Double Domain Swapping in Human γ C and γ D Crystallin Drives Early Stages of Aggregation

Balaka Mondal

*Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore,
Karnataka 560012, India*

Jayashree Nagesh

*Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore,
Karnataka 560012, India*

Govardhan Reddy*

*Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore,
Karnataka 560012, India*

Abstract

Human γ D (H γ D) and γ C (H γ C) are double domain crystallin (Crys) proteins expressed in the nucleus of the eye lens. Structural perturbations in the protein often trigger aggregation, which eventually leads to cataract. To decipher the underlying molecular mechanism, it is important to characterize the partially unfolded conformations of Crys proteins. Using coarse grained protein models and molecular dynamics simulations, we studied the role of on-pathway folding intermediates in the early stages of aggregation. The multi-dimensional free energy surface revealed at least three different folding pathways with the population of partially structured intermediates. The two dominant pathways confirm sequential folding of the N-terminal [Ntd] and the C-terminal domains [Ctd], while the third, least favored pathway involves intermediates where both the domains are partially folded. A native like intermediate (I^*), featuring the folded domains and disrupted inter domain contacts, gets populated in all the three pathways. I^* forms domain swapped dimers by swapping the entire Ntds

*Corresponding author

Email address: greddy@iisc.ac.in (Govardhan Reddy)

and Ctds with other monomers. Population of such oligomers can explain the increased resistance to unfolding resulting in hysteresis observed in the folding experiments of H γ D Crys. An ensemble of double domain swapped dimers are also formed during refolding, where intermediates consisting of partially folded Ntds and Ctds swap secondary structures with other monomers. The double domain swapping model presented in our study provides structural insights into the early events of aggregation in Crys proteins and identifies the key secondary structural swapping elements, where introducing mutations will aid in regulating the overall aggregation propensity.

Introduction

Lens opacity or cataract[1] is a protein aggregation disorder associated with misfolding and subsequent aggregation of eye lens crystallin (Crys) proteins, resulting in partial or complete loss of vision. Detailed structural characterization of the misfolded states and molecular mechanism of cataract formation have remained elusive to date. In humans, surgically removed cataractous lenses contain high molecular weight aggregates of α , β and γ Crys, the three main classes of Crys found in vertebrates[2, 3]. α Crys primarily acts as molecular chaperones[4], while β and γ Crys are responsible for optical functions of the lens[5, 6].

Monomeric γ Crys are the simplest members of the Crys family. The three most abundant γ Crys proteins in human are H γ C, H γ D and H γ S Crys, which are present in the central region of the lens. H γ C and H γ D Crys are expressed in the lens nucleus[7], which is the oldest region of the lens with no protein turnover. Since these proteins are only synthesized once, the native state of these proteins should be stable and functional for human lifetime. Any misfolding significantly perturbs the equilibrium concentrations of H γ C and H γ D Crys in the lens.

H γ C and H γ D Crys are both 173 residue long proteins with ≈ 71 % sequence similarity. The native structure of γ Crys is topologically complex with

a N-terminal domain (Ntd) and a C-terminal domain (Ctd), which are highly homologous. Each domain consists of two intercalated greek-key motifs, which are complex supersecondary structures formed by four antiparallel β strands (Figure 1A,2A). Intercalated greek key motifs termed M_1 and M_2 constitute the Ntd, while M_3 and M_4 constitute the Ctd [8] (Figure 1B). A short peptide linker connects the two domains, which form a highly conserved hydrophobic interface necessary for the long term stability of the proteins [9]. The topology of the native state of γ Crys leads to complex folding behavior as observed in experiments[10, 11, 12, 9] and computer simulations[13, 14, 15].

Due to their implications in cataract, biophysical properties of γ Crys have been studied extensively using both experimental and computational techniques. Studying folding in multi-domain proteins is challenging since the domain-domain interactions add to the complexity as they can influence folding in addition to the intra-domain interactions[16, 17, 18, 19]. Denaturant induced unfolding experiments[10, 20, 21, 11] demonstrated a three state unfolding transition in γ Crys indicating the presence of a partially structured intermediate state. This intermediate state is structurally characterized for H γ D Crys, which consists of a folded Ctd and an unfolded Ntd[22, 9], in agreement with the simulations[13].

Hysteresis is another interesting property observed in the denaturant induced equilibrium unfolding/refolding experiments[10, 23, 24]. Native states of H γ D Crys are observed to be stable at higher denaturant concentrations where unfolded states did not refold. Increased temperature or equilibration time led to diminished hysteresis causing a shift in the unfolding transition only, indicating a high kinetic barrier to unfolding as a possible reason for hysteresis. Further, increased stability of the inter domain interface led to enhanced hysteresis, which is absent in the truncated domains[11]. Thus, the inter domain interactions are attributed to the high kinetic barriers in the early stages of unfolding[24, 25]. However, the origin of hysteresis is yet to be explored.

Experiments[10, 26, 27] further revealed a competitive protein aggregation pathway during the refolding of H γ D and H γ C Crys at $[\text{GdmCl}] < 1$ M due to

population of partially folded intermediates[10]. Domain swapping[28, 29, 30, 31, 32, 33] plays an important role in the aggregation of globular proteins such as prion proteins[34], cystatin C[35, 36], c-Src SH3[37], β 2-microglobulin[38] etc., where it is believed to drive the early stages of aggregation. In proteins such as RNase A, which can swap secondary structures from both Ntd and Ctd, double domain swapping is observed leading to a variety of differently assembled oligomeric states[39, 40, 41, 42]. To elucidate the structure of aggregation prone intermediates and underlying aggregation mechanism, several studies have investigated the dimer formation in H γ D Crys, which is the simplest oligomer[14, 12, 43]. A transient intermediate state with an extruded N terminal β hairpin loop is identified in single molecule atomic force microscopy (AFM) experiments[12] and computer simulations[43], which undergoes dimerization yielding different dimer structures in experiments and simulations. Although the molecular events leading to aggregation are not fully understood, domain swapping is considered to be the most probable aggregation mechanism[14, 12, 43] in H γ D Crys.

All-atom simulations predicted domain swapped structures where Ntd of one monomer interacts with the three β strands from M_4 of Ctd from another monomer[14], and β_1 from extruded β hairpin loop in Ntd forms antiparallel hydrogen bonds with the second β strand from M_4 of Ctd [43]. These structures remain to be validated experimentally. AFM experiments on H γ D Crys indicated a different domain swapped structure with Ntd β hairpin loops swapped between adjacent monomers[12]. Experiments also suggest that the aggregation prone intermediate is not structurally similar to the unfolding intermediate and consists of a partially folded Ctd[44].

To probe the reasons behind the observed hysteresis during the folding of γ Crys and to gain insight into the aggregation prone misfolded intermediates, we studied the folding thermodynamics of H γ C (protein data bank (PDB) ID: 2NBR)[45] and H γ D (PDB ID: 1HK0)[46] Crys using molecular dynamics simulations and coarse grained self-organized polymer side-chain (SOP-SC) protein model[47, 48]. The folding free energy surface (FES) constructed from low

friction Langevin dynamics simulations show that H γ C and H γ D Crys exhibit kinetic partitioning mechanism (KPM) and folds through at least three distinct pathways populating different intermediate states. One of the intermediates is an early unfolding intermediate with both domains folded and an exposed interface. Through the population of this intermediate, we show that the native proteins form a domain swapped dimer, which is a more stable entity with four folded domains and two interfaces. The population of this dimer structure during the early stages of unfolding can lead to hysteresis by resisting unfolding and requiring relatively higher denaturant concentrations to unfold. Experiments report similar domain swapped dimer structures for human β B2 Crys[49, 50, 51] (PDB ID: 2BB2) and mouse γ S Crys (PDB ID: 6MYG, 6MYH).

