

Some recommendations for the practitioner to improve the precision of experimentally determined protein folding rates and Φ values

Ingo Ruczinski^{1*} and Kevin W. Plaxco²

¹Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, and Institute for Multiscale Modeling of Biological Interactions, Johns Hopkins University, Baltimore, Maryland

²Department of Chemistry and Biochemistry and Interdepartmental Program in Biomolecular Science and Engineering, University of California at Santa Barbara, Santa Barbara, California

ABSTRACT

The mechanism by which proteins fold from an initially random conformation into a functional, native structure remains a major unsolved question in molecular biology. Of particular interest to the protein folding community is the structure that the protein adopts in the folding transition state (the highest free energy state on the pathway from unfolded to folded), as that state forms the barrier that defines the folding pathway. Unfortunately, however, unlike those of the initial, unfolded state and the final, folded state of the protein, the structure in the transition state cannot be directly assessed via experiment. Instead, experimentalists infer the structure of the transition state, often by estimating changes in its free energy by measuring the effects of amino acid substitutions on folding and unfolding rates (Φ -value analysis). In this article we show how to obtain more efficient estimates of these important quantities via improved experimental designs, and how to avoid common pitfalls in the analysis of kinetic data during the extraction of these parameters.

Proteins 2009; 74:461–474.
© 2008 Wiley-Liss, Inc.

Key words: protein folding; transition state; Φ -value; free energy; estimate; standard error.

INTRODUCTION

Proteins are synthesized in the cell as unstructured, random-coil polymer chains that then rapidly and spontaneously fold into their functional, native structures. In the late 1960s, Cyrus Levinthal noted that this process embodies an apparent paradox: the number of configurations available to the unfolded, unstructured chain is so great that a random search to find the single, correct native structure should require greater than the age of the Universe to reach completion.¹ The means by which evolution has solved this so-called “Levinthal Paradox” and folds proteins in a biologically relevant timeframe remains one of the most important unanswered questions in biochemistry; despite more than 40 years effort there is no widely accepted theory of folding kinetics that, for example, can predict even relative protein folding rates (reviewed in Ref. 2).

Because it is difficult to study the folding of proteins as they are synthesized in the cell, the dominant experimental approach to folding kinetics is to harvest already folded proteins, unfold them via the use of chemical denaturants (such as guanidine hydrochloride or urea) and then monitor the rate with which the system re-equilibrates after folding is initiated by dilution of the denaturant (reviewed in Ref. 3). Similarly, it is common to measure the rate with which a folded protein unfolds by rapidly introducing it to high concentration of the denaturant. A limitation of these experiments, however, is that they measure folding and unfolding rates in the presence of denaturant (even the refolding measurement takes place in the presence of denaturant because the denaturant cannot be diluted to zero concentration), and both refolding and unfolding rates are sensitive functions of denaturant concentration. Because of this, it is common to measure multiple re-equilibration rates over a range of denaturant concentrations and then to extrapolate these observations to estimate the rates that would be observed in the absence of denaturant.⁴ Because a plot of the logarithm of the observed re-equilibration rate versus denaturant concentration approximates a “V”, this experimental approach is commonly referred to as “chevron analysis.”

Of particular interest to the community of protein folding aficionados is the structure that the protein adopts in the folding transition state (the state with highest free energy), as that state forms the barrier that ultimately defines the folding pathway. Unlike the structure in the initial, unfolded state and the final,

*Correspondence to: Ingo Ruczinski, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, and Institute for Multiscale Modeling of Biological Interactions, Johns Hopkins University, Baltimore, MD.
E-mail: ingo@jhu.edu

