Published in final edited form as:

Nanoscale. 2014 January 16; 6(3): 1797–1807. doi:10.1039/c3nr03782g.

Dissecting the contributions of β -hairpin tyrosine pair to the folding and stability of long-lived human γ D-crystallins

Zaixing Yang^{1,†}, Zhen Xia^{1,2,†}, Tien Huynh², Jonathan A. King³, and Ruhong Zhou^{1,2,4,*}

¹Institute of Quantitative Biology and Medicine, SRMP and RAD-X & Collaborative Innovation Center of Suzhou Nano Science and Technology; Jiangsu Provincial Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou 215123, China

²Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

³Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

⁴Department of Chemistry, Columbia University, New York, NY 10027

Abstract

Ultraviolet-radiation-induced damage to and aggregation of human lens crystallin proteins are thought to be a significant pathway to age-related cataract. The aromatic residues within the duplicated Greek key domains of γ - and β -crystallins are the main ultraviolet absorbers and are susceptible to direct and indirect ultraviolet damage. The previous site-directed mutagenesis studies have revealed a striking difference for two highly conserved homologous β -hairpin Tyr pairs, at the N-terminal domain (N-td) and C-terminal domain (C-td) respectively, in their contribution to the overall stability of H γ D-Crys, but why they behave so differently still remains a mystery. In this paper, we systematically investigated the underlying molecular mechanism and detailed contributions of these two Tyr pairs with large scale molecular dynamics simulations. A series of different tyrosine-to-alanine pair(s) substitutions were performed in either the N-td, the C-td, or both. Our results suggest that the Y45A/Y50A pair substitution in the N-td mainly affects the stability of the N-td itself, while the Y133A/Y138A pair substitution in the C-td leads to a more cooperative unfolding of both N-td and C-td. The stability of motif 2 in the N-td is mainly determined by the interdomain interface, while motif 1 in the N-td or motifs 3 and 4 in the C-td are mainly stabilized by the intradomain hydrophobic core. The damage to any tyrosine pair(s) can directly introduce some apparent water leakage to the hydrophobic core at the interface, which in turn causes a serious loss in stability of the N-td. However, for the C-td substitutions, it may further impair the stable “sandwich-like” Y133-R167-Y138 cluster (through cation- π interactions) in the wild-type, thus causing the loop regions near the residue A138 to undergo large fluctuations, which in turn results in the intrusion of water into the hydrophobic core of the C-td and induces the C-td to lose its stability. These findings help resolve the “mystery” on why these two Tyr pairs display such a striking difference in their contributions to the overall protein stability despite their highly homologous nature.

*Corresponding author: ruhongz@us.ibm.com.

†These authors contributed equally

Supporting information available

RMSDs (C α) for the wild-type at 310 K; Solvent accessible surface (SAS) area of M43, L44, and L132 residues as a function of simulation time; Snapshots of the wild-type and mutant proteins during their respective representative unfolding trajectories.

Keywords

γ D-Crystallins; Cataract; Ultraviolet radiation; Tyrosine; Alanine; Protein unfolding; Molecular dynamics; Hydrophobic interactions; Ophthalmic disease

Introduction

Human age-onset cataracts are the most common cause of vision loss for elder people as well as the principal cause of blindness in the world. Cataract shares similar pathogenic mechanism with other fatal protein conformational diseases, such as Alzheimer's disease, Huntington disease, and type II diabetes. It is directly associated with the aggregation of partially unfolded or misfolded γ and β -crystallins in the eye lens.¹ The insoluble fibrillar aggregates of misfolded crystallins form a cloudy or opaque area in the eye lens, which obstructs the passage of light to the retina.^{2–8} Epidemiological researches of humans and laboratory experiments on animals suggest that ultraviolet (UV) radiation is one of the primary causal factors in cataractogenesis,^{9–12} which involves the impairment of the folding and stability of H γ D-Crys due to direct and indirect UV damage proceeding through the aromatic residues that are main UV absorbers^{10,13} in the protein. Direct analysis of the crystallin proteins and peptides recovered from surgically removed cataracts reveal a variety of oxidized and photo-oxidized side chains.^{13–16} For aromatic residues, photo-oxidation leads to ring opening and loss of aromaticity.

H γ D-Crys, found in the oldest central part of the lens, is synthesized early in life, and must remain folded and soluble throughout an individual's lifetime to maintain eye lens clarity and transparency. As shown in Fig. 1A, H γ D-Crys consists of two homologous domains, each contains two Greek key motifs forming a β -sandwich like structure with eight intercalated β -strands. The two domains are linked together, creating a well-defined dry hydrophobic interface.¹⁷ H γ D-Crys contains a high proportion of aromatic residues distributed throughout the protein: 14 Tyr, 6 Phe, and 4 Trp. Among these aromatic residues, there are two highly conserved homologous aromatic Tyr pairs, present in the β -hairpin sequences of the Greek keys at the interface (Fig. 1A): the Y45/Y50 pair in the N-terminal domain (N-td), and the Y133/Y138 pair in the C-terminal domain (C-td)¹⁸ respectively. These Tyr pairs exhibit a conservation rate >90% through the crystallin protein family. The β -hairpin aromatic pairs are a remarkable structural feature of the Greek keys in H γ D-Crys. Kong and King used site-directed mutagenesis, together with both thermal and GuHCl denaturation to assess the contributions of these β -hairpin aromatic pairs to the folding and stability of H γ D-Crys.¹⁸ For the N-td substitutions, substitution of the Y45/Y50 pair with an A45/A50 pair (denoted by Y45A/Y50A_N-td mutant), reduced the stability of the N-td but not the C-td. However, for the C-td substitutions, with Y133/Y138 pair substituted with A133/A138 (denoted by Y133A/Y138A_C-td mutant), not only decreased the stability of the C-td but also the N-td, i.e., the entire protein.

