High-Precision, High-Throughput Stability Determinations Facilitated by Robotics and a Semiautomated Titrating Fluorometer†

Marshall Hall Edgell,*,‡,§ Dorothy A. Sims,‡ Gary J. Pielak,§,|| and Fang Yi§

Departments of Microbiology and Immunology, Chemistry, and Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599

Received January 13, 2003; Revised Manuscript Received March 31, 2003

ABSTRACT: The use of statistical modeling to test hypotheses concerning the determinants of protein structure requires stability data (e.g., the free energy of denaturation in H_2O, ΔG_{HOH}) from hundreds of protein mutants. Fluorescence-monitored chemical denaturation provides a convenient method for high-precision, high-throughput ΔG_{HOH} determination. For eglin c we find that a throughput of about 20 min per protein can be attained in a two-channel semiautomated titrating fluorometer. We find also that the use of robotics for protein purification and preparation of the solutions for chemical denaturation gives highly precise ΔG_{HOH} values in which the standard deviation of values from multiple preparations (±0.051 kcal/mol) differs very little from multiple measurements from a single preparation (±0.040 kcal/mol). Since the variance introduced into model fitting by ΔG_{HOH} increases as the square of measurement error, there is a premium on precision. In fact, the fraction of stability behavior explicable by otherwise perfect models goes from 98% to only 50% over the error range commonly reported for chemical denaturation measurements (0.1−0.6 kcal/mol). We have found that the precision of chemical denaturation ΔG_{HOH} measurements depends most heavily on the precision of the instrument used, followed by protein purity and the capacity to precisely prepare the solutions used for titrations.

Site-directed mutagenesis (*1*) is our most powerful tool for testing hypotheses about the determinants of protein structure. However, the complexity of the protein fold problem is such that the application of mutagenesis to test hypotheses, while enormously useful, does not always produce clear answers. This is, in part, because most residues in proteins serve multiple roles. Such complexity means that testing hypotheses by looking at a few tens of mutants can be an anecdotal process. Statistical modeling is often used to extend analyses beyond the anecdotal (*2*). Statistical modeling or regression analysis relies on a decomposition of the observable into component parts and provides a formalism for parametrizing the descriptors in the decomposition. While there is some controversy concerning the theoretical underpinnings associated with decomposing free energy into component parts (*3*−*7*), there seems to be general agreement that the approach has utility for protein studies. For instance. the decomposition of the free energy of protein unfolding has a long history of utility in simulations (*8*−*10*) and in the analysis of the thermodynamic consequences of mutation (*11*−*21*).

Using regression analysis to parametrize decompositions or models with the data from large numbers of mutants has several advantages over the anecdotal approach. Regression analysis provides an estimate of the fraction of behavior accounted for by the hypothesis. It accurately parametrizes the descriptors even in models where important effects are missing from that model, provided that the descriptors are independent of the missing effects (*22*). Regression analysis can accurately parametrize descriptors even when the observable is only weakly related to the hypothesis (*2*). The method is robust even in the face of considerable measurement error (*23*). The cost of these advantages is a need to acquire data from large populations of mutant proteins. The number of protein mutants that need to be characterized to achieve both significance and robustness depends on the number of descriptors to be parametrized and the precision of the measurements but is likely to be in the hundreds.

Hence, to test hypotheses concerning protein stability using the statistical modeling approach, it is necessary to employ high-throughput methods to determine protein stability. As in other areas, robotics and automated instrumentation provide a means to achieve high-throughput without the need for laboratory heroics. Chemical denaturation is the most common approach to determine stability (*24*, *25*). Data from some reporter in the protein collected in various concentrations of denaturant can be fit to a model of the unfolding process (usually a two-state model) to determine the free energy of denaturation in the absence of denaturant (ΔG_{HOH}). We explored fluorescence monitoring of the chemical denaturation of mutant proteins (*26*−*28*) since it provides the most potential for high throughput.

EXPERIMENTAL PROCEDURES

Culture Growth for High-Throughput Protein Purification. Eglin c is purified from 12 mL of culture in twelve 1 mL batches on 3 separate days, 4 mL per day. Archived clones of eglin c mutants in the protein expression plasmid, pET28a (Novagen, Inc.) in *Escherichia coli* NovaBlue, are kept in

† Funding for this work was provided by NIH Grant GM58665, NSF Grant MCB0212939, and the North Carolina Biotechnology Center, Grant 2002-MRG-134.

* To whom correspondence should be addressed. Phone: (919) 962-0147. Fax: (919) 962-8103. E-mail: marshall@med.unc.edu.

‡ Department of Microbiology and Immunology.

§ Department of Biochemistry and Biophysics.

|| Department of Chemistry.