Experiments[10] further show that γ Crys exhibits significant aggregation while refolding at lower denaturant concentrations. To gain insight into this *in vivo* aggregation, we carried out dimerization simulations of the unfolded proteins. During dimerization, secondary structural elements from partially folded Ntd and Ctd of the proteins took part in domain swapping leading to an ensemble of double domain swapped structures. The primary swapping units in H γ C and H γ D Crys were found to be very similar, which can be attributed to their high sequence similarity ($\approx 71\%$). Both in H γ D and H γ C Crys, Ntds undergo aggregation by swapping the $\beta_1\beta_2$ hairpin loop, while Ctds swap the first hairpin loop in motif M_3 ($\beta_6\beta_7$ in H γ D Crys, $\beta_8\beta_9$ in H γ C Crys). Other less common swapping units identified from the dimer structures include strands β_1 , $\beta_3\beta_4$, motif M_1 in the Ntd of H γ D Crys (strands β_1 , β_7 and motif M_1 in the Ntd of H γ C Crys) and strands β_6 , β_{12} in the Ctd of H γ D Crys (β_8 , β_{14} in the Ctd of H γ C Crys). AFM experiments[12] on H γ D Crys also predicted a domain swapped dimer structure where N-terminal β hairpin loops get swapped between the monomers. Interestingly, a large number of cataract associated mutations are found in $\beta_1\beta_2$ loop region in H γ D Crys[24, 52, 53, 54, 55]. Such double domain swapping mechanism is feasible in an end to end fashion[56] among specific partially structured intermediates, giving rise to long chains of interconnected oligomers, which might act as aggregation precursors in cataractous lenses.

Results and Discussion

Folding Thermodynamics of H γ C and H γ D Crys: Folding of γ Crys proteins have been studied using different experimental methods including fluorescence spectroscopy, circular dichroism, AFM etc[12, 44, 10, 22, 11]. However, detailed characterization of the transient intermediates populated in the folding pathways still remains a challenge. We performed low friction Langevin dynamics simulations to study the folding of H γ C and H γ D Crys using coarse grained SOP-SC protein model[47, 48]. A detailed description of the SOP-SC protein model and simulation parameters are in the supporting information (SI). The simulations are performed at temperatures ranging from $T = 300$ K to 500 K and we computed various thermodynamic properties associated with the folding.

Temperature dependent variation in specific heat capacity (C_v) and average potential energy ($\langle E \rangle$) indicate a cooperative two-state folding for H γ D Crys with the melting temperature, $T_M \approx 426$ K (Figure 1C). Whereas a three-state folding with the population of an intermediate is observed for H γ C Crys with two peaks in the C_v plot at $T_M^I \approx 410$ K and $T_M^{II} \approx 420$ K (Figure 2B). Experimentally[57, 26] measured T_M for H γ C and H γ D Crys are 352 K and 357 K, respectively. The deviation in the T_M obtained from experiments and simulations is due to the coarse-grained level description used to model the proteins, which cannot accurately mimic the experimental conditions. Experimental studies on H γ D and H γ C Crys further indicated a three state equilibrium unfolding with the population of an intermediate state[10]. Although the C_v plot associated with H γ C Crys indicates a three state folding/unfolding transition with an intermediate, only two state folding/unfolding transition is visible from the C_v plot of H γ D Crys.

We also estimated the T_M of Ntd and Ctd domains of both the proteins from the full-length protein simulations by computing the variance in the fraction of native contacts in the Ntd (χ_{Ntd}) and Ctd domains (χ_{Ctd}) as a function of T (see SI for details). The T_M of Ntd and Ctd domains approximately corresponds to the peak position in the plots (Figure S1, S2). The Ntd and Ctd of H γ C Crys

exhibit differential domain stability with T_M of Ntd and Ctd ≈ 410 K and ≈ 420 K, respectively. The T_M of Ntd and Ctd domains corresponds to T_M^I and T_M^{II} of H γ C Crys in the C_v plot (Figure 2B), indicating sequential unfolding of Ntd followed by Ctd. Thus the intermediate populated in the C_v plot (Figure 2B) of full-length H γ C Crys consists of a folded Ctd and an unfolded Ntd, evident from the T_M values of the individual domains. Unlike H γ C Crys, truncated Ntd and Ctd of H γ D Crys indicated comparable thermal stabilities (with $T_M \approx 426$ K) in agreement with the C_v plot of full-length H γ D Crys (Figure 1C). However, experiments[27] report a higher thermodynamic stability of the Ctd ($T_M = 349.2$ K) as compared to the Ntd ($T_M = 337.5$ K) of H γ D Crys. This anomaly arises due to the reduced-level description of the model systems, as discussed earlier.

FES of H γ D and H γ C Crys Folding Show Population of Multiple Intermediate States: Although thermodynamic properties such as $\langle E \rangle$ and C_v capture major transitions associated with protein folding/unfolding, they often fail to detect transient intermediates, which occur due to minor structural perturbations that play an important role in protein aggregation. To identify the transient intermediates we constructed the FES associated with the folding of H γ D and H γ C Crys. The FES is projected onto 3 collective variables (CVs): (1) fraction of native contacts in Ntd (χ_{Ntd}), (2) fraction of native contacts in Ctd (χ_{Ctd}) and (3) fraction of native contacts at the interface of Ntd and Ctd (χ_{Int}) (see Methods for details). We employed weighted histogram analysis method (WHAM)[58] to construct the three-dimensional FES, $F(\chi_{Ntd}, \chi_{Ctd}, \chi_{Int})$.

The folding FES of H γ D and H γ C Crys show that for both the proteins, at least 3 distinct minimum energy pathways connect the native (N) and unfolded (U) basins populating different intermediates (Figure 1D,E,2C,D,S3,S4). In H γ D Crys, the most preferred pathway (Path A, shown in black) corresponds to the folding of the Ctd followed by the folding of Ntd (Figure 1D). In this pathway, 4 intermediates are populated: (1) a partially folded Ctd with motif M_4 folded and Ntd unfolded (basin D_1), (2) Ctd completely folded and Ntd

unfolded (basin D_2), (3) Ctd completely folded and motif M_1 of Ntd folded (basin D_3), and (4) both Ctd and Ntd folded with disrupted inter-domain contacts (basin I^*) (Figure S3A,B,C,H,S5A). Intermediate D_2 with a folded Ctd and unfolded Ntd was inferred from experiments[22, 9] and was also observed in the H γ D unfolding simulations[13, 15].