Received 30 January 2008; Revised 19 May 2008; Accepted 20 May 2008

Published online 24 July 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22155

folded state however, the structure in the transition state cannot directly be assessed by experimental means such as X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR).³ Instead researchers typically rely on an indirect approach, termed Φ -value analysis, which provides a means of defining the structure of the transition state using parameters obtained via chevron analysis. In this approach, one (or, rarely, more) amino acid in the naturally occurring, “wild-type” sequence of the protein is mutated. This has the effect of removing atoms from (or, more rarely, adding atoms to) the protein. By measuring the relative folding (and unfolding) rates of the wild type and variant protein, one can define a quantitative measure of the extent to which the ablated atoms contribute to the free energy (and thus, presumably, structure) of the folding transition state. Specifically, the Φ -value is given by the ratio of the change in the relative free energy of the folding transition state ($\Delta\Delta G_{\ddagger}$) versus the change in the equilibrium free energy ($\Delta\Delta G_U$) engendered by the mutation. As such, a Φ value of 1 denotes a side chain that is contributing as much to the stability of the transition state as it is to the well-folded native state, suggesting that this region of the polypeptide chain is well-ordered in the folding transition state. Likewise, a Φ value of 0 is often taken to imply a side chain that is as unstructured in the folding transition state as it is in the unfolded state (but see interpretational caveats listed in, for example, Ref. 5).

Over the last three decades, the folding rates of hundreds of proteins and the Φ -values of thousands of mutations have been characterized and reported using chevron analysis.^{4,6,7} Despite the important role that these experiments have played in our understanding of protein folding, however, neither the experimental approach nor analysis have been optimized in terms of the precision and accuracy with which the relevant kinetic parameters can be estimated.^{7–9} Indeed, as we show here, the currently standard experimental methods lead to unnecessarily large imprecision in these estimates. Here we discuss methods by which the key kinetic parameters can be estimated with improved precision and accuracy, and how to avoid common pitfalls in the analysis of kinetic data during the extraction of these parameters.

RESULTS

As noted earlier, it is generally assumed that, for small proteins (less than about 120 amino acids in length), a plot of the logarithm of the folding and unfolding rates (more precisely the observed relaxation rate, which is the sum of the folding and unfolding rates under the given set of conditions) forms a chevron-shaped curve with linear or near-linear folding and unfolding arms at low and high denaturant concentrations, respectively (Fig. 1). Although the origins of this behavior remains controversial,¹⁰ it is

widely assumed in the community that significant deviations from linearity are the exception,⁴ and the usual method for estimating the rates expected to be observed in the absence of denaturant from experiments necessarily conducted at non-zero denaturant concentrations is to fit observed relaxation kinetics to a linear “chevron relationship”:

$$\ln(k_{\text{obs}}) = \ln[\exp(\ln(k_f) + m_f [\text{denaturant}]/RT) + \exp(\ln(k_u) + m_u [\text{denaturant}]/RT)] \quad (1)$$

where $\ln(k_{\text{obs}})$ represents the experimentally observed relaxation rate at a given denaturant concentration (denoted [denaturant]), and $\ln(k_f)$ and $\ln(k_u)$ represent the estimated folding and unfolding rates in the absence of denaturant, respectively. The final two terms, m_f and m_u , represent the slopes of the folding and unfolding arms of the chevron plot.

The traditional method for performing chevron analysis entails the collection of rate data at numerous, approximately evenly spaced denaturant concentrations and the fitting of these data to Eq. (1) in order to estimate the folding and unfolding rates that would be observed in the absence of denaturant and, from them, the above-described Φ -values are calculated. We find here, however, that the efficiency of this traditional experimental design is far from optimal, and that the precision in the estimates for the kinetic parameters can easily be improved. If the chevron arms are known to be linear, we propose a practical and efficient experimental design that assesses folding and unfolding rates at the extreme values of the range for the denaturant concentrations and the inflection point. If the linearity of the chevron arms is not known and also of interest to the investigator, we propose a second experimental design that loses very little in efficiency, and has sufficient power to detect the departure from linearity. We also find that several commonly employed methods to make inference about the kinetic parameters (e.g., to calculate Φ -values) are not correct and can lead to invalid conclusions. We illustrate this via some simulated examples, and describe simple, valid methods for correct inference.