10.1021/bi034063g CCC: $25.00
Published on Web 05/30/2003

250 μL of growth media plus 10% glycerol at −70 °C in 96-well format microtiter plates. Growth medium is Terrific broth (900 mL containing 12 g of tryptone, 24 g of yeast extract, and 5 mL of glycerol is combined after autoclaving with 100 mL containing 4.6 g of KH_2PO_4 and 25.0 g of K_2HPO_4). Inoculating cultures directly from the archival plate gives considerable variation in lag times and growth rates. Consequently, a starter plate is prepared from the archival plate. Culture preparation is done with the Biomek 2000 liquid handling robot housed in a clean bench sterile workstation. One milliliter of growth media is transferred into each 2.2 mL square well of a deep-well microtiter plate (Marsh Bio Products). Each well is inoculated with 50 μL of culture from the archival plate. The plate is sealed with an AirPore porous adhesive sheet (Qiagen) and grown to saturation (~15 h at 37 °C) using a short radius (0.5 mm) microtiter plate rotator, which is required to generate a good vortex during growth. The next day the robot is used to prepare two replacement archival plates (180 μL of culture and 70 μL of 26% glycerol), to be stored in separate −70 °C freezers. On three successive days 4 mL of culture of each mutant is prepared as follows. The robot is used to add 1 mL of Terrific broth to each well of four 96-deep-well format microtiter plates (square 2.2 mL wells). The robot is then used to inoculate each well in the four plates with 50 μL of culture from the appropriate well of the starter plate. The four deep-well plates are covered with porous adhesive sheets and grown at 37 °C on the short radius shaker for 3.5−4 h. The OD of the cultures from one plate was checked by transferring 20 μL of culture into 180 μL of growth media in a microtiter plate. When an OD_{595nm} of 0.125 of the diluted culture is attained (equivalent to an OD_{595nm} of 0.8 in a 1 cm cuvette), protein expression is induced by using the robot to add 10 μL of 100 mM IPTG. Induction increases protein yield 2−3-fold. The cultures are grown for an additional 4−5 h at 37 °C. To facilitate freeze−thaw lysis, cells are collected by centrifugation in three passes into a single new deep-well culture plate in which the wells are cylindrical and separated from each other. An eight-channel electric pipet is used to transfer 1.3 mL (>1.5 mL gives well-to-well cross-contamination) from the four plates into the collection plate, which is spun at 1500*g* for 15 min in a swinging bucket rotor. The supernatants are decanted by hand, and the plate is blotted. The single collection plate with pellets from 4 mL of growth culture is covered with a 96-well plate mat (Phenix Research Products) and placed in the −70 °C freezer.

Cell Lysis by Freeze−Thaw. Protein preparations are made from 12 mL of growth culture that has been stored at −70 °C as three separate cell pellets in microtiter plates. The three pellets are thawed at room temperature on a support to allow air to reach the bottom of the plates. The pellets are freeze−thawed three times using a dry ice−ethanol bath for freezing and air for thawing. Each of the pellets is resuspended separately in 400 μL of 50 mM tris(hydroxymethyl)-aminomethane (Tris), pH 8.5, 13 units/mL DNase, and 6 μM $MgCl_2$. The lysates are combined and cleared by centrifugation for 30 min at 1500*g* at 4 °C. Each supernatant (1.1 mL) is transferred with the Biomek to a new deep-well plate and centrifuged again at 1500*g* at 4 °C. The supernatants are then transferred to a new Phenix 96 deep-well plate.

High-Throughput Purification of His-Tagged Proteins with Qiagen Ni-NTA Superflow Resin. All pipeting steps, unless otherwise indicated, are performed with the robot at room temperature in a sterile air workstation. Liquid is always removed using the Biomek suction apparatus at ≤150 mbar. A hand-held electronic pipet is used to pipet 800 μL of Qiagen Ni-NTA Superflow resin slurry into each well of a Whatman/Polyfiltronics 800 filter plate. The resin storage buffer is removed from the wells. The resin is washed three times with 400 μL of wash buffer I (50 mM sodium phosphate, 300 mM NaCl, 40 mM imidazole, pH 8.2). The wells are loaded with 400 μL of cleared lysate. Liquid is removed after a 15 min equilibration period. The well loading steps (addition, equilibration, and liquid removal) were repeated with 400 μL of cleared lysate until all of the cleared lysate was processed. The resin in each well was washed with 400 μL of wash buffer I and liquid removed after 1−2 min. The protein-loaded resin was washed five times with 400 μL of wash buffer II (50 mM sodium phosphate, 300 mM NaCl, pH 8.2), removing the liquid after 1−2 min of equilibration. The filtration plate was then rearranged so that it sits on top of a deep-well receiver plate in the Biomek suction apparatus. Four hundred microliters of elution buffer (50 mM sodium phosphate, 300 mM NaCl, 375 mM imidazole, pH 8.2) was added and suction applied after 5 min to transfer the liquid from the filtration plate to the receiver plate. This elution step was repeated three more times. The solution from the receiver plate was transferred to a 96-well dialysis plate (2000 M_w cutoff; Harvard Biosciences) and dialyzed against several changes of 50 mM Tris and 100 mM NaCl, pH 8.5, over a 3 day period. The purified proteins are stored at 4 °C. Protein yield is almost directly proportional to the stability of the protein such that a mutant with a $\Delta(\Delta G_{HOH})$ of −2.5 kcal/mol yields only a few tens of micrograms from 12 mL of culture as contrasted to wild type, which gives ~800 μg. Yields are estimated to be ~75% of the His-tagged protein present in the cleared lysate. Eight consecutive days are required to prepare purified protein from 96 clones. For quality control we include three separate isolates of bacteria expressing wild-type eglin c on each archival plate. In our experience, if $\Delta(\Delta G_{HOH})$ determinations for the wild-type controls are outside of the normal range, the $\Delta(\Delta G_{HOH})$ values from most of the other proteins copurified from that archival plate are also unreliable.

