UNIVERSITY OF CALIFORNIA,
IRVINE

Increased hydrophobic surface exposure in the cataract-related G18V variant of human γS-crystallin

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Chemical and Material Physics

by

Eric Wong

Dissertation Committee:
Professor Douglas Tobias, Chair
Professor Rachel Martin
Professor Ioan Andricioaei

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my collaborators, Domarin Khago and Carolyn Kingsley, for their contribution to the experimental portion of this work as well as insightful discussion on crystallin research. Additionally, I am most grateful for the guidance and support provided by Dr. J. Alfredo Freites, Dr. Douglas J. Tobias, and Dr. Rachel W. Martin. Computational resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant ACI-1053575. Financial support was provided under the National Science Foundation grant DMR-1410415.

I would also like to thank Elsevier for permission to use their published material as content for this thesis.

The text of this thesis/dissertation is a reprint of the material as it appears in "Increased hydrophobic surface exposure in the cataract-related G18V variant of human γS-crystallin". The co-authors listed in this publication directed and supervised research which forms the basis for the thesis.

ABSTRACT OF THE THESIS

Increased hydrophobic surface exposure in the cataract-related G18V variant of human γS-crystallin

By

Eric Wong

Master of Science in Chemical and Material Physics

University of California, Irvine, 2017

Professor Douglas Tobias, Chair

The objective of this study was to determine whether the cataract-related G18V variant of human γS-crystallin has increased exposure of hydrophobic residues that could explain its aggregation propensity and/or recognition by αB-crystallin. We used an ANS fluorescence assay and NMR chemical shift perturbation to experimentally probe exposed hydrophobic surfaces. These results were compared to flexible docking simulations of ANS molecule to the solution-state NMR structures of γS-WT and γS-G18V. γS-G18V exhibits increased ANS fluorescence, suggesting increased exposed hydrophobic surface area. The specific residues involved in ANS binding were mapped by NMR chemical shift perturbation assays, revealing ANS binding sites in γS-G18V that are not present in γS-WT. Molecular docking predicts three binding sites that are specific to γS-G18V corresponding to the exposure of a hydrophobic cavity located at the interdomain interface, as well as two hydrophobic patches near a disordered loop containing solvent-exposed cysteines, all but one of which is buried in γS-WT. Characterization of changes in exposed hydrophobic surface area between wild-type and variant proteins can help elucidate the mechanisms of aggregation propensity and chaperone recognition, presented here in the context of cataract formation.

Chapter 1

Introduction

The solution-state NMR structure of wild-type γS-crystallin has been determined [16], revealing a double Greek key architecture for each of the two domains, consistent with the structures of other $\beta\gamma$-crystallins [13, 21]. The childhood-onset cataract variant G18V (γS-G18V) is structurally similar to γS-WT, but it has dramatically lower thermal stability and solublity [3, 19], as well as strong, specific interactions with αB-crystallin, the holdase chaperone of the lens [16]. Despite the well-documented aggregation propensity and reduced stability of γS-G18V, the particular intermolecular interactions leading to its aggregation are as yet unknown. Protein self-aggregation leading to cataract can occur due to an increase in net hydrophobic interactions, as previously shown in the congenital Coppock-type cataract variant D26G γS-crystallin [15], the cerulean cataract variant P23T γD-crystallin [25], acetylation of the G1 and K2 residues in γD-crystallin [6], and the lamellar cataract variant D140N αB-crystallin [18]. All of these mutations introduce altered conformations that produce lowered solubility by exposure of hydrophobic patches on the surface, even though the structural differences from their wild-type counterparts are relatively subtle. γS-G18V is no exception; the mutation does not cause large-scale unfolding or rearrangement into a misfolded conformation, but rather produces altered intermolecular interactions with

itself and with αB-crystallin [16].

The fluorescent probe 1-anilinonaphthanlene-8-sulfonate (ANS), which has both negatively charged and hydrophobic moieties, is often used to quantify exposed hydrophobic surface patches in proteins by introducing known concentrations of ANS into a protein solution and measuring its emission spectrum [25, 1, 2]. Two types of protein-ANS interactions are required for fluorescence enhancement: hydrophobic interactions between the conjugated ring system of ANS and the protein surface [20], and electrostatic interactions between the sulfonate group and positively charged side chains at the binding site [24]. An increase in fluorescence intensity indicates that either more ANS is binding to the protein surface, or that it is bound more tightly, correlating with higher surface hydrophobicity. This method has been used to characterize exposed hydrophobic surface in a number of protein systems, including the mitochondrial chaperone protein Atp11p, which recognizes its client proteins via hydrophobic interactions [28], and aggregation-prone variants of superoxide dismutase-1 (SOD1), an essential cellular enzyme whose aggregation is associated with amyotrophic lateral sclerosis (ALS) [4, 32]. Despite the utility of ANS binding as a probe of hydrophobic surface exposure, and the sensitivity afforded by using fluorescence as a reporter, this assay is limited by the lack of detailed information about which amino acid residues, or even general regions of the protein, are taking part in the dye-binding interaction. NMR chemical shift perturbation (CSP) mapping can forge a link between fluorescence enhancement upon dye binding and the corresponding changes in the local chemical environment of specific residues in the protein.

Nuclear Magnetic Resonance (NMR) is an analytical technique where structural information is gathered based on the nuclear spin characteristics of the protein of interest. By using a 2-D correlation experiment such as ^{1}H-^{15}N HSQC, chemical shifts are assigned that repre-

sent each N-H correlation within the protein, thereby identifying each residue or sidechain. Of course, there must be an existing protein structure to use the HSQC assignments. The addition of a ligand allows for determination of binding sites within the protein [38]. Using ANS titration experiments, comparisons between wild-type and variant proteins can then be used to compare differences in exposure of hydrophobic residues on the surface under particular solution conditions. CSP mapping is a commonly used technique for investigating protein-protein or protein-ligand interactions and interfaces [38], and is the basis of the "SAR by NMR" methodology that is indispensable in the identification of active pharmaceutical agents [29].

Molecular docking, a computational technique used widely to model the conformation of protein-ligand complexes, enables experimental perturbations to be analyzed in atomistic detail. Bound ligand conformations, or poses, are ranked using an empirical scoring function designed to evaluate intermolecular interactions using minimal computational time. Conventionally, knowledge of the active site is used to guide the pose generation, often in the context of screening large libraries of compounds against known protein structures [31, 10, 30, 34]. However, docking protocols without prior knowledge of the active site (blind docking) [9], have successfully identified putative allosteric binding sites of drugs, leading to the design of novel allosteric modulators [12], and fluorescent dyes [17, 1]. Bis-ANS binding sites found by docking, validated with steady-state and time-resolved fluorescence assays, have been used to identify hydrophobic patches in a lipase from *Bacillus subtilis* [14]. Using ANS binding mapped by NMR in conjunction with molecular docking, we focus on determining whether the cataract-related G18V variant of human γS-crystallin has increased exposure of hydrophobic residues that could explain its aggregation propensity and/or recognition by αB-crystallin.

Chapter 2

Materials and Methods

2.1 ANS fluorescence assay

Wild-type and G18V γS-crystallins were expressed and purified as previously described [3]. Fluorescence spectra were collected as a function of ANS binding for γS-WT and γS-G18V with a F4500 Hitachi fluorescence spectrophotometer. The excitation and emission wavelengths were 390 nm and 500 nm, respectively, with slits set to 5 nm. Protein concentrations for both γS-WT and γS-G18V were approximately 1 mg/mL in 10 mM sodium phosphate buffer and 0.05% sodium azide at pH 6.9. ANS concentrations ranged from 5 μM to 2 mM were measured using ϵ= 4.95 mM^{-1} cm^{-1} at 350 nm [37].

2.2 NMR sample preparation

Purified protein with the 6x-His tag removed was concentrated and supplemented with 2 mM TMSP, 10% D2O, and 0.05% sodium azide. The final concentration of all γS-WT and γS-G18V samples was 0.3 mM. ANS was titrated into the protein samples to give final molar

Center	^{1}H: 799.8056964 MHz	^{13}C: 201.1282461 MHz	^{15}N: 81.0504078 MHz
Offset	^{1}H: -294.932 Hz (4.8 ppm)	^{13}C: -9863.17 Hz (43 ppm)	^{15}N: 2400 Hz (116.7 ppm)

Table 2.1: Final concentrations of γS-WT and γS-G18V

ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS. Spectra were acquired at 25 °C.

2.3 NMR experiments

Experiments were performed on a Varian $^{\text{Unity}}$INOVA spectrometer (Agilent Technologies) operating at 800 MHz and equipped with a ^{1}H–^{13}C–^{15}N 5 mm tri-axis PFG triple-resonance probe, using an 18.8 Tesla superconducting electromagnet (Oxford instruments). Decoupling of ^{15}N nuclei was performed using the GARP sequence [27]. ^{1}H chemical shifts were referenced to TMSP, and ^{15}N shifts were referenced indirectly to TMSP. NMR data were processed using NMRPipe [5] and analyzed using CcpNMR Analysis [35]. Center operating frequencies and (unless otherwise stated) center frequency offsets were as follows:

2.4 Calculation of chemical shift perturbations

^{1}H-^{15}N HSQC spectra of γS-WT and γS-G18V were collected in the presence and absence of ANS at concentration ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS, and resonances were identified and assigned based on chemical shift data previously collected by our group. Resonances showed perturbations that are indicative of ANS binding. The change in chemical shift for each peak in the 2D spectrum upon ANS binding was calculated using the following chemical shift perturbation (CSP) equation:

$$\Delta\delta_{avg} = \sqrt{\frac{(\Delta\delta_N/5)^2 + (\Delta\delta_H)^2}{2}} \tag{2.1}$$

A strong-binding threshold for each set of conditions was set at two times the root mean square (RMS) of the calculated CSP, while the weak-binding threshold was set at half the RMS to determine which residues had strong or weak binding with ANS. The values used for each threshold appear in Supplementary Table A.1.

2.5 Binding site search by rigid receptor docking

Protein coordinates were obtained from the NMR structures of γS-WT and γS-G18V crystallins (PDB ID: 2M3T and 2M3U) [16]. Autodock Tools [23] was used to prepare both the receptor (crystallin) and ligand (ANS) by merging non-polar hydrogens atoms into united heavy atoms. Gasteiger charges[8] were added to each atom. The sulfonic acid group of ANS was deprotonated before processing by Autodock Tools. Molecular docking was performed using Autodock Vina [33]. In order to ensure good coverage of the protein binding surface, 27 search spaces were placed in an overlapping 3 x 3 x 3 grid around the protein (Supplementary Figure A.1). Since Autodock Vina works optimally with search spaces with at most a 27,000 Å^3 volume, a 30 x 30 x 30 Å search space was chosen. The exhaustiveness parameter was set to 20 (over the default value of 8) in order to ensure an extensive search of the protein surface. Docking was performed over each one of twenty solution-state NMR conformations for either γS-WT or γS-G18V. The resulting poses were screened to ensure that both electrostatic and hydrophobic interactions required for ANS fluorescence enhancement upon binding were present. Docked poses that did not include both interactions within the first coordination shell of the ANS-protein radial distribution function were considered non-fluorescent and removed from the docked set. The screened docked set covers most of the protein surface (see Supplementary Figure A.2).

