



Published in final edited form as:

*J Phys Chem B*. 2013 December 12; 117(49): . doi:10.1021/jp405159v.

## Study of the $\gamma$ D-Crystallin Protein Using Two-Dimensional Infrared (2DIR) Spectroscopy: Experiment and Simulation

A. R. Lam<sup>‡</sup>, S. D. Moran<sup>§</sup>, N. K. Preketes<sup>‡</sup>, T. O. Zhang<sup>§</sup>, M. T. Zanni<sup>§</sup>, and S. Mukamel<sup>‡</sup>  
Department of Chemistry, University of California, Irvine, Irvine, CA 92697-2025 USA, and  
Department of Chemistry, University of Wisconsin, Madison, 1101 University Avenue, Madison,  
WI 53706 USA

A. R. Lam: alamng@uci.edu

### Abstract

Cataracts is a misfolding protein disease in which one of its major components is the  $\gamma$ D-crystallin protein. The conformational structure of the aggregated  $\gamma$ D-crystallin and the interactions that cause aggregation are largely unknown. A recent experimental two-dimensional infrared (2DIR) spectroscopy study determined that the C-terminal domain has a high propensity to form  $\beta$ -sheets whereas the N-terminal domain forms a disordered structure in the fiber state. We present a combined computational molecular dynamics (MD) and infrared spectroscopy study of the local dynamics of these domains. The computed 2DIR signals agree remarkably well with experiment. We show that both domains having a Greek key structural fold experience different electrostatic environments, which may be related to the fact that the C-terminal domain is more structurally stable than the N-terminal domain. We correlate the vibrational couplings to known energy dissipation mechanisms and reveal their origin.

### Keywords

Cataracts; Aggregation; Greek Key Motifs; Misfolding; Isotope Labelling; Molecular Dynamics

### Introduction

Many NMR, FTIR and X-ray Crystallography studies have been used to characterize the structures of the Crystallin proteins aggregates (1–5). Cataracts are misfolded protein diseases that cause blindness and affect elderly population over the age of 65 (National Eye Institute). Its treatment is invasive, expensive and performed only when the disease has severely damaged the eye lens. Pathology of human cataractous lenses reveals that cataracts are composed of protein aggregates called Crystallin proteins. These proteins are degenerative during adulthood, and are damaged by radiation exposure and other chemical modifications that obstruct the passage of light through the lens which cause blurred vision that worsens with time (6–9). Even though the causes of cataracts are known, the aggregation mechanism and the three-dimensional structure of these aggregates remain unknown.

There are three classes of Crystallins:  $\alpha$ ,  $\beta$  and  $\gamma$ -Crystallins. Their solubilities and stabilities in their native states are critical to keep the transparency of the lens for a lifetime since

Correspondence to: A. R. Lam, alamng@uci.edu.

<sup>‡</sup>University of California, Irvine

<sup>§</sup>University of Wisconsin, Madison

Crystallin are not regenerated (7, 9, 10).  $\alpha$ -Crystallins are oligomers that exhibit chaperone activity, interact with unfolded proteins and prevent aggregation (11–15).  $\beta$  and  $\gamma$ -Crystallins proteins are believed to be structural proteins involved in aggregation and share similar structures composed of antiparallel  $\beta$ -sheets. Their secondary structures are separated into two domains, labeled as N-terminal domain (N-td) and C-terminal domain (C-td) and further clustered into four subdomains called *Greek Key Motifs* that contain antiparallel  $\beta$ -strand (See Fig. 1A). A key biological difference between  $\beta$  and  $\gamma$ -Crystallin proteins is that  $\beta$ -Crystallin forms domain-swapped dimers while the  $\gamma$ -Crystallins are monomeric in solution (1–3, 16).

Wild-type  $\gamma$ -Crystallins are stable eye lens proteins *in vitro* with higher free-energies of unfolding than the other Crystallin proteins. The two domains of a  $\gamma$ -Crystallin protein interact intermolecularly across a hydrophobic domain interface that helps stabilize the protein (17, 18). Human  $\gamma$ D-Crystallin (H $\gamma$ D-Crys) is the third most abundant  $\gamma$ -Crystallin found in young human eyes. H $\gamma$ D-Crys is a 173 amino-acid protein with a high structural homology with other  $\gamma$ -Crystallins (10, 19). Its secondary structure is composed of antiparallel  $\beta$ -sheets and loops that are grouped in two domains connected by a loop linker. Mutagenesis studies revealed that single amino acid substitutions in the H $\gamma$ D-Crys sequence are linked to aggregation and juvenile-onset cataracts (20–26). The aggregation mechanism in the aging lens differs from that of the juvenile-onset cataracts because they present a large number of modifications such as Methionine and Cysteine residues oxidation, deamidation, and glycation (27–30)

Fluorescence studies in triple tryptophan mutants of H $\gamma$ D-Crys by Kosinski-Collins *et al.* (31) found that the C-td is more structurally stable than the N-td. Studies of unfolding/folding experiments under physiological conditions demonstrated that the protein refolds through sequential structuring of its domains, where the C-td folds first and domain interface residues subsequently mediate the refolding of the N-td. Unfolding/folding kinetics studies by Flaugh *et al.* (17, 18) revealed the formation of partially folded intermediate structure containing the C-td in its native state and the N-td in its random state. They further suggested that hydrophobic residues in the intercore domain are critical in the unfolding/refolding process where the refolded C-td plays the role as a nucleating center for the N-td.

Two-dimensional ultrafast spectroscopy has been used to characterize geometrically ordered structures such as hairpins, amyloid fibrils and helical dimers for many proteins (32–38). One of the advantages of this technique is its high temporal resolution; detailed information about the kinetics of the proteins can be obtained on timescale regimes inaccessible by others techniques. It can also be applied continuously to study amyloid and protein folding kinetics on timescales of months to hours. The QM/MM computation of 2D spectroscopy signals has become a valuable tool in the study of the internal dynamics occurring in biological complexes such as amyloid fibrils (37, 39–43). Often, computations are necessary to fully interpret experimental 2D spectra. A recent study of  $\gamma$ D-Crystallin using isotope labeling of domains and 2DIR spectroscopy by Moran *et al.* (44), determined that the fibril nucleation of the Crystallin occurs exclusively in the C-td. The authors further proposed a structural model of the fibrils composed of a segment of  $\beta$ -sheets formed by the C-td and another segment of random structure formed by the N-td.

In this article, we present a simulation study of 2DIR spectroscopy of natively folded  $\gamma$ D-Crystallin and compared with the experimental results of Moran *et al.* (44). This comparison of experiment and simulation provides new insight into differences in how the native domains are predisposed to misfolding and aggregation.

To the best of our knowledge, this is the first successful comparison of 2DIR simulation with experiment on a protein of this size. We find excellent agreement between the two. Details of the experimental conditions may be found in ref. (44).

## Materials and Methods

### Protein Model

The H $\gamma$ D-Crys crystal structure at 1.25 Å resolution was solved and is available in the Protein Data Bank (PDB ID: 1HKO) (19). Its sequence consists of 173 amino acids which are grouped in the N-td and C-td domains which each has 2 sub-groups called *Greek Key Motif* with the labels: N-td1(2–40), N-td2(41–83), C-td3(88–128), and C-td4 (129–171) as shown in 1A. In order to mirror the experimental studies, the mutant S84C H $\gamma$ D-Crys(S84C-H $\gamma$ D-Crys) was created and the MET residue added in their N-Terminus using the MODELLER (45) package.