ΔG_{HOH} Measurements Using the ATF105 Dual-Channel Titrating Fluorometer (29). The ATF105 dual-channel titrating fluorometer (Protein Solutions, Inc.) uses two solutions per channel to carry out automated chemical denaturation. The two sample cuvettes are filled with 1.8 mL of protein solution in the absence of denaturant and titrated with protein solution at the same concentration in 6.8 M guanidine hydrochloride (GuHCl). Measurements were made at 25 °C. Forty-two fluorescence intensity readings are acquired (excitation at 290 nm, emission at 350 nm). The concentration steps are spaced at 0.2 M GuHCl intervals in the ranges of 0−2.4 and 4.5−6.2 M and at 0.1 M GuHCl intervals from 2.4 to 4.5 M. Protein samples, in a jacketed cell holder held at 25 °C with a thermostat, a thermoelectric heater, and high flow rate coolant, are stirred at a setting of 14 or 15 for 30 s to allow the protein to come to chemical and thermal equilibrium after each change in denaturant concentration. Other proteins may require different times. The bandwidth for the excitation and emission was set at 4 nm.

Protein Solutions for Automated Chemical Denaturation. The robot prepares the two solutions used for titration. Both solutions contain 100 mM NaCl and 50 mM Tris at pH 8.5. The robot prepares a 2.25 mL no-denaturant solution containing 150 μL of the purified protein preparation. The robot also prepares 6.0 mL of a 6.8 M GuHCl solution containing 400 μL of the purified protein preparation.

Calculating ΔG_{HOH}. The data are fit to a six-parameter equation (eq 4) (*30*). The assumptions in that equation are that the fluorescence signals from both pure native and denatured states can be expressed as linear equations:

$$F_{obsd,N} = f_N + s_N[\text{GuHCl}] \qquad F_{obsd,D} = f_D + s_D[\text{GuHCl}] \tag{1}$$

over the entire range of GuHCl concentrations in the experiment and that in the transition region the equilibrium between the native and denatured states is described by

$$K_{N-D} = (F_N - F_{obsd})/(F_{obsd} - F_D) = \exp(-\Delta G_{N-D}/RT) \tag{2}$$

where the free energy of unfolding is

$$\Delta G_{N-D} = \Delta G_{HOH} - m_{N-D}[\text{GuHCl}] \tag{3}$$

Hence the observed fluorescence signal F_{obsd} is

$$F_{obsd} = \{(f_N + s_N[\text{GuHCl}]) + (f_D + s_D[\text{GuHCl}])\exp[m_{N-D}([\text{GuHCl}] - [\text{GuHCl}]_{1/2})/RT]\}/(1 + \exp[m_{N-D}([\text{GuHCl}] - [\text{GuHCl}]_{1/2})/RT]) \tag{4}$$

where f_N is the fluorescence signal from the native state at 0 M GuHCl and $s_N = df_N/d[\text{GuHCl}]$ (i.e., the slope of the native state response); f_D and s_D are the corresponding parameters for the denatured state, R is the gas constant (1.987 cal/mol), T is the absolute temperature of the measurements (302 K), $[\text{GuHCl}]_{1/2}$ is the GuHCl concentration where half of the protein molecules are in the denatured ensemble, and m_{N-D} is the change in free energy with respect to the change in guanidine hydrochloride concentration.

The curve fit option in the KaleidaGraph program (version 3.5; Synergy software, PCS Inc.) was used to obtain values for $[\text{GuHCl}]_{1/2}$ and m_{N-D} with their calculated standard errors for individual measurement. ΔG_{HOH} for each mutant protein is the product of $[\text{GuHCl}]_{1/2}$ and m_{N-D}. The difference in unfolding free energy between each mutant protein and wild type was calculated using $\Delta(\Delta G_{HOH}) = \Delta G^{mut}_{HOH} - \Delta G^{wt}_{HOH}$, where ΔG^{mut}_{HOH} is obtained from a single measurement and ΔG^{wt}_{HOH} is the average value from 19 different wild-type protein preparations.

RESULTS

We choose the ATF105, dual-channel titrating spectrofluorometer (*29*) for chemical denaturation measurements. This is a semiautomated instrument in that the protein in low denaturant is placed in a cuvette and then the instrument, after recording the fluorescence of that solution, uses a programmable syringe to remove small volumes from the cuvette ($\sim$50 μL) and another to replace that volume with a denaturant solution or a solution of mutant protein at high denaturant concentration held at room temperature. The solution in the cuvette is mixed with a magnetic stir bar, and fluorescence intensity readings are acquired after a user-determined interval. The time to collect a single fluorescence reading is the sum of the time for the solution to come to physical homogeneity after the addition of denaturant, the time for the mixture to come to thermal and chemical equilibrium after reaching physical homogeneity, and the time for the instrument to acquire a fluorescence intensity measurement.

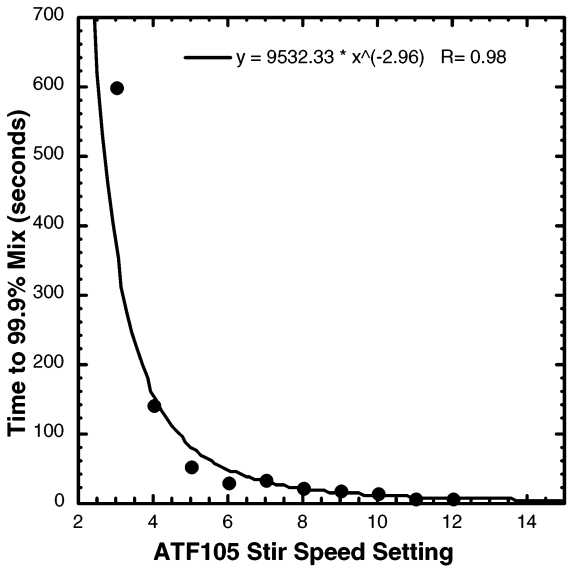

FIGURE 1: Time for solution within the fluorometer cuvette to reach 99% of the equilibrium value at various stirring speeds.

