628 *Biochemistry* **2011,** *50,* 628–639
DOI: 10.1021/bi100978r

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Computational Design and Biophysical Characterization of Aggregation-Resistant Point Mutations for γD Crystallin Illustrate a Balance of Conformational Stability and Intrinsic Aggregation Propensity[†]

Erinc Sahin,[‡,§] Jacob L. Jordan,[∥] Michelle L. Spatara,[‡] Andrea Naranjo,[‡] Joseph A. Costanzo,[∥] William F. Weiss IV,[‡] Anne Skaja Robinson,[‡] Erik J. Fernandez,*[,∥] and Christopher J. Roberts*[,‡,§]

[‡]*Department of Chemical Engineering, and* [§]*Center for Molecular and Engineering Thermodynamics, University of Delaware, Newark, Delaware 19716, United States, and* [∥]*Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22904, United States*

Received June 18, 2010; Revised Manuscript Received December 22, 2010

ABSTRACT: γD crystallin is a natively monomeric eye-lens protein that is associated with hereditary juvenile cataract formation. It is an attractive model system as a multidomain Greek-key protein that aggregates through partially folded intermediates. Point mutations M69Q and S130P were used to test (1) whether the protein-design algorithm RosettaDesign would successfully predict mutants that are resistant to aggregation when combined with informatic sequence-based predictors of peptide aggregation propensity and (2) how the mutations affected relative unfolding free energies ($\Delta\Delta G_{un}$) and intrinsic aggregation propensity (IAP). M69Q was predicted to have $\Delta\Delta G_{un} \gg 0$, without significantly affecting IAP. S130P was predicted to have $\Delta\Delta G_{un} \sim 0$ but with reduced IAP. The stability, conformation, and aggregation kinetics in acidic solution were experimentally characterized and compared for the variants and wild-type (WT) protein using circular dichroism and intrinsic fluorescence spectroscopy, calorimetric and chemical unfolding, thioflavin-T binding, chromatography, static laser light scattering, and kinetic modeling. Monomer secondary and tertiary structures of both variants were indistinguishable from WT, while $\Delta\Delta G_{un} > 0$ for M69Q and $\Delta\Delta G_{un} < 0$ for S130P. Surprisingly, despite being the least conformationally stable, S130P was the most resistant to aggregation, indicating a significant decrease of its IAP compared to WT and M69Q.

Non-native protein aggregation is a long-standing problem during recombinant protein expression, purification, and storage of biotechnology products (*1*, *2*), as well as in the context of a variety of protein-deposition diseases (*3*, *4*). Non-native aggregation is the process of forming protein aggregates in which the constituent protein chains have adopted some degree of non-native secondary structures and/or non-native contacts, including interprotein contacts that stabilize the aggregated state. The most commonly cited non-native structures are inter- and/or intramolecular β-sheet motifs (*1*, *5*, *6*), of which the cross-beta, amyloid structure is among the most well studied due to its affiliation with a number of protein-deposition diseases (*6*, *7*).

Non-native aggregation (hereafter simply termed aggregation) is typically irreversible under the conditions that aggregates form and requires extremes of temperature, pressure, and/or concentrated chemical denaturants in order to dissociate the resulting aggregates (*1*, *5*, *8*). Control of aggregation rates or kinetics, rather than aggregate thermodynamics, therefore is important in many practical situations (*1*, *9*). Aggregation often proceeds through a partially or fully unfolded monomer state (*3*, *5*, *9*) involving multiple stages of folding, aggregate nucleation, and subsequent growth (*10*−*13*). As a result, aggregation rates for initially folded proteins under mildly or strongly native-favoring conditions are typically accelerated by decreasing conformational stability with respect to the aggregation-prone conformer(s). In this context, conformational stability is equivalent to the unfolding free energy (ΔG_{un}) from folded monomer to the unfolded or partially unfolded state (*14*) that is involved in aggregation; i.e., decreased conformational stability corresponds to an increased concentration of aggregation-prone conformers and thus to greater aggregation rates. Changes in conformational stability may be due to altered solution conditions, temperature, pressure, and/or mutations. However, even at a fixed concentration of unfolded conformers, it may be possible to alter the inherent "reactivity" or intrinsic aggregation propensity (IAP)[1] for a given protein sequence and/or conformational state (*7*, *15*−*21*). In practical terms, however, it may be difficult to alter only ΔG_{un} or IAP. For example, a given mutation, change in protonation, or altered disulfide bonding may alter both ΔG_{un} and IAP simultaneously. While ΔG_{un} may be accessible directly in experiment, IAP typically must be inferred indirectly unless one can directly monitor the amount of (partially) unfolded, aggregation-prone intermediate(s) (*9*).

In a general sense, IAP reflects the interactions between protein chains that lead to strong, relatively specific contacts that stabilize aggregates, in many cases through intermolecular β-sheet formation (*1*−*7*, *9*−*11*). Molecular scale models of polypeptide aggregation indicate that the basic driving forces for

[†]Financial support from the National Science Foundation (CBET-0853639) is gratefully acknowledged.

*Corresponding authors. E.J.F.: e-mail, erik@virginia.edu; tel, 434-924-1351; fax, 434-982-2658. C.J.R.: e-mail, cjr@udel.edu; tel, 302-831-0838; fax, 302-831-1048.

[1]Abbreviations: γD-crys, γD crystallin; ΔG_{un}, unfolding free energy; IAP, intrinsic aggregation propensity; LENP, Lumry−Eyring nucleated polymerization; WT, wild type; GuHCl, guanidine hydrochloride.

 Published on Web 12/24/2010

aggregation are the same as those for folding, i.e., maximizing burial of hydrophobic side chains, peptide−peptide hydrogen bonding, and favorable van der Waals contacts, while minimizing steric clashes, electrostatic energy, and loss of chain entropy (*7*, *22*, *23*). Thus, if one seeks to control aggregation via protein engineering, there is an inherent competition between folding and aggregation that in principle must be balanced. To date, the vast majority of rational protein-design algorithms have focused on maximizing the conformational stability of isolated monomers and/or the stability of protein−ligand complexes (*24*−*26*). There have been relatively fewer attempts to computationally design improved conformational stability as a means to slow aggregation.

Application of computational design to control IAP has been primarily restricted to small, natively unfolded polypeptides, where there is little or no contribution of ΔG_{un} to aggregation rates (*1*, *15*−*20*). Even in cases where researchers have been able to rationally engineer resistance to aggregation for folded proteins (*27*, *28*), it is often unclear whether aggregation rates were slowed because of a change in ΔG_{un}, a change in IAP, or both. In addition, it typically is not known whether overall aggregation rates are slowed because of changes in nucleation and/or growth stages. This follows because of one or more of the following (*9*): (i) only overall aggregation rates or extents of reaction are monitored; (ii) only heuristic models are available or employed to quantify nucleation and growth stages based on apparent lag times and sigmoidal aggregation profiles; (iii) the assays employed do not monitor both extent of reaction (monomer loss) and one or more moments of the aggregate size distribution. More robust mechanistic approaches (*12*, *29*−*33*) are available to separately quantify nucleation and growth rates, including the Lumry−Eyring nucleated polymerization (LENP) treatment adopted here and elsewhere for analyzing aggregation kinetics of globular proteins (*29*, *31*, *32*, *34*, *35*) and monoclonal antibodies (*36*, *37*).

Human γD crystallin (γD-crys) is a member of the crystallin family and is important for lens transparency in the eye. The protein is rich in antiparallel β-sheets arranged in Greek-key motifs, similar to the fold in many immunoglobulins (*38*). γD-crys is natively monomeric and highly stable in terms of its conformational stability at neutral pH, yet is susceptible to aggregation during refolding (*11*). The wild-type gene was previously cloned and is readily expressed in *Escherichia coli* as soluble monomers of ca. 20 kDa molecular mass (173 amino acids) (*11*, *39*). Careful refolding studies of the protein near physiological conditions (pH 7.0, 37 °C) showed an unfolding transition midpoint at ≈2.7 M GuHCl, with a partially folded, aggregation-prone state populated at concentrations of GuHCl below 1.0 M (*11*, *39*). Aggregation starting from the folded state occurs in the absence of denaturant at acidic pH. The unfolding free energy is greatly reduced relative to neutral pH, making it easier to populate partially or fully unfolded monomers without requiring temperatures near boiling (*40*). The resulting aggregates in both cases are amyloid, with characteristic cross-β-sheet fiber X-ray diffraction patterns and fibrillar aggregate morphology (*11*, *40*). The relative roles of nucleation and growth in aggregation of wild-type γD-crys or its variants, or the quantitative roles of conformational stability or IAP have not yet been determined.

γD-crys was selected as a model protein for testing the rational design of aggregation resistance within the context of multidomain, initially folded proteins. The Greek-key fold is structurally similar to many of the multidomain, antibody-based systems of interest as pharmaceutical products, for which aggregation is a common problem during drug development (*1*, *2*). Longer term goals include assessing design strategies that target a given domain and the domain−domain interface, as well as isolated domains for γD-crys. As such, γD-crys provides a useful model system without having to deal with multiple interfaces and differential stability of more than two domains. This first report focuses on acidic pH in order to monitor aggregation kinetics on reasonable time scales while remaining at typical temperatures for accelerating aggregation via (partial) unfolding (*1*, *2*, *34*).

