

Replica Exchange Molecular Dynamics Simulations Reveal Self-association Sites in M- Crystallin Caused by Mutations Provide Insights of Cataract

Sunita Patel (✉ sunita.iitb@gmail.com)

UM-DAE Centre for Excellence in Basic Sciences

Ramakrishna V. Hosur

UM-DAE Centre for Excellence in Basic Sciences

Research Article

Keywords: crystallins, cataracts, eye lens crystallin

Posted Date: June 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-644182/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on December 1st, 2021. See the published version at <https://doi.org/10.1038/s41598-021-02728-8>.

Abstract

Crystallins are ubiquitous, however, prevalence is seen in eye lens. Eye lens crystallins are long-lived and structural intactness is required for maintaining lens transparency and protein solubility. Mutations in crystallin often lead to cataract. In this study, we performed mutations at specific sites of M-crystallin, a close homologue of eye lens crystallin and studied by employing replica exchange molecular dynamics with generalized Born solvation model. Mutations were made on the Ca^{2+} binding residues (K34D and S77D) and in the hydrophobic core (W45R) which is known to cause congenital cataract in homologous γD -crystallin. The chosen mutations caused large motion of the N-terminal Greek key, concomitantly break the interlocking Greek keys interactions and perturbed the compact core resulting in several folded and partially unfolded states. Partially unfolded states expose large hydrophobic patches that can act as precursors for self-aggregation. Accumulation of such aggregates is the potential cause of cataract in homologous crystallins.

Introduction

Cataract is the leading cause of blindness that affects millions of people worldwide. It is manifested by opacity in eye lens which can happen due to mutation in the crystallins leading to congenital cataract or due to post-translational modification, oxidative stress, exposure to UV radiation, heat and specific metabolites causing age related cataract. As of 2019, WHO reported 2.2 billion people around the world to have vision impairment of which 1 billion of cases could have been prevented by medical intervention (<https://www.who.int/publications-detail/world-report-on-vision>)^{1,2}. The dominant protein crystallins in the eye lens constitute 90% of the total soluble proteins². High concentration of soluble crystallins in the cytoplasm of the lens fiber cells provide transparency and high refractive index to the eye-lens^{3,4}. Cataract is caused due to aggregation of eye lens crystallins which causes scattering of the light and prevent it from reaching the retina⁴⁻⁶. In several studies the molecular insights of the cause of cataract is elusive.

Although, crystallins are wide spread in microbes and vertebrates, they have certain unique property which is different in these two organism types. The primitive archaeal and microbial crystallins possesses Ca^{2+} binding canonical motifs and are stabilized further by binding to Ca^{2+} ^{7,8} while vertebrate eye-lens crystallins impart greater stability by lowering its affinity for Ca^{2+} ^{9,10}. The structural stability of microbial crystallins can be enhanced by mutating Ser to Arg at 5th position of the canonical motifs while eye lens γD - and γS -crystallins have natural Arg at that place⁹⁻¹². Similar mutation from Ser to Arg in the canonical motif even stabilizes an intrinsically disordered $\beta\gamma$ -crystallin, Hahellin in the absence of Ca^{2+} ¹³. This signify importance of distinct position on the crystallin domain which is responsible for its unusual stability^{9,14}. Mutation(s) in these regions can cause local and/or global change in the structure which in turn can cause aggregation of crystallins^{14,15}.

There are three common families of crystallins in the vertebrate eye lens such as α -, β - and γ -crystallins¹⁶. β - and γ -crystallins form a separate superfamily by having structural similarity and evolutionary relation while α -crystallin acts as chaperone. γ -crystallins are the smallest and simplest member of $\beta\gamma$ -crystallin superfamily which exist as monomer while β -crystallins exist as oligomers. A typical $\beta\gamma$ -crystallin domain consists of two β -sheets formed from eight β -strands arranged in two Greek key motifs. The arrangement of two Greek key motifs takes place in such a way that three of the four strands of a Greek key motif form one β -sheet and the remaining one β -strand pairs with three β -strands from other Greek key motif forming the second β -sheet⁸.

The 3D structure of an archaeal protein, M-crystallin from *Methanosarcina acetivorans* has been solved by X-ray crystallography (PDB ID:3HZ2)¹⁷ as well as by NMR spectroscopy (PDB ID:2K1W)¹⁸. M-crystallin consists of 84 residues and belongs to bg-crystallin superfamily by having two Greek key motifs which fold into a single domain. The primary structure of M-crystallin shows presence of Trp corner and the signature sequence Y/FxxxxF/YxG for two Greek key motifs^{17,18}. M-crystallin displays striking structural similarity with human eye lens bg-crystallins although it is a single domain protein while eye lens crystallins are double domain^{17,18}. Therefore, we consider this structural homologues as a simplistic model to understand the mechanistic insights of cataract. In this study, we performed mutation of the Ca^{2+} binding residues (K34D and S77D) and a hydrophobic core residue, W45R which is known to cause congenital cataract in homologous eye lens, γ D-crystallin and studied by replica exchange molecular dynamics (REMD) simulation using generalized Born (GB) solvation model. The study led us to understand the underlying mechanism of $\beta\gamma$ -crystallin relation to aggregation upon mutation.

Methods

Starting Structure

M-crystallin (PDB id: 3HZ2) from *Mithanosarsina acetavoran* is used to build starting structures of the mutants¹⁷. M-crystallin is a well folded protein and retains its bg-crystallin fold both in presence and absence of Ca^{2+} ¹⁸. Simulations were performed without Ca^{2+} to monitor the effect of mutations. Single mutant of M-crystallin was generated by mutating W45 to R45 and is abbreviated as M-crytallin-SM. Double mutant of M-crytallin was made by mutating K34D and S77D in the Ca^{2+} binding canonical motifs of M-crystallin and is abbreviated as M-crystallin-DM. In the similar way, triple mutant of M-crystallin was made by mutating K34D, W45R and S77D and is referred as M-crytallin-TM. These mutations were made using PyMOL software¹⁹.

Replica exchange molecular dynamics simulation

REMD simulation is an enhance conformational sampling method which was performed for the wild type M-crystallin (M-crystallin-WT) and its three mutants (M-crystallin-SM, M-crystallin-DM and M-crystallin-TM). REMD is an enhance technique because it lowers the free energy barrier at higher temperature by sampling high energy conformations which otherwise might not be accessible at moderate temperature.

In the REMD simulations, simultaneously multiple parallel runs were started at predefined temperatures which were derived from an exponential distribution. The exchange rate were maintained constant across all the replicas by keeping relatively more temperature gap at higher temperatures and lesser gap at lower temperatures. Likewise we considered sixteen replica temperatures spanning from 280 and 340 K which are 281.85, 285.39, 288.98, 292.60, 296.28, 300.00, 303.77, 307.58, 311.44, 315.36, 319.32, 323.33, 327.39, 331.50, 335.66 and 339.88 K. The replica exchange probability was 60% for all the REMD simulations. The conformational coordinates were exchanged between the neighboring replicas following the Metropolis algorithm²⁰. This algorithm was evaluated at every 2 ps in order to facilitate the exchange of coordinates. When the Metropolis algorithm was satisfied, the exchange attempt was considered as successful and the coordinates between the neighbouring replicas were exchanged. The velocity of each atom was then rescaled to the changed target replica temperature and that's how REMD simulations were carried out.