2.6 Calculation of residue contacts

To compare the screened docked set with the residue-based CSP data, ANS-residue contact frequencies were calculated by summing the Boltzmann weights of all the poses in contact with a given residue. The Boltzmann weight of a given docked pose was calculated according to

$$w_i = \frac{exp(-E_i/k_BT)}{\sum_i exp(-E_i/k_BT)} \tag{2.2}$$

where i is the index of the docked pose, E_i is the pose binding energy, k_B is the Boltzmann constant, and T is the absolute temperature. The residue contact frequencies for each protein are shown in Supplementary Figure S3. Following the CSP analysis, to determine which residues had strong or weak binding with ANS, a strong-binding threshold was set at two times the RMS of the calculated ANS-residue contact frequency, while the weak-binding threshold was set at the RMS value. The values used for each threshold appear in Supplementary Table A.1.

2.7 Flexible refinement of binding sites

A flexible docking refinement was performed near all the highly perturbed residues identified using the strong-binding cutoff on the CSP data described in the previous section. Docking search spaces were defined by clustered conformations of ANS from the screened docked set used to calculate the ANS-residue contact frequencies. Using a root-mean-square deviation cutoff (RMSD) of 5.0 Å, clustered poses were grouped into potential binding sites near the experimentally perturbed residues (see Supplementary Figure A.2). Search spaces were defined as boxes surrounding the clustered ligands with an 8 Å padding. The padding was necessary to include flexible side chains within the search space. Residues with an

experimental CSP above the low-binding cutoff were considered as flexible. A total of five potential binding sites were used to dock ANS to either flexible γS-WT or γS-G18V. The resulting poses were clustered again, and the location and interactions of each pose were compared visually.

Chapter 3

Results and Discussion

3.1 ANS fluorescence indicates that the relative surface hydrophobicity of γS-G18V is higher than that of γS-WT

Dye-binding assays were performed on γS-WT and the aggregation prone variant, γS-G18V. The ANS fluroscence measurements for γS-WT and γS-G18V, shown in Figure 3.1, indicate more exposed hydrophobic surface in γS-G18V compared to its wild type counterpart. These data also allow determination of the lowest ANS concentration required to produce the maximum emission before saturation, which was 1.5 mM for γS-WT and 1 mM ANS for γS-G18V. The lower concentration required to saturate γS-G18V is consistent with the observation that it binds ANS more readily than wild-type.

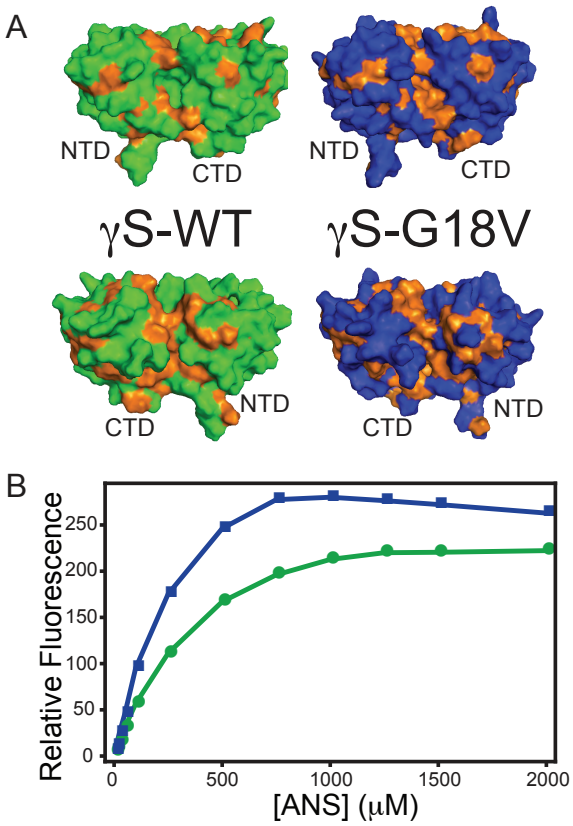

Figure 3.1: A. Molecular surface representation of γS-WT (green) and γS-G18V (blue) based on the solution-state NMR structures (PDB ID 2M3T and 2M3U, respectively). Hydrophobic residues are highlighted in orange. B. Fluorescence spectra representing ANS binding monitored at 500 nm using γS-WT and γS-G18V crystallins. Protein concentrations for both γS-WT and γS-G18V were approximately 1mg/mL. Saturation occurred at 1.5mM ANS for γS-WT and 1 mM ANS for γS-G18V. Higher emission was observed for γS-G18V, indicatingmore hydrophobic surface area exposed than for γS-WT.

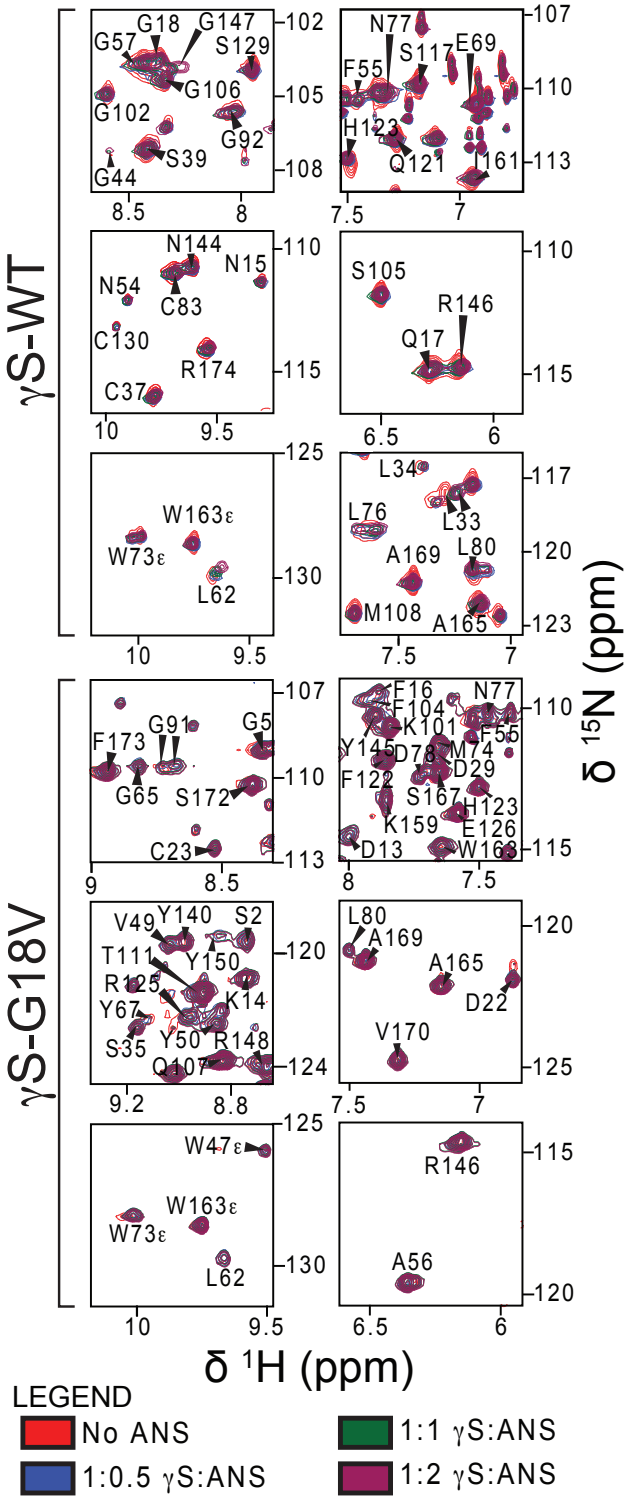

Figure 3.2: Selected portions of the ^{1}H-^{15}N HSQC spectra of γS-WT and γS-G18V. Experiments were carried out using final ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS. Resonances indicative of a change in chemical shift are indicative of multiple ANS binding to specific residues. However the perturbations observed are small due to the low concentrations of γS-crystallin and ANS used. Spectra were acquired at 25 °C with final concentrations of all γS-WT and γS-G18V samples at 0.3 mM.

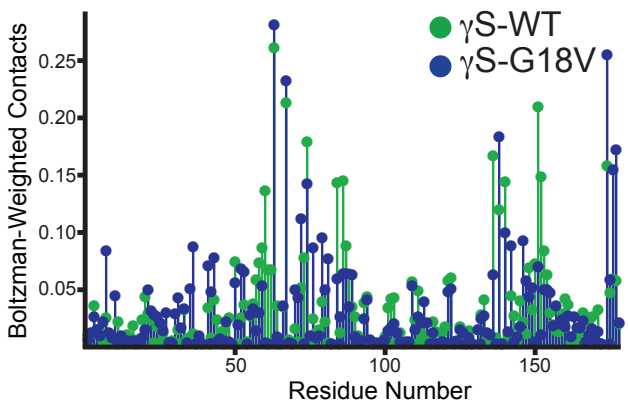

Figure 3.3: Average chemical shift perturbation (CSP) of γS-WT (green) and γS-G18V (blue). Nonspecific binding, with maximum perturbation in the N-terminal domain, is observed in both proteins. However, in γS-G18V, more of the CSPs are localized to the N-terminal domain. Particularly between residues 15 to 50, in the cysteine loop near the mutation site. Inspection of the structures confirms that this region is exposed to solvent in γS-G18V but not in γS-WT

3.2 Chemical shift perturbation mapping reveals the residues involved in ANS binding and the relative strengths of the interactions

Binding interactions between ANS and γS-WT or γS-G18V were measured at concentration ratios of 1:0, 1:0.5, 1:1, and 1:2 of γS:ANS, using CSP mapping via ^{1}H-^{15}N HSQC spectra. Selected regions of the NMR spectra where resonances show perturbations indicative of ANS binding are shown in Figure 3.2. The full NMR spectra can be found in the Supplemental Information (Supplementary Figures S4 and S5). The change in chemical shift for each peak in the 2D spectrum upon ANS binding was calculated using Equation 1. Representative CSP data for γS-WT and γS-G18V upon 1:1 ANS binding is shown in graphical form in Figure 3.3. The complete set of calculated CSP data can be found in the Appendix (Supplementary Figures A.6 and A.7). As shown in Figure 3.3, although nonspecific binding is observed throughout the surfaces of both proteins, γS-G18V binds ANS more strongly in the N-terminal domain (approximately the first 100 residues). The maximum ANS binding occurs

within residues 15 through 50, close to the mutation site. These observations are mapped onto the protein structures in Figure 3.4 (left panel) where the residues exhibiting strong (CSP at least two times the RMS) and weak ANS binding (CSP at least half the RMS) are highlighted. For γS-WT, strong binding residues are highlighted in bright green and weak binding residues in pale green. For γS-G18V, strong binding residues are highlighted in dark blue and weak binding residues in pale blue. Although some strong binding residues are observed in both proteins near the mutation site, (e.g. G18 in γS-WT and D22 in γS-G18V), G18V displays more ANS binding, both strong and weak, in the N-terminal domain. String binding is also seen in the interdomain interface of γS-WT, (residues L62, S82, and H123), and γS-G18V (residues L62, W73, H87, L88, and G91). These results are consistent with the observation that αB-crystallin strongly binds near the N-terminal domain and the interdomain interface in γS-G18V, but not γS-WT[16]. Thus, the ANS-binding data support the hypothesis that the chaperone may be recognizing an exposed hydrophobic patch in this region of γS-G18V.

