLQ  104
gi_3845607_r._norvegicus  60  CPREDQAMSAIQDLQRDSSIQHADLESTKARVRSLESLLHQMTSGGV--TGTQEVQEGLQ  117
gi_62632725_d._rerio      48  CPSPG---------------STPEMEALMSRLGLLEALVARLVGGEAMPESSQSSGSGLQ   92
                              efgabcdefgabcdefgabcdefgabcdefgabcdefgabcdeefga  bcdefgabcde
                              **   ***::*: ****** :   :*:::*: :**:*:::::: :       ::   ***


                                   x x
gi_3065674_h._sapiens    119  RELGTLRRERDQLETQTRELETAYSNLLRDKSVLEEEKKRLRQENENLARRLESSSQEVA  178
gi_15077142_m._musculus  105  GQLGALRRERDQLETQTRDLEAAYNNLLRDKSALEEEKRQLEQENEDLARRLESSSEEVT  164
gi_74356501_b._taurus    105  RELEALRRREREQLETQTQELESAYSNLVRDKSALEEEKRRLQAENEDLARRLESSSQEVA 164
gi_3845607_r._norvegicus 118  GQLGALRRERDQLETQTRDLEVAYNNLLRDKSALEEEKRQLEQENKDLARRLEGSSQEVA  177
gi_62632725_d._rerio      93  DSYNQVMGENAQL--------------------KREKQRLDRQVQDLQQRMEELRQEAE   131
                              fgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgabcdefgab
                              :  :::*: ******:*:** **:**:****:*:::** :*  ::  :*::*:* :::*:


gi_3065674_h._sapiens    179  RLRRGQCPQTRDTA------------RAVP------PGSR-------EVSTWNLDTLAFQ  213
gi_15077142_m._musculus  165  RLRRGQCPSTQYPS------------QDML------PGSR-------EVSQWNLDTLAFQ  199
gi_74356501_b._taurus    165  SLRRGQCPQAHSSS------------QDVP------SGSR-------EVAKWNLENMDFQ  199
gi_3845607_r._norvegicus 178  RLRRGQCPSTHHPS------------QDML------PGSR-------EVSQWNLDTLAFQ  212
gi_62632725_d._rerio     132  RLRSRPCMQQTSSRVPQKDNSFRPGSGHVPSNLASRPGNPQEDKSSLRDPAWQYSNPGYQ  191
                              cdefga
                              :**::*::              ::      :*::     ::  *:: :  :*


                                   x
gi_3065674_h._sapiens    214  ELKSELTEVPASRI--------------  227
gi_15077142_m._musculus  200  ELKSELTEVPASQILKENPSGRPRS---  224
gi_74356501_b._taurus    200  ELKSELTEVPASQI--------------  213
gi_3845607_r._norvegicus 213  ELKSELTEVPASQILK-NQSGHPRSKEG  239
gi_62632725_d._rerio     192  ELTAVVTEVTAPNQDGPAD---------  210
                              **::::***:*: :
```

b

Supplemental Figure S9. Related to Fig. 1. Sequence comparison of myocilins across organisms. (a) Multiple sequence alignment of N-terminal region of myocilin from human, mouse, bovine, rat, and zebrafish (OLF domain excluded). (b) HMM weblogo representation of conservation across the same region as in (a) across myocilin from 75 species.

44

Supplemental Figure S10. Molecular dynamics simulations of mLZ$_{122-171}$. Related to Fig. 4. (a) Root mean squared fluctuation (RMSF) values for individual mLZ$_{122-171}$ residues for simulations conducted at 310 K. (b)/(c) Salt bridges formed between individual coils A and B during 100-ns simulation of mLZ$_{122-171}$ at 400 K. For (c), backbone structures are shown for 0 ns (left) and 100 ns (right) of simulation time with salt-bridge-forming residues shown explicitly. Salt bridges were identified by ≤3.2 Å distance between side chain oxygen and nitrogen atoms. (d) Distances between backbone carbonyl carbon atoms of chains A and B for each residue of mLZ$_{122-171}$ during steered molecular dynamics (SMD) simulations at 310 K. Backbone structures from the SMD1 trajectory are shown above the graph with side chains of heptad positions '*a*' and '*d*' shown explicitly. For all protein structures shown: Individual protein chains are labeled A and B. Protein backbone is shown in cartoon representation colored by secondary structure: (magenta) α-helix and (white) random coil. Explicit side chains are shown in licorice representation colored by residue type: (blue) positively charged, (red) negatively charged, (green) polar, and (white) hydrophobic.