These results reveal the different roles of the β -hairpin Tyr pairs in stabilizing H γ D-Crys, which is surprising due to the fact that these two Tyr pairs are highly homologous and are both situated at the interface. However, why the Tyr to Ala pair mutations in different domains can lead to a different unfolding behavior for the N-td and C-td still remains a mystery. Our other recent experimental and computational studies on the wild-type and H γ D-Crys mutants suggested a sequential unfolding pathway (N-td unfolds first, followed by the C-td) and a “domain swapping” mechanism for the H γ D-Crys aggregation, in which the largely folded C-td of one monomer acts as the nucleation site, and the fully unfolded N-td from the other monomer “refolds” to the nucleation core, thus forming a chain-like structure.^{2, 19–23} Similar to the loss of aromaticity caused by UV damage, previous

experimental studies show that Tyr to Ala substitution can be a good model system to evaluate the effect of broken aromatic rings to those conserved Tyr pairs.¹⁸ Therefore, by comparing the stability of HyD-Crys protein with and without Tyr-pairs, we might be able to better understand the key triggers for the protein unfolding, such as these β -hairpin Tyr-pairs, which might eventually lead to deeper insights to the HyD-Crys unfolding pathways and development of therapies for the prevention and treatment of cataracts.

In this study, we perform extensive all-atom molecular dynamics (MD) simulations to dissect the contributions and the underlying mechanisms of the two homologous β -hairpin Tyr pairs at the interface from both domains to investigate the folding and stability of the long-lived HyD-Crys. In addition to the wild-type and the two mutants Y45A/Y50A_N-td and Y133A/Y138A_C-td, we also investigate the mutant where all four Tyr residues (i.e., both Y45/Y50 and Y133/Y138 pairs) are substituted by the Ala (denoted by 4Y>4A mutant). We find that despite their highly homologous nature, the two β -hairpin Tyr pairs in N-td and C-td behave very differently – for the N-td substitutions the N-td unfolds faster but the C-td remains largely intact, whereas for the C-td substitutions both the N-td and C-td unfold more quickly, supporting our previous experimental observations.¹⁸ Moreover, for the first time, our molecular dynamics simulations reveal the atomic-detailed unfolding processes of HyD-Crys due to the Ala substitutions of these two key Tyr pairs, thus resolving the mystery on why these two highly homologous β -hairpin Tyr pairs behave so differently in stabilizing HyD-Crys.

System and Methods

The initial structure of the wild-type HyD-Crys determined by Basak et al¹⁷ has been taken from the x-ray crystal structure in the Protein Data Bank (PDB code: 1HK0, 173 residues, see Fig. 1A for details). The protein was immersed in a pre-equilibrated 8M aqueous urea solution with a box size of 73.1 Å × 73.1 Å × 73.1 Å (1,920 urea and 8,192 water molecules; following our previous similar protocols^{24–28}). The wild-type HyD-Crys and 8 M urea mixture system was then equilibrated with a 1,000-ps simulation at constant temperature and pressure (310 K and 1 atm), with the final box size converged to $\approx 72.9 \text{ Å} \times 72.9 \text{ Å} \times 72.9 \text{ Å}$. Five configurations were randomly extracted from the aforementioned trajectories to use as initial state for production runs for the wild-type simulations. To pinpoint the contributions of the β -hairpin Tyr pairs (Fig. 1A) at the interface in each domain to the folding and stability of HyD-Crys, three mutant systems have been set up: N-td substitutions, with Y45/Y50 pair substituted with A45/A50 (Fig. 1B); C-td substitutions, with Y133/Y138 pair substituted with A133/A138 (Fig. 1C); and double pair substitutions, with both Y45/Y50 and Y133/Y138 pairs substituted by Ala (Fig. 1D). Hereafter we refer to the three mutant HyD-Crys systems mentioned above as “Y45A/Y50A_N-td”, “Y133A/Y138A_C-td”, and “4Y>4A” mutant systems, respectively. In order to validate/test the protein stability, a 200 ns simulation was performed for the wild-type in pure water at 310K as a control (see Fig. S1). The initial configurations of the mutant HyD-Crys were generated by the simple replacements of residues Tyr to Ala, from the above-described wild-type systems and then re-equilibrated with a 1,000-ps simulation at constant temperature and pressure (310 K and 1 atm). Therefore, for both the wild-type and mutant HyD-Crys (three of them), five trajectories starting from different initial configurations were run for each for better statistics.