### Simulation Protocol

Molecular dynamics (MD) trajectories of the S84C-H $\gamma$ D-Crys were obtained by using the NAMD Package 2.6 (46) with the CHARMM27 force field (47). Simulations were carried out in aqueous solutions (TIP3P water model) (48) with cubic periodic boundary conditions. The particle-mesh Ewald protocol was used to calculate the long-range electrostatic interactions and a 12 cutoff distance was used for nonbonded interactions. 20 ns MD simulations were performed after 2 ns equilibration with a time step of 1 fs at T=300K. 100 MD snapshots with a time interval of 2 ps from the last 2 ns were harvested to compute the signals. The simulated linear absorption (LA) and the 2DIR signals were computed using the SPECTRON package (49, 50).

A 2DIR experiment uses four input pulses  $\mathbf{k}_1$ ,  $\mathbf{k}_2$ ,  $\mathbf{k}_3$  with  $\mathbf{k}_s$  satisfying  $\mathbf{k}_s = \pm\mathbf{k}_1 \pm\mathbf{k}_2 \pm\mathbf{k}_3$  that interact with the protein sample with delays  $t_1$ ,  $t_2$  and  $t_3$ . The signal is defined as the transmission change of the  $\mathbf{k}_s$  beam is recorded. The signals were computed using the Fluctuating Vibrational-Exciton Hamiltonian (FVE) (51) which is based on the Frenkel Exciton Model. The FVE Hamiltonian,  $H_{FVE}$ , has the form:

$$\hat{H}_{FVE} = \hat{H}_S + \hat{H}_F, \quad (1)$$

where  $H_S$  is the hamiltonian of the system:

$$\hat{H}_S = \sum_m \varepsilon_m \hat{B}_m^\dagger \hat{B}_m + \sum_m J_{mn} \hat{B}_m^\dagger \hat{B}_n - \frac{1}{2} \sum_m \Delta_m \hat{B}_m^\dagger \hat{B}_m^\dagger \hat{B}_m \hat{B}_m, \quad (2)$$

and  $H_F$  is the interaction with the optical field,  $E(t)$ :

$$\hat{H}_F = -\mathbf{E}(t) \cdot \sum_m \mu_m (\hat{B}_m^\dagger + \hat{B}_m). \quad (3)$$

$\hat{B}_m^\dagger$  and  $\hat{B}_m$  are the bosonic annihilation and creation operators for the  $m$ -th amide I mode,  $\varepsilon_m$  is the frequency,  $\Delta_m$  is the anharmonicity,  $J_{mn}$  are the harmonic intermode couplings between modes  $m$  and  $n$ , and  $\mu_m$  is the transition dipole moment.  $J_{mn}$  remains the same for all isotopomers, as known from experiments on H $\gamma$ D-Crys.

$H_S$  and  $H_F$  were evaluated for each snapshot using the fluctuating parameters for each amide I unit using the vibrational map of Hayashi et al. (52). Vibrational maps connect the frequencies and transition dipoles of molecular vibrations with the fluctuating local

electrostatic environment.  $\Delta_m = 16 \text{ cm}^{-1}$  were used from the anharmonicities (53). The intermode coupling,  $J_{mm}$ , is calculated using the transition dipole coupling scheme (54). Once the Hamiltonian is constructed and diagonalized, the  $S_{\nu_4\nu_3\nu_2\nu_1}^{\mathbf{k}_s}$  signal is computed using equations 115, 119 and 121 in ref. (50):

$$S_{\nu_4\nu_3\nu_2\nu_1}^{\mathbf{k}_s}(\Omega_3, t_2=0, \Omega_1) = 2i \sum_{e_4 \dots e_1} \langle \mu_{e_4}^{\nu_4} \mu_{e_3}^{\nu_3} \mu_{e_2}^{\nu_2} \mu_{e_1}^{\nu_1} \rangle_0 I_{e_1}^*(-\Omega_1) I_{e_4}(\Omega_3) \times \\ \Gamma_{e_4 e_3 e_2 e_1}(\Omega_3 + \varepsilon_{e_1} + i\gamma_{e_1}) \mathcal{J}_{e_3 e_2}(\Omega_3 + \varepsilon_{e_1} + i\gamma_{e_1}), \\ I_e(\Omega) = \frac{i}{\Omega - \varepsilon_e + i\gamma_e}, \\ \mathcal{J}_{e e'}(\omega) = \frac{i}{\omega - \varepsilon_e - \varepsilon_{e'} + i(\gamma_e + \gamma_{e'})}.$$

where  $I_e(\Omega)$  and  $\mathcal{J}_{ee'}$  are the one and two-exciton Green's functions, respectively. In this study, we set the broadening factor  $\gamma_e = 5.5 \text{ cm}^{-1}$ .  $\Gamma_{e_4 e_3 e_2 e_1}(\Omega)$  is the exciton scattering matrix with two incoming ( $e_1$  and  $e_2$ ) and two outgoing excitons ( $e_3$  and  $e_4$ ) as defined in ref. (50), equation 325.

The signals are plotted by using two-dimensional Fourier transform of the times  $t_1 \rightarrow \Omega_1$  and  $t_3 \rightarrow \Omega_3$ . The signals can be either non-chiral or chiral depending on the polarization configuration of the laser pulses. In the dipole approximation, there are three independent signals to leading zeroth order in  $\mathbf{k}$ :  $xyxy$ ,  $xyyx$ , and  $xyyy$  (where x and y are the polarization of the laser pulses). The absorptive all-parallel signal,  $xxxx$ , is given by the linear combination  $xxxx = xyxy + xyyx + xyyy$ . Some combinations of non-chiral induced signals can be used to enhance desired features. The  $xyxy$ - $xyyx$  combination suppresses the population pathways and reveals the coherent quantum dynamics through crosspeaks (50). The combination  $xyyy$ - $xyxy$  highlights the population dynamics and eliminates the coherent pathways in stimulated emission. For  $t_2 = 0$ , we have  $xyxy = xxyy$ , and the combination  $xyxy$ - $xyyx$  would only reveal excited state absorption.

We have simulated the absorptive 2DIR signals  $S^{\mathbf{k}_I} + S^{\mathbf{k}_{II}}$ . Here  $\mathbf{k}_I = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$  and  $\mathbf{k}_{II} = \mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3$ . We focused on the amide I band, which corresponds to the backbone carbonyl vibrational motions. The  $xxxx$ , the  $xxyy$  signals are compared with experiment. We further predict the combination  $xyxy$ - $xyyx$  signal. The 2D non-chiral spectra used to compare with the experimental signals are plotted using a linear scale.

## Results

### Linear Absorption Spectra of the $\gamma$ D-Crys Protein

The linear absorption (LA) signals in the IR regime for the  $\gamma$ D-Crys protein with its respective labeled domains are displayed in 1B. The labeling protocol was implemented in this study as described by Moran *et al.* (44): One domain is  $^{13}\text{C}$  isotopically labeled and the other domain remains  $^{12}\text{C}$  intact. In the LA signal, the unlabelled protein displays a peak centered at  $1638 \text{ cm}^{-1}$ . When one of the domains is isotope labeled, the signal acquires a bimodal shape with two peaks around at  $1595 \text{ cm}^{-1}$  and  $1638 \text{ cm}^{-1}$  (1B). The LA signal for the labeled Ct-d shows that the peak at  $1638 \text{ cm}^{-1}$  moves slightly to the higher frequency making the bimodal shape more pronounced than that of the labeled Nt-d.