Time To Reach Physical Homogeneity. To estimate this time, for various stir speeds, a small volume (100 μL) of a concentrated fluorescein solution in 50% glycerol was placed underneath 1.7 mL of buffer in a cuvette, the cuvette was placed in the fluorometer, and fluorescence emission was then measured for 600 s. The time to reach 99% of the equilibrium value as a function of stir speed was determined (Figure 1). The time to reach physical homogeneity at the stir speed used for the chemical denaturation measurements (3−4 s) is estimated by curve fitting data because the highest speeds generate bubbles that interfere with the intensity measurements (stirring can be turned off during measurements in a titration experiment but not during a kinetic experiment).

Time To Reach Chemical Equilibrium. We measured the time for the wild-type eglin c protein to reach chemical equilibrium after a transition from 0 to 3.5 M GuHCl by stopped flow (Figure 2) as $\sim$5 s. The final concentration of 3.5 M GuHCl is approximately the C_m, the concentration that requires the longest time to reach chemical equilibrium.

Verification of the Measurement Interval. Given the times to acquire a fluorescence measurement with the ATF (5 s) and to achieve physical and chemical equilibrium, we choose 30 s as the interval between measurements during the chemical denaturation titration. To verify this choice, we carried out separate measurements with the same preparation of eglin c but different measurement intervals. The values derived from the chemical denaturation data remain unchanged with measurement intervals of 24 s or longer (Figure 3).

We also tested the proposed 30 s measurement interval using it for the first half of the transition and 5 min for the

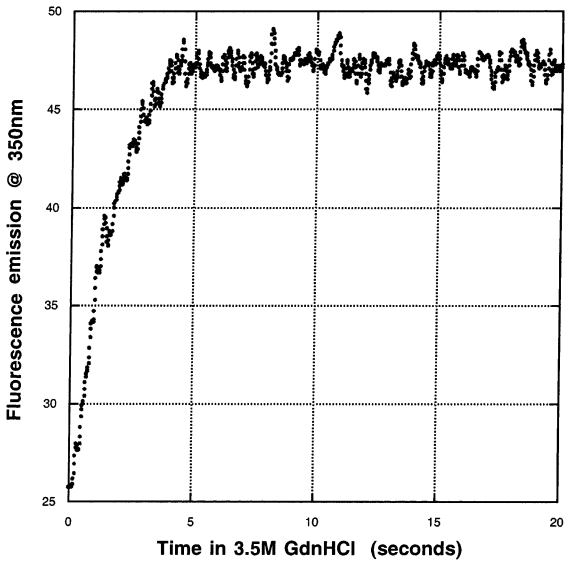

FIGURE 2: Stopped-flow determination of the time to reach chemical equilibrium at the C_m of wild-type eglin c.

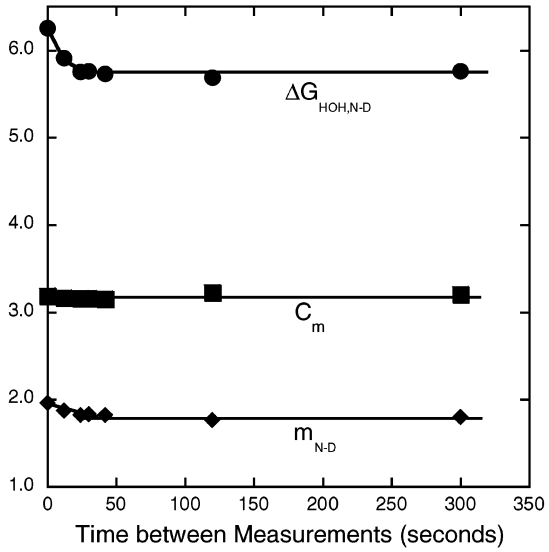

FIGURE 3: Chemical denaturation parameters obtained with various intervals between the changes in denaturant concentrations. The units of *y*-axis representation depend on which response is being plotted. Parameter: ΔG_{HOH}, unfolding stability (kcal/mol), ●; C_m, midpoint of the denaturation transition (M), ■; m_{N-D}, change in ΔG_{HOH} as a function of denaturant concentration (kcal mol^{-1} M^{-1}), ◆.

second half (*31*). We observed no discontinuity at the point of measurement interval change, indicating that the 30 s interval is sufficient (data not shown) to bring the solution to thermal and chemical equilibrium under the conditions of these measurements. If the conditions were changed significantly, new tests would need to be performed.

Reproducibility of the ΔG_{HOH} Measurements. Mutant proteins are purified in 96-well plate format, and three wild-type isolates are included on each 96-well plate to assess the quality of the protein purification for that plate. The average ΔG_{HOH} for 19 individually grown and purified wild-type protein samples is 6.12 ± 0.05 kcal/mol. The ΔG_{N-D} value for repeated measurements from a single wild-type protein preparation is not significantly different, 6.14 ± 0.04 kcal/mol, indicating that the freeze−thaw/His-tag protein purification scheme is quite reproducible. To further assess