Simulation details

The REMD simulations were run using AMBER 14 molecular modeling package²¹ and AMBER FF99SB force field^{22,23}. Generalized Born Onufriev, Bashford and Case implicit solvent model (GB-OBC) with *igb=5* option was used to treat solvent²⁴. The advantage of using implicit solvent is that the system size and solvent viscosity is significantly reduced due to exclusion of physical water molecules making the conformational sampling faster²⁵⁻²⁹. Certain GB implicit solvent model although associated with limitation³⁰, combined use of GB implicit solvent model with REMD simulations provides an optimum balance between speed and accuracy and the results obtained are in agreement with experimental findings³¹⁻³⁴. All the simulations were done under NVT ensembles³⁵. The force-field parameters, topology and coordinates were generated using *tLeap* program of AMBER²¹. The side-chains of the amino acid residues which are polar and charged were adjusted to physiological pH. Energy minimization of the starting structures were carried out for 2000 cycles of which steepest decent energy minimization was performed for the first 1000 cycles, followed by conjugate gradient minimization for next 1000 cycles to minimize any atomic overlap. The possibility of unwanted rotation at high temperature was desisted by generating chirality constraints on the minimized structure. SHAKE algorithm was used to constrain bonds stretching freedom which involves hydrogen atoms³⁶. The non-bonded van der Waals potential cut-off was kept at 16 Å. Langevin thermostat was used to maintain the replica temperature by weak coupling with a collision frequency of 1 ps⁻¹. Equilibration molecular dynamic run was performed on the system for 200 ps. During this the temperature of each replica was increased gradually from 0 to the corresponding target temperature. Following equilibration, REMD simulation was started for the sixteen systems with 2 fs as integration time step. The exchange attempts between the neighboring replicas were made at every 2 ps. The REMD simulation were performed using *Verlet* integration algorithm. *Multisander* program of AMBER molecular modelling package was used to run the simulations²¹. A time step of 2 ps was used to write the coordinate and output files. The trajectory corresponding to each replica temperature was filtered using *cpptraj* program and backbone Rg and C^a

RMSD analyses were determined for all (Figure S1)²¹. The trajectory corresponding to the physiological temperature of 300 K was used for the data analysis. The REMD simulations were performed in the high performance computing facility instituted at Tata Institute of Fundamental Research, Hyderabad and the National PARAM Supercomputing Facility (NPSF) of C-DAC.

Convergence

REMD simulations were performed for M-crystallin-WT, M-crystallin-SM, M-crystallin-DM and M-crystallin-TM for a total 0.8 μ s. Backbone Rg and C ^{α} RMSD were monitored for all trajectories corresponding to each temperature show excellent convergence as no drastic change has been observed after about 70 ns suggesting acquisition of equilibration (Figure S1 and Figure S2). Additionally, we performed convergence measurements following Sawle and Ghosh³⁷ where number of clusters and cluster entropy were estimated as a function of time to ensure adequate conformational sampling. The cluster entropy was determined using $-\sum P_j \log(P_j)$, where P_j is the probability of j^{th} observed cluster. The C ^{α} RMSD obtained at 300 K from each simulation shows a clear sampling of steady value after 70 ns till 200 ns therefore 70-200 ns was considered as equilibrated region for further analysis. The stretch was then divided into 20 ns overlapping time segments. Number of clusters and the distribution of cluster entropy were determined for each time segments and plotted as shown in Figure S2. For clustering the structures, GROMOS clustering algorithm³⁸ with a cut-off of 5 Å was used. The number of clusters as a function of time showed a marginal increase for M-crystallin-TM while for the rest of the simulations, it was almost constant. Further, the distribution of cluster entropy was steady for all the REMD simulations indicating convergence. Therefore, the trajectories corresponding to 70 to 200 ns at 300 K were used for further analyses.

Analyses

For the analysis, we monitored several conformational parameters such as radius of gyration (Rg) of backbone atoms, root mean square deviation (RMSD) of C ^{α} atoms with respect to X-ray structure of M-crystallin (3HZ2), C ^{α} atom's root-mean-square-fluctuation (RMSF), network cluster layout, native contact analysis, distribution of surface charge potential, principal component analysis, absolute entropy calculated for all the simulations following Schlitter's method³⁹ which make use the covariance matrix of C ^{α} atomic fluctuations, hydrophobic solvent accessible surface area per residue (rhSASA), ionic interaction and hydrogen bond following the criteria defined by Visual Molecular Dynamics (VMD)⁴⁰. Native contacts were determined by an in-house program where C ^{α} -C ^{α} atoms in X-ray structure of M-crystallin separated at least by 3 residues (non-local contact) come closer than 6.5 Å were considered as reference native contacts and how many of these contacts were present in a given conformation represents the number of native contacts. The percentage of native contacts was computed by taking average over a given cluster. The ionic interactions were calculated by determining all the possible distances between the positive (NE, NH1, NH2) and negative charged (OD1 and OD2) sidechain atoms of the interacting residues and an ionic interaction was considered to be present if any one of the six

distances was less than 4Å at a given time. The standard errors in ionic and hydrogen bonding interactions were determined by dividing the equilibrium segments into intervals of 10 ns from which mean and standard deviation were computed. Trajectories visualization were done using VMD software. Data analyses were done using VMD *tcl* script, Matlab⁴¹ and graphs were prepared using Xmgrace⁴².

Network cluster layout

Network cluster layouts were generated based on the pairwise C^α RMSD matrix where a pairwise C^α RMSD cut-off value was used to group the structures into clusters. It consists of nodes and links. Each node represents a conformation and the links connecting the nodes were built based on the cut-off value described by Ahlstrom et al.⁴³. Profuse-force-directed algorithm⁴⁴ implemented in Cytoscape version 3.5.1⁴⁵ was used to construct the layout. Average linkage algorithm implemented in MATLAB was used to generate centroid structure of each cluster⁴⁶. The centroid structure so obtained represents the central conformation of a given cluster.

Results And Discussion

Choosing specific mutations in M-crystallin

Crystallins are well known for their unusual stability. Stability comes from specific arrangement of the double Greek key motifs. bg-crystallins from microbial origin undergo further stabilization upon binding to Ca²⁺ while eye lens crystallins are stable without Ca²⁺. Microbial bg-crystallins possess canonical motifs N/D-N/D-X₁-X₂-S/T-S to coordinate with Ca²⁺. In M-crystallin the canonical motifs are ³²NDKISS³⁷ and ⁷⁵DNSSIS⁸⁰ in the N- and C-terminal Greek keys respectively. The X₁ position of canonical motifs (K34 and S77) provide direct coordination sites for Ca²⁺ through its mainchain carbonyl. In order to understand the effects of the mutations at X₁ position in M-crystallin, we carried out K34D and S77D mutations. Several βγ-crystallins possess a buried Trp residues in the core surrounded by aromatic and hydrophobic residues as in M-crystallin (Figure S3) which is conserved in many crystallins. Studies reveal that the Trp corner/Tyr corner is important for the βγ-crystallin domain stability⁸. In human γC and γD crystallins, there are four Trp residues, W42 and W68 in the N-terminal domain and W130 and W156 in the C-terminal domain buried in the core and are surrounded by several other aromatic residues and provide enormous stability to the crystallins^{14,47}. W42R mutation in γD crystallin causes congenital cataract⁴⁸. Therefore, we mutate the conserved W45R in M-crystallin (W45 here corresponds W42 in gD Crystallin) to unravel the effect of mutation on βγ-crystallin stability and aggregation propensity (Figure 1A, B).

Highly flexible regions in the mutants of M-crystallin

Root mean square fluctuation measures the average fluctuations of each C^α atoms over time with respect to the reference structure. Thus, it indicates flexibility of a given residue and identifies the highly flexible regions. In the present REMD simulations, the average RMSF for M-crystallin-WT is 1.4 Å for most of the residues except for the N- and C-terminal ends, loop1 between β₂ and β₄ strands and loop2 between α₁

and β_8 (Figure 1C). The RMSF values are mapped onto the starting structure of M-crystallin-WT and represented as sausage plot which clearly indicates that loop1 and loop2 are highly flexible (Figure 1D). In M-crystallin-SM the average RMSF value is 1.9 Å which is marginally higher than that of M-crystallin-WT. Sausage plot identifies similar regions as highly flexible region in M-crystallin-SM and follows almost similar trend as M-crystallin-WT. The only difference observed in M-crystallin-SM w.r.t. M-crystallin-WT is that the N-terminal end is highly flexible (Figure 1C, D). M-crystallin-DM shows relatively higher fluctuations compared to M-crystallin-WT and M-crystallin-SM across the entire sequence with an average value of ~ 2.4 Å. For loop1 and loop2 the fluctuation is even higher (Figure 1C). The M-crystallin-TM shows significantly large fluctuation with an average value of ~ 4.2 Å across the entire polypeptide sequence except β_1 , β_5 , β_6 and β_8 strands. Here, the N-terminal end and the loop1 show dramatic increase in RMSF (Figure 1C). Thus, overall higher flexibility is observed for loop1 and loop2 of all simulations which is remarkably large in M-crystallin-TM. Similar findings are observed as described in SI (Figure S4) from principal component analysis. The absolute entropy calculated using Schlitter method³⁹ are 70.6, 76.4, 80.0 and 86.3 $\text{J M}^{-1} \text{K}^{-1}$ for M-crystallin-WT, M-crystallin-SM, M-crystallin-DM, M-crystallin-TM suggesting M-crystallin-TM display large conformational change, followed by M-crystallin-DM and then M-crystallin-SM while least conformational change is observed for M-crystallin-WT. There is a good correlation between RMSF, entropy and eigenvector analysis which unambiguously suggest that flexibility is more in mutants especially for N-terminal Greek key.