### 2DIR Signals

The 2DIR signals for the S84C-H $\gamma$ D-Crys and their respective labeled domain samples were computed at  $t_2 = 0$ . The experimental and simulated all parallel  $xxxx$  spectra of the mutant protein and its respective labeled domains are compared in 2. For the unlabeled and labeled full protein (2,  $^{12}\text{C}$  Nt/Ct and  $^{13}\text{C}$  Nt/Ct, respectively), a pair of negative and positive peaks

are observed in each spectrum: At  $\Omega_3 = 1638 \text{ cm}^{-1}$  for the unlabeled protein and at  $\Omega_3 = 1595 \text{ cm}^{-1}$  for the labeled protein. These frequencies correspond to a structure with antiparallel  $\beta$ -sheets (55). The negative peak results from ground state bleach and excited state emission pathways while the positive peak results from excited state absorption. The broadening of these peaks is due to contributions of other secondary structures formed in the protein. These peaks are reproduced in the simulations (2, right panels). These pairs are shifted around by  $40 \text{ cm}^{-1}$  which is consistent with the effective mass replacement of the carbon. The node slopes (44) were calculated for both of these states to compare the contributions of inhomogeneous versus homogeneous broadening (Table S1). A node slope of one represents a perfectly inhomogeneously broadened peak while a value of infinity represents a perfectly homogeneously broadened peak. The node slope of the unlabeled peak pair is in good agreement with experiment, while the node slope for the labeled full protein shows a 15% deviation from experiment. This deviation may be because while the isotope labeling was accounted for in the Hamiltonian (equation 2), the MD simulations used for conformational sampling did not use a heavy carbon, which may have a slight influence on the dynamics.

For the proteins with labeled domain, two pairs of positive and negative peaks appear in the spectra. The experimental spectra of the labeled domains (2, Nt-d and Ct-d) show two pairs of peaks near  $\Omega_3 = 1640 \text{ cm}^{-1}$  and  $\Omega_3 = 1605 \text{ cm}^{-1}$  in each spectrum. Differences between the experimental signals can be seen. The peak pairs in the labeled Nt-domain signal are less intense and more separated than those of the labeled Ct-domains (see Table S1), where the high frequency pair in the labeled Ct-domain are elongated and broader than those of in the labeled Nt-domain signal. For the low frequency pair, the labeled Nt-domain signal shows difused peaks, especially the negative peak, while the positive peak is well defined and centered at  $\Omega_3 = 1605 \text{ cm}^{-1}$ . In the labeled Ct-domain signal, the low frequency peak pair is intense and are centered at  $\Omega_3 = 1605 \text{ cm}^{-1}$ .

Some of these features are reproduced in the simulated labeled domain signals (2, right panels third and fourth rows). Both show a similar profile with two pairs of peaks separated by  $40 \text{ cm}^{-1}$ . In the labeled Nt-domain signal, the negative peaks are less broad than those of in the labeled Ct-domain signal. The peak pairs in the labeled N-td signal are more separated than those of the labeled C-td signal, as observed in experiment and the simulated linear absorption spectra. Comparing the simulated and experimental node slopes, we see that the node slopes attributed to the unlabeled domains are in good agreement with experiment while those attributed to the labeled domains show a larger deviation from experiment. Again, we believe this larger deviation from experiment may be because all MD simulations used for conformational sampling used  $^{12}\text{C}$  in the amide backbone.

We next turn to the  $xyxy$  signal (3, left panels). All spectra share similar profile as in the  $xxxx$  spectra but with less intense peaks at  $\Omega_3 = 1640 \text{ cm}^{-1}$  and  $\Omega_3 = 1605 \text{ cm}^{-1}$ . This pattern is observed in the simulated data as well. However, the peak pairs in the labeled full protein are stronger and broader than in the unlabeled full protein which indicates that the N-td follows different dynamics where its eigenstates exhibit a wider frequency range despite the high homology in the backbone geometry. We note that in the simulated spectra, we are able to better resolve the low intensity cross peaks better than in experiment because of the lack of experimental (white) noise in the simulations.

The simulated combination  $xyxy$ - $xyyx$  spectrum (4) for labeled domains share the same pairs of peaks as observed in the  $xxxx$  combination. However, the N-td shows more elongated positive peaks along the diagonal than the positive peaks in the C-td. The spectra also show cross-peaks at  $\Omega_1 = 1610, 1660 \text{ cm}^{-1}$  that are more diffuse in the Ct-Domain than the N-td.

Both experiment and simulation show clear differences in the 2DIR spectra when the N-t or C-t domain are labeled, despite their high structural homology. The experimental and simulated spectra show some weak cross peaks which suggest couplings between the backbone amide states. These couplings depend strongly on the geometry of the protein. The domains are connected by a flexible linker loop and one domain is essentially the reflection of the other domain. The secondary structure propensity for the protein displayed in 5 showed that reflection among domains especially where the propensity to form  $\beta$ -strand is high except in the residues 40–41 and 60–61 from the N-td2 motif, and their corresponding reflections in the C-td2 motif (residues 129–130 and 147–148). Other differences are observed in the turn propensities where the loop between N-td1 and N-td2 motifs are higher than that of among C-td1 and C-td2 motifs. These differences in secondary structure indicate that the interactions among the sub-domains may affect the overall stability of each domain.

The coupling matrix (shown in 6) may help reveal how the various residues are relevant in the folding and stability of the domains because the couplings are sensitive to distance and orientation. Couplings are observed from intramolecular interactions between residues that form  $\beta$ -sheets structures (anti-parallel to the diagonal). Residues in the N-td (residue indexes 1 to 83) have weaker couplings when compared with the corresponding section in the C-td (residue indexes 88 to 168). In the N-td, R36 is strongly coupled to T4 and K2. These surface residues are solvent-exposed and it is known that mutant R36S is linked to congenital juvenile onset-cataracts (56). Interestingly, Q54 and R58 show a strong coupling with those residues in the segment where R36 is located. The image residue for Q54 in the C-td is Q143. Q54 and R58 are solvent-exposed residues that form one of the  $\beta$ -strands in their respective domains and belong to the set of residues in the interface core. Deamidation of Q54 and Q143 introduce a negative charge at physiological pH into the domain interface and destabilizes H $\gamma$ D-Crys, lower the unfolding kinetic barrier and leads to the formation of partially unfolded intermediate with structured C-td and a unstructured N-td (29). The R58H mutation, which results in aculeiform cataract (20), and the R36S mutant have lower solubilities and crystallize more readily than the wild-type (19).

A cluster of couplings seen in the map corresponds to the hydrophobic interdomain region located between Nt-d2 (residues M43, Q54, F56, R79, I81) and Ct-d4 (V132, Q143, L145, and M147). These residues play a role in the stabilization of the domains and a decrease of the refolding rate of the N-td as indicated in alanine screening experiments (17).