Simple experimental design considerations can lead to improved estimates

Selecting denaturant concentrations

The traditional experimental approach to obtaining chevron plots is to monitor relaxation kinetics at 10–20 denaturant concentrations (over a range from near zero to 6 to 8 or even 10M, the latter two values representing near saturated solutions of guanidine hydrochloride and urea, respectively) approximately evenly spaced along the gradient of denaturant concentrations.⁴ We have simulated the effects of non-uniform distributions of data points, however, and find that the traditionally employed

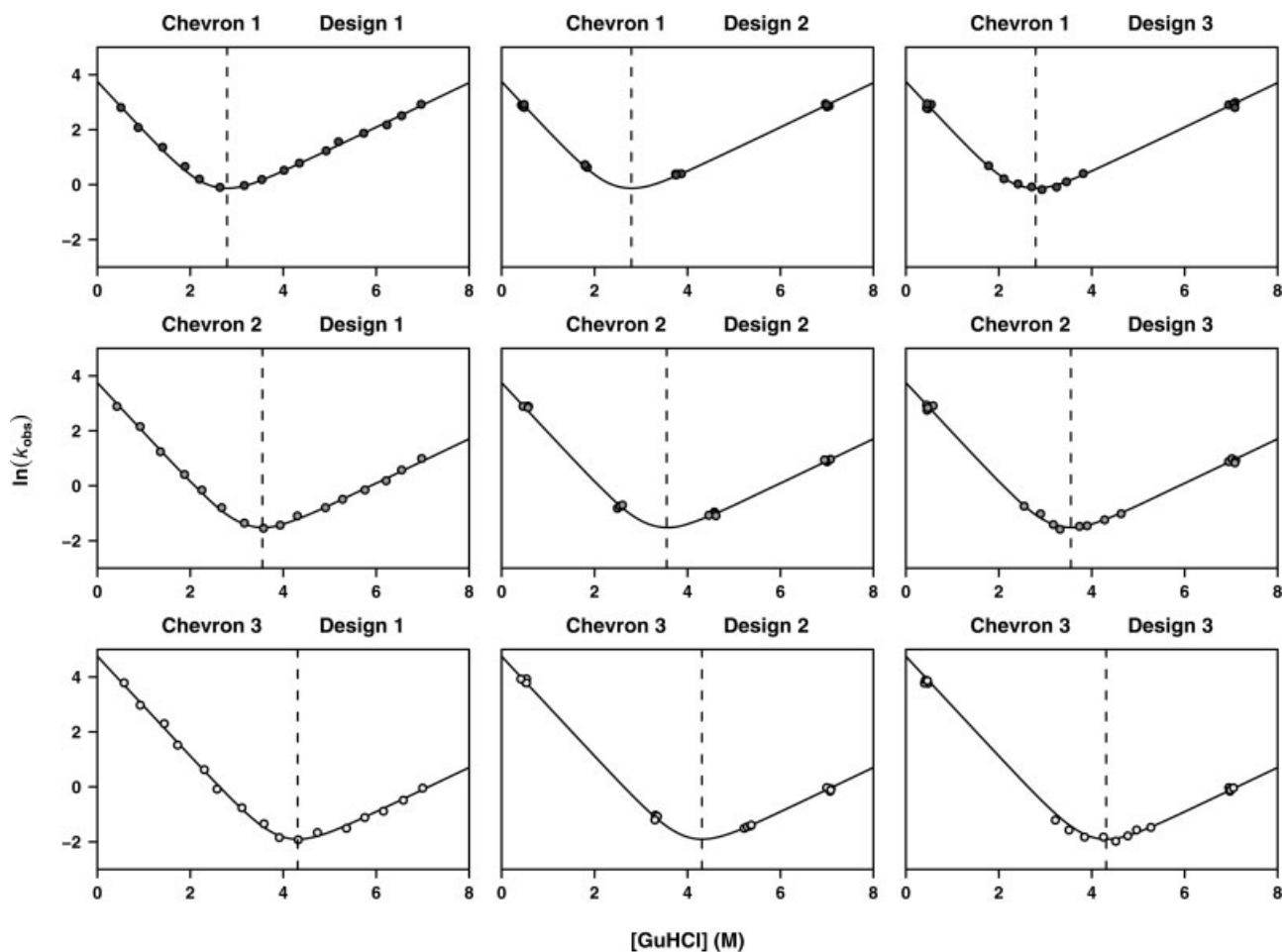


Figure 1

Three different experimental designs (columns 1–3) to gather kinetic chevron data for three different chevrons (rows 1–3), typical of small two-state folding proteins. Qualitatively, the chevrons differ mainly in the location of the inflection point. In the first design, the relaxation rates are monitored at evenly spaced denaturant concentrations, between 0.5 and 7M denaturant concentration. In the second design, an equal number of design points are placed at the extremes of the folding and unfolding arms, defined here as 1M denaturant concentration away from the location of the inflection point, and 0.5 and 7M, respectively. In the third design, the design points that were near the inflection points in the previous design are spread around the respective inflection points. Per panel, 16 kinetic measurements are simulated, using Gaussian error with mean zero, and standard deviation typical for such experiments.