The NAMD2 MD program²⁹ was utilized for our MD simulations, with the all-atom CHARMM (parameter set c32b1) force field³⁰ for proteins and urea. A slightly modified TIP3P water model was used for water.³¹ All production MD simulations were performed under NPT ensemble, with both temperature and pressure controlled by the Berendsen methods. Considering that the HyD-Crys is extremely stable at physiological

temperature,^{18, 20} the simulation was performed at high temperature (425 K) and 1 atm. Bonds were constrained by SHAKE.³² A time step of 2 fs was used. The particle-mesh Ewald (PME) method³³ was applied to treat the long-range electrostatic interactions, and a typical cutoff of 12 Å was used for van der Waals interactions. The periodic boundary conditions (PBC) were applied in all directions. The time step for all production runs was 2.0 fs. The simulation lengths were at least 200 ns for each simulated system. The total aggregated simulation time for this study is about 6.0 μs.

Results and Discussion

The role of the highly conserved β -hairpin Tyr pairs (Fig. 1 A) in the folding and stability of the long-lived HyD-Crys is assessed with large-scale MD unfolding simulations. HyD-Crys formed *in utero* represents one of the oldest and most stable proteins in the human body. Previous experiments have shown that its half life time of the initial unfolding step is about ~19 years (wild-type),³⁴ its thermal denaturation midpoint is up to 83.8 °C,²³ and it is resistant to even high concentration of urea denaturant.² Therefore, it is very difficult to directly study its unfolding behavior and structural stability under normal conditions. A very stable trajectory was shown in a control run for the wild-type at 310K in water (Fig. S1). In this paper, following our previous chemical/thermal denaturation studies for the wild-type HyD-Crys,^{19, 20} we have performed our unfolding simulations in 8 M urea, at 425 K and 1 atm. We have also recently simulated the effect of replacing the buried tryptophans with kynurenes, as a model of UV-induced photodamage.³⁵

Faster unfolding speed in mutant HyD-Crys proteins than the wild-type

We first study the overall unfolding process of the wild-type and all the mutant proteins to access the influence of the various site-specific mutations. We also compare the unfolding speed of each individual domain. The unfolding process can be illustrated via plotting the fraction of native contacts Q of each domain with respect to the simulation time. Here, a native contact between two residues is counted if any heavy atom of residue i is within 6.5 Å of any heavy atom of another residue j ($j - i > 3$) in the crystal structure of the protein. For the folded state, $Q \approx 1$, and for the fully stretched state, $Q \approx 0$, thereby the time evolution of Q value for each domain reflects its corresponding unfolding dynamics.

As shown in Fig. 2, for all the mutants, the N-td unfolds faster than that in the wild-type. For the wild-type, the timescale for Q (N-td) to decrease to < 0.3 (N-td loses most of its structures) is ~150 ns. Whereas, for the Y45A/Y50A_N-td, Y133A/Y138A_C-td and 4Y>4A mutant proteins, it takes ~70, ~50, and ~40 ns for Q (N-td) to decrease to < 0.3 , respectively, which is much quicker than that in the wild-type. Different trajectories show consistent results (more below). Meanwhile, we observe that the effect of the N-td substitutions (Y45A/Y50A_N-td mutant) on the unfolding of the C-td is negligible (C-td remains largely intact). However, for the Y133A/Y138A_C-td and 4Y>4A mutants, the C-td also unfolds much faster as compared to the wild-type and Y45A/Y50A_N-td mutant (see Fig. 2 C and D red curve).

Fig. 3 shows another perspective of the fraction of native contacts by plotting Q (N-td) and Q (C-td) against each other, for all five independent simulation trajectories of each system. The results indicate that for each simulated system (wild-type HyD-Crys or mutants), the unfolding pathways of the N-td or C-td for all five independent trajectories are very similar, though the exact time for each event could be slightly different. Overall, our current simulation results clearly show that: the Y45A/Y50A_N-td mutant accelerates the unfolding of the N-td, but keeps the C-td intact, while the Y133A/Y138A_C-td mutant not only accelerates the unfolding of the N-td, but also the C-td. As for the 4Y>4A mutant, it results in additional disruption to both the C-td and N-td, with an even faster unfolding speed than

the Y133A/Y138A_C-td mutant. Moreover, we observe a clear “two-stage” unfolding process for the wild-type and all mutants (with the N-td unfolds first, followed by the C-td) (see Figure 2 and 3), which implies the formation of partially unfolded intermediate structures with intact C-td and unfolded N-td. These results are fully consistent with previous site-directed mutagenesis studies.¹⁸

Similar unfolding pathway for the wild-type and mutant proteins

In Fig. 4B we compare the root-mean-square fluctuation (RMSF) of each residue (represented by C_{α}) for the wild-type and the three mutants in the early stage of unfolding, i.e., before the β -sheets start to break up (≤ 30 ns in our simulations). The RMSF results show that the loop regions are most flexible and the mutant proteins undergo much larger fluctuations than the wild-type, notably in the loop regions of the mutation sites. This is because Y45/Y50 or Y133/Y138 pair in the wild-type packs better with the hydrophobic core at the interface from being “attacked” by water than the corresponding Ala pair [bottom left figures in Fig. 1 (A–D)], and also because the local stable sandwich-like Y45-R79-Y50 or Y133-R167-Y138 cluster formed by the cation- π interactions is disrupted due to the Ala substitutions [see Fig. 1 (A–D) right column. We note that it has been reported that the CHARMM force field used here can properly model the cation- π interactions even though it is non-polarizable force field³⁶]. These results indicate that the unfolding process for the wild-type and all the mutants may start from these unstructured loop regions.