A growing body of work documents the importance of amino acid interactions across domain−domain interfaces for conformationally stabilizing multidomain proteins (*41*−*46*). This includes systematic investigations of (de)stabilizing effects of single-point mutations for proteins such as γD-crys (*45*−*47*) and a variety of single-chain antibodies (*13*). For multidomain proteins, it is not uncommon that at least one domain is significantly less stable than the other(s) (*48*, *49*), and formation of the domain−domain interface stabilizes that otherwise unstable domain (*13*, *50*). If domain−domain binding is sufficiently strong, folding of one domain provides a sufficiently high-affinity surface to drive folding and association of the other domain, and apparent two-state equilibrium is recovered (*13*, *43*, *44*, *48*, *51*). In the context of protein aggregation and the role of conformational stability in controlling aggregation rates, the stability of the domain−domain interface is therefore anticipated to be important if the partially unfolded intermediate is aggregation prone, as is the case for wild-type γD-crys.

One computational approach successfully used to stabilize globular proteins in the context of protein folding (in the absence of aggregation) is the design module of the molecular modeling software Rosetta. Originally developed for protein structure prediction (*52*), Rosetta has been built upon for a variety of applications such as increasing folding stability (*25*), protein docking (*24*), and loop modeling (*26*). The design module of Rosetta, RosettaDesign, utilizes an energy score incorporating a Lennard-Jones potential, the Lazaridis−Karplus implicit solvation model (*53*), orientation-dependent hydrogen-bonding terms (*54*), torsion potentials based on backbone and rotamer alignment (*55*), a reference value unique to each of the twenty amino acids, and finally a term accounting for electrostatic interactions between charged residues. To estimate low energy sequences, RosettaDesign uses Monte Carlo optimization with simulated annealing. Given a target protein structure, RosettaDesign searches for amino acid sequences that pack well, bury hydrophobic atoms, and satisfy the hydrogen-bonding potential. Amino acid substitutions or rotamer alterations are accepted or rejected based on the Metropolis criterion (*56*). Recently, RosettaDesign has been rather comprehensively tested for improved conformational stability (*57*, *58*), as well as improving binding affinity (*59*) and specificity (*60*, *61*).

There are a number of primary-sequence-based, aggregation "calculators" that have been shown with reasonably short polypeptides to predict "hot spots", i.e., stretches of amino acid sequence that are highly aggregation prone (*15*−*21*). The majority of these algorithms utilize some combination of physical properties, phenomenological or informatic analysis of structural databases, and/or molecular simulation to statistically rank the relative probability of small polypeptides sequences (∼5−10 residues) to aggregate. Almost all are based on multivariate statistical regression against training-set databases of experimental aggregation rates for small polypeptides. The training- and testing-set

databases are typically composed of polypeptides that do not stably fold to form secondary or tertiary structures as isolated monomers but do form β-sheet structures upon aggregation. The resulting algorithms therefore provide a statistical or structural informatic approach to possibly predict the effects of point mutations on the IAP of short amino acid sequences in foldable proteins, assuming that those sequences are solvent exposed and sterically accessible within (partially) unfolded states (*15*−*21*).

To date, no algorithm has been shown to provide clearly superior performance in head to head comparison, and such algorithms have only been tested against experimental aggregation behavior of full-length, foldable proteins in the case of relatively simple/small proteins or fragments (*62*). To the best of our knowledge, no results have been reported in the context of experimental aggregation of multidomain proteins or used in an analysis of the relative effects of ΔG_{un} vs IAP in controlling aggregation rates of foldable proteins. In one recent review (*63*), a subset of these algorithms gave predictions regarding aggregation "hot spots" in a set of marketed therapeutic monoclonal antibodies, but no experimental data were available to test the validity of those predictions.

The goal of the present study is a first-stage evaluation of whether variants of γD-crys that are selected with the aid of computational design algorithms are able to show differences in experimental aggregation rates and/or pathways, based on whether a given mutation leads to a change in monomer conformational stability vs a change in intrinsic aggregation propensity or both. Prospective mutants were screened with a combination of RosettaDesign (*25*) and a subset of the available aggregation-propensity "calculators" summarized above. The "calculators" chosen for this work were AGGRESCAN (*16*), TANGO (*17*), and PASTA (*21*). These were selected to span a reasonable range of alternative approaches, different training- and testing-set databases and because all are available in the public domain and do not focus solely on one specific type of aggregate structure. A detailed examination of all available algorithms was beyond the scope of the present study. As a first study, mutants were selected such that they were predicted to have a significant improvement in IAP, with small or negligible change in ΔG_{un}, or vice versa.

Selected mutants and wild-type (WT) γD-crys were expressed and characterized in terms of their folding stability and aggregation behavior. The results here show a number of differences between WT and the selected variants and lay a foundation for future work to address a number of fundamental questions regarding the molecular mechanisms of aggregate formation and the roles of conformational stability vs intrinsic aggregation propensity in determining aggregation pathways and kinetics.

MATERIALS AND METHODS

Computational Predictions of Point Mutants for ΔG_{un} and IAP. RosettaDesign 2.1 was used to carry out fixed backbone-based design of the wild-type γD crystallin crystal structure (PDB code: 1HK0 (*64*)). During fixed backone design, only the side chains are permitted to move which results in the need for additional rotamer libraries and dihedral-space conformers to sufficiently sample Lennard-Jones potentials (*65*). Both point mutation scans and interfacial binding energy scans were performed. In silico point mutations were generated for each of the 20 naturally occurring amino acids at every residue position provided in the crystal structure. Mutations were evaluated based on their total Rosetta energy score, solvent-accessible surface area packing score, and hydrogen-bonding energy score relative to the wild-type protein. For interfacial binding energy scans, the crystal structure was separated at residue 86 into two separate chains. RosettaDesign treats all ionizable amino acids as existing in the charge state that dominates at neutral pH; as such, experimental pH is not an input to the calculations.

AGGRESCAN, PASTA, and TANGO were used to predict sequences within wild-type γD crystallin regions with the highest tendency for aggregation. Default parameters were employed for all predictions. The consensus region (V126−Y134) that was predicted by all three predictors to be an aggregation hot spot was selected for evaulations of potential point mutations, in an effort to balance differences between the underlying algorithms. S130P was selected as a variant that was predicted by RosettaDesign to have minimal effects on conformational stability but caused V126−Y134 to no longer be a predicted "hot spot" for aggregation in the aggregation algorithms above.

Among the computational tools utilized for our study, only TANGO provided the option to enter solution parameters (pH and ionic strength) in addition to sequence information, allowing our experimental conditions to be used as input parameters (pH 3.0, 22 mM ionic strength) for the prediction listed in Results.

Molecular Cloning for Generation of Mutants. *E. coli* DH5α bacterial strain was used for genetic manipulation and *in vivo* amplification of the mutated plasmids. Bacteria were transformed with pQE-1 that carried the 6-His-tagged human wild-type γD crystallin gene (pQE1-gDc-WT), a gift from Jonathan King (MIT). Transformed cells were cultured in Luria−Bertani (LB) liquid media or agar plates, both supplemented with carbenicillin (Sigma, St. Louis, MO) at a final concentration of 100 μg/mL. Mutations were introduced into pQE1-gDc-WT using QuickChange II XL site-directed mutagenesis kits purchased from Stratagene/Agilent Technologies (La Jolla, CA) as described in the manufacturer's manual. Primers used for site-directed mutagenesis (M69Q, CCACCAGCAGTGGCAGGGCCTCAGCGAC; $M69Q_{reverse}$, GTCGCTGAGGCCCTGCCACTGCTGGTGG; $S130P_{forward}$, GTGCT GGAGGGCCCCTGGGTCCTCT; $S130P_{reverse}$, AGAGGACCCAGGGGCCCTCCAGCAC; substitution mutations are underlined) were purchased from Integrated DNA Technologies (Coralville, IA).

The WT and mutant γD crystallin genes were each subcloned into a pET15b plasmid (Novagen) for expression. Genes were PCR amplified using PfuUltra polymerase (Stratagene). The forward primer (CGACGACATATGGGGAAGATCACC) contained an *Nde*I site, and the reverse primer (GACGCCGGATCCAAATTAAGAAACAAC) contained a *Bam*HI site. Following digestion of the PCR product and the pET15b plasmid with *Nde*I and *Bam*HI (New England Biolabs), the PCR product was inserted into the pET15b plasmid with T4 ligase (New England Biolabs). The final expression constructs were verified by full gene sequencing.

Protein Expression and Purification. Frozen stocks of *E. coli* BL21(DE3) strains carrying pET15-gDc-WT, pET15-gDc-M69Q, and pET15-gDc-S130P plasmids were grown overnight on LB media containing 15% (w/v) agar. Single colonies of individual γ-D crystallin variants were selected and grown overnight as primary cultures in 100 mL of super broth (SB) liquid media in 250 mL flasks at 37 °C and 300 rpm. Primary cultures were diluted into fresh 1.3 L SB liquid media in 4 L flasks to obtain secondary cultures with a final optical density of 0.05 at 600 nm (OD_{600} = 0.05). Secondary cultures were grown at 30 °C

and 300 rpm until $OD_{600} \approx 0.8$, at which time isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to yield a final concentration of 1 mM. The IPTG-induced culture was grown further at 30 °C and 300 rpm for 24 h. All solid and liquid media contained 100 μg/mL carbenicillin to maintain selective pressure.

E. coli cells were harvested via centrifugation at 5000*g* for 30 min at 4 °C. Pellets were weighed and frozen at −80 °C for storage. Frozen pellets were resuspended in 100 mL of chilled buffer comprised of 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, adjusted to pH 8.0. Following resuspension, the cells were lysed via passing the suspensions through a Microfluidics M-110 L microfluidizer five times at 6000 psi pressure. To reduce viscosity of the lysate, in some cases DNase was added to a final concentration of 20 μg/mL in 10 mM magnesium chloride, followed by a 30 min incubation at 20 °C.