“even distribution” approach is far from optimal and that basic experimental design considerations can significantly improve the precision of the estimates of the kinetic parameters.

To study the effect of experimental design on the precision with which kinetic parameters are estimated we generated three synthetic chevrons (rows 1–3), typical of small proteins and differing mainly in the location of the inflection point of the chevron. Three different designs (columns 1–3) were considered for these three simulated proteins. The first design, in which relaxation rates are monitored at evenly spaced denaturant concentrations, is typical of experiments reported in the literature.⁴ The second design is derived from standard linear model theory: as the folding and unfolding arm are in essence linear (and curvature only arises near the inflection

point), more precision for the kinetic parameters should be achievable by placing an equal number of design points at the extremes of the folding and unfolding arms. However, as the point of inflection might not be known a priori (but the possible extremes of the denaturant concentrations are limited by experimental considerations and are generally well-defined), our third design places observations points at the extrema of the possible denaturant concentrations, and infer the point of inflection using a small range of likely denaturant concentrations.

For each of the nine scenarios (three synthetic chevrons times three experimental designs), we simulated 10,000 kinetic data sets around the “true” chevron using a Gaussian error with mean zero and standard deviation 0.05 (a value typical of those observed in the literature⁴;) and estimated the variability in the fitted kinetic param-

Table 1

Variance Ratios (Efficiencies) of the Estimates Obtained from the Traditional Design (Design 1) Over the Variances of the Estimates Obtained from the Other Designs (Design 2 and 3) and Reported for all Chevrons in this Study (Chevrons 1–3)

	$\log(k_f)$	m_f	$\log(k_u)$	m_u	m_{eq}	ΔG
Chevron 1 : design 1 versus design 2	2.6	2.1	1.3	1.5	1.8	1.2
Chevron 1 : design 1 versus design 3	2.6	1.7	1.2	1.5	1.9	1.4
Chevron 2 : design 1 versus design 2	2.2	2.1	1.3	1.4	1.5	1.1
Chevron 2 : design 1 versus design 3	2.2	1.5	1.4	1.7	1.7	1.4
Chevron 3 : design 1 versus design 2	2.0	1.9	1.2	1.3	1.2	1.1
Chevron 3 : design 1 versus design 3	1.9	1.3	1.5	1.7	1.9	1.4

For each chevron/design combination, 10,000 replicates of kinetic data were simulated according to the scheme in Figure 1, and the kinetic parameters were estimated in each simulation. The variances between the simulations were recorded, and are reported here as fractions, comparing the traditional design to the other designs. As all values are larger than one, this indicates that both alternative designs have smaller variability in the parameter estimates as the traditional design.

ters for each data set. In doing so we find that the precision in the estimates of the kinetic parameters can be substantially improved relative to the precision obtained using the traditional experimental approach of evenly spaced denaturant concentrations (Table 1). In specific, the efficiency (defined as the ratio of the variances of the estimates) obtained using experimental designs 2 and 3 is always improved, for all kinetic parameters, over that obtained using the traditional approach, with improvements in precision of up to 2- to 3-fold being observed in several cases. As a measure of the importance of such large changes in efficiency, we note that a variance ratio of 2.25 implies a ratio of standard errors of 1.5, and thus the precision of a kinetic parameter defined using the “traditional” experimental design is 50% poorer than that produced by the competing design.