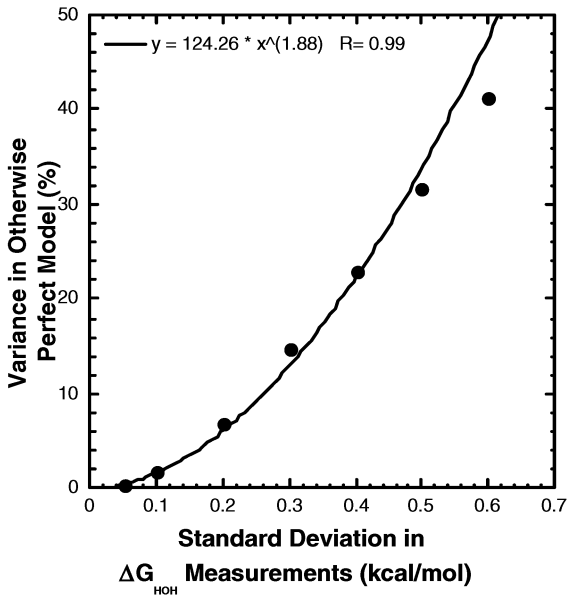

FIGURE 4: Proportion of the variation in measured values, in otherwise perfect models, coming from ΔG_{HOH} measurement error. Simulated ΔG_{HOH} errors were generated by drawing numbers from normal distributions with means of zero and various standard deviations. These errors were added to 85 real ΔG_{HOH} measurements taken from the library of eglin c mutant proteins. These values were then fit to the non-error-containing values using a least-squares method which gives the proportion of variation in the measured values due to terms in the model (ΔG_{HOH}) and to random error (the errors we added). The values for the proportion of variation in the measurements due to measurement error in the plot are the average from nine separate simulations.

reproducibility, we have collected ΔG_{HOH} measurements from four different preparations of 14 mutant eglin c proteins (*22*). The average standard deviation in ΔG_{HOH} for these 14 mutant proteins was ±0.087 kcal/mol.

DISCUSSION

The precision of ΔG_{HOH} measurements in the literature range, in general, from ±0.1 to ±0.6 kcal/mol (*30*, *32*−*34*). There are three kinds of errors that could be reported. First is the error in ΔG_{HOH} determination reported by the curve fitting process, which is usually quite small (∼0.01 kcal/mol). Second is the measurement error determined by the reproducibility of the determination from a single protein preparation. Third is the error determined from the reproducibility of the determination from different protein preparations. The latter value is the largest, suggesting that protein impurities are a significant source of variability, but this is the value that most closely reflects one's capacity to "know" the ΔG_{HOH} value of a particular mutant protein.

How good do these measurements need to be? Our rationale for developing high-throughput protocols for ΔG_{HOH} measurements is to collect enough data to apply formal hypotheses testing to models concerning the determinants of protein structure. Such analyses provide values for the weights of the descriptors used in the hypothesis and allow one to assess the completeness of the hypothesis tested. A quantitative assessment of the completeness of the model, which determines whether new effects need to be added to fully predict behavior, has been difficult to obtain with the more anecdotal form of using mutagenesis for hypothesis testing. To show the consequences of ΔG_{HOH} measurement

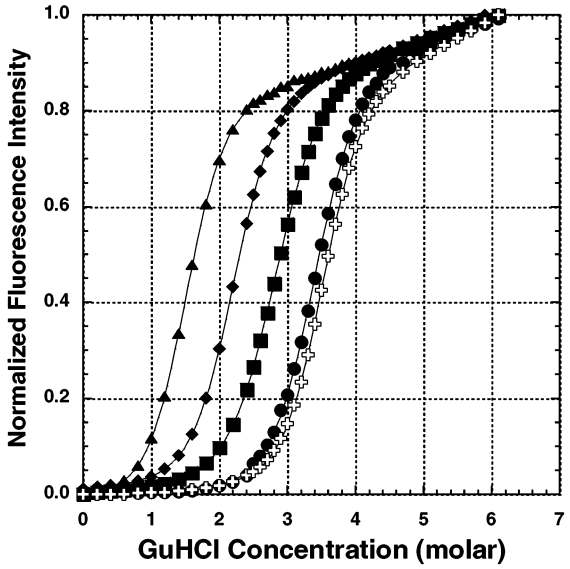

FIGURE 5: Chemical denaturation curves for eglin c proteins. The curves for the wild-type protein and four eglin c mutant proteins have been normalized to have an intensity of zero in 0 M GuHCl and 1 in 6.2 M GuHCl to show the nature of the pre- and posttransition properties of eglin denaturation in guanidine hydrochloride. The solid lines are the best fit of the data to a two-state process assuming that the pre- and posttransition behaviors represent the behavior of pure native and denatured protein over the entire range of guanidine hydrochloride concentrations. Curves: eglin wt, ●; R21A,E22K,T25Q,L26Q, ✣; R21E,E22D,T25E,L26Q, ■; R21N,E22Q,T25D,L26A, ◆; R21E,E22G,T25Q,L26G, ▲.

error of various magnitudes on the capacity to assess the completeness of models, we carried out a simulation using real ΔG_{HOH} data on eglin c mutant proteins to which were added various amounts of error. In this simulation the "model" is set up to predict the values we actually measured. If there was no measurement error, all of the differences in the measurements would be predicted by the model. As the amount of error added increases, the fraction of the variation in the measured values predicted by the otherwise perfect model drops. As expected, the fraction of the variation in the measured values attributable to the measurement error goes up as the square of that error (Figure 4). While a measurement error of ±0.10 kcal/mol leaves only 1.6% of the variation in measured values for an otherwise perfect model unexplained by the model (here attributable to measurement error), an error of ±0.4 kcal/mol leaves 16% of the variation unexplained by the model, and an error of ±0.6 kcal/mol leaves 48% unexplained. Hence data sets with measurement standard deviations of ~±0.5 kcal/mol would be weak data sets for deciding a model's completeness even if one were not using regression analysis. With that level of uncertainty in ΔG_{HOH}, a large fraction (~50%) of the measured behavior of the mutant proteins would be unaccounted for by even perfect models. Note that the consequences of measurement error go from negligible to large over the same range as the errors commonly reported in the literature (±0.1 to ±0.6 kcal/mol). Given this requirement for high precision, we wanted to assess the contribution to error from each part of the measurement protocol.