To clarify the differences in the unfolding behavior between the wild-type and the three mutants, we then examine the time evolution of the structures during the unfolding process for each system. Since HyD-Crys consists of mainly β -sheets (β -sheets content $> 80\%$), we use the fraction of native contacts Q of each β -sheet to assess their respective persistence (or stability) as time progresses. In Fig. 4C–F, the top panel of each figure shows the secondary structural components of each β -sheet as function of time for the wild-type and three mutants, respectively.

For the wild-type (Fig. 4C), the major β -sheets in the C-td are fairly stable till the very end of the simulation, while those in the N-td are less stable, with both β -sheet 2 and 3 in motif 1 losing their secondary structures completely at ~ 70 and ~ 180 ns, respectively, and β -sheets in motif 2 starting to fully unfold at ~ 200 ns (Fig. 4C top). Interestingly, motif 2 shows very similar unfolding trend with the hydrophobic core at the interface formed by motifs 2 and 4 (Fig. 4C top and middle), while motif 1 shows very similar unfolding process with the hydrophobic core of the N-td (Fig. 4C top and bottom red curve). These results imply that motif 4 in the C-td contributes significantly to the stability of motif 2, and the hydrophobic core within the N-td also makes important contribution to the stability of motif 1.

Similar to the wild-type, the stability of the C-td in the Y45A/Y50A_N-td mutant (Fig. 4D), is largely unaffected. However, the stability of the N-td is noticeably weakened, especially for motif 2. We also note that in this mutant, the hydrophobic core at the interface is quickly disrupted at the beginning of the simulation and becomes mostly breached at ~ 60 ns (faster than that in the wild-type). The same is true for motif 2, which is largely destroyed after only ~ 40 – 60 ns, again faster than the wild-type. Interestingly, motif 1 and the hydrophobic core within the N-td also show a highly correlated unfolding pathway, but with their unfolding speed comparable to the wild-type. These results confirm that the hydrophobic core within the N-td and at the interface largely contribute to the stability of motif 1 and motif 2, respectively.

For the Y133A/Y138A_C-td mutant (Fig. 4E), the stability of both N-td and C-td is remarkably reduced, albeit it is more stable for the C-td than the N-td. In the N-td, motif 2 and the hydrophobic core at the interface, as well as motif 1 and the hydrophobic core within

the N-td, are again found to unfold synchronously. Here, we find that the unfolding speed is much faster than both the wild-type and Y45A/Y50A_N-td mutant. In the C-td, as the hydrophobic core of the domain loses integrity gradually, both β -sheets in motifs 3 and 4 slowly disappear correspondingly as well. These results suggest that the stability of motifs 3 and 4 highly correlates with the hydrophobic core within the C-td. The higher stability of the C-td with respect to the N-td is mainly caused by C-td's more compact hydrophobic core.

As for the 4Y>4A mutant (Fig. 4F), most of the unfolding features are very similar to those of Y133A/Y138A_C-td mutant. For example, motif 2 and the hydrophobic core at the interface unfold simultaneously; however, the unfolding speed in this mutant is even faster, indicating more disruptions to both C-td and N-td.

From these structural analyses for the wild-type and all the three mutants, we again reveal an asynchronous unfolding behavior for the N-td and C-td, namely, the N-td unfolds first, followed by the C-td. Moreover, we also notice that the Y45A/Y50A_N-td mutant only accelerates the unfolding of the N-td, whereas the Y133A/Y138A_C-td mutant accelerates both the N-td and C-td unfolding. The 4Y>4A mutant results in additional disruption to both the N-td and C-td, with even faster unfolding speed. These findings indicate that the stability of motif 2 in the N-td is heavily influenced by the interdomain hydrophobic interface, while motif 1 in the N-td, and motifs 3 and 4 in the C-td, are mainly stabilized by their respective intradomain hydrophobic cores. It is interesting to note that maintaining a dry and close packed hydrophobic core is crucial for protein stability and integrity, as shown in many previous protein folding and peptide self-assembly studies.^{37–40} It has also been recently shown that the nanoscale dewetting (drying) within the hydrophobic core provides a significant driving force for the protein folding and is important for protein stability;^{28, 41–47} while the intrusion of water (or other solvents) into the protein interior usually acts as the initial step for protein unfolding.^{48–54}

The Tyr-to-Ala pair mutations disrupt the hydrophobic core at the interdomain interface

To further investigate how the Tyr-to-Ala pair mutations in both domains give rise to the aforementioned different unfolding behavior for the N-td and C-td, we examine the structural details of the interdomain interface of the wild-type protein. We observe that a well-defined hydrophobic core (blue wireframe in Fig. 1A left column) is shielded by a loosely connected hydrophilic “residual circle” (bold bond in Fig. 1A left column). The Y45, Y50, Y133, and Y138 residues are members belonging to this hydrophilic “residual circle”. More importantly, their bulky planar phenol side-chain can tightly contact with their neighboring hydrophobic residues, and thus act as the solvent blocker for the interdomain interfacial hydrophobic core (Fig. 5 A, and C–E). The amphiphilic nature of the Tyr side chain and its relatively large size can make favorable interactions with both polar residues/waters (outside the dry cavity) and non-polar residues (inside the dry cavity), thus smoothly bridging the polar (outside) and non-polar (inside) regions. When these pairs, Y45/Y50, or Y133/Y138, or both, are substituted with the Ala pairs, apparent water leakages to the interdomain hydrophobic core emerge due to the much smaller side chain of alanines, as shown in Fig. 5 (A, B and F–H).