Protein folding and unfolding rates can, in turn, be used to calculate a number of other parameters of interest to the aficionado. These include Φ -values (which are, as described above, a measure of the structure of the folding transition state) and the equilibrium free energy of folding. Using kinetic parameters, the free energy of folding is defined as:

$$\Delta G_U = RT \times (\ln(k_f) - \ln(k_u)) \quad (2)$$

where $\ln(k_f)$ and $\ln(k_u)$ represent the estimated folding and unfolding rates of the protein, respectively, R is the gas constant and T is the temperature in kelvins (note that under typical experimental conditions the scaling factor RT equals approximately 0.6 kcal/mol). From these parameters we can also define the Φ -value associated with a given mutation by:

$$\Phi = \frac{\Delta \Delta G_{\ddagger}}{\Delta \Delta G_U} \quad (3)$$

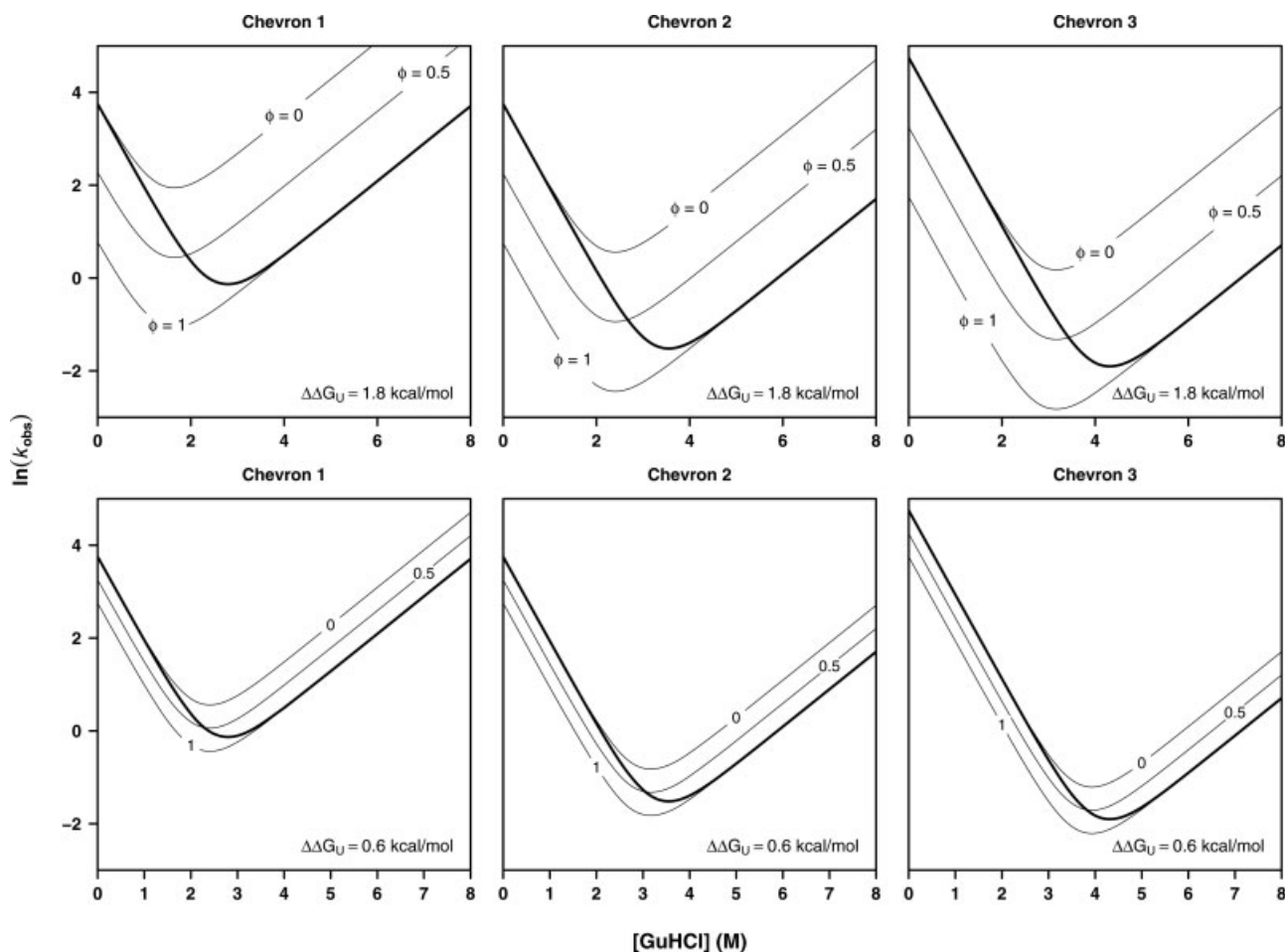
where $\Delta \Delta G_{\ddagger} = RT \times (\ln(k_f) - \ln(k'_f))$ represents the change that the mutation imparts on the relative free energy of the folding transition state, and $\Delta \Delta G_U = RT \times (\ln(k_f) - \ln(k_u) - \ln(k'_f) + \ln(k'_u))$ denotes the mutation-linked change in the equilibrium unfolding free