H $\gamma$ D-Crys protein has, on the other hand, a large number of aromatic residues (14 tyrosine, 6 phenylalanine, and 4 tryptophan), few of them are lone residues and the rest are distributed in pairs or set of three residues that maintain the protein structure (57). In the coupling map, we indicate two pairs of tryptophan of interest: W42 and W68 in the N-td and their image residues, W131 and W157, in the C-td. W42 and its image W131 are buried at the bottom of their respective domains. The map indicates that W42 couples with Y28, and W131 couples with D114 but the pair (W42,Y28) displays a weaker coupling than that of the pair (W131,D114). W42 and its image W131 have coupling with W68 and W157, respectively. (W42,W68) has a weaker coupling intensity than (W131,W157) where the latter is surrounded by strong couplings generated by other residues. The other two tryptophan, W68 and its image W157, located at the top of the structure, exhibit some couplings. For W68, we have the pairs (W68,Y6), (W68,Y28) and (W68,Y50) and for W157, we have (W157,H88), (W157,F117), and (W157,W132). It is well established that UV radiation may contribute to the origin of cataract due to the absorption by aromatic residues (57). These tryptophan residues, W42, W68, W131 and W157, have been studied using fluorescence quenching (31) and determined that the local environment around W68 and W157 originate the native state quenching of the protein fluorescence emission. During the refolding/unfolding process, W42 and W68 unfolded at lower concentrations of GdnHCl than their

images residues, W131 and W157, respectively. Further analysis of unfolding and refolding of the tryptophan mutants determined an intermediate state along the wt-H $\gamma$ D-Cryst folding pathway with a structured C-td and unstructured N-td which could be a suitable conformational model for cataract formation.

Even though the two domains are geometrically homologous, they make different contributions to the 2DIR signals. The large separation between densities of states for the pair Q54 and Q143 could be due to their exposure to the solvent and their location in the interface domain which could change the strengths of the interactions with other residues. The tryptophan residues considered here are buried in their respective domains. Although both domains share basically identical backbone geometry, the frequency shifts upon domain labelling suggest that the kinetics of these domains are different where the folding in the N-td depends on the transitions occurring in the C-td, which agrees well with experimental reports (17, 30, 59)

## Discussion and Conclusions

Moran *et al.* used a combined approach of isotope labeling and 2DIR spectroscopy (44) to study the  $\gamma$ D-Crystallin protein. They demonstrated that the use of this combined tool is valuable for the characterization of the kinetics and structure in proteins that help to obtain additional information not available from standard methods. In their studies, they expressed labeled and unlabeled domains of the protein separately, and the full length of  $\gamma$ D-Crystallin protein was constituted by protein ligation using a covalent linker residue located in the connecting loop. They determined that the C-td of the  $\gamma$ D-Crystallin protein is the leading domain to form amyloid fibrils and the use of this covalent linkage between domains plays a role in determining the fibril structure which is a striking result considering that each individual domain can form fibrils on its own (58).

We note that our simulations are limited to the computation of 2DIR spectra of the monomeric form of H $\gamma$ D-Cryst, and do not offer a direct comparison with the 2DIR kinetics experiments. We used a computational approach to complement these results at the level of the native state of the mutant S84C- $\gamma$ D-Crystallin monomers. The computed spectra are in excellent agreement with experiment. Although the N-td and C-td domains are geometrically similar with highly conserved residues in their sequences, the average coupling map analysis shows an asymmetric contribution of some relevant residues and their corresponding image residue to the signals when one of the domains is isotopically labeled. Experiments made by Mills *et al.* (59) concluded that the isolated N-td was less stable than the full-length proteins and less stable than the C-td. A computational study by Das *et al.* (60) confirmed that the stability and the folding mechanism of the N-td are regulated by the interdomain interactions.

Our findings are consistent with the conclusions that Moran *et al.* stated about the role of the covalent linkage between domains in determining the fibril structure and the prevalent dynamics that the C-td has over the N-td in the fibril formation. It will be interesting to use 2DIR spectroscopy to analyze the coupling between domains at different environmental conditions and mutations.

As stated in the introduction, we can consider H $\gamma$ D-Cryst in terms of two coupled domains. In a very simple approximation, it is possible to think of this as a coupled two-mode exciton system:

$$H \approx \begin{pmatrix} \varepsilon_i & J_{ij} \\ J_{ij} & \varepsilon_j \end{pmatrix}$$

where  $\varepsilon_i$  is the mean site-energy of domain  $i$  and  $J_{ij}$  is the coupling between modes  $i$  and  $j$ . The eigenvalues are:

$$\varepsilon_{\pm} = \frac{1}{2} \left[ \varepsilon_i + \varepsilon_j \pm \sqrt{4J_{ij}^2 + (\varepsilon_i - \varepsilon_j)^2} \right]$$

When neither domain is isotope labeled or both domains are labeled,  $\varepsilon_i \approx \varepsilon_j$ , and we have (roughly) a linearly coupled homodimer. When either the N-td or the C-td is isotope labeled, we then have a linearly coupled heterodimer. In this approximation, we expect to see two peaks at  $\varepsilon_+$  and  $\varepsilon_-$  in the linear absorption spectrum. These peaks are separated by

$\Delta\omega = \sqrt{4J_{ij}^2 + (\varepsilon_i - \varepsilon_j)^2}$ . If the system is symmetric, then  $(\varepsilon_i - \varepsilon_j)^2 = (\varepsilon_j - \varepsilon_i)^2$  and the N-td and C-td labeled isotopomers will produce equivalent linear absorption spectra (61). However, as seen from the calculated linear absorption spectra, this is not the case as the peaks in the N-td labeled isotopomer are less separated compared to the peaks in the C-td labeled isotopomer. Since  $J_{ij}$  is the same for the two isotopomers, this indicates that  $(\varepsilon_i - \varepsilon_j)^2$  is greater for the C-td labeled isotopomer. This is likely related to the fact that the C-td has a better defined structure, which leads to a lower approximate site energy

While the diagonal elements of the Hamiltonian are decreased by isotopic substitution, the off-diagonal elements remain the same because the transition dipoles of the individual amide I modes, which determine the coupling, do not change upon isotopic substitution. This was verified experimentally by comparing the diagonal cross-sections of the 2DIR spectra of the unlabeled and uniformly labeled  $^{13}\text{C}$  proteins.

However, as noted by Fang *et. al* (62), the cross peaks are a direct measure of the off-diagonal an-harmonic coupling. Since we assume zero mechanical off-diagonal anharmonicity, all off-diagonal anharmonicity must result from electrostatic coupling. In the weak-coupling limit, which is appropriate here, this off-diagonal anharmonicity can be approximated as:

$$\Delta_{ij} = -\frac{4\Delta_{ii}}{(\varepsilon_i - \varepsilon_j)^2}$$

Where  $\Delta_{ij}$  is the off-diagonal anharmonicity and  $\Delta_{ii}$  is the diagonal anharmonicity ( $16 \text{ cm}^{-1}$ ). Therefore, systems with a smaller  $(\varepsilon_i - \varepsilon_j)^2$  will have stronger cross peaks. In our case, the N-td labeled isotopomer has the smaller value and therefore has stronger cross peaks.

Using this very simple model, we conclude that the asymmetry in the linear absorption and 2DIR spectra is caused by the differences in the site-energies in the two domains. The differences in these site-energies are caused by the differing local electrostatic environments of residues within these two domains and is likely related to the secondary structure propensity in each domain.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



## Acknowledgments

We gratefully acknowledge the support of the National Institutes of Health (Grant GM059230 and GM091364), and the National Science Foundation (Grant CHE-1058791). NKP is supported by a National Science Graduate Research Fellowship. MTZ acknowledges NIH DK79895.