We then studied the local residual contact network of residue-sites 45, 50, 133, and 138 for both the wild-type and mutants, by calculating their local native contact probability with other residues in the first 30 ns, the period when most of the secondary structural components of the proteins are largely preserved. We observe that for the Y45A/Y50A_N-td mutant, the local contacts between A45 with L144 and M146, as well as A50 with S20, D21, H22, P23, N24, M43, S77, and C78 disappear with time. Whereas for the Y133A/Y138A_C-td mutant, the contacts between A133 with F56, as well as A138 with D107, C108, S109, C110, V131, Q164, L166, and R168 also disappear. Likewise, for the 4Y>4A

mutant, the contact probabilities of A45 and A50 with their neighbor residues are lost, similar to those in Y45A/Y50A_N-td mutant, and the contact probabilities of A133 and A138 are also lost just like those in Y133A/Y138A_C-td mutant, which implies that there are additive leakages generated. These results clearly show that the functions of Y45/Y50 and Y133/Y138 pairs in protecting the well-defined interdomain hydrophobic core in the wild-type are disrupted by the Tyr-to-Ala pair(s) mutations. In the previous section, we have demonstrated that motif 2 in the N-td is mainly stabilized by the dry and well-defined hydrophobic core at the interface. Thereby, this explains why the Tyr-to-Ala pair(s) mutations can significantly increase the unfolding speed of motif 2.

Moreover, we also observe that for the wild-type, the contacts among Y50, R79, and Y45, and contacts among Y138, R167, and Y133, are very well kept. However, in the Y45A/Y50A_N-td mutant, the contact probabilities among A50, R79, and A45 decrease sharply, particularly those between A50 and R79, and A50 and A45. On the other hand, in the Y133A/Y138A_C-td mutant, the contact probabilities among A138, R167, and A133 also reduce significantly, with only the contact probability between A133 and R167 remains somewhat unchanged. In addition, the contact probabilities among Y50, R79 and Y45 also decrease largely. Finally, for the 4Y>4A mutant, we find that the contact probabilities among A50, R79, and A45, and among A138, R167, and A133 all reduce significantly. The main reason for this is the loss of the stable “sandwich-like” clusters, Y45-R79-Y50, or Y133-R167-Y138, or both, due to the alanine substitutions. These clusters in the wild-type are formed through the so-called cation- π interactions, similar to the previous findings in protein lysozyme where Arg-Trp-Arg sandwich clusters contributed significantly to the protein stability.^{24, 55} The loss of these tightly bound clusters makes the loop regions near the residue A50 or A138 or both more flexible (more discussion below). These findings are also consistent with the insights gained from above analyses of the flexibility and fluctuations (see Figure 4B).

Water “leakages” to the dry hydrophobic cores due to Tyr-to-Ala mutations

We then monitored the water solvation dynamics near the hydrophobic cores at the interdomain interface (interdomain hydrophobic core) and within each domain (intradomain hydrophobic core) of the protein, for both the wild-type and the three mutants. L44 and L132 were selected as our target residues to illustrate the dynamics of water intrusion into the intradomain hydrophobic cores of the N-td and C-td, respectively, since they are deeply buried in the very central regions of the respective hydrophobic cores (Fig. 7A). Meanwhile, residue M43 was chosen as the target for studying the penetration of water into the interdomain hydrophobic core (Fig. 7A). A water molecule is considered to be in contact with a target residue if its oxygen atom is within 4.0 Å of any heavy atom in the side chain of the target residue.

Fig. 7 B–D shows the solvation dynamics of residues M43, L44, and L132, respectively, for both the wild-type and the mutants. The timescale for residue M43 to be fully hydrated is about 200, 60, 40, and 20 ns for the wild-type, Y45A/Y50A_N-td, Y133A/Y138A_C-td, and 4Y>4A mutants, respectively (Fig. 7 B). These results clearly show that Tyr-to-Ala pair(s) mutant can directly induce water (or other solvent) more readily to intrude into the interdomain hydrophobic core. Interestingly, the timescale for residue L44 (representing the intradomain hydrophobic core in N-td) to be fully hydrated is comparable with those of residue M43 for all cases (Fig. 7 C). This may be because the loss of the hydrophobic core at the interface destabilizes motif 2 of the N-td, which subsequently results in the exposure of L44 to water (it can also be seen from their close positions in sequence). This simultaneous co-occurrence of the water intrusion into the interdomain and N-td’s intradomain hydrophobic cores also explains why the N-td is less stable and unfolds once the interdomain hydrophobic core is disrupted. On the other hand, the target residue L132 in the