In the second preferred pathway (Path B, shown in red), Ntd folds prior to the folding of Ctd (Figure 1D). In this pathway, 4 intermediates are populated: (1) a partially folded Ntd with motif M_1 folded and Ctd unfolded (basin D_4), (2) completely folded Ntd and unfolded Ctd (basin D_5), (3) completely folded Ntd and partially folded Ctd with motif M_4 folded (basin D_6) and (4) intermediate I^* , which is also common to Path A (Figure S3D,E,F). AFM experiments on H γ D Crys provide evidence for the population of an intermediate with a folded Ntd and unfolded Ctd, which is similar to the intermediate, D_5 . Interestingly, kinetic unfolding experiments on H γ D Crys in the absence of denaturants indicated higher kinetic stability of Ntd compared to Ctd[11].

In Paths A and B, sequential folding of the domains is observed. Whereas in the third folding pathway (Path C, shown in green), which is the least favored, intermediate with both the domains partially folded are observed. An intermediate with partially folded Ntd with motif M_1 folded and partially folded Ctd with motif M_4 folded (basin D_7) is populated in addition to D_4 , D_3 and I^* (Figure S3G,S5A).

Similar to H γ D Crys, the preferred path connecting the N and U basins in H γ C Crys (Path A shown in black) involves stepwise folding of the Ctd prior to the folding of Ntd. The 4 intermediates populated in this path are: (1) partially folded Ctd with motif M_4 folded and unfolded Ntd (basin C_1), (2) completely folded Ctd and unfolded Ntd (basin C_2), (3) completely folded Ctd and partially folded Ntd with M_1 folded (basin C_3), and (4) both Ctd and Ntd folded with disrupted inter-domain contacts (I^*) (Figure 2C,S4A,B,C,H,S5B). In the second preferred path (Path B shown in red), Ntd folds prior to Ctd. The 4 intermediates populated in this path are (1) partially folded Ntd with folded motif M_1 and unfolded Ctd (basin C_4), (2) completely folded Ntd and unfolded

Ctd (basin C_5), (3) completely folded Ntd and partially folded Ctd with motif M_4 folded (basin C_6), and (4) intermediate I^* (Figure S4D,E,F,S5B). The least favored path (Path C shown in green) involves an intermediate with partially structured Ntd with motif M_1 folded and partially structured Ctd with motif M_4 folded (basin C_7) in addition to C_3 , C_4 and I^* (Figure S4G,S5B).

The intermediate, I^* , which is common to all the folding pathways of H γ D and H γ C Crys is populated due to structural fluctuations in the native state ensemble. A hydrophobic patch at the domain interface of Ntd and Ctd gets exposed when I^* is populated and this makes it an ideal candidate for domain swapping. Human βB_2 Crys, which belongs to the $\beta\gamma$ Crys superfamily, forms such domain swapped dimers by swapping the Ntd and Ctd while the linker region acts as the hinge loop[51]. Similar domain swapped oligomers are also found in mouse γS Crys, which is a member of γ Crys family. Although no such domain swapped structures are reported for H γ D and H γ C Crys, intermediate I^* can potentially lead to such domain swapped oligomers. A recent computational study on W42R mutant of H γ D Crys showed the presence of I^* intermediates with exposed interdomain hydrophobic interface, which led to large scale aggregation involving Ntd-Ntd interactions[59].

Rest of the partially structured intermediate states populated in the folding pathways of H γ D and H γ C Crys can act as potential precursors to aggregation as well. Formation of sticky hydrophobic patches on the protein surface has been related to aggregation in cataract associated mutants of H γ D Crys[53, 60, 61]. Experimental studies on H γ D Crys indicated the presence of intermediates with either Ctd folded (basin D_2)[22, 9] or Ntd folded (basin D_5)[12, 11]. However, the role of these intermediates in aggregation remains unclear. Domain swapped dimer structure of H γ D Crys was predicted in AFM experiments, where $\beta_1\beta_2$ strands from Ntd get swapped between neighboring monomers[12]. Further, fluorescence spectra of chaperone (αB Crys) bound complex of H γ D Crys showed that both the domains were partially unstructured in the substrate[44], unlike the experimentally observed folding intermediates, which consist of a folded domain. Intermediates such as D_7 , C_7 populated in the FES of H γ D and H γ C

Crys, consist of partially unfolded Ntd and Ctd, and meet the requirements for possible aggregaion prone precursors. These intermediates or structurally related intermediates with exposed hydrophobic interface might play an important role in the aggregation of γ Crys proteins.

Dimerization in H γ D and H γ C Crys: Surgically removed cataractous lenses contain high molecular weight aggregates of partially folded γ Crys[62, 26, 44]. However, the specific conformations of the proteins in the aggregates remain elusive due to the insoluble nature of the aggregates. A detailed understanding of the misfolded conformations is crucial to elucidate the underlying molecular mechanism of cataract formation and design essential therapeutics. To understand the initial molecular interactions that advances to cataract formation in later stages, we studied the formation of dimers, which are the simplest possible oligomers, using symmetric Go potential to model inter-monomer interactions[32, 63] (Figure S9).

Dimerization from Native State: γ Crys are generally monomeric in solution. However, octameric structures of mouse γ S Crys are published in the PDB, where Ntds and Ctds are swapped between the neighboring monomers with linker region acting as hinge loop (PDB ID: 6MYG, 6MYH), indicating oligomerization in the native state of the protein. Similar domain swapping is also observed in human β B₂ Crys[51], which belongs to the $\beta\gamma$ Crys superfamily. To elucidate whether such oligomerization is also feasible in the native states of H γ D and H γ C Crys, we performed 50 independent Brownian dynamics simulations starting from two folded monomers (say monomers A and B) separated by 20 Å ($\approx R_g$ of a monomer) using a harmonic restraint with a spring constant, $k = 2.0$ kcal/mol/Å² at $T = 300$ K. To track the progress of dimerization, we computed the number of contacts Q_{AB} formed between the monomers A and B as a function of simulation time (Figure 3A,B). A two step dimerization process was observed in both the proteins yielding unique dimer structures ($Dim_{\gamma D}$ and $Dim_{\gamma C}$).

In the final dimer structure, both Ntd and Ctd of the two monomers remain folded (fraction of native contacts in monomer A, $\chi_A \approx 1$ and in monomer B, $\chi_B \approx 1$), however the interdomain contacts present between Ntd and Ctd of each monomer get disrupted ($\chi_{A,Int} \approx 0$, $\chi_{B,Int} \approx 0$) (Figure 3A,B). Instead, two new interfaces form between Ntd of monomer A and Ctd of monomer B ($\chi_{A,Ntd;B,Ctd} \approx 1$), and Ctd of monomer A and Ntd of monomer B ($\chi_{A,Ctd;B,Ntd} \approx 1$). This reveals a domain swapping mechanism, where neighboring monomers exchange Ntd and Ctd with each other forming two new interfaces, while the linker region acts as hinge loop. Dimers $Dim_{\gamma D}$ and $Dim_{\gamma C}$ obtained for H γ D and H γ C Crys, respectively, are very similar to the domain swapped structure obtained for mouse γ S Crys (PDB ID: 6MYG, 6MYH). NMR studies[64] further show that amorphous aggregates of cataract associated P23T mutant of H γ D Crys retains the native fold, indicating native-like conformations taking part in aggregation.