Errors Due to Differences in Protein Concentrations in the Solutions Used for Titration. Schwehm and Stites (*31*) suggest that the most accurate way to carry out fluorescence-monitored chemical denaturation is to start the titration with a protein solution in low denaturant and a half-filled cuvette and then add denaturant to that solution while correcting the fluorescence for dilution. This method imposes a limit on the range of denaturant that can be utilized. However, the quality of ΔG_{HOH} measurements derived from chemical denaturation depends on the capacity to define the pre- and posttransition behavior of the protein (Figure 5), and hence one would want to center that limited range on the midpoint of the denaturation transition. To use the one-protein solution approach, one usually determines the approximate midpoint of the denaturation transition (C_m) in preliminary experiments and then tailors the measurement to the properties of the mutant. In high-throughput mode this method is too expensive in terms of time and protein.

An alternative is to titrate by removing aliquots of low denaturant protein solution and replacing the removed solution with protein in high concentrations of denaturant. This approach causes measurement error from inequalities in the protein concentrations in the low and high denaturant solutions. How accurately can the two solutions be made and what are the consequences of errors in the solution preparation?

We use robotic liquid handling equipment (Beckman Biomek 2000) to prepare the mutant protein solutions. The variation in protein concentrations was measured by instructing the robot to prepare eight pairs of solutions and determining the volumes dispensed by weighing the tubes after each series of liquid additions. Remarkably, the Biomek robot achieves a coefficient of variation (100 × the standard deviation of the mean) of ±0.1% for any individual addition and a final average coefficient of variation of ±0.4% in the protein concentrations of the eight pairs (Table 1). This represents a standard deviation of ±0.5 to ±0.6 μL in the total volumes (2.5 or 6.0 mL) of the two solutions used in the titrations! The robot may be even better than this since the calibration tool we used is not 10 times more precise than the observable being measured as required for most accurate calibration.

Table 1: Performance of the Biomek Robot at Preparing Solutions for Chemical Denaturation

tube	$[protein]_{lG}$[a]	$[protein]_{hG}$[b]	$[protein]_{hG}$[b]/ $[protein]_{lG}$[a]	$[GuHCl]_{measd}$/ $[GuHCl]_{target}$[c]
1	0.0668	0.0670	1.003	1.001
2	0.0667	0.0668	1.001	1.000
3	0.0664	0.0668	1.006	1.000
4	0.0666	0.0669	1.005	0.999
5	0.0665	0.0671	1.009	0.999
6	0.0664	0.0670	1.009	0.998
7	0.0663	0.0671	1.012	0.998
8	0.0663	0.0670	1.011	0.998
mean	0.0665	0.0670	1.008	0.999
SD	0.0002	0.0001	0.004	0.001
SD as vol (μL)	0.5	0.6		
coeff of variation[d] (%)	0.3	0.2	0.4	0.1

[a] The protein concentration of the low denaturant titration solution (lG) expressed as protein volume to total volume. [b] The protein concentration of the high denaturant titration solution (hG) expressed as protein volume to total volume. [c] Ratio of the measured volume of GuHCl solution dispensed into the high denaturant solution tube to the targeted volume. [d] Standard deviation of the values from the eight preparations over the mean × 100.

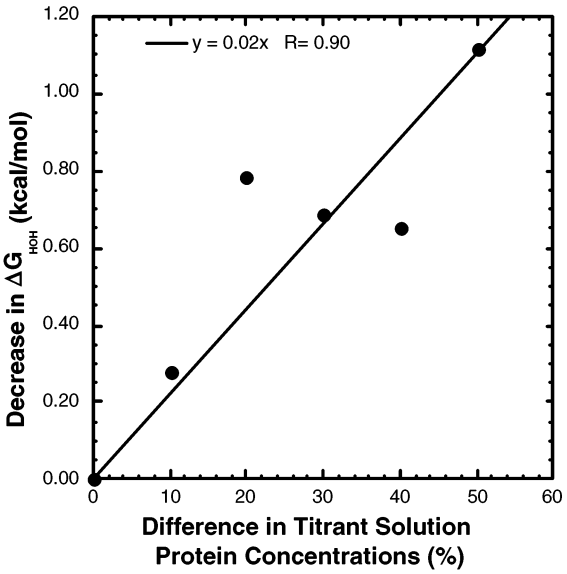

FIGURE 6: Decrease in ΔG_{HOH} generated by increases in the protein concentration of the high denaturant titrant solution relative to the low denaturant solution. High denaturant titrant solutions were prepared with concentrations of egg white lysozyme that were higher than the protein concentration in the low denaturant solution. ΔG_{HOH} values were then determined with solutions having the indicated level of protein concentration difference.