C-td's intradomain hydrophobic core keeps dry till the very end of the simulation in both the wild-type and Y45A/Y50A_N-td mutant. It begins to fully hydrate at about 100 and 50 ns for the Y133A/Y138A_C-td and 4Y>4A mutants, respectively. To further confirm the water solvation dynamics of these key residues M43, L44, and L132, we also analyzed their solvent accessible surface (SAS) areas as a function of simulation time (Fig. S2). Our results show that the patterns in the time evolution of these SAS areas for the WT and mutants are nearly identical to their water solvation dynamics (contacting water numbers, Figure 7). For example, the characteristic times for these residues starting to be largely exposed to water, or being fully hydrated, are nearly identical in these two plots (see Fig. S2 and Fig. 7 for details), indicating both are good metrics for investigating the wetting/drying phenomena. As further shown in supporting figures Fig. S3 C and D, at 80 and 40 ns, some apparent water leakages to the C-td's intradomain hydrophobic core emerge near the A138 loop regions. At 120 and 60 ns, a few water molecules reach to the residue L132. And at 140 and 80 ns, residue L132 becomes fully hydrated in the Y133A/Y138A_C-td and 4Y>4A mutants, respectively. These results indicate that the replacement of Y133/Y138 with A133/A138 causes the relative stable "sandwich-like" Y133-R167-Y138 cluster to disappear, thus making the loop regions near the residue A138 more flexible. As a result, the larger fluctuations in the loop regions near A138 facilitate the intrusion of water into the intradomain hydrophobic core, which initiates the unfolding of the C-td.

Conclusion

Both experimental and theoretical considerations identified two homologous β -hairpin Tyr pairs at the domain interface as contributing to the folding and stability of the long-lived H γ D-Crys.¹⁸ Here we use extensive all-atom explicit solvent molecular dynamics simulations to assess the contributions of the two tyrosine pairs. Our results show that for the N-td substitutions, with Y45/Y50 pair substituted by A45/A50, the unfolding of the N-td is accelerated, but the C-td remained largely intact. While for the C-td substitutions, with Y133/Y138 pair substituted with A133/A138, both N-td and C-td unfolded much faster. An asynchronous unfolding pathway is revealed, with the N-td unfolding first, followed by the C-td. These findings are in excellent agreement with previous experimental results.¹⁸ Furthermore, detailed structural analyses suggest that the stability of motif 2 in the N-td is largely determined by the interdomain interface, while motif 1 in the N-td, motifs 3 and 4 in the C-td, are mainly stabilized by their respective intradomain hydrophobic cores.

Local contact analyses also show that the bulky planar phenol side-chains of Y45/Y50 and Y133/Y138 pairs can tightly bind with their neighbor residues, and act as water blocker for the interdomain hydrophobic core. The amphiphilic nature of the Tyr side chain and its large size can make favorable interactions with both polar residues/waters and non-polar residues, thus bridging smoothly the polar (outside) and non-polar (inside) environments. Damages to any tyrosine pair(s) can directly introduce apparent water leakages to the interdomain hydrophobic core, which in turn causes the N-td to lose its stability. The simultaneous water intrusion into both the interdomain and N-td's intradomain hydrophobic cores explains why the N-td is less stable and unfolds first in general, once the interdomain hydrophobic core is disrupted. On the other hand, for the C-td substitutions, the stable "sandwich-like" Y133-R167-Y138 cluster in the wild-type through the strong cation- π interactions is further destroyed, which results in larger fluctuations in the loop regions near A138. This larger loop fluctuation then gives rise to the water intrusion into C-td's intradomain hydrophobic core and finally promotes the C-td to fully unfold. This finding is similar to what we observed when tryptophans were replaced by kynurenines, with water entering into the buried cores.³⁵ Overall, these results from current large-scale MD simulations complement recent experimental results, and provide new insights into the strikingly different roles of

these two β -hairpin Tyr pairs at the interface from both domains in stabilizing H γ D-Crys, despite their high homologous nature.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Peng Xiu, Xiaowei Tang, Weiqiu Zhu, Bo Zhou, and Bruce Berne for useful discussions and insightful comments. This work is partially supported by National Natural Science Foundation of China (grant Nos. 11374221, 21320102003, and 30870593), and the China Postdoctoral Science Foundation (grant No. 2012M511351). J.K. acknowledges financial support from NIH EY015834. R.Z. acknowledges the financial support from the IBM BlueGene Science Program. A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