Although both proteins display ANS binding throughout both the N- and C-terminal domains, γS-G18V has additional ANS binding residues, mostly in the N-terminal domain. Although some strong binding residues exhibited in the N-terminal domain for both proteins near the mutation site, (e.g. G18 in γS-WT and D22 in γS-G18V), G18V displays more ANS binding, both strong and weak, in the N-terminal domain. Strong binding is also seen in the interdomain interface of γS-WT, (residues L62, S82, and H123), and γS-G18V (residues L62, W73, H87, L88, and G91). Similar perturbations at the interdomain interface (residues G65, Y67, S82, S85, and G91) were shown for γS-G18V in the presence of αB-crystallin, the primary holdase chaperone protein of the eye lens [16]. αB-crystallin only weakly interacts with γS-WT at the surface of the protein (residues S35, W47, E66, G92, F122, and H123) [16]. The full list of residues interacting with both ANS and αB-crystallin for both γS-WT and γS-G18V is tabulated in Table 2 for comparison. Notably, αB-crystallin strongly binds near the interdomain interface in γS-G18V in but not γS-WT, consistent with

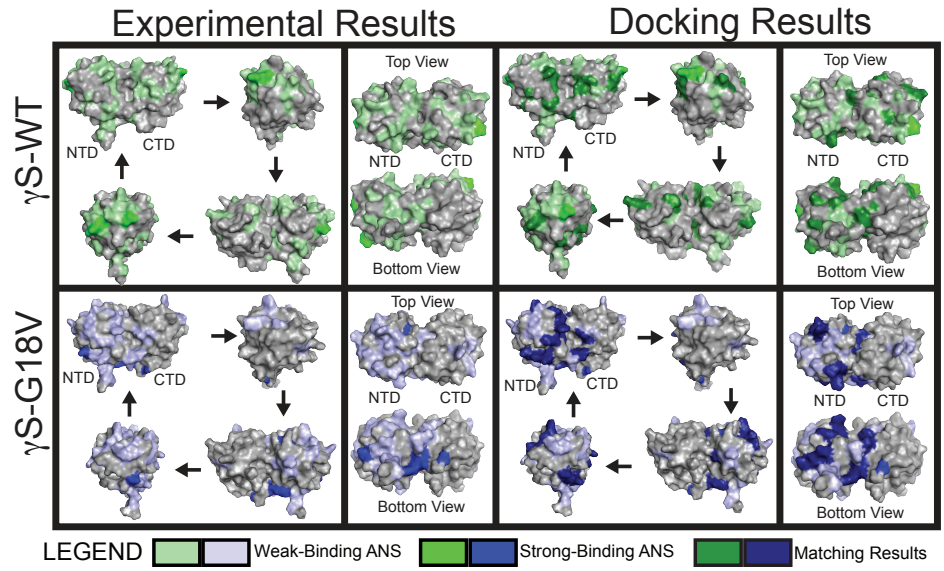

Figure 3.4: ANS interactions with γS-WT and γS-G18V. The strong-binding threshold and weak-binding threshold were defined as two times the RMS and half the RMS, respectively. Experimental CSP values indicate that ANS binding occurs throughout the N- and C-terminal domains for γS-WT, (strong binding residues in green and weak binding residues in pale green), while in γS-G18V ANS binding mainly occurs at the N-terminal domain (strong binding residues in blue and weak binding residues in pale blue). Some strong binding is observed in the N-terminal domain for both proteins near the mutation site, e.g.G18 in γS-WT and D22 in γS-G18V. However, γS-G18V displays more ANS binding (both strong and weak) overall in the N-terminal domain. Strong binding is also observed in the interdomain interface of γS-WT, residues L62, S82, and H123, and γS-G18V, residues L62,W73, H87, L88, and G91. G18V exhibits more binding (strong and weak) within that interdomain interface suggesting that this variant has higher surface hydrophobicity localized to the N-terminal domain near the mutation site and the interdomain interface. Coverage of both strong and weak binding residues are nearly identical between experimental and docking results, highlighted in dark green for γS-WT and dark blue for γS-G18V, indicating that the docking results are in good agreement with the experimental data.

the hypothesis that the chaperone is recognizing an exposed hydrophobic patch in this region of γS-G18V.

In order to character the aggregation states of γS-WT and γS-G18V, dynamic light scattering (DLS) data were acquired for both proteins under the same solution conditions used in the NMR experiments (data shown in Supplementary Figure A.8). As observed in previous studies of γS-G18V [16, 3], the γS-WT solution contains only monomers, while γS-G18V shows a slightly broader range of sizes consistent with transient formation fo dimers and potentially other small oligomers. However, the NMR spectra rule out the presence of a significant stable population of large aggregates; the linewidths for representative peaks in the HSQC spectra are comparable to the corresponding peaks in γS-WT. Linewidth comparisons for representative peaks in the spectra of γS-WT and γS-G18V are tabulated in the Supplemental Information (Table A.2). Despite the presence of transient oligomers in the γS-G18V sample, consistent with its increased aggregation propensity, it is clear from the linewidth data that the chemical shift changes upon addition of ANS are due to dye binding rather than a change in aggregation state. Although the oligomerization states of the starting solutions were not identical, this is accounted for by the chemical shift differences between γS-WT and γS-G18V in the absence of ANS, while the chemical shift perturbations reflect binding of each protein to ANS. If stable, large complexes were present in the NMR samples, the increased aggregation would be expected to cause significant line broadening and disappearance of signals, as was observed for mixtures of γS-G18V with αB-crystallin[16],where large complexes were formed and TROSY techniques were required to observe the NMR signals. Although it is possible to prepare purely monomeric samples of γS-G18V at low pH, for the current study, neutral pH was chosen in order to investigate intermolecular interactions under more physiologically realistic conditions

In order to investigate whether ANS changes the oligomerization states of γSαB complexes and interferes with binding of αB-crystallin to γS-G18V, we performed gel filtration chro-

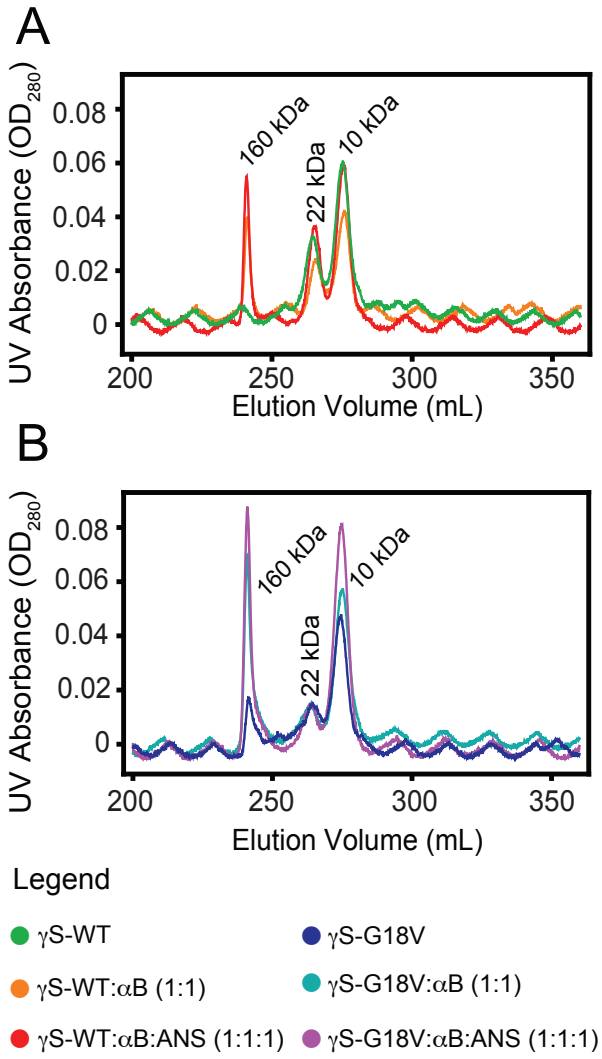

Figure 3.5: Gel filtration chromatograms for γS-WT (A) and γS-G18V (B) in the presence and absence of αB-crystallin and ANS. (A) For γS-WT alone (green), the sample is mostly monomeric (10 kDa) with a small amount of dimers (22 kDa). Upon addition of αB-crystallin (orange), a population of larger oligomers (160 kDa) appears, at the expense of the populations of both the monomeric and dimeric states. Addition of ANS to this mixture (red) slightly increases the proportion of large aggregates. (B) For γS-G18V alone (blue), much of the sample is monomeric, although small populations of dimers and large oligomers exist. The peak at 160 kDa is much broader than in γS-WT, suggesting greater polydispersity. In the presence of αB-crystallin (cyan), the main effect is a dramatic narrowing of the peak corresponding to large oligomers, indicating a more uniform population. Addition of ANS to this mixture (purple) produces both further narrowing and an increase in the population of monomers, suggesting that interaction with ANS does disrupt the αB-γS complex formation.

matography (Fig. 5). Sam- ples of γS-WT and γS-G18V were prepared at 1mg/mL and compared to equivalent samples in the presence of αB-crystallin (1:1) and both αB-crystallin and ANS(1:1:1). For γS-WTalone, the sample is mostly monomeric with a small amount of dimers. Upon addition of γB-crystallin, a population of larger oligomers at about 160 kDa appears, at the expense of the populations of both th emonomeric and dimeric states. Addition of ANS to this mixture slightly increases the proportion of large aggregates. In the case of γS-G18V, both the initial oligomerization states and the effect of adding ANS is different. Initially, although much of the sample is monomeric, small populations of dimers and large oligomers exist. The peak at 160 kDa is much broader than in γS-WT, suggesting greater polydispersity. In the presence of αB-crystallin, the main effect is a dramatic narrowing of the peak corresponding to large oligomers, indicating a more uniform population. Addition of ANS to this mixture produces both further narrowing and an increase in the population ofmonomers, suggesting that interaction with ANS does disrupt the αBγScomplex formation. The full chromatogram including molecular weight standards is provided in the Supplemental Information (Fig. A.9).