Conformational clusters in mutants and wild type M-crystallin

The C ^{α} RMSD distribution determined for all the simulations illustrate that one narrow distribution for wild type while wider distributions are observed for mutants (details in SI, Figure S5). Free energy landscape shows more numbers of distinctive minima for mutants than the M-crystallin wild type (details in SI, Figure S6). In the network analysis we observed one cluster accounting 100 % in M-crystallin wild type. The centroid structure of the cluster is similar to the X-ray structure of M-crystallin suggesting attainment of $\beta\gamma$ -crystallin fold (Figure 2). In M-crystallin-SM where W45 residue is mutated to R45, the number of clusters observed are four. Out of which cluster 1 is the major cluster with 92% of population, cluster 2 is about 6 % while the rest two are minor clusters having 1.4 and 0.6 % of population (Figure 2). Cluster 1 has $\beta\gamma$ -crystallin-like fold whereas cluster 2 has deformed loops but has $\beta\gamma$ -crystallin-like topology. The two minor clusters sample partially unfolded conformations where the central compact core of $\beta\gamma$ -crystallin fold is lost. In M-crystallin-DM, the network layout also shows four conformational clusters. Of which, cluster 1 is the major cluster having 93 % of population while cluster 2 is only 4 % and the rest two are minor clusters having 2 % and 1 % population, respectively (Figure 2). Cluster 1 and 2 are having $\beta\gamma$ -crystallin-like fold while cluster 3 and 4 are showing partially unfolded conformations where both Greek keys are separated. M-crystallin-TM shows seven conformational clusters. Of which, cluster 1, 2 and 5 are populated to 73, 9 and 3%, respectively. These clusters retain $\beta\gamma$ -crystallin like fold while cluster 3, 4, 6 and 7 are sampled for 8%, 5%, 1.4% and 0.6% of times respectively are mostly in partially unfolded states (Figure 2). Some of the conformational ensemble such as cluster 3 and 4 of M-crystallin-SM look similar to cluster 4 of M-crystallin-TM while cluster 3 and 4 of M-crystallin-DM are similar to

cluster 3 and 7 of M-crystallin-TM. Thus, the partially unfolded conformations are not very diverse but limited to only certain species. Overall, we observed $\beta\gamma$ -crystallin-like ensemble for M-crystallin-WT while mixture of folded and partially unfolded conformational states for mutants.

Native contact analysis segregates conformational clusters into folded and unfolded states

Native contact percentage is used to identify whether a given cluster is folded like a $\beta\gamma$ -crystallin or unfolded. If the percentage of native contact of a given cluster is more than 60%, it is considered as folded or else it is unfolded and are labelled as F and U respectively as shown Figure 3A. The single cluster of M-crystallin-WT possesses $76.5\pm 4.3\%$ of native contacts suggesting all the conformations are well folded and have $\beta\gamma$ -crystallin-like fold (Figure 3A). In M-crystallin-SM, out of four clusters, cluster 1 and 2 have 72.4 ± 5.7 and $62.1\pm 2.8\%$ of native contacts and therefore are considered as folded while cluster 3 and 4 have 49.4 ± 2.2 and $41.3\pm 2.1\%$ of native contacts thus are considered as unfolded (Figure 3A). Similarly, in M-crystallin-DM out of four clusters cluster 1 and 2 are folded having 77.1 ± 3.8 and $61.2\pm 2.8\%$ of native contacts while cluster 3 and 4 are unfolded having 49.4 ± 2.2 and $41.3\pm 1.1\%$ of native contacts. In the M-crystallin-TM cluster 1, 2 and 5 possess 69.1 ± 5.7 , 60.4 ± 3.2 and $62.2\pm 2.2\%$ of native contacts respectively and therefore are folded while the clusters 3, 4, 6 and 7 are having 41.5 ± 1.6 , 57.5 ± 2.3 , 47.4 ± 2.8 and $47.6\pm 2.3\%$ respectively and therefore are unfolded clusters. Native contact analysis reveals that the observed unfolded states possess ~ 40 - 50% of native contacts and thus are not completely unfolded like random coil but are in partially unfolded states.

Structural insights into the folded and partially unfolded states

The X-ray structure of M-crystallin has $\beta\gamma$ -crystallin fold which has two Greek key motifs formed by two β -sheets. In this arrangement, three (β_5 , β_6 and β_8) of the four β -strands of C-terminal Greek key form one β -sheet and the remaining one, β_7 -strand pairs with three β -strands (β_1 , β_2 and β_4) of N-terminal Greek key forming the second β -sheet which are tightly linked by several interactions (Figure 3B). β_3 -strand is not found in M-crystallin instead a long loop1 is present in the N-terminal Greek key. There is a second loop2 in the C-terminal Greek key which extend and include unwound α -helix and β_7 -strand in the unfolded states (Figure 3B). There are several interlocking Greek key interactions holding loop1 and β_8 and loop1 and loop2 similarly holding β_7 and β_4 -strands together (Figure 3B). Besides, it has several intra- β -sheet interactions in addition to the strong hydrophobic core holding the two opposite β -sheets (Figure S3 and Figure 3B). The interlocking Greek key interactions are of prime importance as unlocking these interactions can unwind the interlocked Greek keys. Although, there are several unfolded states in the mutants broadly they can be divided into two main types. In one type, the two interlocked Greek keys are widely separated by losing all the interlocking interactions. It has only β_1 , β_4 -strands and loop1 from N-terminal Greek key and β_5 , β_6 -strands and loop2 from C-terminal Greek key (Figure 3B, bottom left). In the second type of unfolded state, the central hydrophobic compact core is lost however it is held by loop1 and β_8 -strand and β_4 and β_7 -strands interlocking Greek key interactions and intra- β -sheet interactions. It

can be noted that M-crystallin-SM resulted second type of while M-crystallin-DM resulted first type and M-crystallin-TM resulted both types of partially unfolded states.

Destabilizing interactions drives towards partially unfolded states

In order to identify the destabilizing interactions, all possible interactions are determined. Surface charge potential is generated as shown in Figure S7 which reveals that M-crystallin-WT, M-crystallin-SM display smaller while M-crystallin-DM and M-crystallin-TM displays large negative surface potential on the loop regions which is an important site for interlocking Greek key interactions. Large repulsive interaction can break the interconnected loops and lead to partially unfolded states as observed in M-crystallin-DM and M-crystallin-TM. Further, total number of hydrogen bonds in each cluster are determined. However, we did not find any correlation between unfolded cluster and reduced number of hydrogen bond because certain clusters which are unfolded have more number of hydrogen bonds than the folded one. Such situation can arise only when there are more number of non-native hydrogen bonds than the native one. Further, we determined hydrogen bonds and ionic interactions contributed by the residues undergone mutations (Table S1). Out of seven native interactions only three are observed in M-crystallin-SM, one is observed in M-crystallin-DM and none of the native interactions are observed in M-crystallin-TM suggesting non-native interactions can also contribute in unfolding and drive towards partially unfolded states.