In the first step of dimerization, proteins break their inter domain contacts exposing the hydrophobic pocket present at the interface, populating the intermediate I^* . In I^* , the linker region adopts an extended conformation, which facilitates domain swapping (Figure 3C,D). In the next step, monomers swap either Ntd or Ctd with each other leading to half swapped dimers $Dim_{int,\gamma D}$ and $Dim_{int,\gamma C}$. Finally the remaining domains are swapped between the monomers resulting in fully swapped dimers $Dim_{\gamma D}$ and $Dim_{\gamma C}$. Intermediate I^* , which is key to dimerization in native state gets populated during the initial stages of unfolding of both the proteins (Figure 1D,2C). Low denaturant concentrations, which introduce small scale perturbations in the native structure, can enhance the population of I^* . I^* states can further take part in domain swapping with neighboring monomers leading to long chain of oligomeric structures. Population of such dimeric or oligomeric moieties with multiple inter protein interfaces can resist unfolding to a relatively higher denaturant concentrations and lead to hysteresis as observed in experiments[10, 23, 24].

Dimerization During Refolding: Aggregation is also observed experimentally for H γ D and H γ C Crys when [GdHCl] is lowered from high values to

less than 1 M[10]. Aggregation suppression experiments on γ Crys using chaperone α B Crys further identified partially unfolded substrates of H γ C, H γ D and H γ S Crys in the aggregates[44]. It was further concluded that both Ntd and Ctd of H γ D Crys must be partially unfolded for *in-vitro* aggregation[44]. To identify conformations of H γ D and H γ C Crys populated during refolding that result in aggregation, we studied dimerization of the proteins during refolding from unfolded conformations at $T = 300$ K. We performed 100 independent Brownian dynamics simulations starting from different unfolded conformations of the monomers separated by ≈ 20 Å. Dimerization was observed in $\approx 60\%$ of the trajectories in H γ D and H γ C Crys.

An ensemble of domain swapped dimer structures were identified, where at least in 45% of the structures of H γ D Crys ($\approx 40\%$ of H γ C Crys), secondary structural units were swapped from both Ntd and Ctd resulting in double domain swapped dimers. In rest of the structures, either Ntd ($\approx 40\%$ in H γ D Crys, $\approx 37\%$ in H γ C Crys) or Ctd ($\approx 15\%$ in H γ D Crys, $\approx 23\%$ in H γ C Crys) took part in swapping, while the other domain folded to its native structure. In both the proteins, most populated double domain swapped dimers are formed by swapping the secondary structural units from both Ntd and Ctd. The primary swapping element in the Ntds is the first hairpin loop $\beta_1\beta_2$ from motif M_1 (Figure 4A,C, S6A,C). Swapping of $\beta_1\beta_2$ strands was predicted in an AFM study of H γ D Crys[12]. In Ctds (Figure 4B,C, S6B,C), the primary swapping units are the first hairpin loop of motif M_3 ($\beta_6\beta_7$ in H γ D Crys and $\beta_8\beta_9$ in H γ C Crys).

In $\approx 10\%$ of the cases, we observed domain swapped structures similar to $Dim_{\gamma D}$, $Dim_{\gamma C}$ where both the domains are completely swapped between the monomers. Other less common swapping units (with population $< 5\%$) were identified in H γ D and H γ C Crys, which led to both single and double domain swapped dimer structures (Figure S7,S8). Ntd swapping units include β_1 , $\beta_3\beta_4$, motif M_1 in H γ D Crys, and β_1 , β_7 , motif M_1 in H γ C Crys. Ctd swapping units include β_6 , β_{12} in H γ D Crys, and β_8 , β_{14} in H γ C Crys. Some of the intermediates populated in the folding FES of H γ D (D_3 , D_4 , D_7) and H γ C

Crys (C_3 , C_4 , C_7) with a folded motif M_1 should facilitate swapping of M_1 between the Ntds as observed in the ensemble of swapped dimer structures.

In the most populated dimers, domain swapping took place after formation of the β hairpin loops $\beta_1\beta_2$, $\beta_6\beta_7$ (in H γ D Crys) and $\beta_8\beta_9$ (in H γ C Crys) (Figure 4, S6). In the final domain swapped structures, β_1 strand makes contacts with β'_3 strands from the other monomer and vice versa. While in Ctds, β_6 strand in H γ D Crys (β_8 strand in H γ C Crys) makes contacts with β'_8 strand (β'_{10} strand in H γ C Crys) from the other monomer and vice versa. Interestingly, in H γ D Crys, highest number of pathogenic mutations are located in $\beta_1\beta_2$ hairpin loop[24, 52, 53, 54, 55]. The double domain swapping mechanism reported in this study can be used to explain higher order oligomer formation in γ Crys where instead of reciprocal domain swapping, secondary structures can be exchanged with neighboring monomers in a chain-like manner.

Conclusions and Outlook

Cataract associated misfolding and aggregation mechanism of eye lens Crys proteins remains largely unknown. Using coarse-grained models and molecular dynamics simulations, we explored the folding pathway and dimerization of H γ D and H γ C Crys. The multi-dimensional FES shows three competing folding pathways with distinct intermediates. In the two dominant pathways, the domains sequentially fold, populating intermediates with either Ntd or Ctd folded. The third, least favored pathway, populates intermediates with partially structured Ntd and Ctd, which can facilitate aggregation from both the domains. Interestingly, fluorescence spectra of chaperone (α B Crys) bound complex of H γ D Crys indicated that both the domains are partially structured in the aggregation prone conformations[44]. All three pathways populate a native-like intermediate state (I^*), with two folded domains and disrupted inter-domain contacts. Under native-like conditions, I^* forms domain swapped dimers by exchanging folded domains with other monomers. Such oligomerization, if takes

place during early stages of unfolding, may lead to hysteresis in the folding experiments. Similar domain swapped structures are reported for human β B2 Crys[49, 50, 51] and mouse γ S Crys.

We have also identified an ensemble of domain swapped dimers from the dimerization simulations performed using unfolded conformations of H γ D and H γ C Crys, which mimic aggregation during refolding. The Ntd swapping units ($\beta_1\beta_2$) for H γ D Crys were already hinted in previous studies[12, 43, 65] and contains the highest number of pathogenic mutations. We also identified swapping units in the Ctds ($\beta_6\beta_7$ in H γ D and $\beta_8\beta_9$ in H γ C Crys), which in conjunction with Ntd swapping units led to double domain swapped dimers. Other less populated dimers involved swapping of different secondary structures from the Ntds and the Ctds, leading to heterogeneity in the dimer structures. The double domain swapping mechanism presented in this study provides structural insights to the early oligomeric assemblies in H γ D and H γ C Crys. Knowledge of the swapping units present in the two domains can be exploited to design preventive measures and essential therapeutics for cataract.

The coarse-grained model used in this study has been successfully applied to elucidate folding[48, 66, 67] and aggregation of other proteins[68, 69, 70]. However, the native-centric nature of this model limits the incorporation of non-native interactions that might play a crucial role in protein aggregation. Due to the low resolution of coarse-grained model, atom level description of protein-protein and solvent-protein interactions are also beyond the scope of this model. In the SOP-SC model used in this work, we have not incorporated features that allow the formation of non-native disulfide bonds. However, such disulfide bonds have been proven to be very important in the formation of crystallin aggregates.