energy of the protein. We find that, not surprisingly, the improvement seen for the precision of the kinetic chevron parameters also translates into an improvement in the precision of the changes in both of these important measurements.

To test the effects of experimental design on the precision with which Φ -values and folding free energies are defined, we generated sets of “mutant” chevron curves based on our initial set of three chevron curves (Fig. 2). These mutant chevron curves differ from the original (wild-type) chevron curve by being displaced such that they result in Φ -values of 0, 0.5, or 1. We consider two cases: one in which the change in free energy of the folded state ($\Delta \Delta G_U$) is relatively large (1.8 kcal/mol; Fig. 2, upper row) and one in which the change in free energy is relatively small (0.6 kcal/mol; Fig. 2, lower row). For each of these settings, we simulated 10,000 kinetic data sets around the chevrons using a random Gaussian error with mean zero and standard deviation typical of those observed in the literature. Using these 10,000 simulated data sets we estimated the variability in the fitted kinetic parameters, the changes in free energy, and the resulting Φ -values for each of these settings.

In general we find that the precision available via the traditional experimental approach (design 1) is much poorer than that obtained using our modified experimental approaches (designs 2 and 3) for the estimates of the kinetic parameters, the estimates of the relative change in free energy, and the estimates of Φ (Tables II and III). We also find that the precision of the kinetic parameters depends on the range of denaturant concentrations selected in the experimental design. In particular, improved precision is obtained by setting lower limiting denaturant concentration as close to zero as is practical, and the upper limiting concentration as high as is practical (see supplementary material at <http://biostat.jhsph.edu/~iruczins/publications/sm/>).

We have demonstrated that the traditional approach to experimental chevron analysis earlier, which is to measure folding and unfolding kinetics over a range of more-or-less equally spaced denaturant concentrations,

**Figure 2**

For each of the three chevrons, we generate shifted chevron curves with parallel folding and unfolding arms, corresponding to three different Φ values (0, 0.5, and 1). Two different shifts are considered, corresponding to small (0.6 kcal/mol, lower row) and large (1.8 kcal/mol, upper row) changes in free energy of the folded state ($\Delta\Delta G_U$).

Table II

Variance Ratios (Efficiencies) of the Kinetic Parameter Estimates, Obtained from Simulations as Indicated in Figure 2

	$\Phi = 0$				$\Phi = 0.5$				$\Phi = 1$			
	$\log(k_f)$	m_f	$\log(k_u)$	m_u	$\log(k_f)$	m_f	$\log(k_u)$	m_u	$\log(k_f)$	m_f	$\log(k_u)$	m_u
$\Delta\Delta G_U = 1.8$ kcal/mol												
Chevron 1 : design 1 versus design 2	2.4	2.3	1.4	1.8	2.1	1.7	1.4	1.4	2.2	2.1	1.3	1.5
Chevron 1 : design 1 versus design 3	3.0	2.2	1.2	1.5	2.7	2.1	1.3	1.5	3.0	2.0	1.0	1.2
Chevron 2 : design 1 versus design 2	2.0	2.3	1.6	1.6	1.7	2.2	1.5	1.5	2.3	2.6	1.7	1.9
Chevron 2 : design 1 versus design 3	2.3	2.1	1.1	1.2	1.8	1.7	1.2	1.3	2.4	1.8	1.3	1.4
Chevron 3 : design 1 versus design 2	2.0	2.1	1.1	1.3	2.4	2.2	0.9	1.0	2.3	2.6	1.4	1.6
Chevron 3 : design 1 versus design 3	1.9	1.5	1.0	1.2	2.3	1.5	0.9	1.1	2.6	1.7	1.5	1.7
$\Delta\Delta G_U = 0.6$ kcal/mol												
Chevron 1 : design 1 versus design 2	2.4	1.8	1.3	1.5	2.5	1.9	1.3	1.5	2.5	1.9	1.3	1.5
Chevron 1 : design 1 versus design 3	2.8	1.8	1.1	1.4	2.7	1.8	1.2	1.5	2.8	1.8	1.1	1.5
Chevron 2 : design 1 versus design 2	2.4	2.1	1.2	1.4	2.4	2.1	1.3	1.5	2.4	2.2	1.3	1.5
Chevron 2 : design 1 versus design 3	2.4	1.6	1.3	1.6	2.4	1.7	1.3	1.6	2.4	1.7	1.3	1.5
Chevron 3 : design 1 versus design 2	2.1	2.1	1.3	1.4	2.1	2.1	1.2	1.4	2.1	2.1	1.3	1.4
Chevron 3 : design 1 versus design 3	2.0	1.4	1.5	1.7	2.1	1.4	1.5	1.7	2.1	1.5	1.5	1.7