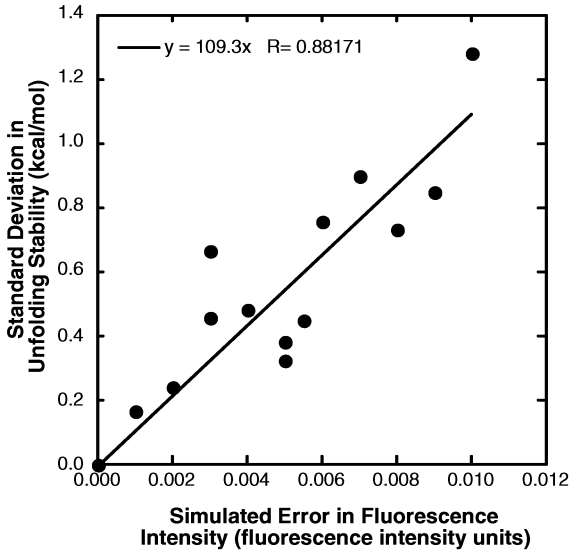

FIGURE 7: Variance in ΔG_{HOH} generated by simulated errors in fluorescence intensity. Fluorescence measurement errors were generated by drawing numbers from normal distributions with means of zero and various standard deviations. These errors were added to perfect measurement values (real values plus the residuals for the curve fit) for data from chemical denaturation of egg white lysozyme. These error-containing values were then used to generate ΔG_{HOH} values as described. Nine sets of error-containing data were used to obtain the standard deviation in ΔG_{HOH} values for each measurement error level.

Measurements of ΔG_{HOH} for egg white lysozyme made with and without intentional differences in protein concentration between the initial solution and the titrant (Figure 6) indicate that the measurements are modestly resistant to such differences (0.02 kcal/mol per percent mismatch). The Biomek generates protein solutions with protein concentration errors of ±0.4% (Table 1); hence we can expect the errors in ΔG_{HOH} data from solution mismatch to have a standard deviation of ±0.008 kcal/mol or 9% of the total variance of ±0.087 kcal/mol.

Variance in ΔG_{HOH} Measurements Generated by Fluorescence Measurement Error. To assess the impact of fluorescence intensity measurement error, we took "perfect" fluorescence intensity data for the chemical denaturation of egg white lysozyme and added simulated errors drawn from normal distributions with means of zero and various standard deviations. A significant level of instrument precision is needed (Figure 7) for precise ΔG_{HOH} measurements. We determined that our instrument gives a standard deviation of ±0.00054 fluorescence unit over 30 s. An examination of Figure 7 shows that such errors lead to a standard deviation of ±0.059 kcal/mol, which accounts for 68% of the observed standard deviation in our measurements (±0.087 kcal/mol) and is the major source of error in our ΔG_{HOH} determinations.

Variance in ΔG_{HOH} Generated by Titration Dispensing Errors. To assess the impact of dispensing errors associated with the titrator, we generated a simulation in which the errors in removed and added volumes were drawn from numbers from normal distributions with various standard deviations. The protein and GuHCl concentrations were then calculated from these volumes with the titration errors propagating throughout the titration using the fluorescence intensity values from a normalized perfect response curve for eglin c. ΔG_{HOH} values calculated from these simulated data sets indicate that the error in ΔG_{HOH} is ±0.0216 kcal/mol times the percent error in titration volume. The Hamilton syringe system has a precision of better than 0.5% when the volume dispensed is greater than 10% of the syringe (conditions met by our protocol) and hence contributes about ±0.011 kcal/mol variance to the ΔG_{HOH} measurements or about 13% of the observed standard deviation (±0.087 kcal/mol) in our ΔG_{HOH} measurements.

Variance in ΔG_{HOH} Arising from Impurities in the Protein Preparations. Purity of protein preparations is a major potential source of variance in ΔG_{HOH} measurements. To assess the effects of protein contamination on ΔG_{HOH}, we collected denaturation data from mixtures of eglin c and egg white lysozyme (both have denaturation profiles that go from low to high fluorescence) and fit the resulting data to a two-state model. The apparent ΔG_{HOH} for eglin c increases by 0.06 kcal/mol per percent of the mixture that is egg white lysozyme (data not shown). Making mixtures of a protein (CheY) that shows the more traditional high to low transition during chemical denaturation gave similar results. When we used EWL to break open the cells, we would often see trace levels of EWL in the purified preparations. Use of a freeze−thaw protocol (*35*) removes that source of contamination and produces a lysate with a lower level of contamination from *E. coli* proteins. The standard deviation of ΔG_{HOH} for eglin c increases by ±0.011 kcal/mol when measurements from several preparations are compared to repeated measurements from the same preparation, implying that the high-throughput protein purifications achieve purity levels of ~99.8%.

The instrumentation described provides a means to obtain very high precision measurements of ΔG_{HOH} from large numbers of mutant proteins. The quality of those measurements of course depends on the extent to which the individual mutant proteins retain the properties of the wild type that are necessary for the measurements, such as fast folding and two-state behavior.