3.3 Docking of ANS on the protein surface predicts more binding sites on γS-G18V than γS-WT and allows interpretation of the CSP data

Rigid receptor docking resulted in a total of 4860 docked poses (27 search spaces 20 NMR conformations 9 poses/search space). After screening for poses consistent with ANS fluorescence enhancement upon binding, 3423 poses and 3367 poses remained for γS-WT and γS-G18V, respectively (see Supplementary Information Fig. A.2A). Filtered poses covered nearly the entire surface of the protein and exhibit a broad range of scores (from 2 kcal/mol

to 7 kcal/mol, with a mean of 4.5 kcal/mol).Due to the pocket-like shape of the interdomain interface, ANS preferentially bound to the large hydrophobic pocket between the N- and C-terminal domains. However, sites were identified near all highly perturbed residues with comparable binding scores (see Supplementary Information Fig. A.3). Flexible docking poses located near the highly perturbed residues according to the CSP data had binding scores between 4.5 kcal/mol and 6.0 kcal/mol, consistent with a stronger preference for ANS to bind near the perturbed residues. A total of ten binding sites were found for γS-G18V and nine binding sites for γS-WT using flexible docking.Most of these binding sites were very similar in both γS-WT and γS-G18V. However, three binding modes were unique to γS-G18V. The first and most populated binding mode is located in the hydrophobic cavity at the interface between the N- and C-terminal domains, shown in Fig. 3.6A and 3.6D [7, 22]. Although this binding site was found in both γS-WT and γS-G18V, the presence of the R84-D153 salt-bridge blocks the exposure of the hydrophobic surface in γS-WT. In contrast, γS-G18V lacks this salt-bridge interaction, which exposes the interdomain hydrophobic cavity and allows the entry of ANS into the interdomain binding site. This is consistent with the experimental NMR data,which indicate that chemically perturbed residues, H87 and L88, located near the interdomain pose, interact strongly with ANS only in γS-G18V (Fig. 3.6). The second and third binding sites are located close to res- idues 2030,which includes a loop region containing three cysteine residues (C23, C25, and C27). As a result of the G18V mutation, C23 and C27 become solvent exposed, suggesting possible formation of intermolecular disulfide bridges, consistent with the observation that an excess of reducing agents abrogates the formation of small oligomers [22]. Previous studies suggested that the exposure of these cysteines results from a disruption in secondary structure due to the burial of the V18 sidechain[16]. As a result of this cysteine exposure and concomitant structural changes, a new hydrophobic pocket is uncovered as the second ANS binding site. Although ANS binds this Cys loop in γS-WT after flexible docking refinement, it is not in direct contact with any hydrophobic surface, suggesting that the pose may not be consistent with and

enhancement in ANS fluorescence (Fig. 3.6B). In contrast, when the hydrophobic pocket is exposed, as it is inγS-G18V,ANS becomes buried deep within the pocket (Fig. 3.6E). This conformation provides both the hydrophobic interactions necessary for fluorescence as well as reduced quenching due to water exposure [20]. In addition to cysteine exposure, the third binding site shows additional hydrophobic surface exposure due to the cysteine loop separating from the main Greek key motif. This binding site was not found in γS-WT using the same docking search space, indicating that this hydrophobic patch is a unique characteristic of γS-G18V (Fig. 6F). Additionally, the CSP data shows local perturbation of the backbone amides of the residues involved in these three binding sites only for γS-G18V. The presence of these γS-G18V-specific binding sites can ex- plain the higher ANS fluorescence intensity of the variant protein over WT, and they also identify exposed hydrophobic patches which may potentially serve as proteinprotein interfaces in crystallin aggregates, and which can be targeted in future mutagenesis studies.

The CSP data and the ANS-residue contact data from the docking simulations show generally good agreement in that the same protein regions were observed to bind ANS (see Fig. 4). In some cases, the specific residues classified as strong binding vary between experimental and docking results, but coverage of both strong- and weak-binding residues are nearly identical (highlighted in dark green for γS-WT, and dark blue γS-G18V in the right panel of Fig. 4). This outcome is to be expected because the docking scoring function is more effective at identifying binding sites than distinguishing more subtle changes in binding energy: the standard error of the Autodock Vina scoring function [33] is larger than the variation among scored poses. The agreement between rigid protein docking results and experimental ANS binding results suggests that there is no major change in protein conformation upon binding of ANS, supporting the hypothesis that hydrophobic patches on the surface are involved in intermolecular interactions. Good agreement between the experimental and docking results further confirms that ANS binding is localized near the mutation site in the N- terminal domain for γS-G18V, consistent with the CSP data. Experimental and simulation results

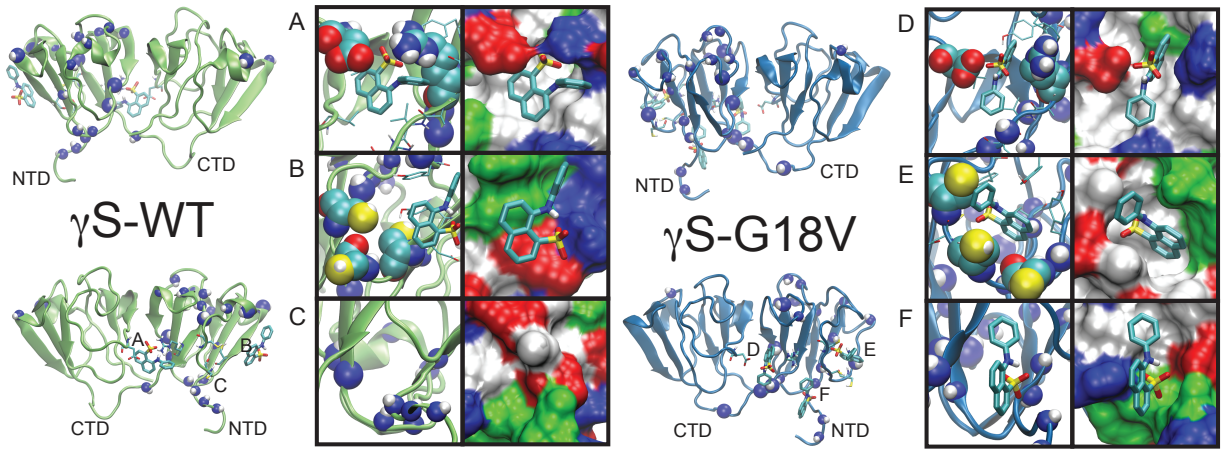

Figure 3.6: Docking poses of ANS bound to a flexible γS-WT and γS-G18V receptor. Three binding sites were found to be unique to γS-G18V. The protein surface was generated with MSMS [26]; red, blue, green, and white correspond to negative, positive, polar and hydrophobic regions, respectively. ANS is shown in licorice representation. The R84-D153 salt bridge and cysteine residues (C23, C25, and C27) critical to the hydrophobic patch availability are shown in space-filling representation. In the left-hand panels, residues defined as strong/weak binding by CSP data have their backbone amides represented as spheres. Large spheres represent strongly-binding residues, and small spheres represent weakly-binding residues. Atoms are colored by element (carbon, cyan; nitrogen, blue; oxygen, red; hydrogen, white; sulfur, yellow). (A & D) At site 1, the R84-D153 salt bridge separates to expose the hydrophobic cavity at the interdomain interface. Although a pose is generated in both proteins, the lack of perturbed residues at the binding site, according to the CSP data, indicates that the binding site is inaccessible to ANS in γS-WT. (B & E) At site 2, the docked pose of ANS shifts from a polar surface in γS-WT to inserting into a hydrophobic cavity in γS-G18V. Due to specific backbone torsions propagating from the G18V mutation site that keep the V18 buried, C23 and C27 become solvent exposed and reveal a hydrophobic cavity. This pose is located near the largest perturbed residue in γS-G18V according to the CSP data. (C & F) At site 3, the disordered cysteine loop separates from the Greek key motif of the N-terminal domain resulting in exposure of an additional hydrophobic patch. No equivalent pose could be generated on the γS-WT structure.

are also consistent on the binding of ANS to the exposed interdomain hydrophobic surface located in the interdomain interface between the two domains due to the breaking of the R84-D153 salt bridge in γS-G18V. Exposure of this hydrophobic patch facilitates ANS binding and may be involved in hydrophobic protein-protein interactions.

Chapter 4

Conclusion

In this study, we have used ANS fluorescence, solution-state NMR chemical shift perturbation mapping, and molecular docking to investigate the differences in exposed hydrophobic surface between human γS-crystallin and its cataract-related γS-G18V variant. The experimental results indicated that both proteins have a fairly high level of nonspecific binding, but both the fluorescence and NMR measurements indicate more ANS binding to γS-G18V, particularly in the N-terminal domain near the mutation site. The docking studies, in agreement with the NMR data, found three binding modes that are unique to γS-G18V that were not found in γS-WT: one in the exposed hydrophobic patch in the interdomain interface and two binding modes in the exposed hydrophobic pocket formed when the cysteine loop becomes solvent ex- posed. Using docking and binding assays, hydrophobic surface patches were identified that may be responsible for some of the intermolecular interactions between crystallins that promote aggregation in the lens. The results may also guide the design of future mutagenesis and drug-binding studies to further investigate the importance of such intermolecular interactions in mediating γS-crystallin solubility and aggregation resistance in the healthy eye lens.

Bibliography

[1] P. R. Banerjee, S. S. Puttamadappa, A. Pande, A. Shekhtman, and J. Pande. Increased hydrophobicity and decreased backbone flexibility explain the lower solubility of a cataract-linked mutant of γD-crystallin. *Journal of Molecular Biology*, 412(4):647–659, 2011.

[2] S. V. Bharat, A. Shekhtman, and J. Pande. The cataract-associated V41M mutant of human γS-crystallin shows specific structural changes that directly enhance local surface hydrophobicity. *Biochem. Biophys. Res. Commun.*, 443:110—114, 2010.

[3] W. D. Brubaker, J. A. Freites, K. J. Golchert, R. A. Shapiro, V. Morikis, D. J. Tobias, and R. W. Martin. Separating instability from aggregation propensity in γS-crystallin variants. *Biophys. J.*, 100(2):498–506, January 2011.

[4] A. B. Christian Münch. Exposure of hydrophobic surfaces initiates aggregation of diverse als-causing superoxde dismutase-1 mutants. *Journal of Molecular Biology*, 399:512—525, 2010.

[5] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfiefer, and A. Bax. Nmrpipe: A multidimensional spectral processing system based on unix pipes. *Journal of Biomolecular NMR*, 6:277—293, 1995.

[6] M. A. DiMaurro, S. K. Nandi, C. T. Raghavan, R. K. Kar, B. Wang, A. Bhuania, R. H. Nagaraj, and A. Biswas. Acetylation of gly1 and lys2 promotes aggregation of human gammad-crystallin. *Biochemistry*, 53:7269—7282, 2014.

[7] S. L. Flaugh, M. S. Kosinski-Collins, and J. King. Interdomain side-chain interactions in human γD crystallin influencing folding and stability. *Protein Science*, 14(8):2030–2043, aug 2005.

[8] J. Gasteiger and M. Marsili. A new model for calculating atomic charges in molecules. *Tetrahedron Letters*, 19(34):3181–3184, 1978.

[9] C. Hetényi and D. van der Spoel. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein science : a publication of the Protein Society*, 11(7):1729–1737, 2002.

[10] S. Huang and X. Zou. Efficient molecular docking of nmr structures: Application to hiv-1 protease. *Protein Science*, 16:43—51, 2007.

[11] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, feb 1996.

[12] B. Iorga, D. Herlem, E. Barré, and C. Guillou. Acetylcholine nicotinic receptors: Finding the putative binding site of allosteric modulators using the "blind docking" approach. *Journal of Molecular Modeling*, 12(3):366–372, 2006.

[13] R. Jaenicke and C. Slingsby. Lens crystallins and their microbial homologs: Structure, stability, and function. *Critical Reviews in Biochemistry and Molecular Biology*, 36(5):435—499, 2001.