Supplemental Figure S11. Biochemical characterization of glaucoma-associated NTD$_{33\text{-}227}$ variants. Related to Fig. 4. (a) SDS-PAGE analysis of purified variants (~ 30 kDa monomer). (bh) Comparison of CD spectra before (initial) and after (post) thermal melt shows a high level of reversibility for six variants studied.

a

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
157159481	molecular chaperone DnaK [Escherichia coli HS]	510.06	54.23	566	11	28	100	638	69.1	4.97
16763402	molecular chaperone DnaK [Salmonella enterica subsp. enterica serovar Typhimurium str. LT2]	304.82	36.52	552	1	18	62	638	69.2	4.97
16130190	fused UDP-L-Ara4N formyltransferase/UDP-GlcA C-4'-decarboxylase [Escherichia coli str. K-12 substr. MG1655]	76.90	19.70	22	9	9	16	660	74.2	6.87
160961491	keratin, type II cytoskeletal 1 [Pan troglodytes]	76.64	20.88	46	6	8	18	637	65.4	7.81
157376541	molecular chaperone DnaK [Shewanella sediminis HAW-EB3]	54.60	6.09	519	1	3	11	640	69.0	4.73
114667511	PREDICTED: keratin, type I cytoskeletal 10 isoform 3 [Pan troglodytes]	51.02	17.85	15	6	6	11	577	58.2	5.16
47132620	keratin, type II cytoskeletal 2 epidermal [Homo sapiens]	50.72	14.24	30	2	4	11	639	65.4	8.00

b NTD$_{33-227}$

c

d

Annotated Sequence	# PSMs
[G].RVTGTQEAQEGLQGQLGALR.[R]	11
[G].RVTGTQEAQEGLQGQLGALR.[R]	10
[R].VTGTQEAQEGLQGQLGALR.[R]	159
[R].VTGTQEAQEGLQGQLGALRR.[E]	25
[R].VTGTQEAQEGLQGQLGALRRER.[D]	8
[R].VTGTQEAQEGLQGQLGALRRE.[R]	1
[R].VTGTQEAQEGLQGQ.[L]	1
[R].VTGTQEAQEGLQGQLGALR.[R]	152
[R].VTGTQEAQEGLQGQLGALRR.[E]	25
[R].VTGTQEAQEGLQGQLGALRRER.[D]	6
[R].VTGTQEAQEGLQGQLGALRRE.[R]	1
[V].TGTQEAQEGLQGQLGALR.[R]	7
[V].TGTQEAQEGLQGQLGALR.[R]	7
[T].GTQEAQEGLQGQLGALR.[R]	12
[T].GTQEAQEGLQGQLGALR.[R]	13
[G].TQEAQEGLQGQLGALR.[R]	13
[G].TQEAQEGLQGQLGALR.[R]	14
[T].QEAQEGLQGQLGALR.[R]	2
[T].QEAQEGLQGQLGALR.[R]	2
[Q].EAQEGLQGQLGALR.[R]	13
[Q].EAQEGLQGQLGALR.[R]	14
[A].QEGLQGQLGALR.[R]	4
[A].QEGLQGQLGALR.[R]	4
[Q].EGLQGQLGALR.[R]	10
[Q].EGLQGQLGALR.[R]	13
[E].GLQGQLGALR.[R]	3
[E].GLQGQLGALR.[R]	3
[L].QGQLGALR.[R]	1
[R].ERDQLETQTR.[D]	2
[R].ERDQLETQTRDLEAAYNNLLR.[D]	1
[R].ERDQLETQTRDLEAAYNNLLR.[D]	4
[R].ERDQLETQTR.[D]	1

Annotated Sequence	#PSMs
[R].DQLETQTRDLEAAYNNLLR.[D]	4
[R].DQLETQTRDLEAAYNNLLR.[D]	3