Partially unfolded states display large hydrophobic solvent accessible surface area

We calculated hydrophobic solvent accessible surface area per residue in five known single domain bg-crystallin proteins (M-crystallin, Hahellin, Flavollin, Clostrillin and Ciona) from protein data bank (Figure 4). In spite of sequence difference, the common patterns consistently observed in these bg-crystallins are that the completely buried hydrophobic residues are located at the junction of the two Greek key motifs (Figure 4). Secondly, the hydrophobic residues located in the N-terminal Greek key motif are relatively more exposed than the C-terminal Greek key motif. These findings clearly indicates that there is a characteristic pattern of hydrophobic burial in bg-crystallin fold which is probably necessary for the bg-crystallin domain stability. Alteration in the characteristic pattern of rhSASA might cause unfolding of the bg-crystallin domain. Subsequently, the rhSASA is estimated for folded (F) and unfolded (U) conformational states of a given simulation (Figure 5A). They are grouped so based on their native contact percentage. In M-crystallin-SM, the rhSASA of the residues V68, A71, I73 and F81 are relatively more exposed in unfolded state compared to its folded counterpart (Figure 5A). These residues are located in the loop2 region of C-terminal Greek key. Similarly, in M-crystallin-DM hydrophobic exposure is observed in the similar region corresponding to the residues L60, V68, A71, I73, I78 and F81. In addition, there are two more regions corresponding to the residues I35, I38, V40 and F47 located at the junction of N- and C-terminal Greek keys and residues A3 and V5 of N-terminal Greek key showing higher values of rhSASA in the partially unfolded states compared to that of folded state. Similarly, in M-crystallin-TM similar regions as observed in M-crystallin-DM are exposed (Figure 5). These residues showing higher values of rhSASA in the unfolded state are mapped onto the centroid structure of folded and unfolded clusters as shown in Figure 5B. There is a remarkable difference in the hydrophobic surface exposure of a

folded and unfolded conformations. In all the folded conformations from wild type, mutants and X-ray structure, a significant burial of hydrophobic residues is observed while on the partially unfolded states there is conspicuous exposure of hydrophobic residues which form continuous patches of the size ~ 500 to 700 \AA^2 (Figure 5C). These sites can act as attachment sites for self-aggregation into higher molecular weight aggregates and thus can lead to cataract in homologous eye-lens crystallins. In a recent study, γ S-crystallin-G18V mutation is reported to cause cataract. In that study, ANS fluorescence assay and NMR method are used to probe hydrophobic exposure of γ S-crystallin-G18V mutant where non-specific ANS binding to the mutant is observed indicating significant hydrophobic exposure⁴⁹.

Trp residue plays a key role in bg-crystallin stability

Trp corner/Tyr corner are known to provide stability to the $\beta\gamma$ -crystallin fold. Generally, it is located at the beginning or end of an anti-parallel β -strand which is hydrogen bonded in case of Tyr and stabilizes the β -barrel structure of the Greek key motif^{14,50}. In M-crystallin, W45 forms the Trp corner, is located in the hydrophobic core and is surrounded by several aromatic and hydrophobic residues forming a compact core. In the homologous γ C and γ D-crystallins four Trps (positions 42, 68, 130 and 156) are buried and are surrounded by Tyr residues. W42R mutation in human γ D-crystallin is associated with congenital cataract⁴⁸. W42R causes minimal changes in the structure w.r.t. wild type γ D-crystallin as revealed by X-ray but produces small population of partially unfolded states which are in chemical exchange with folded state as revealed by NMR^{48,51}. In line with this, the conserved W45R mutation in M-crystallin generates folded and partially unfolded states in our simulation. It is presumed to be driven by perturbation of the hydrophobic core and formation of non-native interactions by R45 based on our findings. The partially unfolded states could be the potential candidates for aggregation. In another multiscale atomistic simulation of γ D-crystallin having W42R mutation, adopted a distinct conformation in solution where Greek key domains are more or less intact while a large perturbation is observed in the inter-domain interface region which shows large hydrophobic exposure and can act as primary sites for aggregation⁵². In another simulation study, W42 is mutated to polar residues such as Lys and Arg which denatures the Greek key domain as solvent enter into the core and causes hydrophobic exposure of core residues⁵³. In all these studies W42 mutation leading to perturbation in the hydrophobic core accompanied by hydrophobic exposure is the common event irrespective of different independent studies.

Indirect relation between highly flexible region of M-crystallin and the region of highest hydrophobic exposure

The mutants of M-crystallin show high flexibility for the N-terminal motif as evident from RMSF and eigenvector analysis. However, this highly flexible region does not correlate directly with the highest hydrophobic exposure. Most of the hydrophobic residues at the N-terminal Greek key do not show any difference in rhSASA between folded and unfolded conformations (Figure 5) while significant difference is seen for the C-terminal Greek key motif as most of the hydrophobic residues in this region are exposed in partially unfolded states while they are buried in the folded state. The flexibility in the N-terminal is

enhanced by breakage of several interactions connecting loop1 and β_8 -strand and loop1 and loop2, and this in turn causes exposure of the C-terminal Greek key hydrophobic residues which potentially can drive intermolecular aggregation. Recently, NMR study and dynamics light scattering (DLS) data reveal formation of oligomer in M-crystallin as a result of lowering the temperature which is accompanied by large dynamics in the Ca^{2+} binding sites¹⁵. There is an intricate relation between unfolding and aggregation^{4,54,55}. V75D, W42R mutations in γ D-crystallin^{51-53,56}, G75V mutation in γ S-crystallin⁵⁶, S228P in β B1-crystallin⁵⁷ associated with partial unfolding leading to hydrophobic exposure. Our study affirms that partially unfolded species are formed due to mutation which could serve as precursors for aggregation and can develop cataract.

Conclusion

In our REMD simulations of M-crystallin wild type and its three mutants, the dynamic nature of loop1 of N-terminal Greek key is clearly noticeable. In M-crystallin-WT no dramatic conformational change is observed because of optimal packing of hydrophobic residues in the core and presence of several interlocking Greek key interactions between loop1 and loop2, also between loop1 and β_8 -strand and β_4 -strand and β_7 -strand and intra- β -sheet interactions. On the other hand in all mutants of M-crystallin formation of heterogeneous mixture of folded or partially unfolded states are observed. The partially unfolded states are mainly of two types. In one type, the two Greek key motifs fall apart without any significant interactions between the two Greek keys while in second type, the central hydrophobic core is largely perturbed and the compact core is lost but still held by several interactions from both the Greek key motifs especially between loop1 and loop2 and between loop1 and β_8 -strand. In either types of unfolded states, the buried hydrophobic residues are exposed giving rise to large hydrophobic patches. These hydrophobic patches are mainly contributed by the hydrophobic residues located at the junction of both Greek key motifs and at the C-terminal Greek key motif. These hydrophobic patches can provide attachment sites for association into higher order molecular aggregates. Thus, hydrophobic patches on the partially unfolded crystallin are the main determinant of $\beta\gamma$ -crystallin aggregation. Interactions restricting the loop dynamics and promoting the strength of hydrophobic core would reduce the hydrophobic exposure and would prevent the aggregation of crystallin which is the prime cause of cataract.

Declarations

Acknowledgments

S.P. acknowledges support from DST for WOS-A with Grant No. SR/WOS-A/CS-143/2017 (G). SP thanks NPSF PARAM HPC supercomputing facility of C-DAC.

Author contributions

S.P. designed and performed the research, analyzed the results, and wrote the paper. R.V.H. provided his critical comments and participated in the discussion.

Supporting Information paragraph

Seven figures, one table and part of the results are available in Supporting Information.