[14] M. Z. Kamal, J. Ali, and N. M. Rao. Binding of bis-ANS to Bacillus subtilis lipase: a combined computational and experimental investigation. *Biochimica et biophysica acta*, 1834(8):1501—1509, Aug. 2013.

[15] S. Karri, R. B. Kasetti, V. P. R. Vendra, S. Chandani, and D. Balasubramanian. Structural analysis of the mutant protein d26g of human gammas-crystallin, associated with coppock caraeact. *Mol. Vis.*, 19:1231—1237, 2013.

[16] C. N. Kingsley, W. D. Brubaker, S. Markovic, A. Diehl, A. J. Brindley, H. Oschkinat, and R. Martin. Preferential, specific binding of human αB-crystallin to a cataract-related variant of γS-crystallin. *Structure*, 21:2221–2227, 2013.

[17] T. Konuma, Y.-H. Lee, Y. Goto, and K. Sakurai. Principal component analysis of chemical shift perturbation data of a multiple-ligand-binding system for elucidation of respective binding mechanism. *Proteins*, 81:107–118, 2013.

[18] Y. Liu, X. Zhang, L. Luo, M. Wu, R. Zeng, G. Cheng, B. Hu, B. Liu, J. J. Liang, and F. Shang. A novel alphab-crystallin mutation associated with autosomal dominant congenital lamellar cataract. *Invest. Ophthalmol. Vis. Sci.*, 47(3):1069—1075, 2006.

[19] Z. Ma, G. Piszczek, P. Wingfield, Y. Sergeev, and J. Hejtmancik. The G18V CRYGS mutation associated with human cataracts increases γs-crystallin sensitivity to thermal and chemical stress. *Biochemistry*, 48:7334–7341, 2009.

[20] D. Matulis and R. Lovrien. 1-anilino-8-naphthalene sulfonate anion-protein binding depends primarily on ion pair formation. *Biophysical Journal*, 74(1):422–429, 1998.

[21] I. A. Mills, S. L. Flaugh, M. S. Kosinski-Collins, and J. A. King. Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γD-crystallin and γS-crystallin. *Protein Science*, 16(11):2427–2444, 2007.

[22] B. G. Mohr, C. M. Dobson, S. C. Garman, and M. Muthukumar. Electrostatic origin of in vitro aggregation of human γ-crystallin. *The Journal of Chemical Physics*, 139(12):121914, sep 2013.

[23] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, Dec. 2009.

[24] J. J. Ory and L. J. Banaszak. Studies of the ligand binding reaction of adipocyte lipid binding protein using the fluorescent probe 1, 8-anilinonaphthalene-8-sulfonate. *Biophysical Journal*, 77(2):1107–1116, 1999.

[25] A. Pande, K. S. Ghosh, P. R. Banjeree, and J. Pande. Increase in surface hydrophobicity of the cataract-associated P23T mutant of human gammad-crystallin is responsible for its dramatically lower retrograde solubility. *Biochemistry*, 49:6122—6129, 2010.

[26] M. F. Sanner, A. Olsen, and J. Spehner. Fast and robust computation of molecular surfaces. In *Proceedings of the 11th ACM symposium on computational geometry*, pages C6—C7, 1995.

[27] A. J. Shaka, P. B. Barker, and R. J. Freeman. Computer-optimized decoupling scheme for wideband applications and low-level operation. *Journal of Magnetic Resonance*, 64(547—552), 1985.

[28] D. Sheluho and S. H. Ackerman. An Accessible Hydrophobic Surface Is a Key Element of the Molecular Chaperone Action of Atp11p. *Journal of Biological Chemistry*, 276(43):39945–39949, 2001.

[29] S. B. Shuker, P. J. Hajduk, R. P. Meadows, and S. Fesik. Discovering high affinity ligands for proteins: Sar by nmr. *Science*, 274:1531—1534, 1996.

[30] J. L. Stark and R. Powers. Application of nmr and molecular docking in structure-based drug discovery. *Topics in Current Chemistry*, 326:1—34, 2011.

[31] T. ten Brink, C. Aguirre, T. E. Exner, and I. Krimm. Performance of protein-ligand docking with simulated chemical shift perturbations. *Journal of Chemical Information and Modeling*, 55:275—283, 2015.

[32] A. Tiwari, A. Liba, S. H. Sohn, S. V. Seetharanab, O. Bilsel, C. R. Matthews, P. J. Hart, J. S. Valentine, and L. J. Hayward. Metal deficiency increases aberrant hydrophobicity of mutant superoxide dismutases that cause amyotophic lateral sclerosis. *Journal of Biological Chemistry*, 284(27746—27758), 2009.

[33] O. Trott and A. J. Olson. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 2010.

[34] A. J. van Dijk, R. Boelens, and A. M. Bonvin. Data-driven docking for the study of biomolecular complexes. *FEBS Journal*, 272:293—312, 2005.

[35] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pjon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The ccpn data model for nmr spectroscopy: Development of a software pipeline. *Proteins*, 59(4):687—696, 2005.

[36] G. L. Warren, C. W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry*, 49(20):5912–5931, 2006.

[37] G. Weber and L. B. Young. Fragmentation of bovine serum albumin by pepsin. *The Journal of Biological Chemistry*, 239(5):1415—1423, 1964.

[38] M. P. Williamson. Using chemical shift perturbation to characterize ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73:1–16, 2013.

Appendix A

Supplementary Figures

Table A.1: Root mean square (RMS) values used for ANS-residue contact frequency calculations and CSP calculations. The Boltzmann weighted-contact frequencies had a strong-binding threshold set at two times the RMS of the calculated ANS-residue contact frequency, while the weak-binding threshold was set at half the RMS value. Following the CSP calculations, a strong-binding threshold for each set of conditions of γS:ANS was set at two times the RMS of the calculated CSP, while the weak-binding threshold was set at half the RMS.

		Contact Frequencies	1:0.5	1:1	1:2
WT	strong-binding	0.02699	0.02420	0.02699	0.03382
	weak-binding	0.01349	0.00605	0.00675	0.00846
G18V	strong-binding	0.02345	0.02342	0.02345	0.02345
	weak-binding	0.01172	0.00586	0.00586	0.00586

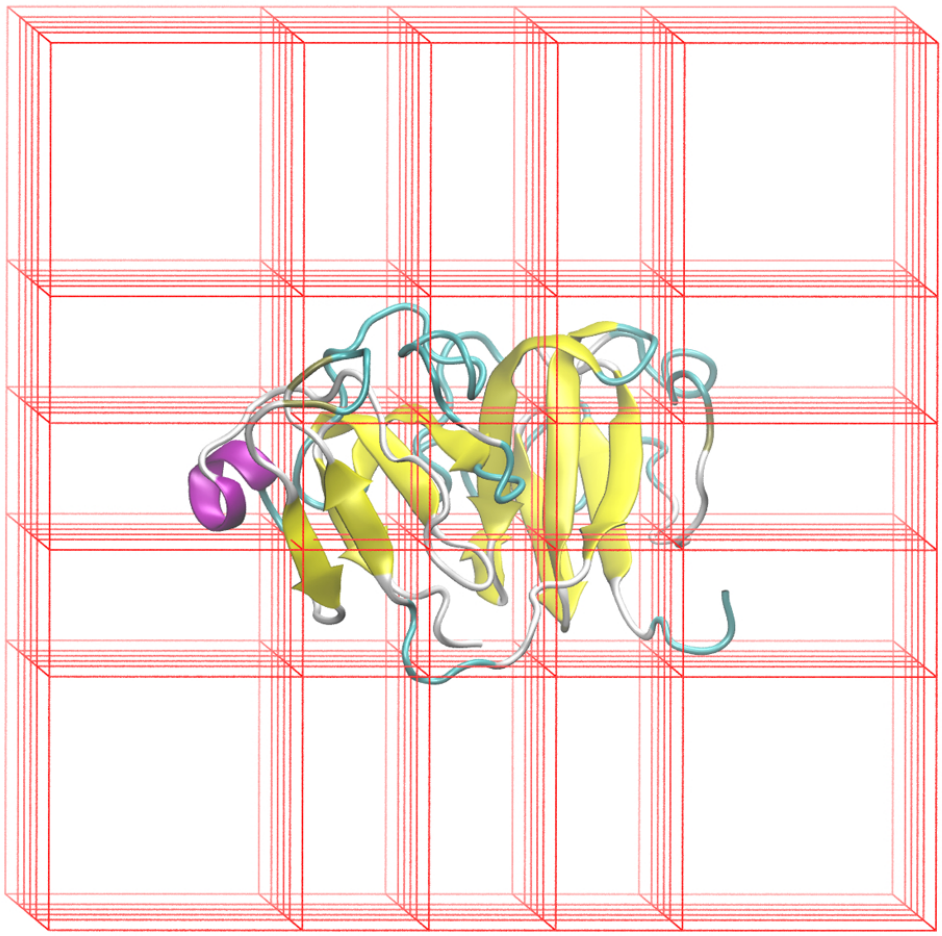

Figure A.1: Search space boundaries used in initial rigid receptor docking to NMR conformations of γS-crystallin. A total of 27 docking runs were performed for each NMR structure. Search spaces were placed in a 3 x 3 x 3 grid with a 10 $\mathring{(A)}$ overlap between search spaces. Boxes were sized so enough space was available to sample ligand conformations on the entire surface of the protein.

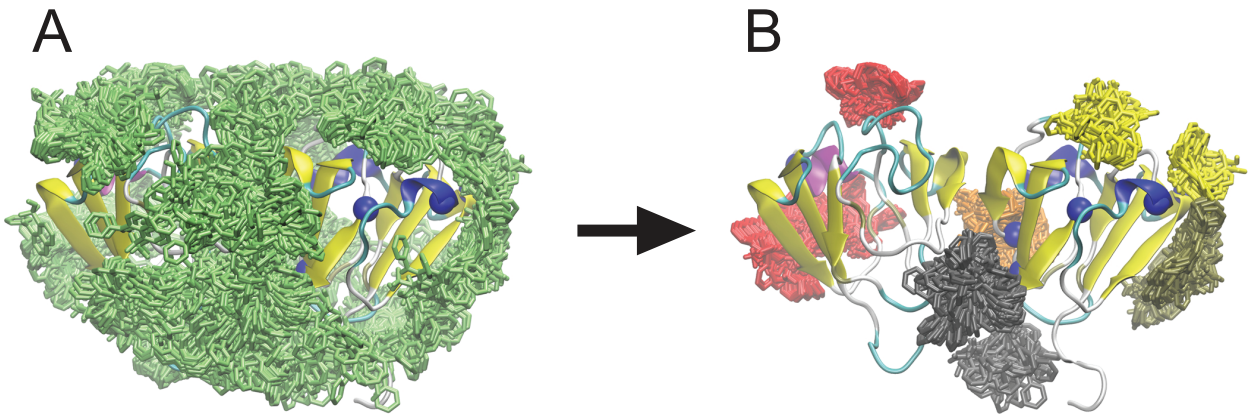

Figure A.2: Clustering of rigid docking results to define flexible binding sites (γS-WT shown as example). (A) Set of poses containing both hydrophobic and electrostatic contacts necessary for fluorescence. The resulting set covers nearly the entire surface of the protein. (B) Clustering of the pose set resulted in 20 binding sites (several of the clusters were overlapping). Pose clusters near highly perturbed residues (shown with blue VDW spheres) were picked to define search spaces for flexible docking. Clusters are color coordinated for each search space.