1. Oyster, CW. The Human Eye: Structure and Function. Sinauer Associates; Sunderland, MA: 1999.
2. Kosinski-Collins MS, King J. Protein Sci. 2003; 12:480–490. [PubMed: 12592018]
3. Papanikolopoulou K, Mills-Henry I, Tho SL, Wang Y, Gross AAR, Kirschner DA, Decatur SM, King J. Mol Vis. 2008; 14:81–89. [PubMed: 18253099]
4. Benedek GB. Invest Ophth Vis Sci. 1997; 38:1911–1921.
5. Lerman S, Borkman R. Ophthalmic Res. 1976; 8:335–353.
6. Yu NT, Barron BC, Kuck JFR. Exp Eye Res. 1989; 49:189–194. [PubMed: 2767167]
7. Palme S, Slingsby C, Jaenicke R. Protein Sci. 1997; 6:1529–1536. [PubMed: 9232654]
8. Smith MA, Bateman OA, Jaenicke R, Slingsby C. Protein Sci. 2007; 16:615–625. [PubMed: 17327390]
9. McCarty CA, Taylor HR. Dev Ophthalmol. 2002; 35:21–31. [PubMed: 12061276]
10. Davies MJ, Truscott RJW. J Photochem Photobiol B Biol. 2001; 63:114–125.
11. Aarts HJM, Lubsen NH, Schoenmakers JGG. Eur J Biochem. 1989; 183:31–36. [PubMed: 2753045]
12. Lampi KJ, Shih M, Ueda Y, Shearer TR, David LL. Invest Ophth Vis Sci. 2002; 43:216–224.
13. Hains PG, Truscott RJW. Biochim Biophys Acta. 2008; 1784:1959–1964. [PubMed: 18761110]
14. Hains PG, Truscott RJW. J Proteome Res. 2007; 6:3935–3943. [PubMed: 17824632]
15. Lampi KJ, Ma ZX, Hanson SRA, Azuma M, Shih M, Shearer TR, Smith DL, Smith JB, David LL. Exp Eye Res. 1998; 67:31–43. [PubMed: 9702176]
16. Takata T, Oxford JT, Demeler B, Lampi KJ. Protein Sci. 2008; 17:1565–1575. [PubMed: 18567786]
17. Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J. J Mol Biol. 2003; 328:1137–1147. [PubMed: 12729747]
18. Kong F, King J. Protein Sci. 2011; 20:513–528. [PubMed: 21432932]
19. Das P, King JA, Zhou R. Protein Sci. 2010; 19:131–140. [PubMed: 19937657]
20. Das P, King JA, Zhou RH. Proc Natl Acad Sci U S A. 2011; 108:1051410519.
21. Flaugh SL, Kosinski-Collins MS, King J. Protein Sci. 2005; 14:569–581. [PubMed: 15722442]
22. Flaugh SL, Kosinski-Collins MS, King J. Protein Sci. 2005; 14:2030–2043. [PubMed: 16046626]
23. Flaugh SL, Mills IA, King J. J Biol Chem. 2006; 281:30782–30793. [PubMed: 16891314]
24. Eleftheriou M, Germain RS, Royyuru AK, Zhou RH. J Am Chem Soc. 2006; 128:13388–13395. [PubMed: 17031950]
25. Huang XH, Hagen M, Kim B, Friesner RA, Zhou RH, Berne BJ. J Phys Chem B. 2007; 111:5405–5410. [PubMed: 17439169]
26. Li X, Li JY, Eleftheriou M, Zhou RH. J Am Chem Soc. 2006; 128:12439–12447. [PubMed: 16984193]

27. Zhou RH, Berne BJ, Germain R. *Proc Natl Acad Sci U S A*. 2001; 98:14931–14936. [PubMed: 11752441]
28. Zhou RH, Huang XH, Margulis CJ, Berne BJ. *Science*. 2004; 305:1605–1609. [PubMed: 15361621]
29. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. *J Comput Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
30. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. *J Comput Chem*. 2009; 30:1545–1614. [PubMed: 19444816]
31. Neria E, Karplus M. *J Chem Phys*. 1996; 105:10812–10818.
32. Ryckaert JP, Ciccotti G, Berendsen HJC. *J Comput Phys*. 1977; 23:327–341.
33. Darden TA, York DM, Pedersen LG. *J Chem Phys*. 1993; 98:10089–10092.
34. Mills-Henry, IS. PhD. Massachusetts Institute of Technology; 2007.
35. Xia Z, Yang ZX, Huynh T, King JA, Zhou RH. *Sci Rep*. 2013;3.
36. Woolf TB, Grossfield A, Pearson JG. *Int J Quantum Chem*. 1999; 75:197–206.
37. Dill KA. *Biochemistry*. 1990; 29:7133–7155. [PubMed: 2207096]
38. Wei GH, Shea JE. *Biophys J*. 2006; 91:1638–1647. [PubMed: 16766615]
39. Wolynes PG, Onuchic JN, Thirumalai D. *Science*. 1995; 267:1619–1620. [PubMed: 7886447]
40. Dobson CM, Karplus M. *Curr Opin Struct Biol*. 1999; 9:92–101. [PubMed: 10047588]
41. Young T, Hua L, Huang XH, Abel R, Friesner R, Berne BJ. *Proteins: Struct, Funct, Bioinf*. 2010; 78:1856–1869.
42. Levy Y, Onuchic JN. *Annu Rev Biophys Biomol Struct*. 2006; 35:389–415. [PubMed: 16689642]
43. Berne BJ, Weeks JD, Zhou RH. *Annu Rev Phys Chem*. 2009; 60:85–103. [PubMed: 18928403]
44. Krone MG, Hua L, Soto P, Zhou RH, Berne BJ, Shea JE. *J Am Chem Soc*. 2008; 130:11066–11072. [PubMed: 18661994]
45. Yang ZX, Shi BY, Lu HJ, Xiu P, Zhou RH. *J Phys Chem B*. 2011; 115:11137–11144. [PubMed: 21863898]
46. Liu P, Huang X, Zhou R, Berne BJ. *Nature*. 2005; 437:159–162. [PubMed: 16136146]
47. Wu Y, Vadrevu R, Kathuria S, Yang X, Matthews CR. *J Mol Biol*. 2007; 366:1624–1638. [PubMed: 17222865]
48. Collins MD, Quillin ML, Hummer G, Matthews BW, Gruner SM. *J Mol Biol*. 2007; 367:752–763. [PubMed: 17292912]
49. Rasaiah JC, Garde S, Hummer G. *Annu Rev Phys Chem*. 2008; 59:713–740. [PubMed: 18092942]
50. Collins MD, Hummer G, Quillin ML, Matthews BW, Gruner SM. *Proc Natl Acad Sci U S A*. 2005; 102:16668–16671. [PubMed: 16269539]
51. Day R, Garcia AE. *Proteins: Struct, Funct, Bioinf*. 2008; 70:1175–1184.
52. Hua L, Huang XH, Liu P, Zhou RH, Berne BJ. *J Phys Chem B*. 2007; 111:9069–9077. [PubMed: 17608515]
53. Roche J, Caro JA, Norberto DR, Barthe P, Roumestand C, Schlessman JL, Garcia AE, Garcia-Moreno B, Royer CA. *Proc Natl Acad Sci U S A*. 2012; 109:6945–6950. [PubMed: 22496593]
54. Jha SK, Udgaonkar JB. *Proc Natl Acad Sci U S A*. 2009; 106:12289–12294. [PubMed: 19617531]
55. Zhou RH, Eleftheriou M, Royyuru AK, Berne BJ. *Proc Natl Acad Sci U S A*. 2007; 104:5824–5829. [PubMed: 17389393]