References

1. Resnikoff, S. *et al.* Global data on visual impairment in the year 2002. *Bull. World Health Organ.* **82**, 844-851, (2004).
2. Pescosolido, N., Barbato, A., Giannotti, R., Komaiha, C. & Lenarduzzi, F. Age-related changes in the kinetics of human lenses: Prevention of the cataract. *Int. J. Ophthalmol. Clin. Res.* **9**, 1506-1517, (2016).
3. Zhao, H., Magone, M. T. & Schuck, P. The role of macromolecular crowding in the evolution of lens crystallins with high molecular refractive index. *Phys Biol* **8**, 046004-046004, (2011).
4. Roskamp, K. W., Paulson, C. N., Brubaker, W. D. & Martin, R. W. Function and aggregation in structural eye lens crystallins. *Acc. Chem. Res.* **53**, 863-874, (2020).
5. Mahler, B. *et al.* Characterization of a transient unfolding intermediate in a core mutant of γ s-crystallin. *J. Mol. Biol.* **405**, 840-850, (2011).
6. Serebryany, E. & King, J. A. The $\beta\gamma$ -crystallins: Native state stability and pathways to aggregation. *Prog. Biophys. Mol. Biol.* **115**, 32-41, (2014).
7. Barnwal, R. P., Jobby, M. K., Devi, K. M., Sharma, Y. & Chary, K. V. Solution structure and calcium-binding properties of m-crystallin, a primordial betagamma-crystallin from archaea. *J. Mol. Biol.* **386**, 675-689, (2009).
8. Mishra, A., Krishnan, B., Srivastava, S. S. & Sharma, Y. Microbial $\beta\gamma$ -crystallins. *Prog. Biophys. Mol. Biol.* **115**, 42-51, (2014).
9. Suman, S. K., Mishra, A., Ravindra, D., Yeramala, L. & Sharma, Y. Evolutionary remodeling of $\beta\gamma$ -crystallins for domain stability at cost of Ca^{2+} binding. *J. Biol. Chem.* **286**, 43891-43901, (2011).
10. Suman, S. K., Mishra, A., Yeramala, L., Rastogi, I. D. & Sharma, Y. Disability for function: Loss of Ca^{2+} binding is obligatory for fitness of mammalian $\beta\gamma$ -crystallins. *Biochemistry* **52**, 9047-9058, (2013).
11. Mishra, A., Suman, S. K., Srivastava, S. S., Sankaranarayanan, R. & Sharma, Y. Decoding the molecular design principles underlying Ca^{2+} binding to $\beta\gamma$ -crystallin motifs. *J. Mol. Biol.* **415**, 75-91, (2012).
12. Srivastava, S. S., Mishra, A., Krishnan, B. & Sharma, Y. Ca^{2+} -binding motif of $\beta\gamma$ -crystallins. *J. Biol. Chem.* **289**, 10958-10966, (2014).

13. Patel, S., Krishnan, B., Hosur, R. V. & Chary, K. V. R. Mechanistic insights from replica exchange molecular dynamics simulations into mutation induced disordered-to-ordered transition in hahellin, a $\beta\gamma$ -crystallin. *J. Phys. Chem. B* **123**, 5086-5098, (2019).
14. Vendra, V. P., Khan, I., Chandani, S., Muniyandi, A. & Balasubramanian, D. Gamma crystallins of the human eye lens. *Biochim Biophys Acta* **1860**, 333-343, (2016).
15. Barnwal, R. P., Devi, K. M., Agarwal, G., Sharma, Y. & Chary, K. V. Temperature-dependent oligomerization in m-crystallin: Lead or lag toward cataract, an nmr perspective. *Proteins* **79**, 569-580, (2011).
16. Graw, J. Genetics of crystallins: Cataract and beyond. *Exp. Eye Res.* **88**, 173-189, (2009).
17. Aravind, P. *et al.* The $\beta\gamma$ -crystallin superfamily contains a universal motif for binding calcium. *Biochemistry* **48**, 12180-12190, (2009).
18. Barnwal, R. P., Jobby, M. K., Devi, K. M., Sharma, Y. & Chary, K. V. R. Solution structure and calcium-binding properties of m-crystallin, a primordial $\beta\gamma$ -crystallin from archaea. *J. Mol. Biol.* **386**, 675-689, (2009).
19. DeLano, W. L. The pymol molecular graphics system, version 1.8 schrödinger, llc. Home page. [Http://www.Pymol.Org/](http://www.Pymol.Org/) (accessed june 15, 2013).
20. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151, (1999).
21. Case, D. A. *et al.* *Amber10, university of california,*. (2008).
22. Ponder, J. W. & Case, D. A. Force fields for protein simulations. *Adv Protein Chem* **66**, 27-85, (2003).
23. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999-2012, (2003).
24. Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Struct. Funct. Bioinform.* **55**, 383-394, (2004).
25. Shang, Y., Nguyen, H., Wickstrom, L., Okur, A. & Simmerling, C. Improving the description of salt bridge strength and geometry in a generalized born model. *J. Mol. Graph. Model.* **29**, 676-684, (2011).
26. Vorobjev, Y. N. Advances in implicit models of water solvent to compute conformational free energy and molecular dynamics of proteins at constant ph. *Adv. Protein Chem. Struct. Biol.* **85**, 281-322, (2011).

27. Kleinjung, J. & Fraternali, F. Design and application of implicit solvent models in biomolecular simulations. *Curr. Opin. Struct. Biol.* **25**, 126-134, (2014).
28. Anandkrishnan, R., Drozdetski, A., Walker, R. C. & Onufriev, A. V. Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.* **108**, 1153-1164, (2015).
29. Onufriev, A. V. & Case, D. A. Generalized born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* **48**, 275-296, (2019).
30. Zhou, R. & Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? *Proc Natl Acad Sci U S A* **99**, 12777-12782, (2002).
31. Ganguly, D. & Chen, J. Atomistic details of the disordered states of kid and pkid. Implications in coupled binding and folding. *J. Am. Chem. Soc.* **131**, 5214-5223, (2009).
32. Ganguly, D. & Chen, J. Modulation of the disordered conformational ensembles of the p53 transactivation domain by cancer-associated mutations. *PLOS Comput. Biol.* **11**, e1004247, (2015).
33. Lee, K. H. & Chen, J. Optimization of the gbm2 implicit solvent force field for accurate simulation of protein conformational equilibria. *J. Comput. Chem.* **38**, 1332-1341, (2017).
34. Patel, S., Ramanujam, V., Srivastava, A. K. & Chary, K. V. Conformational propensities and dynamics of a betagamma-crystallin, an intrinsically disordered protein. *Phys. Chem. Chem. Phys.* **16**, 12703-12718, (2014).
35. Sanbonmatsu, K. Y. & Garcia, A. E. Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins* **46**, 225-234, (2002).
36. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341, (1977).
37. Sawle, L. & Ghosh, K. Convergence of molecular dynamics simulation of protein native states: Feasibility vs self-consistency dilemma. *J. Chem. Theory Comput.* **12**, 861-869, (2016).
38. Daura, X. *et al.* Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* **38**, 236-240, (1999).
39. Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **215**, 617-621, (1993).
40. Humphrey, W., Dalke, A. & Schulten, K. Vmd - visual molecular dynamics. *J. Mol. Graphics.* **14**, 33-38, (1996).