Age related nuclear cataract often results from the oxidative stress built up in aged lens due to gradual depletion in the glutathione (GSH) levels[71, 72]. Due to oxidative stress, Cys residues form disulfide links, which are found commonly in mature cataractous lenses[73, 74, 75, 76]. Consequently, impact of disulfide bonds in aggregation of Crys proteins, especially in Cys rich γ Crys is the sub-

ject of intense investigation[77, 78, 79, 43, 65]. A recent experiment[65] provided evidence for dynamic transfer of disulfide bonds from oxidized WT H γ D Crys to a mutant W42Q promoting aggregation in the latter. Another experimental study[43] in conjunction with simulations identified an internal non-native disulfide bond (Cys32-Cys41), which traps an aggregation prone conformation, essential for oxidative aggregation of mutant variants (W42R, W42Q) of H γ D Crys. The trapped intermediate obtained from the mutant simulations, features a detached N-terminal β -hairpin. This intermediate is structurally similar to the intermediate predicted in AFM experiments[12] of WT H γ D Crys and also consistent with the domain swapped dimer structures predicted in this work for the WT protein. This shows that aggregation prone intermediates, which are predominantly populated by mutants, are transiently populated in WT proteins as well. In our future computational studies, we will probe the role of non-native disulfide bonds in the population of aggregation-prone structures by γ Crys proteins using coarse-grained protein models[80, 81], which allow the formation of such bonds.

Materials and Methods

Molecular dynamics simulations are performed using coarse-grained self-organized polymer side-chain (SOP-SC)[47, 48] description of the human γ Crys proteins. Coarse-grained structures of the proteins are built from their PDB structures with PDB IDs 2NBR[45] (H γ C Crys) and 1HK0[46] (H γ D Crys). The Hamiltonian of a protein conformation in the SOP-SC model represented by a set of coordinates $\{\mathbf{r}\}$ is,

$$E_{CG}(\{\mathbf{r}\}) = E_B + E_{NB}^N + E_{NB}^{NN} + E_{el}, \quad (1)$$

where E_B is the energy due to bonded interaction present between the covalently bonded beads, E_{NB}^N and E_{NB}^{NN} are the energy due to non-bonded native and non-native interactions, respectively, and E_{el} is the energy due to electrostatic

interactions present between the charged residues in the protein. Native interactions exist between two beads, if they are separated by at least three bonds and are within a cut-off distance ($R_c = 8 \text{ \AA}$) in the SOP-SC model of the PDB structure while any other non-covalent interactions are considered as non-native. Symmetric Go potential[32, 63] (Figure S9) is used to model the interactions between two protein monomers which is discussed in details in the supporting information (SI). Langevin dynamics simulations are used to compute equilibrium thermodynamic properties. Weighted histogram analysis method is used to construct the three-dimensional free energy landscape $F(\chi_{Ntd}, \chi_{Ctd}, \chi_{Int})$ from the Langevin dynamics simulations performed at various temperatures. Brownian dynamics[82] simulations are performed to study dimerization kinetics of the proteins. A detailed description of the protein models, simulation methodology and parameters are provided in the SI.

Acknowledgments

G.R. acknowledges funding from the National Supercomputing Mission (NSM) through the grant MeitY/R&D/HPC/2(1)/2014. J.N. thanks DST-SERB for funding under Ramanujan Faculty Fellowship. B.M. acknowledges research fellowship from Indian Institute of Science-Bangalore. The computations are performed using the TUE and Cray XC40 cluster at IISc.

References

- [1] K. L. Moreau, J. A. King, Protein misfolding and aggregation in cataract disease and prospects for prevention, *Trends Mol. Med* 18 (5) (2012) 273–282.
- [2] K. Lampi, Z. Ma, M. Shih, T. Shearer, J. Smith, D. Smith, L. David, Sequence analysis of β A3, β B3, and β A4 crystallins completes the identification of the major proteins in young human lens, *J. Biol. Chem.* 272 (4) (1997) 2268–2275.
- [3] N. E. Robinson, K. J. Lampi, J. P. Speir, G. Kruppa, M. Easterling, A. B. Robinson, Quantitative measurement of young human eye lens crystallins by direct injection Fourier transform ion cyclotron resonance mass spectrometry, *Mol. Vis.* 12 (78-79) (2006) 704–711.
- [4] J. Horwitz, α -crystallin can function as a molecular chaperone, *Proc. Natl. Acad. Sci. U. S. A.* 89 (21) (1992) 10449–10453.
- [5] M. Delaye, A. Tardieu, Short-range order of crystallin proteins accounts for eye lens transparency, *Nature* 302 (5907) (1983) 415–417.
- [6] A. Stradner, G. Foffi, N. Dorsaz, G. Thurston, P. Schurtenberger, New insight into cataract formation: enhanced stability through mutual attraction, *Phys. Rev. Lett.* 99 (19) (2007) 198103,1–4.
- [7] R. Siezen, J. Thomson, E. Kaplan, G. Benedek, Human lens γ -crystallins - isolation, identification, and characterization of the expressed gene-products, *Proc. Natl. Acad. Sci. U. S. A.* 84 (17) (1987) 6088–6092.
- [8] C. Slingsby, N. Clout, Structure of the crystallins, *Eye* 13 (3B) (1999) 395–402.
- [9] S. Flaugh, M. Kosinski-Collins, J. King, Interdomain side-chain interactions in human γ D crystallin influencing folding and stability, *Protein Sci.* 14 (8) (2005) 2030–2043.

- [10] M. Kosinski-Collins, J. King, In vitro unfolding, refolding, and polymerization of human γ D crystallin, a protein involved in cataract formation, *Protein Sci.* 12 (3) (2003) 480–490.
- [11] I. M. Henry, M. Kosinski-Collins, S. Thol, E. Serebryany, J. A. King, Kinetic stability of long-lived human γ D and γ S lens crystallins, derived in part from their domain interfaces, may protect against cataract, *Biophys. J.* 116 (3, 1) (2019) 321A–322A.
- [12] S. Garcia-Manyes, D. Giganti, C. L. Badilla, A. Lezamiz, J. Perales-Calvo, A. E. M. Beedle, J. M. Fernandez, Single-molecule force spectroscopy predicts a misfolded, domain-swapped conformation in human γ D-crystallin protein, *J. Biol. Chem.* 291 (8) (2016) 4226–4235.
- [13] P. Das, J. A. King, R. Zhou, β -strand interactions at the domain interface critical for the stability of human lens γ D-crystallin, *Protein Sci.* 19 (1) (2010) 131–140.
- [14] P. Das, J. A. King, R. Zhou, Aggregation of γ -crystallins associated with human cataracts via domain swapping at the C-terminal β -strands, *Proc. Natl. Acad. Sci. U. S. A.* 108 (26) (2011) 10514–10519.
- [15] M. Sikora, M. Cieplak, Mechanical stability of multidomain proteins and novel mechanical clamps, *Proteins* 79 (6) (2011) 1786–1799.
- [16] K. Itoh, M. Sasai, Cooperativity, connectivity, and folding pathways of multidomain proteins, *Proc. Natl. Acad. Sci. U. S. A.* 105 (37) (2008) 13865–13870.
- [17] S. Batey, A. A. Nickson, J. Clarke, Studying the folding of multidomain proteins, *HFSP J.* 2 (6) (2008) 365–377.
- [18] K. Itoh, M. Sasai, Multidimensional theory of protein folding, *J. Chem. Phys.* 130 (14) (2009) 145104,1–21.