Table A.2: A table of selected line widths taken at half height for several representative residues in the HSQC spectra of γS-WT and γS-G18V for the 1:1 γS:ANS mixtures. The line widths (reported in Hz) are comparable for both proteins.

	γS-WT		γS-G18V	
Residue	^{1}H (Hz)	^{15}N (Hz)	^{1}H (Hz)	^{15}N(Hz)
C37	33.782	36.110	32.238	31.454
L62	36.003	44.881	37.979	35.617
W73ϵ	32.349	48.174	30.599	37.302
G102	32.392	33.081	32.082	37.695
F122	31.787	34.271	31.279	34.540
A165	29.652	30.582	31.966	28.605

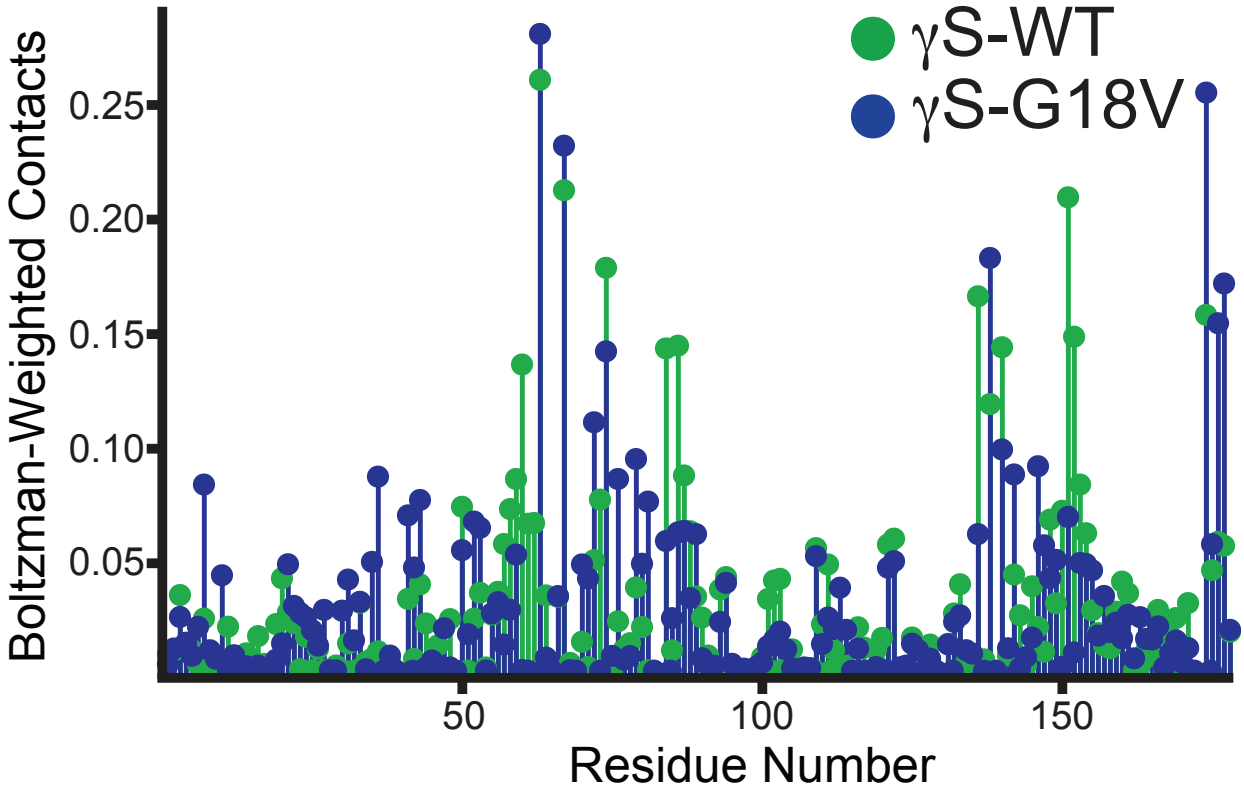

Figure A.3: ANS-residue contact frequencies for γS-WT (green) and γS-G18V (blue) from docking simulations. Although non-specific binding is observed for both proteins, the contact frequencies show more ANS binding for γS-G18V than γS-WT, with maximum binding localized near the interdomain interface

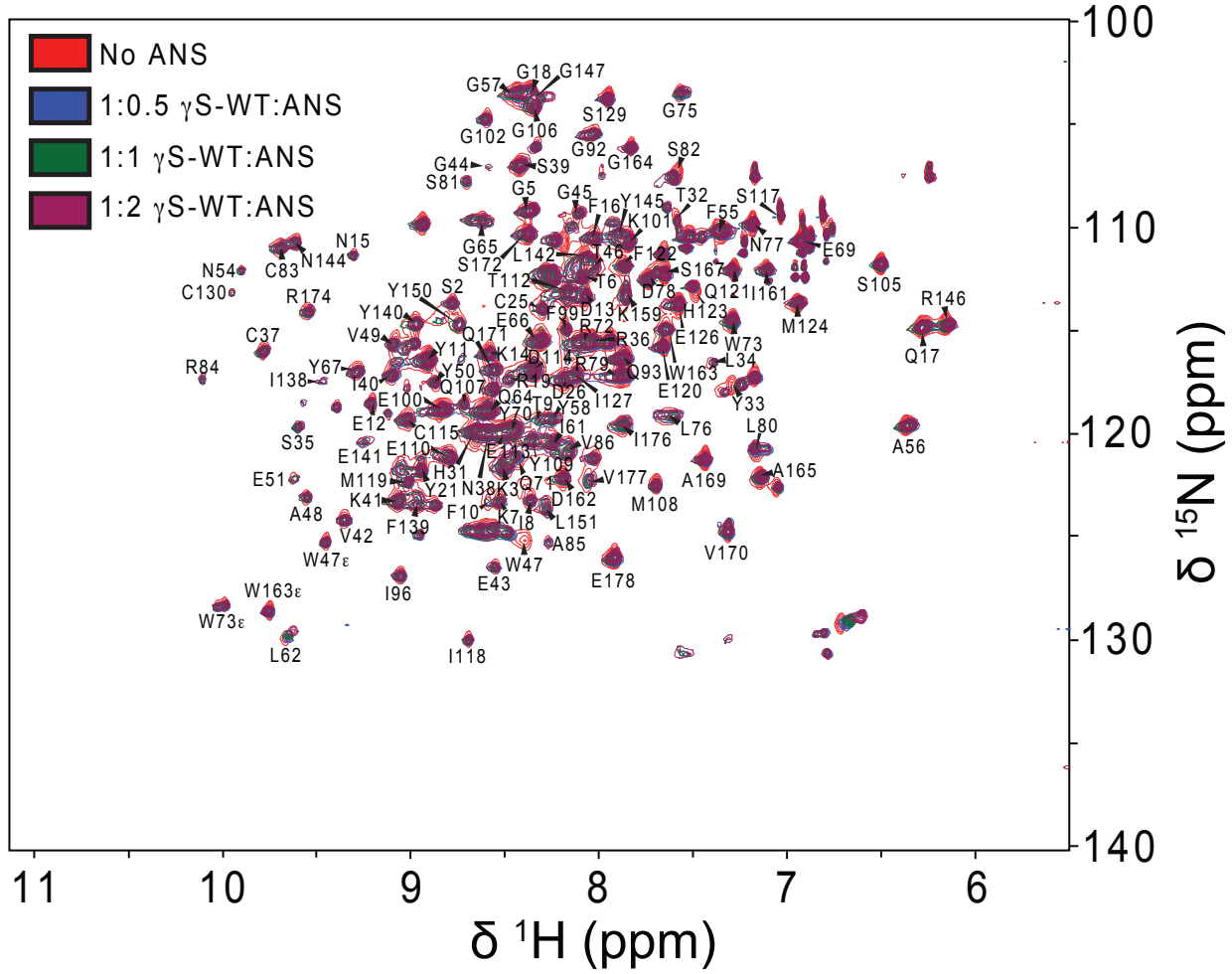

Figure A.4: ^{1}H-^{15}N HSQC spectra of ^{15}N labelled γS-WT with increasing concentrations of ANS. Ratios of γS:ANS were at 1:0, 1:0.5, 1:1, and 1:2 where the concentration of protein was approximately 0.3 mM. Spectra were acquired at 25 °C. Residues were assigned based on previous assignments of γS-WT [3].

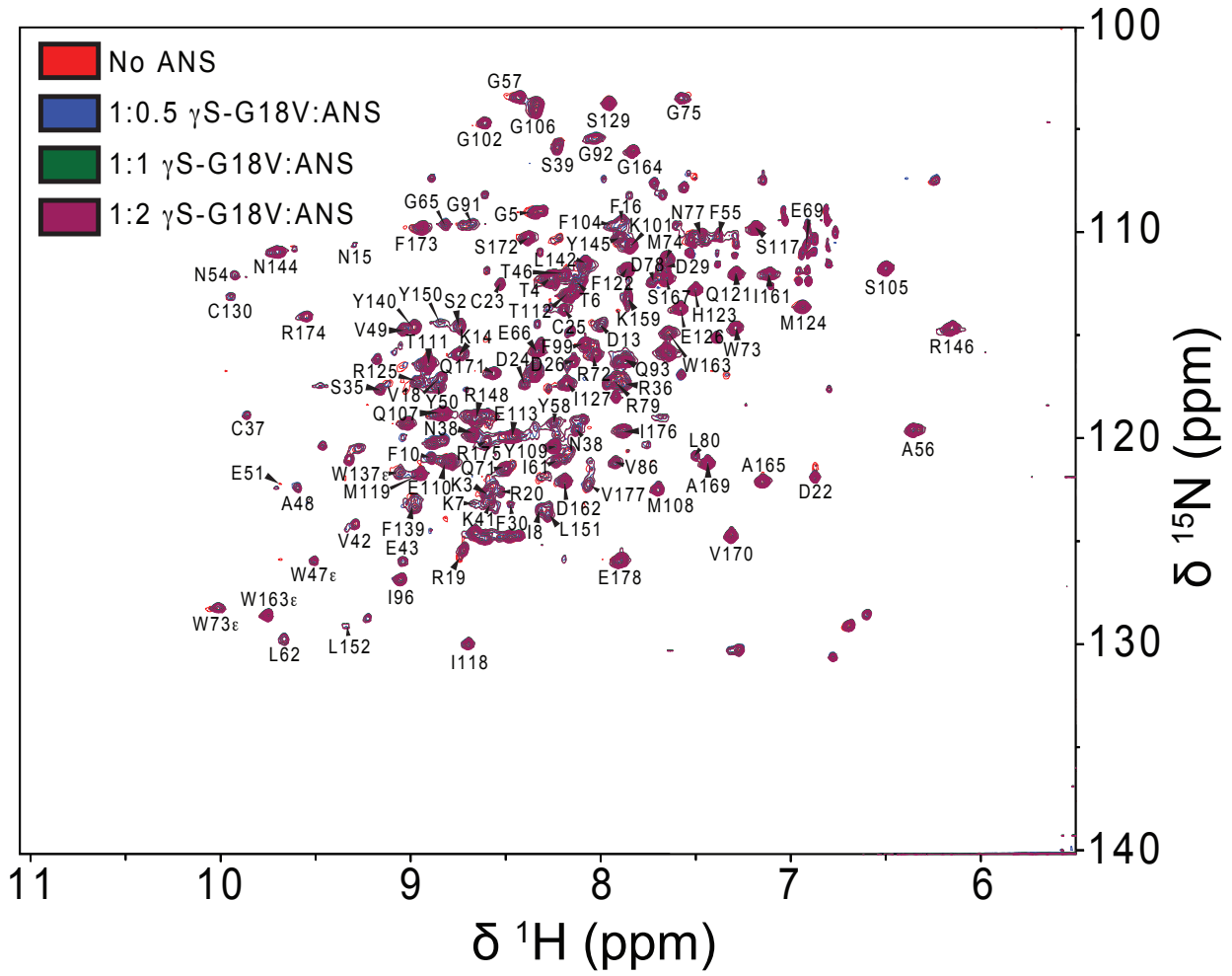

Figure A.5: ^{1}H-^{15}N HSQC spectra of ^{15}N labelled γS-G18V with increasing concentrations of ANS. Ratios of γS:ANS were at 1:0, 1:0.5, 1:1, and 1:2 where the concentration of protein was approximately 0.3 mM. Spectra were acquired at 25 °C. Residues were assigned based on previous assignments of γS-G18V [3].