Figure 1.

The overall structure of the wild-type (A), Y45A/Y50A_N-td mutant (B), Y133A/Y138A_C-td mutant (C), and 4Y>4A mutant (D) H γ D-Crys protein. The backbone of the proteins is represented as yellow transparent cartoon, the hydrophobic core at the interface is shown with blue wireframe, and the hydrophilic residues at the interface are shown with bold bond.

Figure 2.

The time evolution of the fraction of native contacts Q for one representative runs of the wild-type (A), and mutants Y45A/Y50A_N-td (B), Y133A/Y138A_C-td (C) and 4Y>4A (D), respectively. See the main text for the definition of the native contact between any two residues.

Figure 3.

The fraction of native contacts Q for the two domains are plotted against each other, data from all five independent simulation trajectories for each type system. Each data point on plots is colored from blue to red according to its time sequence during unfolding. (A) wild-type, (B) Y45A/Y50A_N-td, (C) Y133A/Y138A_C-td, and (D) 4Y>4A. The color bar

Figure 4. (A) The definition of the serial number of β -sheets within each motif of the N-td and C-td. The two motifs within each domain are divided by the black line. (B) RMSF of each residue for four types of simulated systems at the first 30 ns. The time evolution of fraction of native contacts Q of each β -sheet (C–F, top panel), the number of contacts between hydrophobic residues at the interface (C–F, middle panel), and the number of contacts between hydrophobic residues within each domain (C–F, bottom panel) for the wild-type protein (C), and mutants Y45A/Y50A_N-td (D), Y133A/Y138A_C-td (E), and 4Y>4A (F).

Figure 5.

The sectional view [detailed structural analyses] of the interdomain interface for the wild-type (A), and 4Y>4A mutant proteins (B). The overall surface is shown in light **green** (N-td) and light **cyan** (C-td), residues Y45, Y50, Y133 and Y138 are shown in red, other hydrophobic residues are shown in **blue**. Top view (C), side view from the direction of Y45/Y55 (D), and side view from the direction of Y133/Y138 for the wild-type protein. Top view (F), side view from the direction of A45/A50 (G), and side view from the direction of A133/A138 for the 4Y>4A mutant protein. Residues Y45, Y133, A45, and A133 are shown in blue transparent surface; residues Y50, Y138, A50, and A138 are shown in green transparent surface. Residue M43 together with V131 in the center region of hydrophobic core at the interface are shown in black; while the semi-transparent red regions are the possible water leakage areas introduced by the replacement of Tyr with Ala.

Figure 6. Local residue contact probability of Y45, Y50, Y133, and Y138 for the wild-type system (A); A45, A50, Y133, and Y138 for the Y45A/Y50A_N-td system (B); Y45, Y50, A133, and A138 for Y133A/Y138A_C-td system (C); A45, A50, A133, and A138 for 4Y>4A system (D), at the first 30 ns.

Figure 7.

Locations of “target” residues M43 (**black**), L44 (**yellow**), and L132 (**magenta**) in the protein (A). The number of water in contact with residues M43 (B), L44 (C) and L132 (D), respectively. A contact between water and a residue is defined if oxygen atom of water is within 4.0 Å of any heavy atom in the side chain of that residue. (All data are smoothed over the near 50 data points)