- [19] N. Nagarajan, G. Kumar, R. Sowdhamini, N. Srinivasan, S. Vishwanath, Fold combinations in multi-domain proteins, *Bioinformation* 15 (5) (2019) 342–350.
- [20] L. Fu, J. Liang, Unfolding of human lens recombinant β B2- and γ C-crystallins, *J. Struct. Biol.* 139 (3) (2002) 191–198.
- [21] L. Fu, J. Liang, Spectroscopic analysis of lens recombinant β B2- and γ C-crystallin, *Mol. Vis.* 7 (25) (2001) 178–183.
- [22] M. Kosinski-Collins, S. Flaugh, J. King, Probing folding and fluorescence quenching in human γ D crystallin greek key domains using triple tryptophan mutant proteins, *Protein Sci.* 13 (8) (2004) 2223–2235.
- [23] S. Flaugh, M. Kosinski-Collins, J. King, Contributions of hydrophobic domain interface interactions to the folding and stability of human γ D-crystallin, *Protein Sci.* 14 (3) (2005) 571–581.
- [24] E. Serebryany, J. A. King, The $\beta\gamma$ -crystallins: Native state stability and pathways to aggregation, *Prog. Biophys. Mol. Biol.* 115 (1) (2014) 32–41.
- [25] E. Sahin, J. L. Jordan, M. L. Spataro, A. Naranjo, J. A. Costanzo, W. F. Weiss, A. S. Robinson, E. J. Fernandez, C. J. Roberts, Computational design and biophysical characterization of aggregation-resistant point mutations for γ D crystallin illustrate a balance of conformational stability and intrinsic aggregation propensity, *Biochemistry* 50 (5) (2011) 628–639.
- [26] S. L. Flaugh, I. A. Mills, J. King, Glutamine deamidation destabilizes human γ D-crystallin and lowers the kinetic barrier to unfolding, *J. Biol. Chem.* 281 (41) (2006) 30782–30793.
- [27] I. A. Mills, S. L. Flaugh, M. S. Kosinski-Collins, J. A. King, Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γ D crystallin and γ S crystallin, *Protein Sci.* 16 (11) (2007) 2427–2444.

- [28] M. Bennett, M. Schlunegger, D. Eisenberg, 3D domain swapping - a mechanism for oligomer assembly, *Protein Sci.* 4 (12) (1995) 2455–2468.
- [29] Y. Liu, D. Eisenberg, 3D domain swapping: as domains continue to swap, *Protein Sci.* 11 (6) (2002) 1285–1299.
- [30] N. M. Mascarenhas, S. Gosavi, Protein domain-swapping can be a consequence of functional residues, *J. Phys. Chem. B* 120 (28) (2016) 6929–6938.
- [31] N. M. Mascarenhas, S. Gosavi, Understanding protein domain-swapping using structure-based models of protein folding, *Prog. Biophys. Mol. Biol.* 128 (2017) 113–120.
- [32] S. Yang, S. Cho, Y. Levy, M. Cheung, H. Levine, P. Wolynes, J. Onuchic, Domain swapping is a consequence of minimal frustration, *Proc. Natl. Acad. Sci. U. S. A.* 101 (38) (2004) 13786–13791.
- [33] P. C. A. van der Wel, Domain swapping and amyloid fibril conformation, *Prion* 6 (3) (2012) 211–216.
- [34] K. Knaus, M. Morillas, W. Swietnicki, M. Malone, W. Surewicz, V. Yee, Crystal structure of the human prion protein reveals a mechanism for oligomerization, *Nat. Struct. Biol.* 8 (9) (2001) 770–774.
- [35] R. Janowski, M. Kozak, M. Abrahamson, A. Grubb, M. Jaskolski, 3D domain-swapped human Cystatin C with amyloidlike intermolecular β -sheets, *Proteins* 61 (3) (2005) 570–578.
- [36] M. Nilsson, X. Wang, S. Rodziejewicz-Motowidlo, R. Janowski, V. Lindstrom, P. Onnerfjord, G. Westermarck, Z. Grzonka, M. Jaskolski, A. Grubb, Prevention of domain swapping inhibits dimerization and amyloid fibril formation of cystatin C - use of engineered disulfide bridges, antibodies, and carboxymethylpapain to stabilize the monomeric form of cystatin C, *J. Biol. Chem.* 279 (23) (2004) 24236–24245.

- [37] J. Bacarizo, S. Martinez-Rodriguez, J. Manuel Martin-Garcia, M. Andujar-Sanchez, E. Ortiz-Salmeron, J. Luis Neira, A. Camara-Artigas, Electrostatic effects in the folding of the SH3 domain of the c-Src Tyrosine Kinase: pH-dependence in 3D-domain swapping and amyloid formation, *PLoS One* 9 (12) (2014) e113224.
- [38] C. Liu, M. R. Sawaya, D. Eisenberg, β 2-microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages, *Nat. Struct. Mol. Biol.* 18 (1) (2011) 49–55.
- [39] A. Pica, A. Merlino, A. K. Buell, T. P. J. Knowles, E. Pizzo, G. D'Alessio, F. Sica, L. Mazzarella, Three-dimensional domain swapping and supramolecular protein assembly: insights from the X-ray structure of a dimeric swapped variant of human pancreatic RNase, *Acta Crystallogr. Sect. D-Struct. Biol.* 69 (10) (2013) 2116–2123.
- [40] C. Ercole, J. Pedro Lopez-Alonso, J. Font, M. Ribo, M. Vilanova, D. Picone, D. V. Laurents, Crowding agents and osmolytes provide insight into the formation and dissociation of RNase A oligomers, *Arch. Biochem. Biophys.* 506 (2) (2011) 123–129.
- [41] Y. Liu, G. Gotte, M. Libonati, D. Eisenberg, A domain-swapped RNase A dimer with implications for amyloid formation, *Nat. Struct. Biol.* 8 (3) (2001) 211–214.
- [42] F. Rousseau, J. Schymkowitz, H. Wilkinson, L. Itzhaki, Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues, *Proc. Natl. Acad. Sci. U. S. A.* 98 (10) (2001) 5596–5601.
- [43] E. Serebryany, J. C. Woodard, B. V. Adkar, M. Shabab, J. A. King, E. I. Shakhnovich, An internal disulfide locks a misfolded aggregation-prone intermediate in cataract-linked mutants of human γ D-crystallin, *J. Biol. Chem.* 291 (36) (2016) 19172–19183.