Table IIIVariance Ratios (Efficiencies) of the Estimates, for the Changes in Free Energy and Φ , Obtained from Simulations as Indicated in Figure 2

	$\Phi = 0$			$\Phi = 0.5$			$\Phi = 1$		
	Φ	$\Delta\Delta G_{\ddagger}$	$\Delta\Delta G_U$	Φ	$\Delta\Delta G_{\ddagger}$	$\Delta\Delta G_U$	Φ	$\Delta\Delta G_{\ddagger}$	$\Delta\Delta G_U$
$\Delta\Delta G_U = 1.8$ kcal/mol									
Chevron 1 : design 1 versus design 2	2.1	2.2	1.4	1.9	2.3	1.4	1.2	2.5	1.3
Chevron 1 : design 1 versus design 3	2.9	2.8	1.4	2.0	3.6	1.8	1.1	2.9	1.8
Chevron 2 : design 1 versus design 2	2.4	2.2	1.2	2.3	1.6	1.2	2.0	1.9	1.1
Chevron 2 : design 1 versus design 3	2.5	2.5	1.3	1.4	2.1	1.4	1.2	2.1	1.0
Chevron 3 : design 1 versus design 2	1.8	1.9	1.0	1.6	2.2	1.1	1.4	2.1	1.4
Chevron 3 : design 1 versus design 3	1.8	1.8	1.2	1.2	1.9	1.2	2.1	2.3	1.9
$\Delta\Delta G_U = 0.6$ kcal/mol									
Chevron 1 : design 1 versus design 2	2.5	2.5	1.3	1.9	2.5	1.3	1.3	2.5	1.3
Chevron 1 : design 1 versus design 3	2.7	2.7	1.5	1.6	2.7	1.5	1.2	2.7	1.5
Chevron 2 : design 1 versus design 2	2.3	2.3	1.1	1.7	2.3	1.2	1.3	2.3	1.1
Chevron 2 : design 1 versus design 3	2.3	2.2	1.3	1.7	2.2	1.4	1.4	2.3	1.3
Chevron 3 : design 1 versus design 2	2.0	2.0	1.1	1.4	1.9	1.1	1.2	2.0	1.1
Chevron 3 : design 1 versus design 3	2.0	1.9	1.4	1.7	2.0	1.4	1.7	2.0	1.4

exhibits poorer efficiency than other, easy to implement alternatives. We note, however, that we do not imply optimality here—for this, a criterion for optimization must be established and, at a minimum, the location of the inflection point must be known. But knowledge of the location of the inflection point is typically lacking, and thus, adaptive designs such as the ones proposed here are needed. Even if the inflection point was known, it is unclear what criterion for optimality should be chosen. Ultimately, we are interested in minimizing the standard error on the Φ -value estimate, which depends on *two* chevrons (one for the wild-type protein, and one for the mutant). To select an optimality criterion that guides the choice of denaturant concentrations for a single chevron, a somewhat arbitrary choice about a function of the errors in the kinetic parameter estimates had to be made. A function to be minimized could for example be the term $(\sigma_f)^2 + (\sigma_u)^2 - 2\rho\sigma_f\sigma_u$, which is the contribution of the individual chevron to the variability in the estimate for $\Delta\Delta G_U$, see Eq. (A13). What we have demonstrated is that the traditional design is far from optimal, and that simple alterations to standard experimental designs can significantly improve the precision in the estimates of the kinetic parameters.