After centrifugation at 20000*g* at 4 °C for 30 min, the collected supernatant was incubated with 20 mL of a 50% slurry of Ni-NTA agarose (Qiagen) at 4 °C for a minimum of 2 h. Ni-NTA agarose beads were washed five times with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 20 mM imidazole, adjusted to pH 8.0 with NaOH) via centrifugation (3200*g* at 4 °C) and resuspension steps. Ni-NTA agarose matrix was resuspended and transferred into Bio-Rad Econo-Pac gravity chromatography columns. γD-crys variants were eluted with 3 column volumes of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole, adjusted to pH 8.0 with NaOH). The protein-rich second column volume of eluate was immediately dialyzed using 8 kDa molecular mass cutoff Spectra/Por 7 dialysis tubing against 4 × 0.5 L of 100 mM sodium phosphate buffer, pH 7.4 at 4 °C, resulting in some precipitation. Soluble proteins were separated from the precipitate via centrifugation at 3200*g* at 4 °C. The protein-rich fraction was applied to a Superdex75 column and eluted isocratically with 10 mM phosphate buffer, pH 7.0, to obtain homogeneous monomer solutions, excluding low molecular mass aggregates that can potentially act as seeds in aggregation. Protein was stored at 4 °C for up to 1 week or aliquoted and frozen at −20 or −80 °C for longer times, neither of which resulting in detectable levels of aggregation. Prior to experiments, γD-crys variants were dialyzed into 50 mM citrate buffer, pH 3.0, adjusted with NaOH, and used within 2 days of dialysis. Quality of protein solutions was verified via visual inspection for formation of precipitates and size exclusion chromatography for presence of soluble aggregates prior to each experiment. Protein concentration was determined by absorbance at 280 nm using an extinction coefficient of 41040 M^{-1} cm^{-1}.

Equilibrium Unfolding/Refolding. All chemical unfolding and refolding experiments were carried out using ultrapure urea (MP Biomedicals). For unfolding experiments, a solution of 0.1 mg/mL wild-type γD-crys or protein variant was prepared in 50 mM citrate buffer, pH 3. The solution was then diluted 10-fold into samples with urea concentrations ranging from 0 to 6.2 M, which were calculated gravimetrically (molal basis) and converted to a molar basis using the density of urea−water solutions as a function of urea concentration (*80*). For refolding experiments, protein was denatured at a concentration of 0.1 mg/mL in 7 M urea with 50 mM citrate, pH 3, for 24 h at 20 °C. The denatured protein solution was then diluted 10-fold into refolding samples with urea concentrations between 0.7 and 6.1 M. Both unfolding and refolding samples were incubated at 20 °C for 24 h to reach equilibrium. Fluorescence data were collected at 20 °C using a spectrofluorometer (ISS, IL). Excitation and emission slit widths of 1.0 and 2.0 mm, respectively, were used, and the excitation wavelength was set to 295 nm. The ratio of the baseline-corrected emission intensities at 360 and 320 nm was used for analysis. To determine thermodynamic parameters (ΔG_{un} and *m*-values), equilibrium unfolding data were globally regressed to a two-state folding model (*66*) using Matlab software. A three-state model with an intermediate species (*67*) was also tested, but no improvement in the magnitude or randomness of residuals was observed.

Differential Scanning Calorimetry (*DSC*). Thermal transitions of WT, M69Q, and S130P γD-crys variants were determined via DSC. Measurements were performed using a VP-DSC instrument (Microcal, Northhampton, MA) with 0.5 mg/mL protein in 50 mM citrate buffer, pH 3.0. At least five buffer−buffer thermograms were collected to obtain the average background. Sample thermograms were obtained immediately after buffer−buffer thermograms so as to maintain a common thermal history. All scans were collected between 25 and 80 °C, with a 60 °C/h scan rate. To test for overall reversibility, scans were performed up to the end of any observable endotherms, cooled to 25 °C at nominally 6 °C/min, and immediately reheated to 80 °C without removing sample from the cell. Baseline-subtracted data were converted to excess partial specific heat capacity (c_p) by subtracting the interpolated pre- and posttransition baselines, as no exotherms were observed for any conditions tested (results not shown).

Isothermal Aggregation. Isothermal aggregation was monitored as a function of incubation time at elevated temperature using samples with an initial protein concentration (c_0) of 1 mg/mL in 50 mM citrate buffer adjusted to pH 3.0, a solution condition which allows aggregates to grow without formation of insoluble species. Incubation temperature was selected as 50 °C for WT and both mutants. Incubation was performed in upright, hermetically sealed, deactivated borosilicate glass HPLC vials (Waters, Milford, MA) in circulating water baths with less than ±0.2 °C variability, as verified by an analog thermometer. Samples were removed at selected time intervals and quenched immediately on an ice−water bath. No measurable changes in monomer concentration or increase in aggregate size were detected for samples on the time scales within which analytical characterization experiments were performed for each batch (data not shown).

Size Exclusion Chromatography with In-Line Multi-angle Light Scattering (*SEC-MALS*). Molecular weight distributions of native and aggregated protein samples were analyzed via SEC-MALS. Mobile phase (0.5% phosphoric acid, adjusted to pH 2.7) was operated at 1 mL/min flow rate. Injections of 100 μL and 1 mg/mL protein samples were performed with a Waters (Milford, MA) Protein-Pak 125 (7.8 × 300 mm) column connected to Waters Alliance 2695 separation module and Waters 2996 photodiode array detector for separation and quantitative detection at 280 nm. Peak areas were calculated with respect to external standards. SEC was coupled to a Wyatt Technologies (Santa Barbara, CA) DAWN HELEOS II MALS detector (658 nm laser) and Wyatt Optilab rEX differential refractive index (RI) detector. The peak areas and weight-average molecular weight (M_w) of each peak and each 1 s slice of SEC peaks were calculated using Empower (Waters) and ASTRA V (Wyatt) software packages connected to HPLC and MALS instruments, respectively. As described elsewhere (*32*), SEC-MALS is capable of obtaining a weight-average molecular weight (M_w) for each slice of a chromatogram, and these slices can then be averaged across an entire SEC peak. The peak does not need to be a single species, and there is no need for calibration against

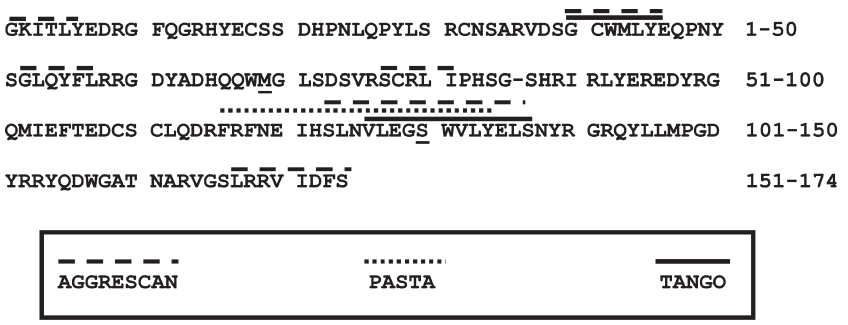

FIGURE 1: Aggregation-prone sites on wild-type γD crystallin based on AGGRESCAN (dashed line), PASTA (dotted line), and TANGO (solid line). The consensus region of the three predictors (V126−Y134) was recognized as a site of interest for increasing colloidal stability by destabilizing aggregation. Locations of experimentally characterized mutations (M69Q and S130P) are underlined.

SEC molecular weight standards. If one combines M_w across all aggregate peaks (*32*), the result is the weight-average molecular weight of the (pooled) aggregates M_w^{agg}; if one instead averages across all peaks (monomer and aggregates), one obtains M_w^{tot}, and this is the equivalent of what is obtained from batch light scattering as long as one does not lose significant protein to the SEC column (*31*, *32*).

LENP Model Regression. Monomer loss and aggregate growth as a function of time were analyzed using a moment form of the LENP model of non-native aggregation (eq 1) that is applicable at early times, before significant aggregate−aggregate coalescence occurs (*31*, *32*, *68*). Lack of coalescence was confirmed as described previously (*31*, *32*, *68*), based on a linear scaling of aggregate molecular weight with monomer loss (not shown).

$$\frac{dm}{dt} = -x\tau_n^{-1}m^x - \delta\tau_g^{-1}m^\delta \quad (1a)$$

$$\frac{d\sigma}{dt} = \tau_n^{-1}m^x \quad (1b)$$

$$\frac{d\lambda_2}{dt} = x^2\tau_n^{-1}m^x + \tau_g^{-1}m^\delta(\delta^2\sigma + 2\delta(1-m)) \quad (1c)$$

$$\frac{M_w^{tot}}{M_{mon}} = 1 + \lambda_2 \quad (1d)$$

In eq 1, x is the stoichiometry of the nucleus, τ_n is the characteristic time scale or inverse rate coefficient for nucleation of new aggregates, τ_g is the analogous quantity for growth of aggregates via monomer addition, and δ is the number of monomers per growth step. M_{mon} is the monomer molecular weight, m denotes the fraction of protein (by mass) that is monomer in SEC, and M_w^{tot} denotes the weight-average molecular weight averaged over both monomer and aggregates. x and δ are integers, with $\delta = 1$ being typically assumed (*12*, *68*−*70*). The values of x and δ were regressed by allowing them to take on integer values between 1 and 10, while performing nonlinear least-squares regression for τ_n and τ_g at each (x, δ) as described previously (*31*, *32*). Equation 1 was regressed globally versus both m and M_w^{tot} values as a function of incubation time for each variant at 50 °C. The regressed parameters were τ_n and τ_g (real positive numbers) for a given choice of x and δ (integers only). Only $x = 2$ and $\delta = 1$ gave reasonable fits, and also were consistent with formation of a discernible dimer species at early times for M69Q and WT. For reasons discussed in the Results section, fits were not performed for S130P.