- [44] L. Acosta-Sampson, J. King, Partially folded aggregation intermediates of human γ D-, γ C-, and γ S-Crystallin are recognized and bound by human α B-crystallin chaperone, *J. Mol. Biol.* 401 (1) (2010) 134–152.
- [45] K. Dixit, A. Pande, J. Pande, S. P. Sarma, Nuclear magnetic resonance structure of a major lens protein, human γ C-crystallin: role of the dipole moment in protein solubility, *Biochemistry* 55 (22) (2016) 3136–3149.
- [46] A. Basak, O. Bateman, C. Slingsby, A. Pande, N. Asherie, O. Ogun, G. Benedek, J. Pande, High-resolution X-ray crystal structures of human γ D crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract, *J. Mol. Biol.* 328 (5) (2003) 1137–1147.
- [47] C. Hyeon, R. I. Dima, D. Thirumalai, Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins, *Structure* 14 (2006) 1633–1645.
- [48] Z. Liu, G. Reddy, E. P. O’Brien, D. Thirumalai, Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins, *Proc. Natl. Acad. Sci. U. S. A.* 108 (19) (2011) 7787–7792.
- [49] B. Bax, R. Lapatto, V. Nalini, H. Driessen, P. Lindley, D. Mahadevan, T. Blundell, C. Slingsby, X-ray-analysis of β B2-crystallin and evolution of oligomeric lens proteins, *Nature* 347 (6295) (1990) 776–780.
- [50] R. Lapatto, V. Nalini, B. Bax, H. Driessen, P. Lindley, T. Blundell, C. Slingsby, High-resolution structure of an oligomeric eye lens β -crystallin - loops, arches, linkers and interfaces in β B2 dimer compared to a monomeric γ -crystallin, *J. Mol. Biol.* 222 (4) (1991) 1067–1083.
- [51] M. A. Smith, O. A. Bateman, R. Jaenicke, C. Slingsby, Mutation of interfaces in domain-swapped human β B2-crystallin, *Protein Sci.* 16 (4) (2007) 615–625.
- [52] J. Jung, I. J. L. Byeon, Y. Wang, J. King, A. M. Gronenborn, The structure of the cataract-causing P23T mutant of human γ D-crystallin exhibits dis-

- tinctive local conformational and dynamic changes, *Biochemistry* 48 (12) (2009) 2597–2609.
- [53] A. Pande, O. Annunziata, N. Asherie, O. Ogun, G. Benedek, J. Pande, Decrease in protein solubility and cataract formation caused by the Pro23 to Thr mutation in human γ D-crystallin, *Biochemistry* 44 (7) (2005) 2491–2500.
- [54] A. Pande, J. Pande, N. Asherie, A. Lomakin, O. Ogun, J. King, N. Lubsen, D. Walton, G. Benedek, Molecular basis of a progressive juvenile-onset hereditary cataract, *Proc. Natl. Acad. Sci. U. S. A.* 97 (5) (2000) 1993–1998.
- [55] S. Santhiya, M. Manohar, D. Rawley, P. Vijayalakshmi, P. Namperumalsamy, P. Gopinath, J. Loster, J. Graw, Novel mutations in the γ -crystallin genes cause autosomal dominant congenital cataracts, *J. Med. Genet.* 39 (5) (2002) 352–358.
- [56] M. J. Bennett, M. R. Sawaya, D. Eisenberg, Deposition diseases and 3D domain swapping, *Structure* 14 (5) (2006) 811–824.
- [57] X. Q. Li, H. C. Cai, S. Y. Zhou, J. H. Yang, Y. B. Xi, X. B. Gao, W. J. Zhao, P. Li, G. Y. Zhao, Y. Tong, F. C. Bao, Y. Ma, S. Wang, Y. B. Yan, C. L. Lu, X. Ma, A novel mutation impairing the tertiary structure and stability of γ C-crystallin (CRYGC) leads to cataract formation in humans and zebrafish lens, *Hum. Mutat.* 33 (2) (2012) 391–401.
- [58] S. Kumar, R. J.M., D. Bouzida, R. Swendsen, P. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules., *J. Comput. Chem.* 13 (1992) 1011–1021.
- [59] E. K. Wong, V. Prytkova, J. A. Freites, C. T. Butts, D. J. Tobias, Molecular mechanism of aggregation of the cataract-related γ D-crystallin W42R variant from multiscale atomistic simulations, *Biochemistry* 58 (35) (2019) 3691–3699.

- [60] A. Pande, K. S. Ghosh, P. R. Banerjee, J. Pande, Increase in surface hydrophobicity of the cataract-associated P23T mutant of human γ D-crystallin is responsible for its dramatically lower, retrograde solubility, *Biochemistry* 49 (29) (2010) 6122–6129.
- [61] A. Pande, J. Zhang, P. R. Banerjee, S. S. Puttamadappa, A. Shekhtman, J. Pande, NMR study of the cataract-linked P23T mutant of human γ D-crystallin shows minor changes in hydrophobic patches that reflect its retrograde solubility, *Biochem. Biophys. Res. Commun.* 382 (1) (2009) 196–199.
- [62] S. Hanson, A. Hasan, D. Smith, J. Smith, The major in vivo modifications of the human water-insoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage, *Exp. Eye Res.* 71 (2) (2000) 195–207.
- [63] F. Ding, N. Dokholyan, S. Buldyrev, H. Stanley, E. Shakhnovich, Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism, *J. Mol. Biol.* 324 (4) (2002) 851–857.
- [64] J. C. Boatz, M. J. Whitley, M. Li, A. M. Gronenborn, P. C. A. van der Wel, Cataract-associated P23T γ D-crystallin retains a native-like fold in amorphous-looking aggregates formed at physiological pH, *Nat. Commun.* 8 (15137) (2017) 1–10.
- [65] E. Serebryany, S. Yu, S. A. Trauger, B. Budnik, E. I. Shakhnovich, Dynamic disulfide exchange in a crystallin protein in the human eye lens promotes cataract-associated aggregation, *J. Biol. Chem.* 293 (46) (2018) 17997–18009.
- [66] G. Reddy, Z. Liu, D. Thirumalai, Denaturant-dependent folding of GFP, *Proc. Natl. Acad. Sci. U. S. A.* 109 (44) (2012) 17832–17838.
- [67] G. Reddy, D. Thirumalai, Dissecting ubiquitin folding using the self-organized polymer model, *J. Phys. Chem. B* 119 (34) (2015) 11358–11370.

- [68] P. I. Zhuravlev, G. Reddy, J. E. Straub, D. Thirumalai, Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins, *J. Mol. Biol.* 426 (14) (2014) 2653–2666.
- [69] B. Mondal, G. Reddy, Cosolvent effects on the growth of protein aggregates formed by a single domain globular protein and an intrinsically disordered protein, *J. Phys. Chem. B* 123 (9) (2019) 1950–1960.
- [70] B. Mondal, G. Reddy, A transient intermediate populated in prion folding leads to domain swapping, *Biochemistry* 59 (1) (2020) 114–124.
- [71] F. Giblin, Glutathione: a vital lens antioxidant, *J. Ocular Pharmacol. Ther.* 16 (2) (2000) 121–135.
- [72] V. M. Berthoud, E. C. Beyer, Oxidative stress, lens gap junctions, and cataracts, *Antioxid. Redox Signal.* 11 (2) (2009) 339–353.
- [73] P. G. Hains, R. J. W. Truscott, Post-translational modifications in the nuclear region of young, aged, and cataract human lenses, *J. Proteome Res.* 6 (10) (2007) 3935–3943.
- [74] R. Michael, A. J. Bron, The ageing lens and cataract: a model of normal and pathological ageing, *Philos. Trans. R. Soc. B-Biol. Sci.* 366 (1568) (2011) 1278–1292.
- [75] P. G. Hains, R. J. W. Truscott, Proteomic analysis of the oxidation of cysteine residues in human age-related nuclear cataract lenses, *BBA-PROTEINS PROTEOM* 1784 (12) (2008) 1959–1964.
- [76] A. Spector, D. Roy, Disulfide-linked high molecular-weight protein associated with human cataract, *Proc. Natl. Acad. Sci. U. S. A.* 75 (7) (1978) 3244–3248.
- [77] X. Fan, S. Zhou, B. Wang, G. Hom, M. Guo, B. Li, J. Yang, D. Vaysburg, V. M. Monnier, Evidence of highly conserved β crystallin disulfidome that can be mimicked by in vitro oxidation in age-related human cataract and