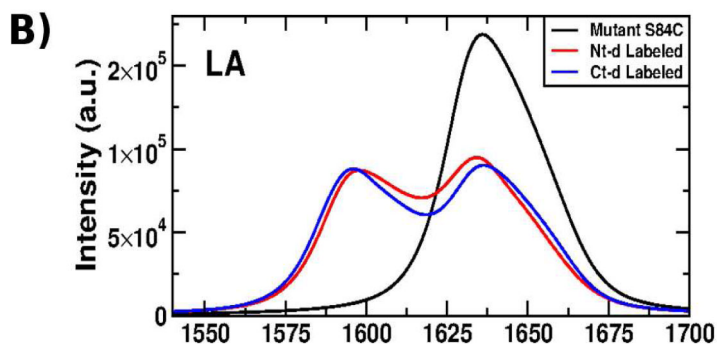
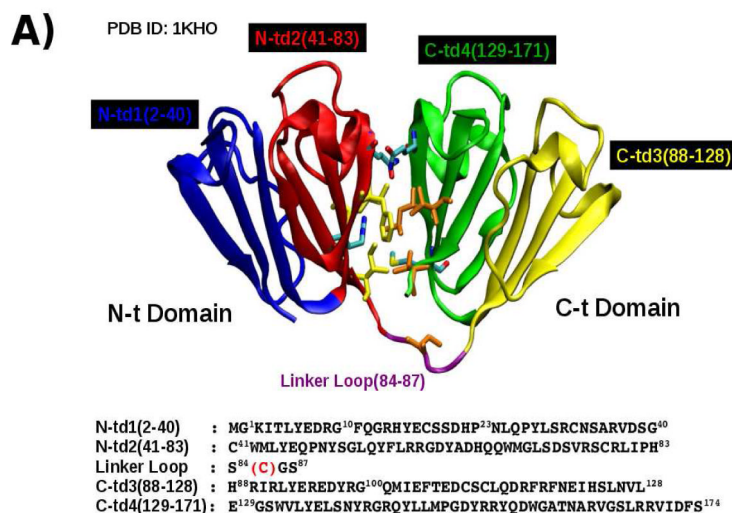
## References

1. Wistow G, Turnell B, Summers L, Slingsby C, Moss D, Miller L, Lindley P, Blundell T. X-ray Analysis of the Eye Lens Protein  $\gamma$ -II Crystallin at 1.9 Å Resolution. *J Mol Biol.* 1983; 170(1):175–202. [PubMed: 6631960]
2. Cooper PG, Carver JA, Aquilina JA, Ralston GB, Truscott RJA. <sup>1</sup>H NMR Spectroscopic Comparison of  $\gamma$ S- and  $\gamma$ B-Crystallins. *Exp Eye Res.* 1994; 59(2):211–20. [PubMed: 7835410]
3. Slingsby C, Norledge B, Simpson A, Bateman OA, Wright G, Driessen HPC, Lindley PF, Moss DS, Bax B. X-ray Diffraction and Structure of Crystallins. *Prog Ret Eye Res.* 1997; 16(1):3–29.
4. Carver JA. Probing the Structure and Interactions of Crystallin Proteins by NMR Spectroscopy. *Prog Retin Eye Res.* 1999; 18(4):431–462. [PubMed: 10217479]
5. Purkiss AG, Bateman OA, Goodfellow JM, Lubsen NH, Slingsby C. The X-ray Crystal Structure of Human  $\gamma$ S-Crystallin C-terminal Domain. 2002; 277(6):4199–4205.
6. Hoenders HJ, Bloemendal H. Lens Proteins and Aging. *J Gerontol.* 1983; 38(3):278–286. [PubMed: 6341444]
7. Slingsby C, Clout NJ. Structure of the Crystallins. *Eye.* 1999; 13(Pt3b):395–402. [PubMed: 10627816]
8. Bloemendal H, de Jong W, Jaenicke R, Lubsen NH, Slingsby C, Tardieu A. Ageing and Vision: Structure, Stability and Function of Lens Crystallins. *Prog in Biophys and Mol Biol.* 2004; 86(3):407–485. [PubMed: 15302206]
9. Graw J. Genetics of Crystallins: Cataract and Beyond. *Exp Eye Res.* 2009; 88(2):173–189. [PubMed: 19007775]
10. Jaenicke R, Slingsby C. Lens Crystallins and Their Microbial Homologs: Structure, Stability, and Function. *Crit Rev Biochem Mol Biol.* 2001; 36(5):435–499. [PubMed: 11724156]
11. Horwitz J.  $\alpha$ -Crystallin Can Function As A Molecular Chaperone. *Proc Natl Acad Sci USA.* 1992; 89(21):10449–10453. [PubMed: 1438232]
12. Groenen PJ, Merck KB, de Jong WW, Bloemendal H. Structure and Modifications of the Junior Chaperone  $\alpha$ -Crystallin. From Lens Transparency to Molecular Pathology. *Eur J Biochem.* 1994; 225(1):1–19. [PubMed: 7925426]
13. Clark JI, Muchowski PJ. Small Heat-Shock Proteins and Their Potential Role in Human Disease. *Curr Opin Struc Biol.* 2000; 10(1):52–59.
14. Andley UP. Effects of  $\alpha$ -Crystallin on Lens Cell Function and Cataract Pathology. *Curr Mol Med.* 2009; 9(7):887–892. [PubMed: 19860667]
15. Kumar PA, Reddy GB. Modulation of  $\alpha$ -Crystallin Chaperone Activity: A Target to Prevent or Delay Cataract? *IUBMB Life.* 2009; 65(5):485–495. [PubMed: 19391162]
16. Cooper PG, Carver JA, Truscott RJ. <sup>1</sup>H-NMR Spectroscopy of Bovine Lens  $\beta$ -Crystallin. The Role of the  $\beta$ B2-Crystallin C-Terminal Extension in Aggregation. *Eur J Biochem.* 1993; 213(1):321–328. [PubMed: 8477704]
17. Flaugh SL, Kosinski-Collins MS, King J. Interdomain Side-Chain Interactions in Human  $\gamma$ D Crystallin Influencing Folding and Stability. *Protein Sci.* 2005; 14(8):2030–2043. [PubMed: 16046626]
18. Flaugh SL, Kosinski-Collins MS, King J. Contributions of Hydrophobic Domain Interface Interactions to the Folding and Stability of Human  $\gamma$ D-crystallin. *Protein Sci.* 2005; 14(3):569–581. [PubMed: 15722442]
19. Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J. High-Resolution X-Ray Crystal Structures of Human  $\gamma$ D Crystallin (1.25 Å) and the R58H Mutant (1.15 Å) Associated with Aculeiform Cataract. *J Mol Biol.* 2003; 328(5):1137–1147. [PubMed: 12729747]