RESULTS

Selection of Mutants. Mutants were selected in an attempt to test alternative approaches to imbue aggregation resistance. The first variant was chosen with significantly improved conformational stability predicted via RosettaDesign, without significantly altered intrinsic aggregation propensity (IAP) predicted by TANGO, PASTA, and AGGRESCAN. In contrast, the second variant was chosen to significantly decrease the IAP but have little or no change on conformational stability. Figure 1 shows the IAP predictions from the three algorithms above. The only consensus sequence they predicted was V126−Y134.

From the 3460 possible simulated point mutations evaluated by Rosetta, 785 structures had energy scores greater than 1 kcal/mol, indicating stabilization relative to wild-type protein. Of these, 36 structures resulted in an improvement of SASApack and hydrogen bond energy scores. Fifteen of these mutations were located in the N-terminal domain, while the other 21 mutations were located in the C-terminal domain.

Mutations from the interface binding energy scan were filtered as previously described (*59*). Apparent binding energies were calculated by subtracting the energy score of the complex from the energy scores of the individual domains. Mutations that stabilized binding by less 0.5 kcal/mol or destabilized folding of a single domain by greater than 0.5 kcal/mol were eliminated from consideration. Of the 14 resulting mutations, 10 were in the N-terminal domain. Three of these 10 structures were located at residue M69. M69F, M69W, and M69Q were all found to stabilize the interface as well as the N-terminal domain via RosettaDesign. Of the 10 best candidates, M69Q had the largest total stabilization by the three metrics in Table 1 and was consequently selected as a variant of interest.

Of the mutations resulting from the interfacial binding energy scan, S130P was the only mutation identified in the 126−134 region. This mutation eliminated the tendency for this region to be aggregation prone in the TANGO predictor. In addition, on the individual amino acid aggregation propensity scales inherent in a number of the aggregation calculators (*16*−*18*, *21*, *71*), proline is the least likely of all amino acids to form amyloid aggregates. In contrast to M69Q, Rosetta predicted that S130P would marginally stabilize the interface and slightly destabilize the C-terminal domain. Therefore, M69Q and S130P were chosen as two appropriate mutations to examine these two contrasting approaches to imbue aggregation resistance.

M69 lies in the N-terminal domain (Ntd), near the interface between Ntd and C-terminal domain (Ctd), while S130 lies in the Ctd. Both M69 and S130 are adjacent to tryptophans that are

Table 1: Predicted Effects of Mutations at Selected Residues on (Interdomain) Binding and (Individual Domain) Folding Stabilities Calculated via RosettaDesign

	predicted $\Delta\Delta G_{binding}$[a] (kcal/mol)	predicted $\Delta\Delta G_{folding,Ntd}$ (kcal/mol)	predicted $\Delta\Delta G_{folding,Ctd}$ (kcal/mol)
M69F	−1.3	−2.5	0
M69Q	−0.6	−5.1	0
M69W	−0.9	−2.9	0
S84Y	−0.7	0.3	0
G85A	−1.2	0.3	0
G85D	−1.6	0.1	0
G85K	−1.5	0.5	0
G85N	−1.4	0.1	0
G85T	−0.5	−0.4	0
G85 V	−0.9	0.1	0
H88W	−0.6	0	−1.3
S130P	−0.6	0.0	0.5
S166D	−0.7	0	−0.1
S166Q	−0.5	0	−0.1

[a] $\Delta\Delta G_{binding} = \Delta G_{binding,WT} - \Delta G_{binding,mutant}$; $\Delta\Delta G_{folding} = \Delta G_{folding,WT} - \Delta G_{folding,mutant}$.

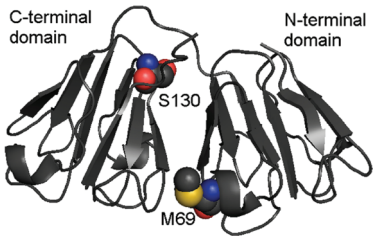

FIGURE 2: Crystal structure of γD crystallin (PDB ID: 1HK0) showing the locations of mutated residues in space-filling mode. For the residues represented by space-filling mode, dark gray, red, blue, and yellow colors represent carbon, oxygen, nitrogen, and sulfur, respectively.

buried in the folded state for WT γD-crys. As a point of clarification in the numbering convention here and in the PDB crystal structure, position 86 does not correspond to an amino acid and instead is used as the dividing line between N- and C-terminal domains. The locations of M69 and S130 within the wild-type structure are highlighted as space-fills in Figure 2 (PDB ID: 1HK0).

Differential Scanning Calorimetry. The effects of mutations on conformational stability of γD-crys were assessed via DSC (Figure 3) based on differences in the temperature at the maximum of the endotherm (T_m), with wild type and both mutants showing a single endotherm. Wild-type γD-crys had a T_m value of approximately 57 °C, while M69Q displayed a significantly higher T_m (ca. 60 °C) and S130P a much lower T_m (ca. 54 °C). The transitions were not fully reversible in any of the γD-crys variants, precluding analysis to obtain free energies or entropies for unfolding. The endotherms were convoluted with aggregation to some degree, based on post-DSC analysis by SEC (not shown), making additional quantitative analysis inconclusive.

Chemically Induced Equilibrium Unfolding/Refolding. Intrinsic tryptophan fluorescence was used to compare the tertiary conformation of the M69Q and S130P variants to that of wild-type γD-crys at pH 3 and 20 °C and to follow conformational changes during unfolding as a function of [urea]. The excitation wavelength was 295 nm, and the emission spectra of

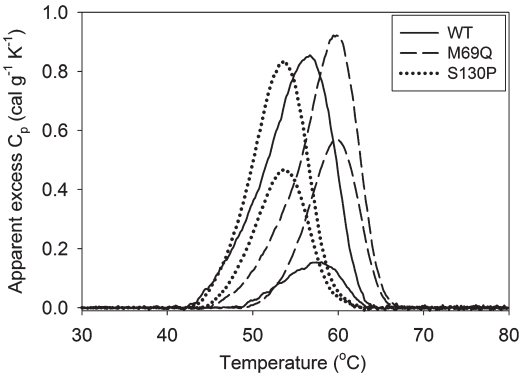

FIGURE 3: Differential scanning calorimetry thermograms of wild type (solid lines), M69Q (dashed lines), and S130P (dotted lines) mutants. Peaks sharing the same symbol with higher and lower amplitudes of apparent c_p belong to first and second scans of the same sample, respectively.

native and urea-denatured wild-type, M69Q, and S130P proteins were collected from 300 to 420 nm, shown in Figure 4A. Upon denaturation with urea, the spectra of all three molecules showed an increase in the emission intensity and a shift of the emission maximum from ~320 nm (native) to ~360 nm (denatured), which is consistent with previously reported behavior for guanidine hydrochloride-denatured γD-crys at pH 7 (*11*, *39*, *45*−*47*). Previous work showed that W68 and W157 are responsible for the anomalous increase in the intensity of tryptophan fluorescence upon exposure to the polar solvent environment, while W42 and W131 display the expected quenching of fluorescence upon unfolding (*39*).

Equilibrium unfolding/refolding experiments were performed in order to compare the conformational stability of the wild-type and protein variants. Unfolding and refolding samples were allowed to equilibrate at 20 °C for 24 h prior to recording fluorescence data. The ratio of the emission intensities at 360 and 320 nm is plotted as a function of urea concentration in Figure 4B−D. An unfolding/refolding hysteresis was observed in all three molecules but is significantly less pronounced in S130P as compared to WT and M69Q. To help ensure that the observed hysteresis was not the result of insufficient equilibration time, selected refolding samples were incubated at 20 °C for an additional 24 h (48 h total), and no significant change in fluorescence was observed (data not shown). All refolding profiles returned to baseline (folded) values at sufficiently low [urea]. The two-state fits for unfolding are shown in Figure 4B−D. Apparent thermodynamic parameters are included in Table 2. As predicted by Rosetta, the M69Q variant showed increased stability over the wild type, based on both an increase in the standard Gibb's free energy of folding (ΔG_{un}) and an increase in the midpoint unfolding concentration of urea (C_m). To be conservative, 95% confidence intervals from the global fits are reported in Table 2, rather than the more common practice of reporting only standard deviations and thereby underestimating the uncertainty of ΔG_{un} when extrapolated to zero denaturant. Differences in conformational stability are more clearly evidenced by differences in C_m, akin to T_m differences in DSC (Figure 3). For the S130P variant, the ΔG_{un} value is only slightly less than that of the wild type. However, interestingly, the mutation causes an increase in the *m*-value, resulting in a significant decrease in C_m. The relative differences in the unfolding profiles are more easily visualized by comparing an overlay of the curves for wild type and each variant in Figure S1D of the