[R].DLEAAYNNLLR.[D]	164
[R].DLEAAYNNLLRDK.[S]	7
[R].DLEAAYNNLLRDKSALEEEK.[R]	3
[R].DLEAAYNNLLRDKSALEEEKR.[Q]	2
[R].DLEAAYNNLLR.[D]	174
[R].DLEAAYNNLLRDK.[S]	9
[R].DLEAAYNNLLRDKSALEEEK.[R]	3
[R].DLEAAYNNLLRDKSALEEEKR.[Q]	2
[D].LEAAYNNLLR.[D]	5
[D].LEAAYNNLLR.[D]	5
[L].EAAYNNLLR.[D]	9
[L].EAAYNNLLR.[D]	9
[R].DKSALEEEKRQLEQENEDLAR.[R]	12
[R].DKSALEEEKRQLEQENEDLAR.[R]	15
[K].SALEEEKRQLEQENEDLAR.[R]	6
[K].SALEEEKRQLEQENEDLAR.[R]	8
[K].SALEEEKR.[Q]	1
[K].RQLEQENEDLAR.[R]	15
[K].RQLEQENEDLAR.[R]	15
[R].QLEQENEDLAR.[R]	93
[R].QLEQENEDLAR.[R]	91
[Q].LEQENEDLAR.[R]	1
[Q].LEQENEDLAR.[R]	1
[L].EQENEDLAR.[R]	6
[L].EQENEDLAR.[R]	5
[E].QENEDLAR.[R]	4
[E].QENEDLAR.[R]	3
[R].RLESSSEEVTR.[L]	11
[R].RLESSSEEVTR.[L]	9
[R].LESSSEEVTR.[L]	1

Supplemental Figure S12. Additional supplemental material accompanying methods. Related to Supplemental Methods. (a) Identification of DnaK by mass spectrometry as main contaminant (~60 kDa) visible after purification on SDS-PAGE. (b) SDS-PAGE analysis of NTD$_{33-227}$ before and after unfolding/refolding procedure. (c) Comparison of Superdex-75 elution profile of two samples in (b). (d) Identification of mLZ$_{93-171}$ as predominant product after cleavage of LZ$_{55-171}$ by Factor Xa.

47

Supplemental Table S1. Related to Fig. 4. Crystallographic statistics.

Wavelength	1 Å
Resolution range	41.88 - 1.922 (1.991 - 1.922)
Space group	C 1 2 1
Unit cell	85.129 48.746 55.798 90 105.965 90
Total reflections	16652
Unique reflections	16650 (1447)
Multiplicity	7.0(5.1)
Completeness (%)	99.0 (89.4)
Mean I/sigma(I)	21.67
Wilson B-factor	28.16
R-meas	0.077(0.339)
R-pim	0.032(0.141)
Reflections used in refinement	16259 (1337)
Reflections used for R-free	1616 (133)
R-work	0.1872 (0.2525)
R-free	0.2354 (0.3237)
Number of non-hydrogen atoms	1852
macromolecules	1670
solvent	182
Protein residues	200
RMS(bonds)	0.006
RMS(angles)	0.79
Ramachandran favored (%)	97.40
Ramachandran allowed (%)	2.08
Ramachandran outliers (%)	0.52
Rotamer outliers (%)	0.00
Clashscore	12.47
Average B-factor	43.62
macromolecules	43.61
solvent	43.71
Number of TLS groups	8

Statistics for the highest-resolution shell are shown in parentheses.

Supplemental Table S2. Related to Supplemental Methods. Primers used in this study.