20. Heon E, Priston M, Schorderet DF, Billingsley GD, Girard PO, Lubsen N, Munier FL. The  $\gamma$ -Drystallins and Human Cataracts: A Puzzle Made Clearer. *Am J Hum Genet.* 1999; 65(5):1261–1267. [PubMed: 10521291]
21. Nandrot E, Slingsby C, Basak A, Cherif-Chefchaoui M, Benazzouz B, Hajaji Y, Boutayeb S, Gribouval O, Arbogast L, et al. Gamma-D Crystallin Gene (CRYGD) Mutation Causes Autosomal Dominant Congenital Cerulean Cataracts. *J Med Genet.* 2003; 40:262. [PubMed: 12676897]
22. Santhiya ST, Manohar MS, Rawlley D, Vijayalakshmi P, Namperumalsamy P, Gopinath PM, Löster J, Graw J. Novel Mutations in the  $\gamma$ -Crystallin Genes Cause Autosomal Dominant Congenital Cataracts. *J Med Genet.* 2002; 9(5):352–358. [PubMed: 12011157]
23. Sandilands A, Hutcheson AM, Long HA, Prescott AR, Vrensen G, Löster J, Klopp N, Lutz RB, Graw J, Masaki S, et al. Altered Aggregation Properties of Mutant  $\gamma$ -Crystallins Cause Inherited Cataract. *The EMBO J.* 2002; 21(22):6005–6014.
24. Evans P, Wyatt K, Wistow GJ, Bateman OA, Wallace BA, Slingsby C. The P23T Cataract Mutation Causes Loss of Solubility of Folded  $\gamma$ D-Crystallin. *J Mol Biol.* 2004; 343(2):435–444. [PubMed: 15451671]
25. Kumar M, Agarwal T, Khokhar S, Kumar M, Kaur P, Roy TS, Dada R. Mutation Screening and Genotype Phenotype Correlation of  $\alpha$ -Crystallin,  $\gamma$ -Crystallin and GJA8 Gene in Congenital Cataract. *Mol Vis.* 2011; 17:693–707. [PubMed: 21423869]
26. Zhang W, Cai HC, Li FF, Xi YB, Ma X, Yan YB. The Congenital Cataract-linked G61C Mutation Destabilizes  $\gamma$ D-Crystallin and Promotes Non-Native Aggregation. *PLoS One.* 2011; 6(5):e20564. [PubMed: 21655238]
27. Stevens A. The Contribution of Glycation to Cataract Formation in Diabetes. *J Am Optom Assoc.* 1998; 69(8):519–530. [PubMed: 9747048]
28. Hanson SR, Hasan A, Smith DL, Smith JB. The Major In Vivo Modifications of the Human Water-Insoluble Lens Crystallins are Disulfide Bonds, Deamidation, Methionine Oxidation and Backbone Cleavage. *Exp Eye Res.* 2000; 71(2):195–207. [PubMed: 10930324]
29. Flaugh SL, Mills IA, King J. Glutamine Deamidation Destabilizes Human  $\gamma$ D-Crystallin and Lowers the Kinetic Barrier to Unfolding. 2006; 281(41):30782–30792.
30. Goulet DR, Knee KM, King JA. Inhibition of Unfolding and Aggregation of Lens Protein Human  $\gamma$ D Crystallin by Sodium Citrate. *Exp Eye Res.* 2011; 93(4):371–381. [PubMed: 21600897]
31. Kosinski-Collins MS, Flaugh SL, King J. Probing Folding and Fluorescence Quenching in Human  $\gamma$ D Crystallin Greek Key Domains using Triple Tryptophan Mutant Proteins. *Protein Sci.* 2004; 13(8):2223–2235. [PubMed: 15273315]
32. Smith AW, Tokmakoff A. Amide I Two-Dimensional Infrared Spectroscopy of  $\beta$ -Hairpin Peptides. *J Chem Phys.* 2007; 126(4):045109-1–10. [PubMed: 17286519]
33. Kim YS, Liu L, Axelsen PH, Hochstrasser RM. Two-dimensional Infrared Spectra of Isotopically Diluted Amyloid Fibrils from A $\beta$ 40. *Proc Natl Acad Sci USA.* 2008; 105(22):7720–7725. [PubMed: 18499799]
34. Strasfeld DB, Ling YL, Shim S-H, Zanni MT. Tracking Fiber Formation in Human Islet Amyloid Polypeptide with Automated 2D-IR Spectroscopy. *J Am Chem Soc.* 2008; 130(21):6698–6699. [PubMed: 18459774]
35. Strasfeld DB, Ling YL, Gupta R, Raleigh DP, Zanni MT. Strategies for Extracting Structural Information from 2DIR Spectroscopy of Amyloid: Application to Islet Amyloid Polypeptide. *J Phys Chem B.* 2009; 113(47):15679–15691. [PubMed: 19883093]
36. Shim S-H, Gupta R, Ling YL, Strasfeld DB, Raleigh DP, Zanni MT. Two-Dimensional IR Spectroscopy and Isotope Labeling Defines the Pathway of Amyloid Formation with Residue-Specific Resolution. *Proc Natl Acad Sci USA.* 2009; 106(16):6614–6619. [PubMed: 19346479]
37. Reddy AS, Wang L, Lin YS, Ling Y, Chopra M, Zanni MT, Skinner JL, Pablo JJD. Solution Structures of Rat Amylin Peptide: Simulation, Theory, and Experiment. *Biophys J.* 2010; 98(3):443–451. [PubMed: 20141758]
38. Remorino A, Korendovych IV, Wu YB, DeGrado WF, Hochstrasser RM. Residue-Specific Vibrational Echoes Yield 3D Structures of a Transmembrane Helix Dimer. *Science.* 2011; 332(6034):1206–1209. [PubMed: 21636774]