41. MATLAB. Matlab and statistics toolbox release 2017a, the mathworks, inc., natick, massachusetts, united states. Home page. [Http://www.Mathworks.Com/](http://www.Mathworks.Com/).
42. Grace. Home page. [Http://plasma-gate.Weizmann.Ac.il/grace/](http://plasma-gate.Weizmann.Ac.il/grace/) (accessed june 15, 2013).
43. Ahlstrom, L. S. *et al.* Network visualization of conformational sampling during molecular dynamics simulation. *J. Mol. Graph. Model.* **46**, 140-149, (2014).
44. Kobourov, S. G. Spring embedders and force-directed graph drawing algorithms *arXiv:1201.3011*, p. 1-23. , (2012).
45. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504, (2003).
46. Shao, J., Tanner, S. W., Thompson, N. & Cheatham, T. E. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory. Comput.* **3**, 2312-2334, (2007).
47. Schafheimer, N. & King, J. Tryptophan cluster protects human γ d-crystallin from ultraviolet radiation-induced photoaggregation in vitro. *Photochemistry and Photobiology* **89**, 1106-1115, (2013).
48. Wang, B. *et al.* A novel crygd mutation (p.Trp43arg) causing autosomal dominant congenital cataract in a chinese family. *Hum Mutat* **32**, E1939-E1947, (2011).
49. Khago, D. *et al.* Increased hydrophobic surface exposure in the cataract-related g18v variant of human γ s-crystallin. *Biochim Biophys Acta* **1860**, 325-332, (2016).
50. Moreau, K. L. & King, J. Hydrophobic core mutations associated with cataract development in mice destabilize human γ crystallin. *J. Biol. Chem.* **284**, 33285-33295, (2009).
51. Ji, F., Jung, J., Koharudin, L. M. I. & Gronenborn, A. M. The human w42r γ d-crystallin mutant structure provides a link between congenital and age-related cataracts. *J. Biol. Chem.* **288**, 99-109, (2013).
52. Wong, E. K., Prytkova, V., Freites, J. A., Butts, C. T. & Tobias, D. J. Molecular mechanism of aggregation of the cataract-related γ d-crystallin w42r variant from multiscale atomistic simulations. *Biochemistry* **58**, 3691-3699, (2019).
53. Aguayo-Ortiz, R. & Dominguez, L. Effects of mutating trp42 residue on γ d-crystallin stability. *Journal of Chemical Information and Modeling* **60**, 777-785, (2020).
54. Moreau, K. L. & King, J. A. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends Mol. Med.* **18**, 273-282, (2012).
55. Serebryany, E. & King, J. A. The betagamma-crystallins: Native state stability and pathways to aggregation. *Prog Biophys Mol Biol* **115**, 32-41, (2014).

56. Zhu, S. *et al.* The cataract-causing mutation g75v promotes γ s-crystallin aggregation by modifying and destabilizing the native structure. *Int. J. Biol. Macromol.* **117**, 807-814, (2018).
57. Qi, L. B. *et al.* Cataract-causing mutation s228p promotes β b1-crystallin aggregation and degradation by separating two interacting loops in c-terminal domain. *Protein Cell* **7**, 501-515, (2016).
58. Samanta, U., Bahadur, R. P. & Chakrabarti, P. Quantifying the accessible surface area of protein residues in their local environment. *Protein Eng.* **15**, 659-667, (2002).

Figures

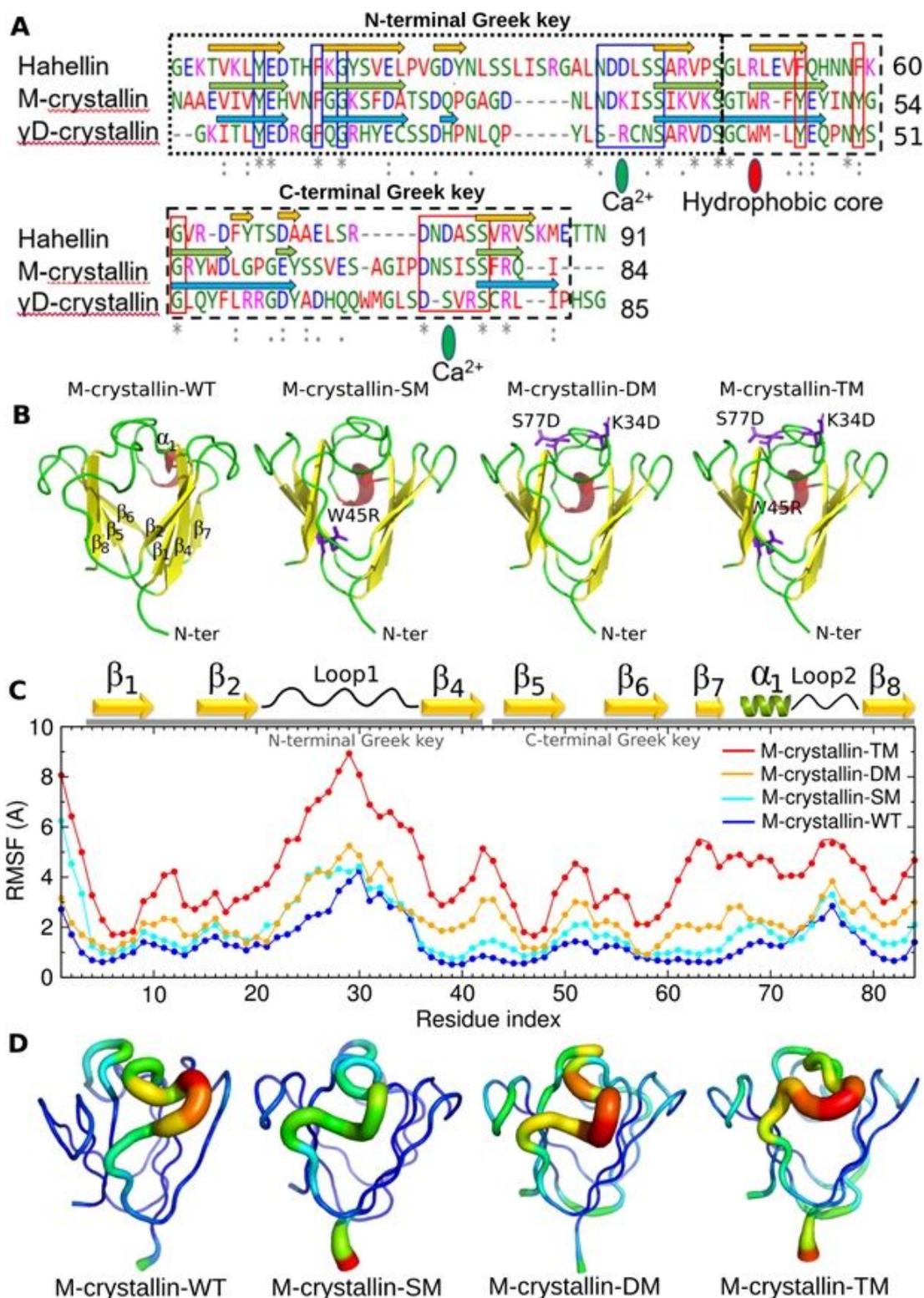


Figure 1

(A) Multiple sequence alignment of Hahellin (2KP5), M-crystallin (3HZ2) and human eye-lens γ D-crystallin (1HK0) (one domain). The rectangular dashed boxes represent N-terminal Greek key (small dashed) and C-terminal Greek key (bigger dashed). The Greek key signature sequence for N-terminal Greek key domain is shown in blue solid box and C-terminal Greek key domain in red solid box. Amino acid code in blue: negatively charged residues; magenta: positively charged residues; red: hydrophobic residues; green:

polar residues. The starting structures for the REMD simulations (B). Mutations are made in the hydrophobic core (W45R) and in the Ca²⁺ binding sites K34D and S77D are shown in purple stick representation. RMSF of Ca atoms of all simulations (C). The relative flexibility are mapped onto the starting structure of each simulation and represented as sausage plots with B-factor colour codes (D). Narrow tube with blue colour indicates rigid structure while wider tube with cyan and green indicate intermediate flexibility and a further wider tube with orange and red indicate high flexibility.