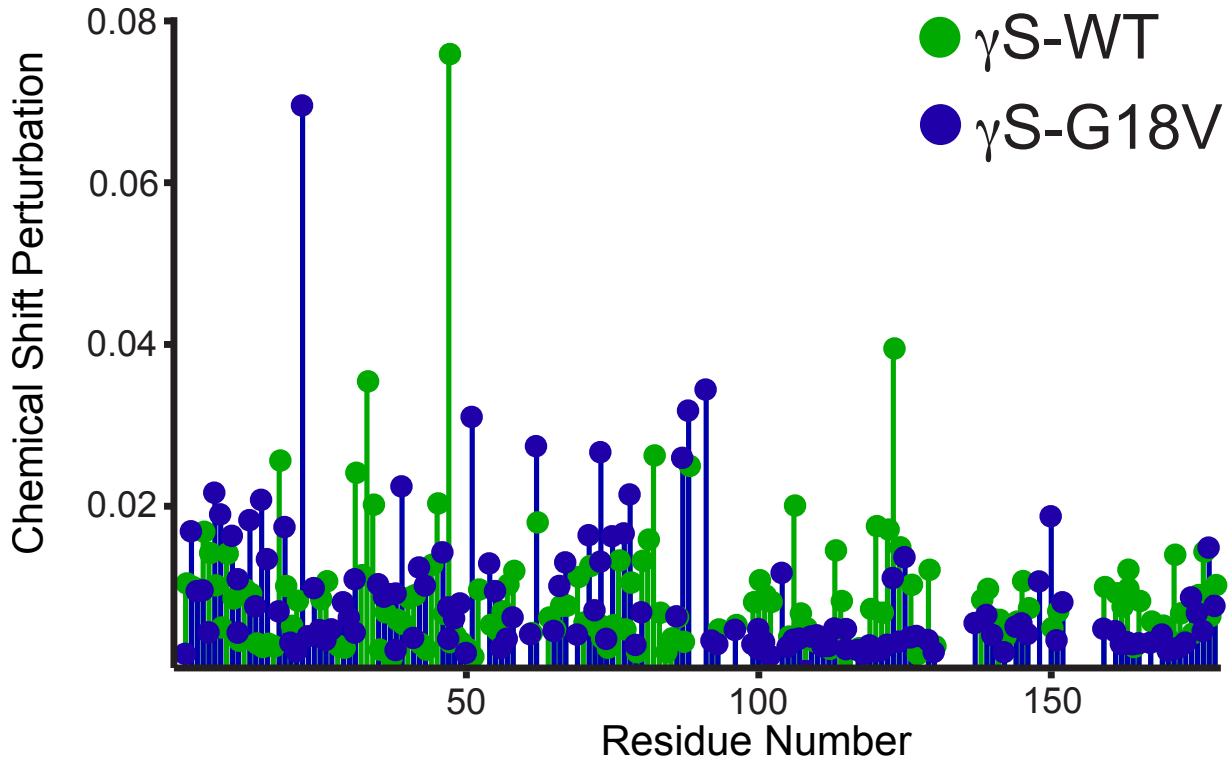

Figure A.6: Average chemical shift perturbation (CSP) of γS-WT (green) and γS-G18V (blue) for 1:0.5 γS:ANS. Maximum perturbation for both proteins is in the N-terminal domain and the interdomain interface

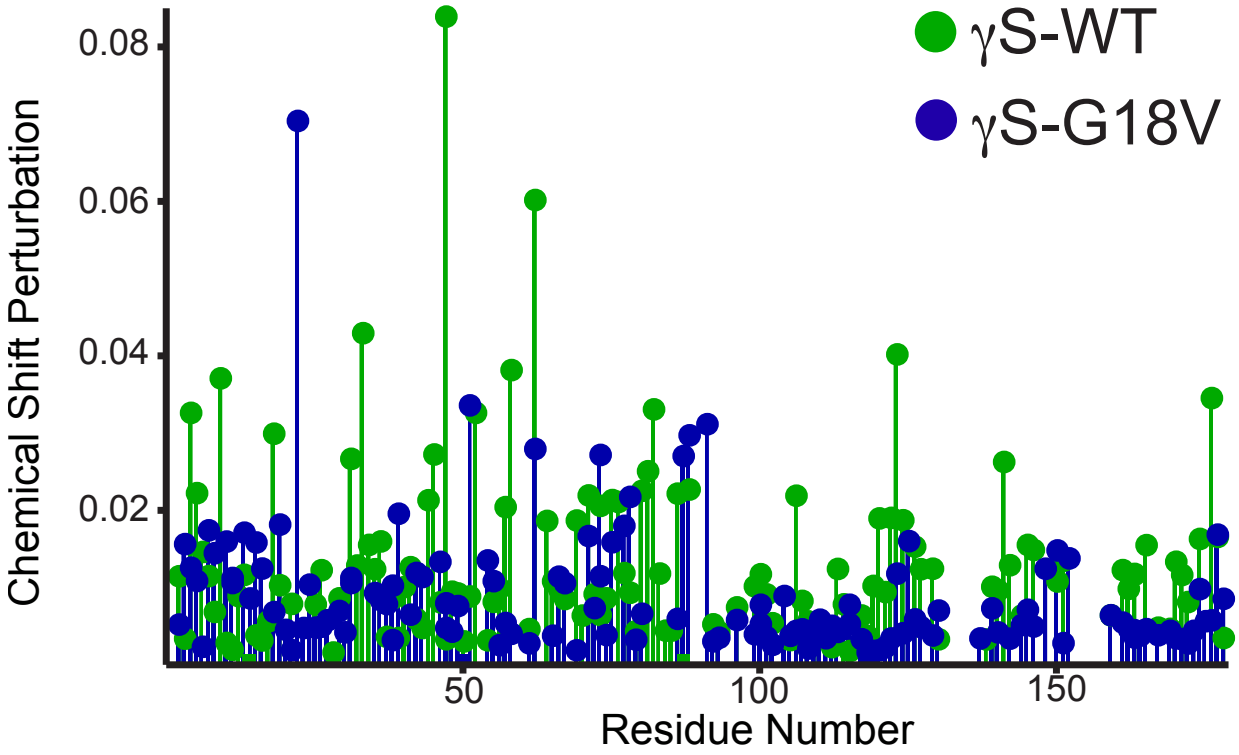

Figure A.7: Average chemical shift perturbation (CSP) of γS-WT (green) and γS-G18V (blue) for 1:2 γS:ANS. Maximum perturbation for both proteins is in the N-terminal domain. At this ANS concentration γS-WT exhibits more perturbation, likely because γS-G18V is already saturated at 1 mM ANS

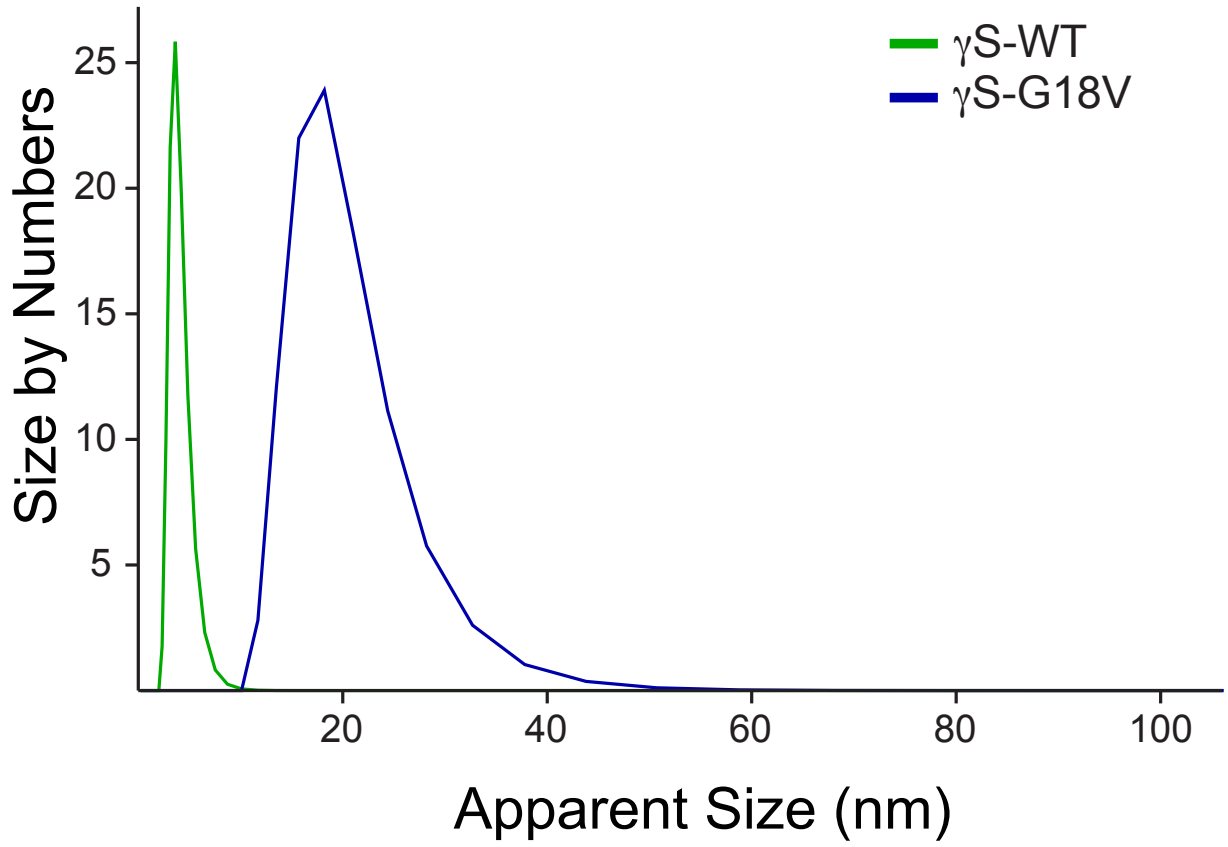

Figure A.8: Dynamic light scattering (DLS) data for γS-WT and γS-G18V in 10 mM phosphate at pH 6.9 displayed as the distribution of particle size by number. The concentrations of both proteins were at 0.3 mM, as in the NMR experiments. γS-WT is fully monomeric under theses conditions, however γS-G18V forms a mixture of transient dimers and small oligomers, as observed in previous studies.