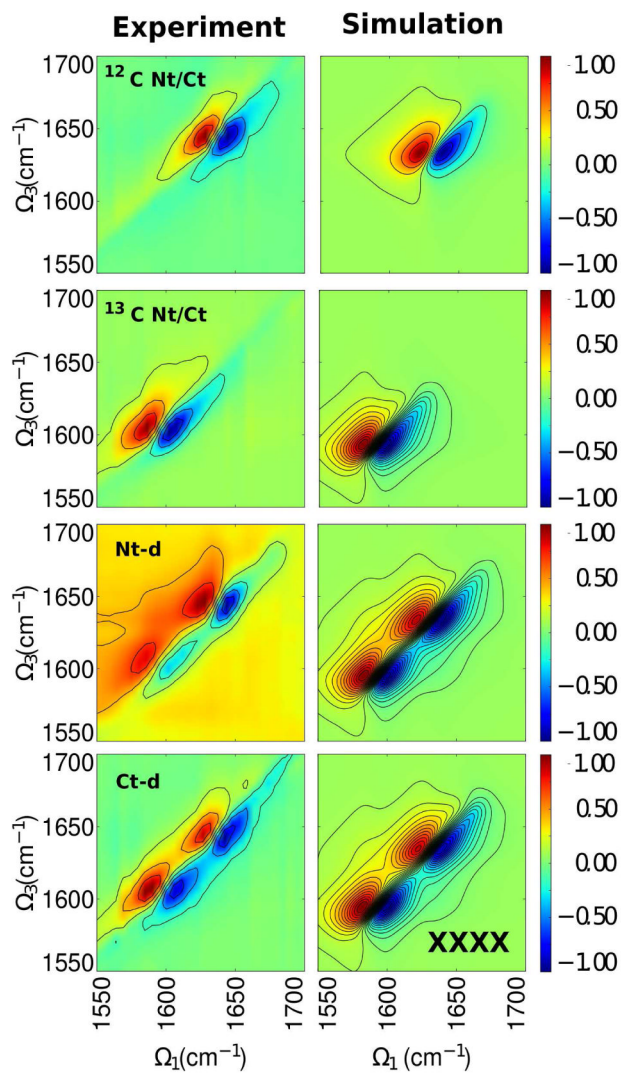
39. Zhuang WDA, Voronine DV, Mukamel S. Simulation of Two-Dimensional Infrared Spectroscopy of Amyloid Fibrils. *Proc Natl Acad Sci USA*. 2007; 104(36):14233–14236. [PubMed: 17675411]
40. Jiang J, Abramavicius D, Falvo C, Bulheller BM, Hirst JD, Mukamel S. Simulation of Two Dimensional Ultraviolet (2DUV) Spectroscopy of Amyloid Fibrils. *J Phys Chem B*. 2010; 114(37):12150–12156. [PubMed: 20795695]
41. Zhuang W, Sgourakis NK, Zhenyu L, Garcia A, Mukamel S. Discriminating Early Stage A $\beta$ 2 Monomer Structure Using Chirality-Induced 2DIR Spectroscopy in a Simulation Study. *Proc Natl Acad Sci USA*. 2010; 107(36):15687–15692. [PubMed: 20798063]
42. Lam A, Jiang J, Mukamel S. Distinguishing Amyloid Fibril Structures in Alzheimer's Disease (AD) by Two-Dimensional Ultraviolet (2DUV) Spectroscopy. *Biochemistry*. 2011; 50(45):9809–9816. [PubMed: 21961527]
43. Wang L, Middleton CT, Singh S, Reddy AS, Woys AM, Strasfeld DB, Marek P, Raleigh DP, de JJ, Pablo, Zanni MT, et al. 2DIR Spectroscopy of Human Amylin Fibrils Reflects Stable  $\beta$ -Sheet Structure. *J Am Chem Soc*. 2011; 133(40):16062–16071. [PubMed: 21916515]
44. Moran S, Woys A, Buchanan L, Bixby E, Decatur S, Zanni M. Two-dimensional IR Spectroscopy and Segmental <sup>13</sup>C Labeling Reveals the Domain Structure of human  $\gamma$ D-Crystallin Amyloid Fibrils. *Proc Natl Acad Sci USA*. 2012; 109(9):3329–3334. [PubMed: 22328156]
45. Fiser A, Sali A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. *Meth Enzymol*. 2003; 374:461–491. [PubMed: 14696385]
46. Phillips JC, Braun RWW, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable Molecular Dynamics with NAMD. *J Comp Chem*. 2005; 26(16):1781–1802. [PubMed: 16222654]
47. MacKerell AD, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Comp Chem*. 1998; 102:3586–3616.
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys*. 1983; 79:926–935.
49. Zhuang W, Abramavicius D, Mukamel S. Dissecting Coherent Vibrational Spectra of Small Proteins into Secondary Structural Elements by Sensitivity Analysis. *Proc Natl Acad Sci USA*. 2005; 102(21):7443–7448. [PubMed: 15894625]
50. Abramavicius D, Palmieri B, Voronine DV, Sanda F, Mukamel S. Coherent Multidimensional Optical Spectroscopy of Excitons in Molecular Aggregates; Quasiparticle versus Supermolecule Perspectives. *Chem Rev*. 2009; 109(6):2350–2408. [PubMed: 19432416]
51. Zhuang W, Abramavicius D, Hayashi T, Mukamel S. Simulation Protocols for Coherent Femtosecond Vibrational Spectra of Peptides. *J Phys Chem B*. 2006; 110(7):3362–3374. [PubMed: 16494351]
52. Hayashi T, Zhuang W, Mukamel S. Electrostatic DFT Map for the Complete Vibrational Amide Band of NMA. *J Phys Chem A*. 2005; 109(43):9747–9759. [PubMed: 16833288]
53. Hamm P, Lim M, Hochstrasser RM. Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy. *J Phys Chem B*. 1998; 102:6123–6138.
54. Torii H, Tasumi M. Model Calculation on the Amide-I Infrared Bands of Globular Proteins. *J Chem Phys*. 1992; 96(5):3379–3387.
55. Ganim Z, Chung HS, Smith AW, DeFlores LP, Jones KC, Tokmakoff A. Amide I two-dimensional infrared spectroscopy of proteins. *Acc Chem Res*. 2008; 41:432–441. [PubMed: 18288813]
56. Kmoch S, Brynda J, Asfaw B, Bezouška K, Novák P, ezá ová P, Ondrová L, Filipec M, Sedlá ek J, Elleder M. Link Between a Novel Human  $\gamma$ D-Crystallin Allele and a Unique Cataract Phenotype Explained by Protein Crystallography. *Hum Mol Gen*. 2000; 9(2):1779–1786. [PubMed: 10915766]
57. Kong F, King J. Contributions of Aromatic Pairs to the Folding and Stability of Long-Lived Human  $\gamma$ D-Crystallin. *Prot Sci*. 2011; 20(3):513–528.
58. Papanikolopoulou K, Mills-Henry I, Thol SL, Wang Y, Gross AAR, Kirschner DA, Decatur SM, King J. Formation of Amyloid Fibrils In Vitro by Human  $\gamma$ D-Crystallin and its Isolated Domains. *Mol Vis*. 2008; 14:81–89. [PubMed: 18253099]

59. Mills IA, Flaugh SL, Kosinski-Collins MS, King JA. Folding and Stability of the Isolated Greek Key Domains of the Long-Lived Human Lens Proteins  $\gamma$ D-Crystallin and  $\gamma$ S-Crystallin. *Prot Sci.* 2007; 16(11):2427–2444.
60. Das P, King JA, Zhou R.  $\beta$ -Strand Interactions at the Domain Interface Critical for the Stability of Human Lens  $\gamma$ D-Crystallin. *Prot Sci.* 2010; 19(1):131–140.
61. Zanni, MT.; Hamm, P. *Concepts and Methods of 2D Infrared Spectroscopy.* Cambridge University Press; 2011.
62. Fang C, Senes A, Cristian L, DeGrado WF, Hochstrasser RM. Amide vibrations are delocalized across the hydrophobic interface of a transmembrane helix dimer. *Proc Natl Acad Sci USA.* 2006; 103:16740–16745. [PubMed: 17075037]