Construct	Primers for pet 30 xa/lic subcloning
NTD_{33-227}	Fwd 5'GGTATTGAGGGTCGCCGCACCGCTCAGCTGCGCAAGGCGAATG Rev 5'AGAGGAGAGTTAGAGCCTTATCTAGATGCAGGAACTTCGGTCAGCTC
CC_{60-185}	Fwd 5' GGTATTGAGGGTCGCTCTTGCCCGGAGCAG Rev 5'AGAGGAGAGTTAGAGCCTTAACACTGACCACGACGCAAGCGTGCGAC
CC_{69-184}	Fwd 5'GGTATTGAGGGTCGCTCGGTGATTCACAATCTGCAGCGTG Rev 5'AGAGGAGAGTTAGAGCCTTACTGACCACGACGCAAGCGTGCGACTTC
$CC_{112-184}$	Fwd 5'GGTATTGAGGGTCGCGAAACTCAGGAAGGCTTGCAACGTG Rev 5'AGAGGAGAGTTAGAGCCTTACTGACCACGACGCAAGCGTGCGACTTC
Construct	**Primers for Site Directed Mutagenesis**
CC_{69-185}[a]	Fwd 5'CTTGCGTCGTGGTCAGTGTTAAGGCTCTAACTCTC Rev 5'GAGAGTTAGAGCCTTAACACTGACCACGACGCAAG
$CC_{112-185}$[a]	Fwd 5'CTTGCGTCGTGGTCAGTGTTAAGGCTCTAACTCTC Rev 5'GAGAGTTAGAGCCTTAACACTGACCACGACGCAAG
CC_{33-111}[b]	Fwd 5'CAAGCTGCGCGTCCGCAGTAAACTCAGGAAGGCTTGC Rev 5'GCAAGCCTTCCTGAGTTTACTGCGGACGCGCAGCTTG
NTD_{33-227} $C_{47}S$[c]	Fwd 5'GACCAGTCGGGCCGTAGTCAGTATACCTTTTC Rev 5'GAAAAGGTATACTGACTACGGCCCGACTGGTC
NTD_{33-227} $C_{61}S$[c]	Fwd 5'GCCCTAATGAGTCCTCTAGCCCGGAGC Rev 5'GCTCCGGGCTAGAGGACTCATTAGGGC
NTD_{33-227} $C_{185}S$[c]	Fwd 5'GCGTCGTGGTCAGAGTCCGCAAACGCG Rev 5'CGCGTTTGCGGACTCTGACCACGACGC
NTD_{33-227} V53A	Fwd 5'TCAGTATACCTTTTCCGCGGCAAGCCCTAATGAGT Rev 5'ACTCATTAGGGCTTGCCGCGGAAAAGGTATACTGA
NTD_{33-227} R82C	Fwd 5'GATTCCAGCACGCAGTGTCTGGACCTGGAAG Rev 5'CTTCCAGGTCCAGACACTGCGTGCTGGAATC
NTD_{33-227} L95P	Fwd 5'GCCTGAGCAGCCCGGAGAGCCTGCT Rev 5'AGCAGGCTCTCCGGGCTGCTCAGGC
NTD_{33-227} R126W	Fwd 5'GTGAATTGGGTACGCTGCGTTGGGAACGTGACCAGCTGGAAAC Rev 5'GTTTCCAGCTGGTCACGTTCCCAACGCAGCGTACCCAATTCAC
NTD_{33-227} R128W	Fwd 5'CGCTGCGTCGTGAATGGGACCAGCTGGAAAC Rev 5'GTTTCCAGCTGGTCCCATTCACGACGCAGCG
NTD_{33-227} L215Q	Fwd 5'CTGGCGTTTCAAGAGCAGAAAAGCGAGCTGACC Rev 5'GGTCAGCTCGCTTTTCTGCTCTTGAAACGCCAG

[a]CC_{69-185} and $CC_{112-185}$ produced by insertion of Cys_{185} to CC_{69-184} and $CC_{112-184}$, respectively.

[b]CC_{33-111} produced by mutation to stop codon after Q_{111} in NTD_{33-227} construct.

[c]NTD_{33-227} double and triple cysteine to serine variants produced by multiple rounds of

mutagenesis.