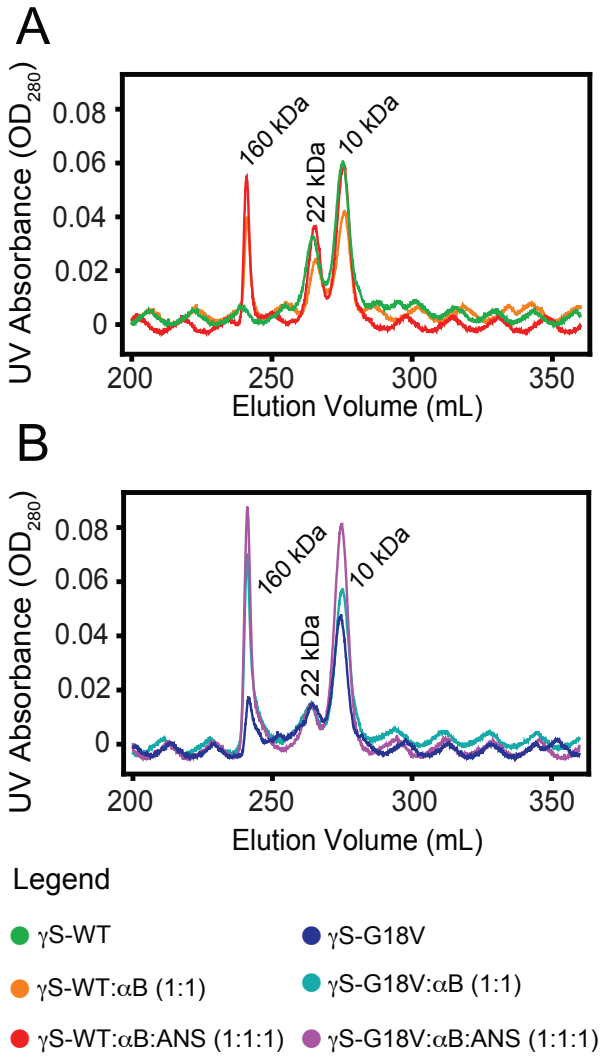

Figure A.9: Gel filtration chromatograms for γS-WT (A) and γS-G18V (B) in the presence and absence of αB-crystallin and ANS. (A) For γS-WT alone (green), the sample is mostly monomeric (10 kDa) with a small amount of dimers (22 kDa). Upon addition of αB-crystallin (orange), a population of larger oligomers (160 kDa) appears, at the expense of the populations of both the monomeric and dimeric states. Addition of ANS to this mixture (red) slightly increases the proportion of large aggregates. (B) For γS-G18V alone (blue), much of the sample is monomeric, although small populations of dimers and large oligomers exist. The peak at 160 kDa is much broader than in γS-WT, suggesting greater polydispersity. In the presence of αB-crystallin (cyan), the main effect is a dramatic narrowing of the peak corresponding to large oligomers, indicating a more uniform population. Addition of ANS to this mixture (purple) produces both further narrowing and an increase in the population of monomers, suggesting that interaction with ANS does disrupt the αB-γS complex formation.

Appendix B

Overview of the Docking Workflow

B.1 Background

Molecular docking can provide important structural information on bound ligands to known structures of receptors. This method often employs a simple energy function and search algorithm in order to quickly generate and rank bound conformations, often for the purpose of processing large libraries of compounds and/or receptors. Though some scoring functions provide binding energies for their poses, the standard error is too large to be predictive[33, 36]. Furthermore, benchmarks of several docking programs found that the scoring functions can reproduce co-crystallographic conformations within a set of top ranked poses, but are unable to identify the correct structure as the top scored pose[36]. Thus, the scoring functions serves best as a metric to rank the most plausible poses generated from the search algorithm, rather than pinpointing the exact binding mode. For this purpose, molecular docking serves as a fast, yet qualitative method to generate sets of top-ranked poses for further inquiry.

In this work, I seek the identify the protein-ligand binding sites. Since virtual screening efforts usually involve screening a particular active site, care should be taken when blindly

docking to an entire protein. To address this, I incorporated a docking protocol involving two stages of docking with two stages of filtering to cluster and remove poses based on prior knowledge from experiments. An outline of the docking protocol is illustrated in Figure B.1.

B.2 Blind docking

For this docking protocol, Autodock Vina[33] was chosen for its relatively fast operation. In the first step, docking simulation search spaces were distributed across the protein's entire surface. Since the search algorithm becomes much less effective with larger search spaces, a grid of 27 smaller 30 x 30 x 30 Å search spaces were used. The second step involves running the docking simulation for each search space, each of the 20 superimposed NMR conformations, and each protein variant, resulting in 1,080 runs for step 2. Since the goal of blind docking is to populate the binding sites for clustering, each run was configured to report the top 20 binding modes, rather than the default top 9 binding modes.

B.3 Define Binding Sites by Clustering

In the third step, the collective pose set is clustered (using RMSD clustering in VMD[11]). The chosen RMSD cutoff of 5 Å was sufficiently large enough to cluster together poses with different orientations, but still small enough to distinguish between binding sites. The resulting 50+ clusters are then filtered for interactions with strongly binding CSP residues. Clusters with no heavy atoms within 8 Å (the interaction cutoff of the Vina scoring function[33]) of a strong binding residues are removed from the set. The remaining clusters form the new search spaces for a flexible refinement.

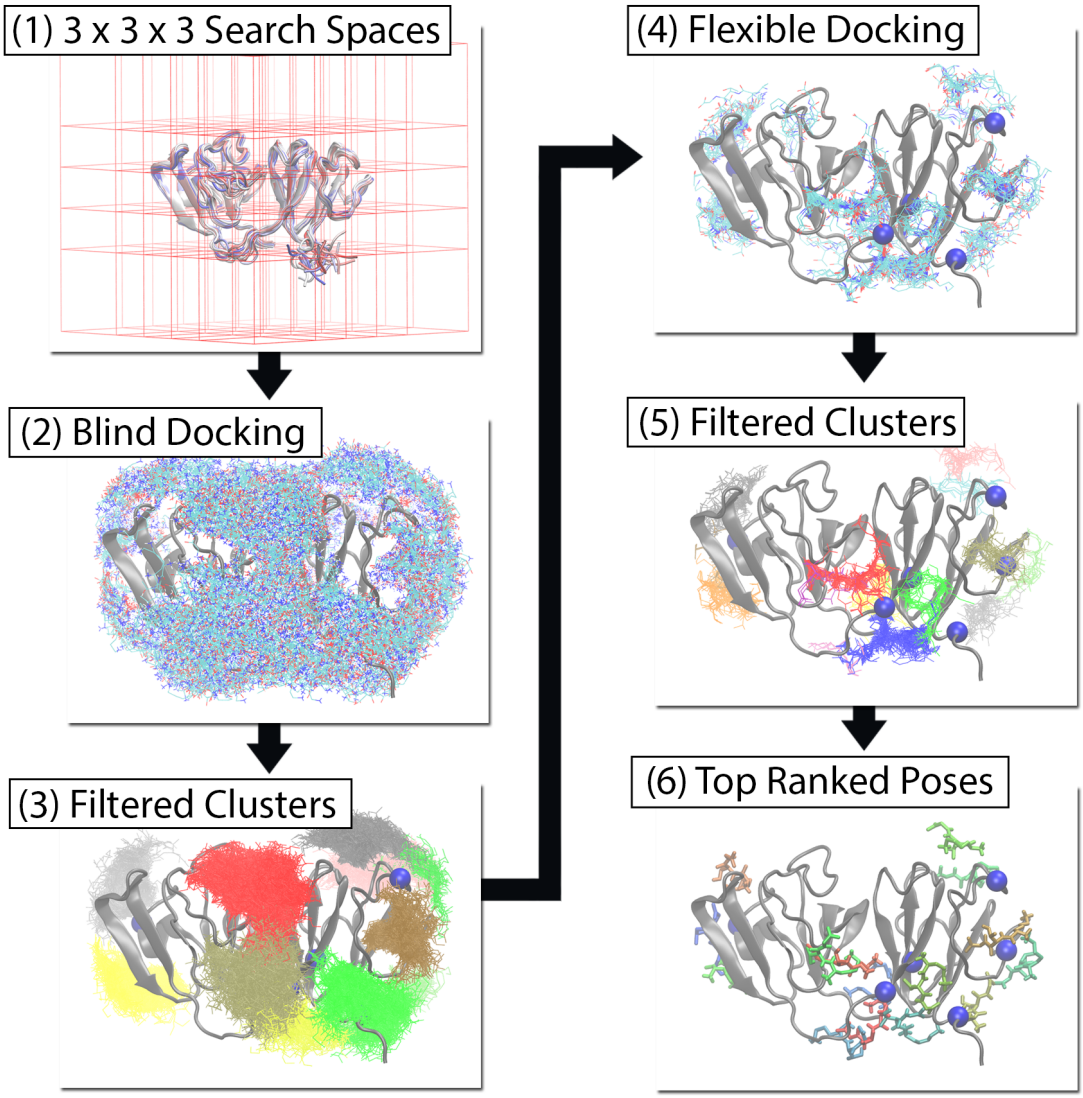

Figure B.1: Illustration of the docking workflow used to identify ANS binding sites. 1) The 27 search spaces shown in red outline around 21 superimposed NMR conformations. 2) Collective set of docked poses to a rigid protein. 3) Resulting pose clusters after filtering for contacts. Strong binding residues from CSP mapping are shown in blue spheres. 4) Docked pose set bound to a flexible protein. 5) Docked poses from flexible receptor docking. 6) The best scored poses for each binding site. These poses are then visually inspected for their contacts.

B.4 Flexible Refinement and Post-Filtering

For flexible docking, residues with strong and weak binding CSP values were set as flexible, and the ligand is redocked to the flexible protein. The pose set is then clustered and refiltered for contacts with strong binding residues. The set is further filtered for the contacts necessary for fluorescence. This requires polar interactions with the ANS sulfonate and hydrophobic interactions with the conjugated rings[20]. The final set of poses include the top ranked poses for each binding site that satisfies all the criteria for a fluorescent pose that contributes to the observed CSP. Each pose is then individually inspected for differences in binding modes between γS-WT and γS-G18V. The three identified binding sites at the site of mutation, behind the cysteine loop, and at the interdomain interface have stronger binding with γS-G18V directly resulting from changes in protein conformation linked to the G18V point mutation.