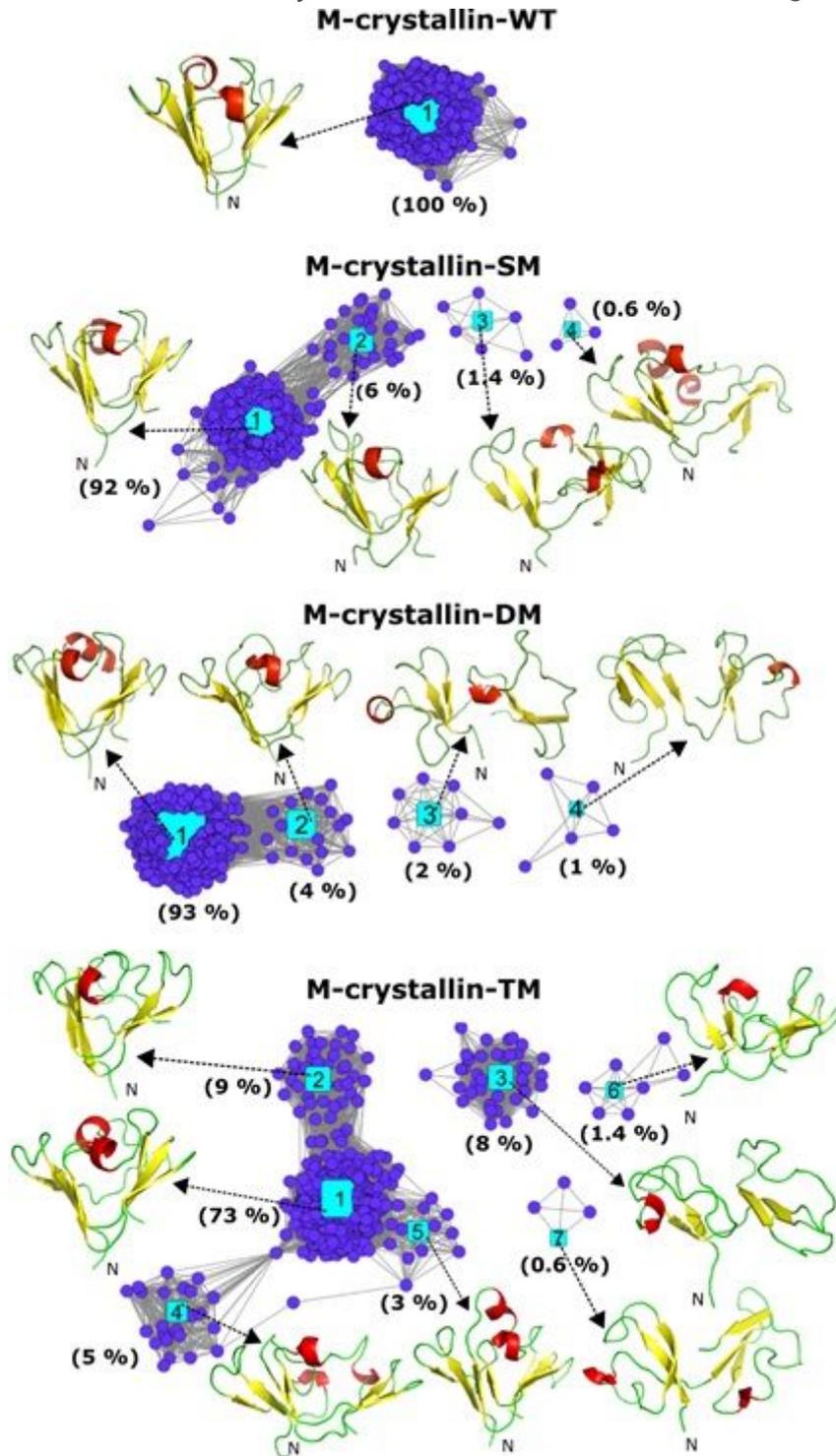


Figure 2

Network cluster layouts of M-crystallin wild type and its mutants. An identical pairwise C RMSD cut-off of 3.5 Å is used for constructing the network layouts. Five hundred frames were taken at a regular interval from the equilibrated region (70-200 ns) of each simulations for this analysis. Each cluster has a centroid structure which is central to the cluster is determined by the average linkage algorithm.

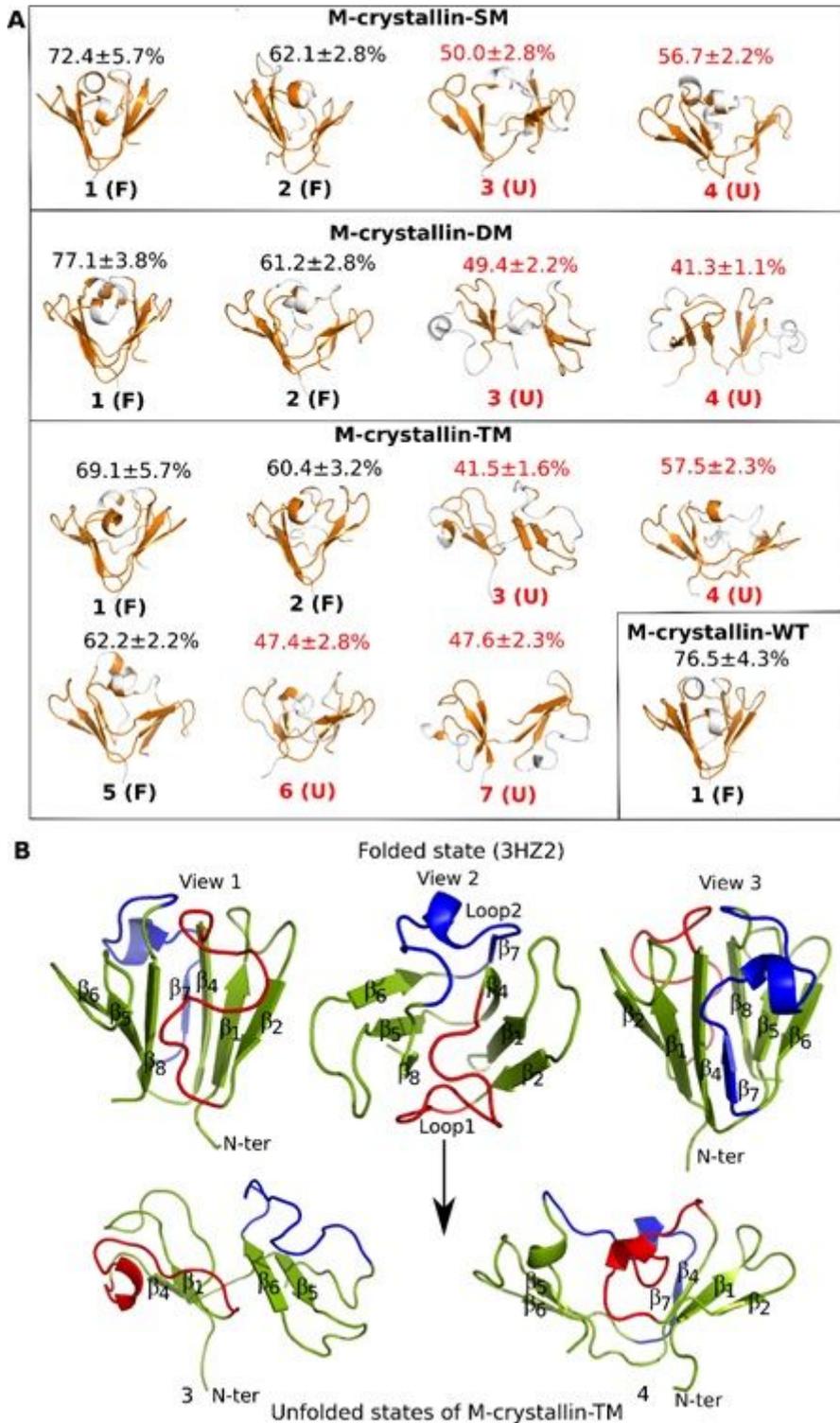


Figure 3

Native contact analysis is performed on the conformational clusters obtained from network cluster analysis. Percentage of native contacts are annotated for a given cluster (A). Any cluster having native

contacts 60% or greater is denoted as folded (F) else it is assigned as unfolded (U). The residues having native contacts are mapped onto the centroid structure and are shown in orange. (B) Folded and unfolded states of M-crystallin in X-ray structure and in mutants.