CONCLUSION

The fraction of the stability behavior of a set of mutants that can be explained by perfect models decreases as the square of the measurement error (Figure 4). In fact, the fraction goes from 98% to only 50% over the range most commonly reported for ΔG_{HOH} measurements in the literature ($\pm$0.1 to $\pm$0.6 kcal/mol). Much of the stability behavior of mutant proteins measured to a precision of $\pm$0.6 kcal/mol would not be explained by even a perfect model. As a consequence, using the formal mathematics of hypothesis testing to assess models for the determinants of protein structure requires high-precision stability measurements on large numbers of mutant proteins if one is to be in a position to assert that there are effects missing from the models tested. To achieve precision levels of $\pm$0.1 to $\pm$0.2 kcal/mol using chemical denaturation, one needs an instrument that can monitor the denaturation process precisely (coefficient of variation in the measurements of $\leq$0.7% of the midpoint of transition signal) as well as protocols to produce quite pure mutant proteins ($\geq$97% purity) and precisely matched protein solutions for the chemical denaturation titration ($\leq$2.5% differences in protein concentrations in the high- and low-denaturant solutions). These criteria can be met in a high-throughput mode using robotics for protein purification and solution preparations (Beckman Biomek 2000) and a semi-automated titrating fluorometer (Protein Solutions Inc. ATF105). Throughputs depend on the time that the protein of interest takes to reach equilibrium during the titration process, and we have been able to collect data from two mutant proteins per 45 min with eglin c. The standard deviations in ΔG_{HOH} values from a single preparation of wild-type eglin c ($\pm$0.040 kcal/mol) and from separate preparations ($\pm$0.051 kcal/mol) are essentially the same, implying that the freeze−thaw/His-tag purification scheme produces very pure proteins.

ACKNOWLEDGMENT

We thank Dr. Charlie Carter for useful discussions during the course of the work.

REFERENCES

1. Hutchison, C. A., III, Phillips, S., Edgell, M. H., Gillam, S., Janke, P., and Smith, M. (1978) *J. Biol. Chem. 253*, 6551−6560.
2. Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. (1996) *Applied Linear Statistical Models*, 4th ed., p 285, McGraw-Hill, New York.
3. Mark, A. E., and van Gunsteren, W. F. (1994) *J. Mol. Biol. 240*, 167−176.
4. Boresch, S., and Karplus, M. (1995) *J. Mol. Biol. 254*, 801−807.
5. Brady, G. P., and Sharp, K. A. (1995) *J. Mol. Biol. 254*, 77−85.
6. Rooman, M., and Wodak, S. (1995) *Protein Eng. 8*, 849−858.
7. Ackers, G. F., and Smith, F. R. (1985) *Annu. Rev. Biochem. 54*, 597−629.
8. Tidor, B., and Karplus, M. (1991) *Biochemistry 30*, 3217−3228.
9. Brooks, C. L., III, and Case, D. A. (1993) *Chem. Rev. 93*, 2487−2502.
10. Boresch, S., Archontis, G., and Karplus, M. (1994) *Proteins: Struct., Funct., Genet. 20*, 24−33.
11. Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M., and Smith, M. (1982) *Nature 299*, 756−758.
12. Pielak, G. J., Mauk, A. G., and Smith, M. (1985) *Nature 313*, 152−154.
13. Fersht, A. R., Shi, J.-P., Knill-Jones, J., Lowe, M. D., Wilkinson, A. J., Blow, D. H., Brick, P., Carter, P., Waye, M. M. Y., and Winter, G. (1985) *Nature 314*, 235−238.
14. Shortle, D., and Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81−89.
15. Matthews, B. W. (1987) *Biochemistry 26*, 6885−6888.
16. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *Biochemistry 29*, 9343−9352.
17. Munoz, V., and Serrano, L. (1995) *J. Mol. Biol. 245*, 275−296.
18. Guerois, R., Nielsen, J. E., and Serrano, L. (2002) *J. Mol. Biol. 320*, 369−387.
19. Carter, C. W., Jr., LeFebvre, B. C., Cammer, S. A., Tropsha, A., and Edgell, M. H. (2001) *J. Mol. Biol. 311*, 625−638.
20. Gilis, D., and Rooman, M. (1996) *J. Mol. Biol. 257*, 1112−1126.
21. Miyazawa, S., and Jernigan, R. L. (1985) *Macromolecules 18*, 534−552.
22. Yi, F., Sims, D. A., Pielak, G. J., and Edgell, M. H. (2003) *Biochemistry 42*, 7594−7603.
23. Lahr, S. J., Broadwater, A., Carter, C. W., Jr., Collier, M. L., Hensley, L., Waldner, J. C., Pielak, G. J., and Edgell, M. H. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 14860−14865.
24. Tanford, C. (1964) *J. Am. Chem. Soc. 86*, 2050−2059.
25. Pace, C. N., and Shaw, K. L. (2000) *Proteins: Struct., Funct., Genet., Suppl. 4*, 1−7.
26. Alexander, S. S., and Pace, C. N. (1971) *Biochemistry 10*, 2738−2743.
27. Greene, R. F., Jr., and Pace, C. N. (1974) *J. Biol. Chem. 249*, 5388−5393.
28. Shortle, D., Sites, W. E., and Meeker, A. K. (1990) *Biochemistry 29*, 8033−8041.
29. Stites, W. E., Byrne, M. P., Aviv, J., Kaplan, M., and Curtis, P. M. (1995) *Anal. Biochem. 227*, 112−122.
30. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry 27*, 8063−8068.
31. Schwehm, J. M., and Stites, W. E. (1998) *Methods Enzymol. 295*, 150−170.
32. Meeker, A. K., Garcia-Moreno, E. B., and Shortle, D. (1996) *Biochemistry 35*, 6443−6449.
33. Lorch, M., Mason, J. M., Clarke, A. R., and Parker, M. J. (1999) *Biochemistry 38*, 1377−1385.
34. Kragelund, B. B., Poulsen, K., Andersen, K. V., Baldursson, T., Kroll, J. B., Neergard, T. B., Jepsen, J., Roepstorff, P., Kristiansen, K., Poulsen, F. M., and Knudsen, J. (1999) *Biochemistry 38*, 2386−2394.
35. Johnson, B. H., and Hecht, M. H. (1994) *Bio/Technology 12*, 1357−1360.

BI034063G