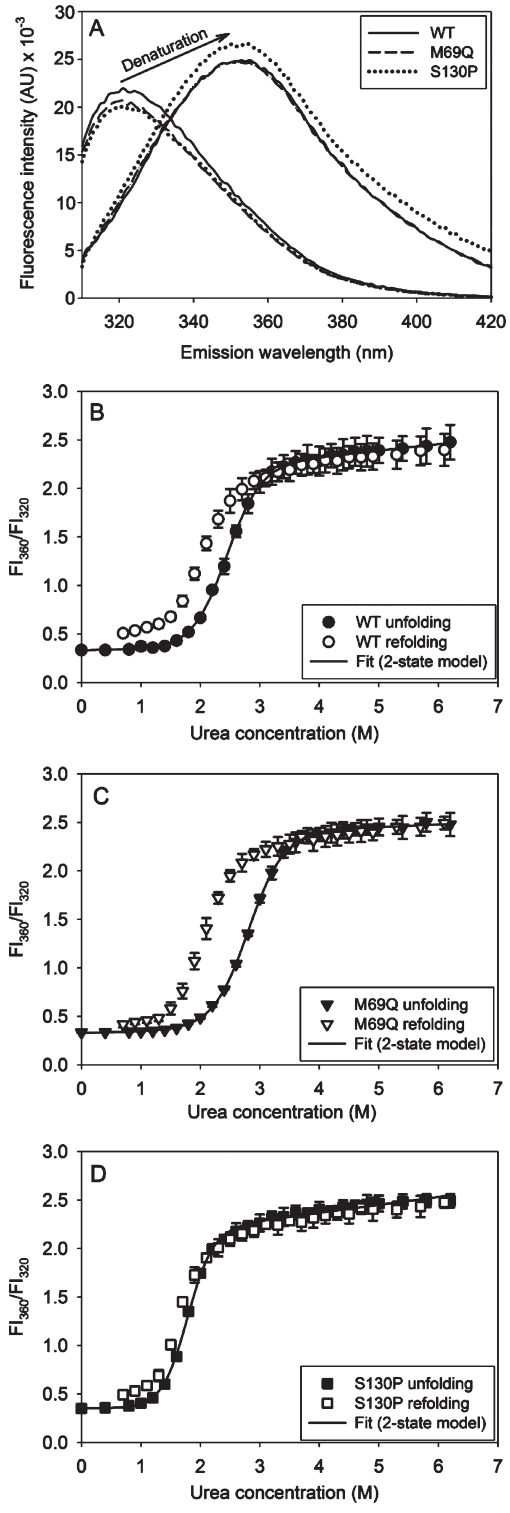

FIGURE 4: (A) Maxima of intrinsic fluorescence spectra shifted from 320 to 360 nm upon urea-induced unfolding (spectra for 0 and 8 M urea are shown; the spectra for intermediate urea concentrations are presented in Supporting Information Figure S1). The ratio of fluorescence intensity at 360 and 320 nm (FI_{360}/FI_{320}) was used to obtain unfolding (filled symbols) and refolding (open symbols) profiles for (B) wild type (circles), (C) M69Q (triangles), and (D) S130P (squares) as a function of urea concentration. Apparent thermodynamic values were determined by a fit of the unfolding data to a model for a single transition between two states. See also Supporting Information for overlay comparison of wild type with both variants.

Supporting Information. From there it is visibly apparent that the unfolding curve for M69Q is shifted slightly but statistically significantly to higher [urea] compared to wild type and that for S130P is shifted to significantly lower [urea] than wild type.

Table 2: Comparison of Conformational Stabilities of γD Crystallin Variants Calculated via a Fit of the Chemical Unfolding Data to a Model for a Single Transition between Two States[a]

	WT	M69Q	S130P
ΔG_{un} (kcal/mol)	5.1 ± 0.3	5.5 ± 0.3	4.8 ± 0.6
m-value [kcal/(mol·M)]	2.1 ± 0.1	2.0 ± 0.1	2.7 ± 0.3
C_{mid} (M)	2.4 ± 0.1	2.8 ± 0.1	1.8 ± 0.2

[a]Uncertainties are 95% confidence intervals from nonlinear regression.

Isothermal Aggregation Analyzed via SEC-MALS. Wild-type, M69Q, and S130P variants of γD-crys were compared with respect to their qualitative and quantitative aggregation, in terms of the monomer mass fraction (*m*) and the weight-average molecular weight of the sample (M_w^{tot}) at a given incubation time (*t*). All aggregates remained soluble, as determined by a lack of visible particulates or haze and based on a constant total area under the combined peaks of the SEC chromatograms, except for the extreme case where S130P was incubated at elevated temperature for 1440 min. Aggregates were generated at elevated temperature (50 °C), and samples representing various time points were collected and analyzed via SEC-MALS, as described in detail in Materials and Methods.

Panels A−C of Figure 5 show illustrative chromatograms for collected time points during aggregation experiments for WT, M69Q, and S130P, respectively. All values of molecular weight are divided by the monomer molecular weight in order to provide an estimate of the average number of monomers per aggregate. Monomer molecular weights were measured via MALS as 20.7 ± 0.2, 20.8 ± 0.2, and 20.9 ± 0.3 kDa for wild type, M69Q, and S130P, respectively. Values of *m* and M_w^{tot}/M_w^{mon} for the early time points of temperature-induced aggregation are shown as a function of incubation time for WT and M69Q in Figure 6. Early time points were used in order to focus on the competition between nucleation and growth, rather than later time phenomena of aggregate−aggregate association that can make distinguishing differences in nucleation rates problematic (*31*, *32*, *68*). In addition, using only earlier time points avoids experimental artifacts with SEC-MALS for very large aggregates (e.g., $M_w/M_{mon} \gg 10^2$) that scatter much more strongly than the monomer and that make monomer concentrations unreliable due to peak overlap and scattering contributions to the apparent UV signal. Fits of the data with the LENP model (eq 1) yielded a nucleus size of a dimer ($x = 2$), with growth by monomer addition ($\delta = 1$). Best fits were determined by the smallest sum of squared residuals, random distribution of residuals, and smallest relative 95% confidence intervals for the fitted values of τ_n and τ_g (see also Table 3).

For S130P, aggregation was too slow for aggregate concentrations and/or size to reach high enough levels to provide accurate M_w^{tot}/M_w^{mon} values. In addition to minimal aggregation, at later time points S130P displayed a split monomer peak (MW verified via MALS). The total monomer area (across both monomer peaks) declined by no more than a few percent throughout the time course shown in Figure 5C, and the lower retention time peak is gradually replaced with the higher retention time peak. Therefore, fits were not applied to *m* and M_w^{tot}/M_w^{mon} values from S130P aggregation time-course data.

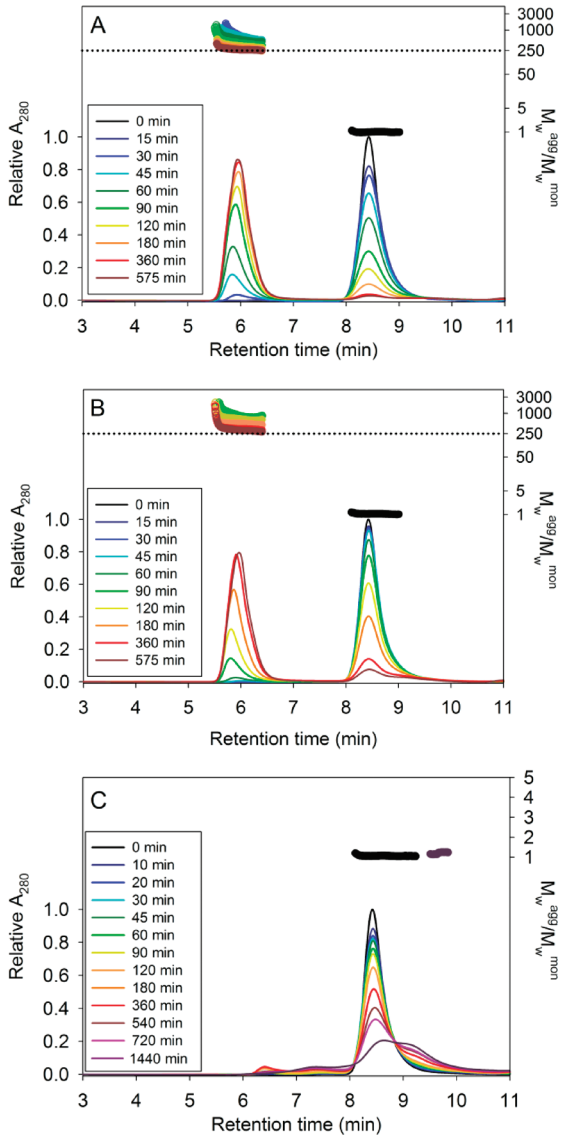

FIGURE 5: Aggregation time course at 50 °C monitored via SEC-MALS. Panels A (wild type) and B (M69Q) show overlaid plots for SEC chromatograms (relative A_{280}, thin lines) and evolution of aggregate size (M_w/M_w^{mon}, thick lines) as a function of incubation time at elevated temperature. Panel C shows only monomer molecular weight data ($M_w/M_w^{mon} = 1$) for S130P, since aggregation was too slow to generate multimers of high enough concentration and size necessary to determine accurate molecular weight values via multi-angle light scattering.