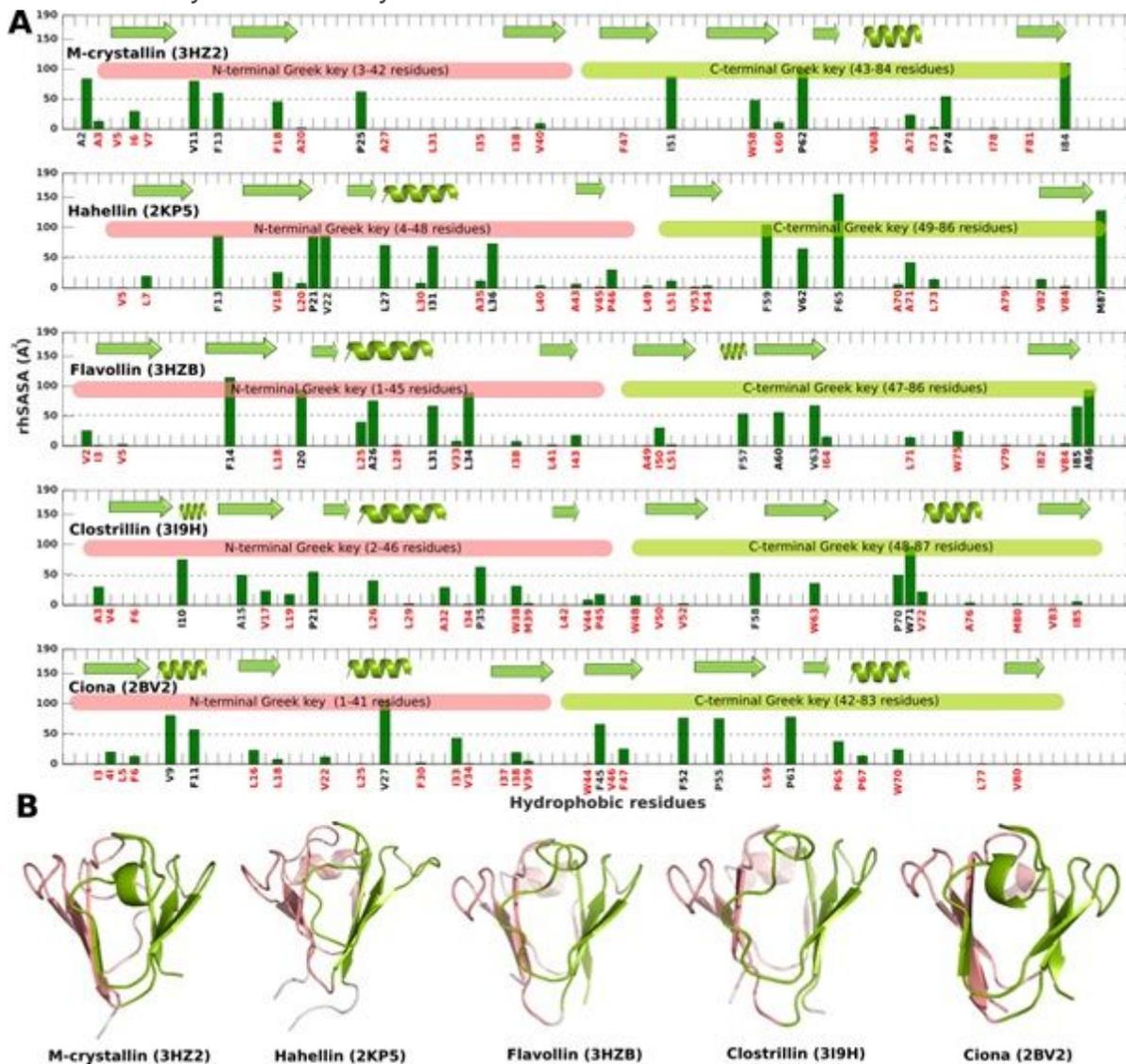


Figure 4

Residue specific hydrophobic solvent accessible surface area of five crystallins. The residues which are having less than 50 Å² of rhSASA are coloured red to indicate that those residues are buried. The reference rhSASA values of hydrophobic residues are Ala (102.7), Leu (184.3), Val (148.5), Ile (176.2), Phe (209.6), Pro (126.8) and Trp (251.8) Å² determined from a tripeptide of the type Ala-X-Ala where X is the residue of interest 58 for relative comparison.

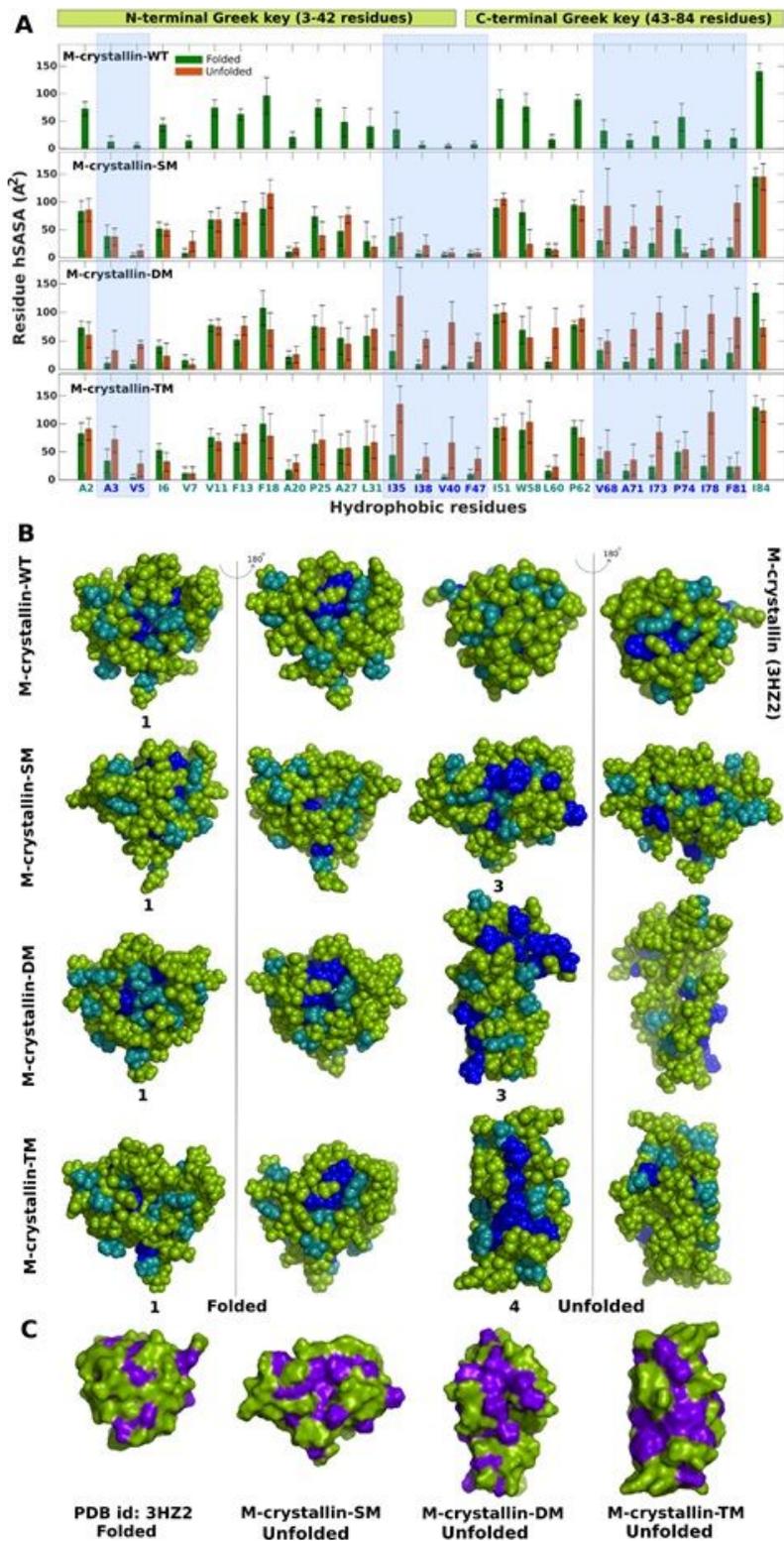


Figure 5

Hydrophobic solvent accessible surface area per residue is determined for the folded or unfolded states in all simulations. For this calculation all folded and unfolded clusters are combined and made into two separate groups. rhSASA of folded and unfolded conformations are shown as bar plots (A). Hydrophobic residues in unfolded conformations having higher exposure than the folded conformations are mapped on the centroid structure in blue van der Waals spheres, the rest hydrophobic residues in cyan and

residues other than the hydrophobic residues are in green color (B). Two color codes used to indicate hydrophobes (purple) and nonhydrophobes (green) to indicate hydrophobic patches (C).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supportinginformation210621.docx](#)