- glutathione depleted mouse lens, *Mol. Cell. Proteomics* 14 (12) (2015) 3211–3223.
- [78] M. Francis Simpanya, R. R. Ansari, V. Leverenz, F. J. Giblin, Measurement of lens protein aggregation in vivo using dynamic light scattering in a guinea pig/uva model for nuclear cataract, *Photochem. Photobiol.* 84 (6) (2008) 1589–1595.
- [79] D. C. Thorn, A. B. Grosas, P. D. Mabbitt, N. J. Ray, C. J. Jackson, J. A. Carver, The structure and stability of the disulfide-linked γ S-crystallin dimer provide insight into oxidation products associated with lens cataract formation, *J. Mol. Biol.* 431 (3) (2019) 483–497.
- [80] A. N. Muttathukattil, P. C. Singh, G. Reddy, Role of disulfide bonds and topological frustration in the kinetic partitioning of lysozyme folding pathways, *J. Phys. Chem. B* 123 (15) (2019) 3232–3241.
- [81] M. Qin, W. Wang, D. Thirumalai, Protein folding guides disulfide bond formation, *Proc. Natl. Acad. Sci. U. S. A.* 112 (36) (2015) 11241–11246.
- [82] D. Ermak, J. Mccammon, Brownian dynamics with hydrodynamic interactions, *J. Chem. Phys.* 69 (4) (1978) 1352–1360.

Figure 1: (A) Contact map of H7D Crys[46] (PDB ID: 1HK0). H7D Crys consists of two domains called N-terminal domain (Ntd, shown in orange) and C-terminal domain (Ctd, shown in red) connected by a short peptide linker. Each domain is further composed of two intercalated greek key motifs. Greek key motifs constituting the Ntd are termed as M_1 (residues Gly1 to Ser39) and M_2 (residues Gly40 to Pro82), while motifs constituting the Ctd are termed as M_3 (residues His88 to Glu128) and M_4 (residues Gly129 to Ile171). The intra domain contacts are colored orange (Ntd) and red (Ctd), while inter domain contacts are colored black. (B) A splay diagram of four ideal interconnected greek key motifs. Four antiparallel β -strands constitute one greek key motif where strands β_1 and β_3 are in contact. (C) Temperature dependent variation in specific heat capacity (C_v) and average potential energy ($\langle E \rangle$) indicate a two state folding transition in H7D Crys. (D) FES, $F(\chi_{Ntd}, \chi_{Ctd}, \chi_{Int})$ projected onto fraction of native contacts in Ntd (χ_{Ntd}), in Ctd (χ_{Ctd}) and between the two domains (χ_{Int}) at $T_M = 426$ K. Three pathways connect the native state (N) and the unfolded state (U) through distinct intermediate states. Representative structures of the intermediates are shown. (E) $F(\chi_{Ntd}, \chi_{Ctd})$ projected onto χ_{Ntd} and χ_{Ctd} . All the intermediate states are visible except for I^* (Ctd and Ntd domains are folded with disrupted inter domain contacts), which is indistinguishable from the N state.

Figure 2: (A) Contact map of H γ C Crys[46] (PDB ID: 2NBR)[45]. H γ C Crys consists of two domains called N-terminal domain (Ntd, shown in green) and C-terminal domain (Ctd, shown in blue) connected by a short peptide linker. Each domain is further composed of two intercalated greek key motifs. Greek key motifs constituting the Ntd are termed as M_1 (residues Gly1 to Ser39) and M_2 (residues Gly40 to Pro82), while motifs constituting the Ctd are termed as M_3 (residues His87 to Glu127) and M_4 (residues Gly128 to Val170). The intra domain contacts are colored green (Ntd) and blue (Ctd), while inter domain contacts are colored black. (B) Temperature dependent variation in C_v and $\langle E \rangle$ indicate a three state folding transition in H γ C Crys. (C) $F(\chi_{Ntd}, \chi_{Ctd}, \chi_{Int})$ projected onto χ_{Ntd} , χ_{Ctd} and χ_{Int} at $T_M^I = 410$ K. Three pathways connect the N state and U state through distinct intermediate states. Representative structures of the intermediates are shown. (D) $F(\chi_{Ntd}, \chi_{Ctd})$ projected onto χ_{Ntd} and χ_{Ctd} at T_M^I . All the intermediate states are visible except for I^* (Ctd and Ntd domains are folded with disrupted inter domain contacts), which is indistinguishable from the N state.

Figure 3: Dimerization from the native states of H γ D and H γ C Crys. (A, B) Time evolution of the fraction of secondary structural contacts (f_{ss}) present in the two monomers A and B, and the total number of native contacts Q_{AB} formed between A and B. Dimerization takes place in two steps. In the first step, only one domain gets swapped ($\text{Dim}_{\text{int}, \gamma D}$, $\text{Dim}_{\text{int}, \gamma C}$) while in the second step both the domains are swapped ($\text{Dim}_{\gamma D}$, $\text{Dim}_{\gamma C}$). Mechanism of dimerization in H γ D (C) and H γ C Crys (D). An intermediate state I^* with disrupted inter domain contacts get populated in the initial stages of dimerization. I^* further swaps the Ntd and Ctd with neighboring monomers in two stages leading to domain swapped dimers.

Figure 4: Multistep dimerization in H7D Crys while refolding at $T = 300$ K. Secondary structural elements from Ntd ($\beta_1\beta_2$) and Ctd ($\beta_6\beta_7$) get swapped leading to a double domain swapped dimer. (A) Time evolution of fraction of secondary structural contacts between $\beta_1\beta_2$ strands in monomer A, B ($\chi^A_{\beta_1\beta_2}$, $\chi^B_{\beta_1\beta_2}$) and between β_1 strand of monomer A and β'_3 strand of monomer B ($\chi_{\beta_1\beta'_3}$) and vice versa ($\chi_{\beta_1\beta'_3}$) along with the number of contacts present between monomer A and B (Q_{AB}). (B) Time evolution of fraction of secondary structural contacts between $\beta_6\beta_7$ strands in monomer A and B ($\chi^A_{\beta_6\beta_7}$, $\chi^B_{\beta_6\beta_7}$) and between β_6 strand of monomer A and β'_8 strand of monomer B ($\chi_{\beta_6\beta'_8}$) and vice versa ($\chi_{\beta_6\beta'_8}$). (C) Contact map of the final domain swapped dimer structure. β_1 strand makes contacts with β'_3 strand of the other monomer and vice versa after swapping of $\beta_1\beta_2$ strands. The first hairpin loop of M_3 ($\beta_6\beta_7$) get swapped between the Ctds where β_6 strand forms contact with β'_8 strand from the other monomer and vice versa.

XXX

Graphical Abstract figure