DISCUSSION

WT and Variants Have Similar Folded Monomer Secondary and Tertiary Structure. Figure 4A shows that WT, M69Q, and S130P have almost identical intrinsic tryptophan fluorescence spectra at 20 °C in folded (zero urea) and unfolded (6 M urea) states. The peak maxima occur at the same wavelength, and the shapes of the spectra were the same within statistical uncertainty. The slight differences in peak intensity are attributed to differences in protein concentration between the samples. Figure 4B shows that the ratio FI_{360}/FI_{320} along the folded (low-urea) and unfolded (high-urea) branches of the denaturation curves is also quantitatively the same within statistical uncertainty. Examination of the full fluorescence spectra as a function of [urea] (Figure 4A and Supporting Information) shows that although none of the variants show

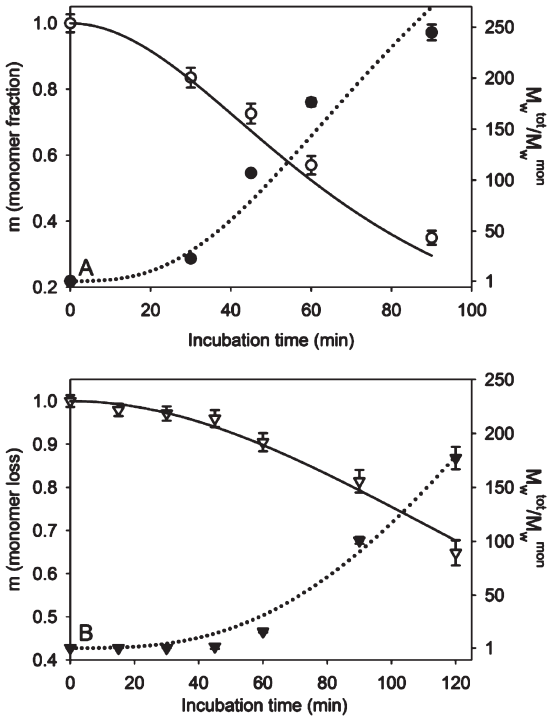

FIGURE 6: Monomer loss (open symbols, solid lines) and evolution of M_w^{tot}/M_w^{mon} (filled symbols, dashed lines) versus incubation time, based on SEC-MALS for (A) wild type (circles) and (B) M69Q (triangles). Data points (symbols) were fit to the LENP model (lines, eq 1) to obtain quantitative characterization of nucleation and growth time scales (Table 3).

Table 3: Growth and Nucleation Time Scales (τ_g and τ_n) As Calculated via Application of LENP Model Fit to Aggregation Early-Time Data[a]

	τ_g (min)	τ_n (min) $\times 10^{-3}$
WT	0.11 ± 0.02	20 ± 9
M69Q	0.13 ± 0.03	120 ± 55
S130P	ND	

[a]Calculations were not performed for S130P due to extremely slow nucleation and growth.

true two-state unfolding/refolding, their starting and ending spectra (low and high [urea]) are quantitatively similar. In addition, the monomer far-UV CD spectra overlay quantitatively for wild type and each variant (Supporting Information, Figure S2). Together, these results indicate no detectable differences in the secondary or tertiary structures of the folded and unfolded monomers for wild-type γD-crys and these two variants.

This does not rule out differences in local tertiary structure that are not reported by Trp fluorescence, as all four Trp residues are buried in the cores of the Ntd or Ctd. Intrinsic fluorescence is therefore not expected to report on differences in domain−domain pairings, such as may lead to altered conformational stability of M69Q (see below). Perhaps surprisingly, the introduction of S130P did not result in detectable differences in far-UV CD or intrinsic fluorescence. In wild-type γD-crys, S130 is relatively solvent exposed and occurs in a short loop between two β-strands (S123 to E128 and W131 to L136). W131 is not solvent exposed in the folded state of wild-type γD-crys and shows a clear red shift upon unfolding at pH 7 (*39*). Given the close proximity of W131 to S130 and the small length of the loop between the two adjacent β-strands, it appears that the mutation S130P does not

significantly disrupt the observable folded state structure of γD-crys.

The fluorescence spectra in high [urea] indicate only that all Trp are similarly solvent exposed in the unfolded states of wild type and the two variants. It is possible that the conformational space of M69Q and/or S130P is altered significantly from wild type in the unfolded state, but steady-state intrinsic fluorescence may not detect those differences. The discussion below highlights clear differences in unfolding free energies between WT, M69Q, and S130P that also support the conclusion that there are differences in the unfolded states of these variants and possibly subtle differences in the folded state structures that do not manifest spectroscopically here.

Wild Type and Variants Exhibit Apparent Two-State Unfolding. Figures 3 and 4 show that unfolding occurs without creating easily detectable unfolding or folding intermediates. Multiple distinct endothermic peaks are not apparent in DSC, although the endotherm for each protein does not have the canonical two-state shape (*51*, *72*). The asymmetric shape and significant lack of reversibility upon repeated scanning are consistent with thermal unfolding convoluted with irreversible aggregation, as seen previously for a number of aggregation-prone proteins (*73*, *74*). Thus, it is possible that unfolding occurs via sequential unfolding of one domain and then the other, but this cannot be resolved experimentally using DSC. Results from DSC suggest increased stability for M69Q (and decreased for S130P), in comparison to wild-type protein. Contrary to the increased stability of M69Q that was predicted via Rosetta-Design, S130P was not calculated to be strongly destabilizing as observed via DSC (Figure 3) and the shift to lower [urea] for chemical unfolding (Figure 4 and Supporting Information). It is notable that S130P was selected as a neutral mutation in terms of intended changes of conformational stability and as a disrupting mutation for initiation and/or propagation of aggregates through the 126−134 aggregation hot-spot region.

Isothermal unfolding as a function of [urea] shows a single sigmoidal transition for wild type and both variants (cf. Figure 4), similar to results at neutral pH for wild type in the presence of guanidine hydrochloride (*11*, *39*, *45*, *46*). Similar results (not shown) are also found at pH 3 using guanidine hydrochloride, but the midpoint denaturant concentration (C_m) for unfolding is shifted to significantly lower values, making it difficult to resolve the folded baseline in unfolding profiles such as in Figure 4B−D. The refolding curves show some hysteresis for wild type and each variant. This is anticipated to be due to slow kinetics of refolding, possibly a result of Pro isomerization. The unfolding or aggregation upon refolding was reversible, as the refolding and unfolding curves overlay at the lowest [urea], and there was no indication of soluble or insoluble aggregates in the refolded samples. Qualitatively, the urea unfolding data at pH 3 also indicated a non-two-state transition, as the full emission spectra showed a decrease in intensity at intermediate urea concentrations and no isosbestic point (Figure S1, Supporting Information). However, regression of the urea unfolding curves to a three-state model showed no increase in the quality of fit as compared to a two-state model (fits not shown). Thus, although a folding intermediate likely exists, it was not possible to quantitatively resolve more than one folding transition in the data, and a two-state fit was used to facilitate comparison of the wild type, M69Q, and S130P (cf. Table 2).

Implications for γD-Crys Aggregation Mechanism(*s*) *and Protein Engineering Strategies*. Figure 6 shows that aggregation monitored quantitatively by both loss of monomer and an increase of weight-average molecular weight are well described by a nucleation and growth mechanism such as that underlying the LENP model in eq 1. The monomer-loss portion of this version of the LENP model is equivalent to a number of previous models (*70*, *75*−*77*); however, those earlier models could not distinguish between nucleation and growth rates because they did not account for changes in molecular weight (*9*).

The nucleation and growth time scales (τ_n and τ_g) obtained via the fits (Table 3) do not show a statistically significant difference in τ_g for WT and M69Q. The observed difference in overall aggregation rate was therefore dominated by the much slower nucleation rates (larger τ_n value) for M69Q when compared to wild-type protein. The large ratio of τ_n to τ_g ($\sim 10^5-10^6$) for both wild type and M69Q is expected for a nucleated polymerization mechanism and results in the characteristic acceleratory kinetic profiles in Figure 6 (*9*, *70*).

τ_n depends implicitly on the equilibrium constant for unfolding (K_{un}), in that higher K_{un} values result in higher concentrations of unfolded, aggregation-prone monomers (*9*, *69*). Assuming a dimer nucleus, $\ln(\tau_n)$ scales roughly as $2\Delta G_{un}/RT$ at temperatures more than a few degrees below T_m (*9*, *69*). Although it was not possible to determine unfolding free energies directly at 50 °C due to convolution with aggregation in the DSC measurements, if one estimates that $\Delta\Delta G_{un}/RT$ between wild type and M69Q are similar at low and high temperature, then this gives an estimated value of $\Delta\Delta G_{un}$ for M69Q of between 0.1 and 0.7 kcal/mol based on the standard deviations for ΔG_{un} in Table 2. The relationship above for τ_n then predicts a ratio of τ_n(M69Q):τ_n(WT) between 2 and 9. Considering that the $\Delta\Delta G_{un}/RT$ values are based on simple two-state fits in Table 2, these results qualitatively and at least semiquantitatively support the conclusion that the decrease in nucleation rates was achieved to a large extent via improved conformational stability of M69Q vs wild type, although some extent of change to IAP cannot be ruled out.

While RosettaDesign did correctly predict that M69Q would result in greater conformational stability, it did not correctly predict that S130P significantly decreases conformational stability. None of the known pathological mutants were predicted by RosettaDesign to have significant conformational perturbations upon mutation from wild type. However, the large majority of those mutants either result in a change in the length of the protein (e.g., by leading to premature termination during expression) or cause a change in the solubility of the folded protein. As such, it is perhaps not surprising that those mutants are not predicted to be detrimental by RosettaDesign, as their role in cataract formation appears to involve a mechanism other than non-native aggregation such as investigated here.