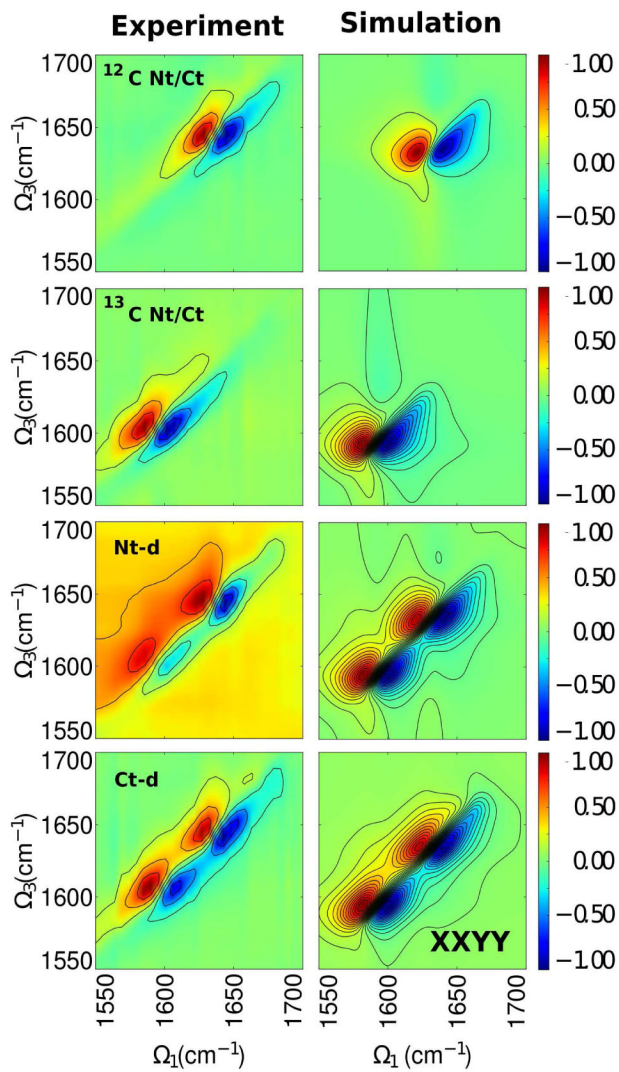


**Figure 1.**

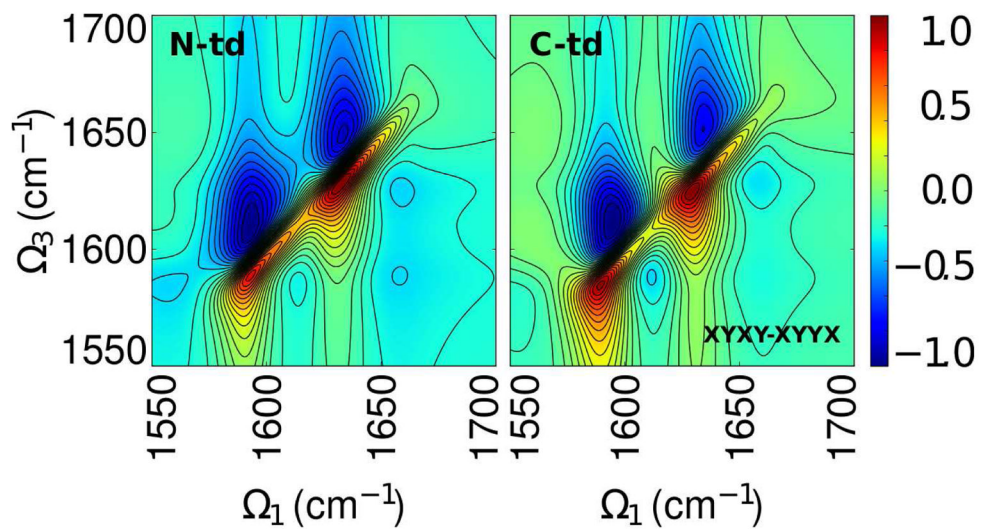
**A:** Structure of the  $\gamma$ D-Crystallin protein showing its four Greek Motifs divided into the two domains Nt (blue and red) and Ct (green and yellow) connected by a loop (purple) and their respective sequence as shown in the crystal structure (PDB ID: 1KHO) with the mutation S84C indicated (red). Residues located in the interdomains (in licorice) are responsible in the stability of the entire structure. **B:** Linear absorption (LA) (top) in the infrared regime of the mutant S84C-H $\gamma$ D-Crystallin for the full sequence (black), N-td labeled (red) and C-td labeled (blue).



**Figure 2.** All-parallel signal  $xxxx S_{k_I} + S_{k_{II}}$  experiment (left) and simulation (right). From top to bottom: full mutant protein ( $^{12}\text{C}$  Nt/Ct), labeled full mutant protein ( $^{13}\text{C}$  Nt/Ct), labeled N-td and the labeled C-td.

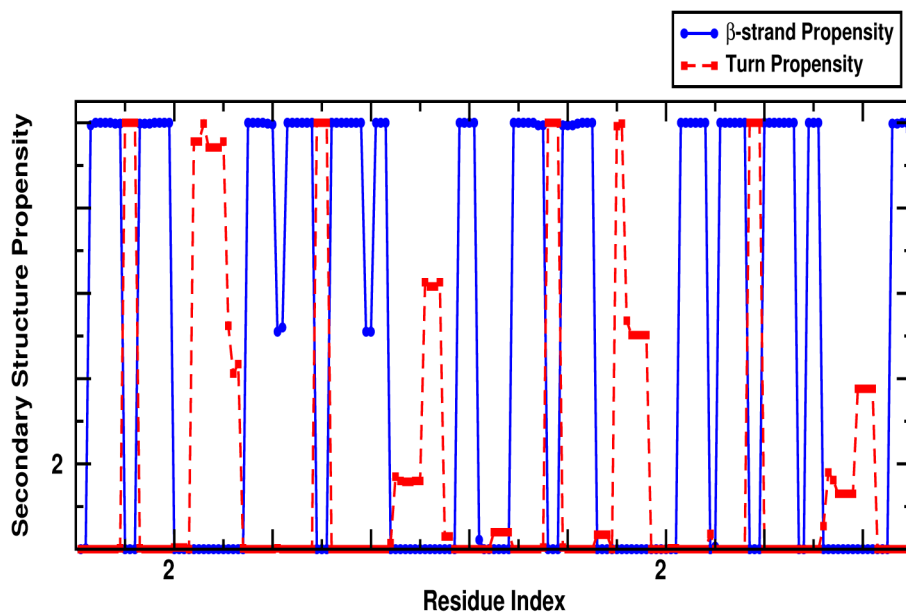


**Figure 3.** Same as in 2 but for polarization configuration  $xyyy$ .

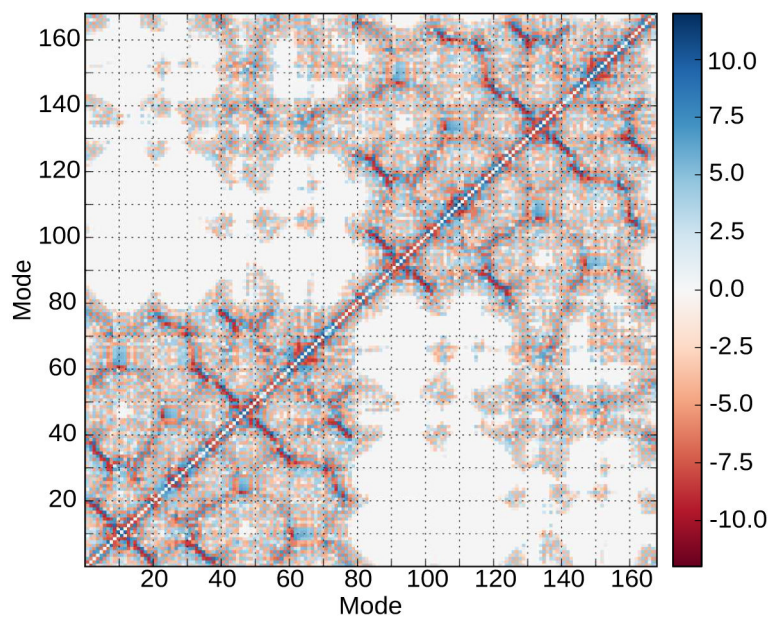


**Figure 4.** Simulated *xyxy-xyyx* signals of the isotopically labeled domains.





**Figure 5.** Secondary structure propensity of the domains of  $\gamma$ D-Crystallin indicates the segments of the sequence that have the probability to form  $\beta$ -strands or turns.



**Figure 6.** Coupling matrix of the  $\gamma$ D-Crystallin protein. Inter-domains interactions are represented by a red-blue color scale indicating their strength.