In addition, S130P aggregation was greatly suppressed compared to both wild type and M69Q, with only a few percent (by mass) of aggregates formed on day-long time scales for S130P at 50 °C. These time scales are comparable to the fitted τ_n values for wild type and M69Q. Given that the T_m for S130P is much lower than those for wild type and M69Q, the concentration of unfolded S130P molecules at 50 °C is much greater than for the other variants. Therefore, the reduced aggregation rates of S130P support the conclusion that S130P results in significantly decreased intrinsic aggregation propensity (IAP); that is, the lower rates cannot be attributed to changes in conformational stability of S130P relative to WT or M69Q.

Although nucleation and growth rates were not quantifiable from the data in Figure 5 for S130P, it is clear that small

aggregates did form slowly (time scales $\sim 10^3$ min or greater), but those aggregates did not grow to high molecular weight. One hypothesis for the lack of significant growth is that the unfolded state of S130P is conformationally altered compared to wild type and M69Q, and the aggregates that form as small oligomers do not present their "hot spots" in a solvent-exposed manner that allows additional monomers to be added easily. This would presumably be in addition to weakening of the key interprotein contacts that putatively are responsible for slowing the creation ("nucleation") of dimers and small oligomers. It appears reasonable that both factors are helping to provide a much lower IAP for S130P than for wild type or M69Q. The large changes in monomer retention time in SEC in Figure 5C are presumably due to interactions with the stationary phase. These may be indicative of unfolded monomer conformations at 50 °C that do not refold upon cooling to room temperature and may support the former hypothesis regarding the mechanism of disruption of aggregate growth for S130P. SEC of post-DSC S130P samples show no aggregation but tailing of the monomer peak in SEC (not shown), also supporting this conclusion.

Previous work at pH 7 and 3 (*11*, *40*) indicated that the aggregation-prone conformational state for wild-type γD-crys is one in which the Ctd is primarily folded and the Ntd is at least partially unfolded, without an intact interface between Ctd and Ntd. S130 is located on the surface of the Ctd and is not directly involved in the interface between Ntd and Ctd. The Ntd and Ctd are able to aggregate if they are expressed and purified separately, making it unclear whether the key contacts in wild-type aggregates involve the Ntd, Ctd, or both (*40*). The reduction of IAP for S130P compared to wild type suggests that the key interprotein contacts in propagating γD-crys aggregates involve a number of possible alternative scenarios: (1) domain swapping between Ntd and Ctd on alternating monomers (*78*); (2) polymerization of the partially unfolded Ntd, with the Ctd simply "carried" along and forced to pack along the periphery of the amyloid spine (*79*); (3) polymerization of the folded Ctd domains with key contacts at the residues that are exposed when the Ntd−Ctd interface is not intact.

Scenarios 1 and 2 could lead to lower aggregation rates for S130P if the Ctd structure in the unfolded state is altered such that it disrupts key Ntd−Ctd contacts or Ntd−Ntd contacts, respectively. Disruption of aggregates via these types of mechanisms is inherently not included in available "aggregation calculators", as those algorithms assume aggregate contacts involve the same hot-spot peptide sequence on each monomer and do not account for situations where two different hot-spot peptides associate from different domains or regions in a protein. Scenario 3 would be consistent with the aggregation of S130P vs WT if S130P causes significant changes in nonlocal structure in the partially or fully unfolded Ctd or if the predicted consensus hot-spot sequence V126−Y134 (Figure 1) is part of the amyloid spine for wild-type γD-crys. Both may be factors, although the fact that the former effects are not part of the aggregation calculators at least suggests the latter is a significant factor.

To the best of our knowledge, S130P represents the first reported case for a multidomain protein in which it was experimentally shown that sequence-only algorithms such as TANGO accurately predicted resistance to aggregation despite reducing the overall conformational stability of the folded monomer to some degree, without requiring such extreme mutations so as to completely destabilize both folded and aggregated states. This success notwithstanding, caution is merited before concluding that such calculators are robust predictors in general for proteins with folded secondary and tertiary structures, without additional tests for a wider range of mutations within predicted hot spots. The available calculators are based on experimental data for short polypeptides and so are not designed to capture nonlocal interactions and changes in the global conformation space and residual structure in the unfolded state for larger polypeptides and proteins.

Finally, it is noteworthy that alternative mutants are predicted to disrupt the consensus hot spot in Figure 1, but most were also predicted to significantly alter the conformational stability. Thus, changing IAP or ΔG_{un} alone is not anticipated to be practical for most proteins, and a lucrative approach will likely incorporate optimization based on both IAP and conformational stability.

ACKNOWLEDGMENT

The authors thank J. King (MIT) for the pQE1-γD-crys-WT vector and for helpful discussion. The authors are also grateful to B. Kuhlman, D. Sammond, and X. Ambroggio for useful discussions regarding Rosetta.

SUPPORTING INFORMATION AVAILABLE

Additional data regarding chemical unfolding and a comparison of monomer and aggregate structure via CD and ThT binding. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Weiss, W. F. I., Young, T. M., and Roberts, C. J. (2009) Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *J. Pharm. Sci. 98*, 1246–1277.
2. Mahler, H. C., Friess, W., Grauschopf, U., and Kiese, S. (2009) Protein aggregation: pathways, induction factors and analysis. *J. Pharm. Sci. 98*, 2909–2934.
3. Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol. 15*, 3–16.
4. Murphy, R. M. (2002) Peptide aggregation in neurodegenerative disease. *Annu. Rev. Biomed. Eng. 4*, 155–174.
5. Fink, A. L. (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding Des. 3*, R9–R23.
6. Rousseau, F., Schymkowitz, J., and Serrano, L. (2006) Protein aggregation and amyloidosis: confusion of the kinds? *Curr. Opin. Struct. Biol. 16*, 118–126.
7. Eisenberg, D., Nelson, R., Sawaya, M. R., Balbirnie, M., Sambashivan, S., Ivanova, M. I., Madsen, A. O., and Riekel, C. (2006) The structural biology of protein aggregation diseases: fundamental questions and some answers. *Acc. Chem. Res. 39*, 568–575.
8. Foguel, D., Robinson, C. R., de Sousa, P. C., Jr., Silva, J. L., and Robinson, A. S. (1999) Hydrostatic pressure rescues native protein from aggregates. *Biotechnol. Bioeng. 63*, 552–558.
9. Roberts, C. J. (2007) Non-native protein aggregation kinetics. *Biotechnol. Bioeng. 98*, 927–938.
10. Andrews, J. M., and Roberts, C. J. (2007) Non-native aggregation of alpha-chymotrypsinogen occurs through nucleation and growth with competing nucleus sizes and negative activation energies. *Biochemistry 46*, 7558–7571.
11. Kosinski-Collins, M. S., and King, J. (2003) In vitro unfolding, refolding, and polymerization of human gammaD crystallin, a protein involved in cataract formation. *Protein Sci. 12*, 480–490.
12. Pallitto, M. M., and Murphy, R. M. (2001) A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state. *Biophys. J. 81*, 1805–1822.
13. Worn, A., and Pluckthun, A. (2001) Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol. 305*, 989–1010.
14. Pace, C. N., Shirley, B. A., and Thomson, J. A. (1989) Measuring the Conformational Stability of a Protein, IRL Press, Oxford.
15. Caflisch, A. (2006) Computational models for the prediction of polypeptide aggregation propensity. *Curr. Opin. Chem. Biol. 10*, 437–444.
16. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007) AGGRESCAN: a server for the prediction

and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinf.,* 8.
17. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol. 22*, 1302–1306.
18. Galzitskaya, O. V., Garbuzynskiy, S. O., and Lobanov, M. Y. (2006) Prediction of amyloidogenic and disordered regions in protein chains. *PLoS Comput. Biol. 2*, 1639–1648.
19. Tartaglia, G. G., Cavalli, A., Pellarin, R., and Caflisch, A. (2005) Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Sci. 14*, 2723–2734.
20. Thompson, M. J., Sievers, S. A., Karanicolas, J., Ivanova, M. I., Baker, D., and Eisenberg, D. (2006) The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. U.S.A. 103*, 4074–4078.
21. Trovato, A., Seno, F., and Tosatto, S. C. E. (2007) The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel. 20*, 521–523.
22. Bratko, D., Cellmer, T., Prausnitz, J. M., and Blanch, H. W. (2007) Molecular simulation of protein aggregation. *Biotechnol. Bioeng. 96*, 1–8.
23. Gsponer, J., and Vendruscolo, M. (2006) Theoretical approaches to protein aggregation. *Protein Pept. Lett. 13*, 287–293.
24. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., and Baker, D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol. 331*, 281.
25. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* (*New York*) *302*, 1364.
26. Rohl, C. A., Strauss, C. E. M., Chivian, D., and Baker, D. (2004) Modeling structurally variable regions in homologous proteins with rosetta. *Proteins: Struct., Funct., Bioinf. 55*, 656.
27. Chennamsetty, N., Voynov, V., Kayser, V., Helk, B., and Trout, B. L. (2009) Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci. U.S.A. 106*, 11937–11942.
28. Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007) Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc. 129*, 10110–10112.
29. Andrews, J. M., Weiss, W. F., IV, and Roberts, C. J. (2008) Nucleation, growth, and activation energies for seeded and unseeded aggregation of alpha-chymotrypsinogen A. *Biochemistry 47*, 2397–2403.
30. Lee, C. C., Walters, R. H., and Murphy, R. M. (2007) Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry 46*, 12810–12820.
31. Li, Y., Ogunnaike, B. A., and Roberts, C. J. (2010) Multi-variate approach to global protein aggregation behavior and kinetics: Effects of pH, NaCl, and temperature for alpha-chymotrypsinogen A. *J. Pharm. Sci. 99*, 645–662.
32. Li, Y., Weiss, W. F., and Roberts, C. J. (2009) Characterization of high molecular-weight nonnative aggregates and aggregation kinetics by size exclusion chromatography with inline multi-angle laser light scattering. *J. Pharm. Sci. 98*, 3997–4016.
33. Lee, C. C., Nayak, A., Sethuraman, A., Belfort, G., and McRae, G. J. (2007) A three-stage kinetic model of amyloid fibrillation. *Biophys. J. 92*, 3448–3458.
34. Roberts, C. J., Darrington, R. T., and Whitley, M. B. (2003) Irreversible aggregation of recombinant bovine granulocyte-colony stimulating factor (bG-CSF) and implications for predicting protein shelf life. *J. Pharm. Sci. 92*, 1095–1111.
35. Weiss, W. F. T., Hodgdon, T. K., Kaler, E. W., Lenhoff, A. M., and Roberts, C. J. (2007) Nonnative protein polymers: structure, morphology, and relation to nucleation and growth. *Biophys. J. 93*, 4392–4403.
36. Brummitt, R. K., Nesta, D. P., Chang, L., Kroetsch, A. M., and Roberts, C. J. (2010) Non-native aggregation of an IgG1 antibody in acidic conditions: 2. Nulceation-and-growth kinetics with competing growth mechanisms (in preparation).
37. Sahin, E., Grillo, A. O., Perkins, M. D., and Roberts, C. J. (2010) Comparative effects of pH and ionic strength on protein-protein interactions, unfolding, and aggregation for IgG1 antibodies, *J. Pharm. Sci. Early View* (doi: 10.1002/jps.22198).
38. Crabbe, M. J., and Goode, D. (1995) Protein folds and functional similarity; the Greek key/immunoglobulin fold. *Comput. Chem. 19*, 343–349.
39. Kosinski-Collins, M. S., Flaugh, S. L., and King, J. (2004) Probing folding and fluorescence quenching in human gammaD crystallin Greek key domains using triple tryptophan mutant proteins. *Protein Sci. 13*, 2223–2235.
40. Papanikolopoulou, K., Mills-Henry, I., Thol, S. L., Wang, Y., Gross, A. A., Kirschner, D. A., Decatur, S. M., and King, J. (2008) Formation of amyloid fibrils in vitro by human gammaD-crystallin and its isolated domains. *Mol. Vis. 14*, 81–89.
41. Mayr, E. M., Jaenicke, R., and Glockshuber, R. (1997) The domains in gammaB-crystallin: identical fold-different stabilities. *J. Mol. Biol. 269*, 260–269.
42. Rudolph, R., Siebendritt, R., Nesslauer, G., Sharma, A. K., and Jaenicke, R. (1990) Folding of an all-beta protein: independent domain folding in gamma II-crystallin from calf eye lens. *Proc. Natl. Acad. Sci. U.S.A. 87*, 4625–4629.
43. Sen, A. C., Walsh, M. T., and Chakrabarti, B. (1992) An insight into domain structures and thermal stability of gamma-crystallins. *J. Biol. Chem. 267*, 11898–11907.
44. Mandal, K., Chakrabarti, B., Thomson, J., and Siezen, R. J. (1987) Structure and stability of gamma-crystallins. Denaturation and proteolysis behavior. *J. Biol. Chem. 262*, 8096–8102.
45. Flaugh, S. L., Kosinski-Collins, M. S., and King, J. (2005) Interdomain side-chain interactions in human gammaD crystallin influencing folding and stability. *Protein Sci. 14*, 2030–2043.
46. Flaugh, S. L., Kosinski-Collins, M. S., and King, J. (2005) Contributions of hydrophobic domain interface interactions to the folding and stability of human gammaD-crystallin. *Protein Sci. 14*, 569–581.
47. Flaugh, S. L., Mills, I. A., and King, J. (2006) Glutamine deamidation destabilizes human gammaD-crystallin and lowers the kinetic barrier to unfolding. *J. Biol. Chem. 281*, 30782–30793.
48. Ramsay, G., and Freire, E. (1990) Linked thermal and solute perturbation analysis of cooperative domain interactions in proteins. Structural stability of diphtheria toxin. *Biochemistry 29*, 8677–8683.
49. Vermeer, A. W., and Norde, W. (2000) The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. *Biophys. J. 78*, 394–404.
50. Palme, S., Slingsby, C., and Jaenicke, R. (1997) Mutational analysis of hydrophobic domain interactions in gamma B-crystallin from bovine eye lens. *Protein Sci. 6*, 1529–1536.
51. Brandts, J. F., Hu, C. Q., Lin, L. N., and Mos, M. T. (1989) A simple model for proteins with interacting domains. Applications to scanning calorimetry data. *Biochemistry 28*, 8588–8596.
52. Das, R., and Baker, D. (2008) Macromolecular modeling with Rosetta. *Annu. Rev. Biochem. 77*, 363–382.
53. Lazaridis, T., and Karplus, M. (1999) Effective energy function for proteins in solution. *Proteins 35*, 133.
54. Kortemme, T., Morozov, A. V., and Baker, D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol. 326*, 1239.
55. Dunbrack, R. L., and Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci. 6*, 1661.
56. Liu, Y., and Kuhlman, B. (2006) RosettaDesign server for protein design. *Nucleic Acids Res. 34*, W235–W238.
57. Dantas, G., Correnti, C., Reichow, S. L., Havranek, J. J., Eletr, Z. M., Isern, N. G., Kuhlman, B., Varani, G., Merritt, E. A., and Baker, D. (2007) High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *J. Mol. Biol. 366*, 1209.
58. Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003) A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol. 332*, 449.
59. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P., and Kuhlman, B. (2007) Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J. Mol. Biol. 371*, 1392.
60. Karanicolas, J., and Kuhlman, B. (2009) Computational design of affinity and specificity at protein-protein interfaces. *Curr. Opin. Struct. Biol. 19*, 458.
61. Sammond, D. W., Eletr, Z. M., Purbeck, C., and Kuhlman, B. (2010) Computational design of second-site suppressor mutations at protein-protein interfaces. *Proteins 78*, 1055.
62. Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F., and Vendruscolo, M. (2008) Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol. 380*, 425–436.
63. Wang, X., Das, T. K., Singh, S., and Kumar, S. (2009) Potential aggregation prone regions in biotherapeutics: a survey of commercial monoclonal antibodies. *mAbs 1*, 254–267.
64. Basak, A., Bateman, O., Slingsby, C., Pande, A., Asherie, N., Ogun, O., Benedek, G. B., and Pande, J. (2003) High-resolution X-ray crystal structures of human gammaD crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J. Mol. Biol. 328*, 1137.

65. Xiang, Z., and Honig, B. (2001) Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol. 311*, 421.
66. Pace, C. N., Shirely, B. A., and Thomson, J. A. (1989) Measuring the conformational stability of a protein, in Protein Structure, a Practical Approach (Creighton, T. E., Ed.) pp 311−330, IRL Press at Oxford University Press, New York.
67. Clark, A. C., Sinclair, J. F., and Baldwin, T. O. (1993) Folding of bacterial luciferase involves a non0native heterodimeric intermediate in equilibrium with the native enzyme and the unfolded subunits. *J. Biol. Chem. 268*, 10773–10779.
68. Li, Y., and Roberts, C. J. (2009) Lumry-Eyring nucleated-polymerization model of protein aggregation kinetics: 2. Competing growth via condensation and chain polymerization. *J. Phys. Chem. B 113*, 7020–7032.
69. Andrews, J. M., and Roberts, C. J. (2007) A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics: 1. Aggregation with pre-equilibrated unfolding. *J. Phys. Chem. B 111*, 7897–7913.
70. Ferrone, F. (1999) Analysis of protein aggregation kinetics. *Methods Enzymol. 309*, 256–274.
71. DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M., and Vendruscolo, M. (2004) Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol. 341*, 1317.
72. Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv. Protein Chem. 33*, 167–241.
73. Remmele, R. L., Jr., Bhat, S. D., Phan, D. H., and Gombotz, W. R. (1999) Minimization of recombinant human Flt3 ligand aggregation at the Tm plateau: a matter of thermal reversibility. *Biochemistry 38*, 5241–5247.
74. Sanchez-Ruiz, J. M., Lopez-Lacomba, J. L., Cortijo, M., and Mateo, P. L. (1988) Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry 27*, 1648–1652.
75. Goldstein, R. F., and Stryer, L. (1986) Cooperative polymerization reactions. Analytical approximations, numerical examples, and experimental strategy. *Biophys. J. 50*, 583–599.
76. Lomakin, A., Teplow, D. B., Kirschner, D. A., and Benedek, G. B. (1997) Kinetic theory of fibrillogenesis of amyloid beta-protein. *Proc. Natl. Acad. Sci. U.S.A. 94*, 7942–7947.
77. Oosawa, F., and Asakura, S. (1975) Thermodynamics of the Polymerization of Proteins, Academic Press, London.
78. Yang, S., Levine, H., Onuchic, J. N., and Cox, D. L. (2005) Structure of infectious prions: stabilization by domain swapping. *FASEB J.* 19.
79. Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M., and Eisenberg, D. (2005) Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure. *Nature 437*, 266–269.
80. Kawahara, K., and Tanford, C. (1966) Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *J. Biol. Chem. 241*, 3228–3232.