A second commonly employed experimental feature also degrades experimental efficiency, and is equally easily rectified. Specifically, although practitioners almost invariably collected multiple observations at a single denaturant concentration, these observations are typically averaged and used as a single data point for chevron analysis. While extremely commonly employed, however, this approach relies on two major assumptions: (1) the same number of rate determinations are performed at each denaturant concentration and (2) the variability in the estimate for the folding rate at any particular non-zero concentration is constant across all experiments.

Although it does not create a bias, the near universal departure of true experimental conditions from these assumptions⁴ reduces the efficiency with which the relevant kinetic parameters are estimated. Of course, while this is rarely done, assumption (1) can be readily enforced by the experimentalist. Assumption (2), in contrast, cannot be enforced and is always violated to varying degrees: in truth, all the folding rate estimates have different standard errors. Both of these issues can be addressed, however, by not averaging the estimates for the folding rates and instead fitting the chevron curves to the “raw” data (for violations of assumption (1), that is if a different number of experiments is run for each denaturant concentration), and by using case weights (defined as the inverse of the standard error of the estimates) in the fitting procedure for the chevron curves (for violations of assumption (2), that is if the variability in the folding rate estimates varies).

Likelihood ratio tests to assess linearity

All of the above considerations assume that the folding and unfolding arms of the chevron are linear. And whilst this generally appears to be true for small proteins, is not universally true and well known departures from linearity have been documented and are thought to report on important aspects of the folding mechanism.¹⁰ In this case, an alternative model using quadratic terms for the folding and unfolding arms has been suggested,⁴ although the physical justification of this model, like that of the linear model, is unclear:

$$\ln(k_{\text{obs}}) = \ln[\exp(\ln(k_f) + m_f [\text{denaturant}])/RT + m_{f_2}[\text{denaturant}]^2) + \exp(\ln(k_u) + m_u[\text{denaturant}]/RT + m_{u_2} [\text{denaturant}]^2)] \quad (4)$$

Given the potential importance of nonlinear behavior, it is clearly desirable to define a test to systematically determine whether or not a departure from linearity is present. The departure from linearity can formally be tested using a likelihood ratio test because the linear model is nested in the quadratic model (i.e., the linear model can be derived by setting the quadratic parameters m_{f_2} and m_{u_2} equal to zero). The linear and quadratic equations are fit to the observed data, and the ratio of the two likelihoods is calculated. If the chevron arms are truly linear, then the test statistic (defined as two times the natural logarithm of this likelihood ratio) follows a χ^2_2 distribution, and P -values that quantify the evidence against the assumption of linearity are calculated using this distribution. For example, the protein CI-2 exhibits linear behavior, whereas the protein U1A clearly does not,⁴ (Fig. 3, upper panels). For these two cases, the likelihood ratio tests for linearity yield P -values of 0.60, and about 10^{-30} , respectively.

When conducting such a kinetic experiment, it may or may not be known a priori whether the data derived for a particular protein will exhibit linear behavior or not: for example, if a study is carried out to replicate some previous experiment, then the shape of the chevrons will be known; conversely the shape of the chevrons for a newly studied protein is typically unknown. If the objective of the experimentalists includes an assessment whether the kinetic data indeed follow a linear chevron or not, we find that it is of advantage to choose a few concentrations on the folding and unfolding arms. Although there is a small price to pay in terms of the efficiency with which the kinetic parameters are defined, the use of such intermediate points greatly increases the power to detect departures from linearity. We compared three experimental designs for a synthetic nonlinear chevron (Fig. 3, lower panel). The chevron was a weighted linear combination of the linear (weight 1.0) and quadratic (weight 0.15) chevrons fits to protein U1A (Fig. 3, upper right), resulting in a slight nonlinear shape of the chevron. In addition to the traditional and previously recommended choices for the denaturant concentrations (designs 1 and 3, respectively), we also considered a design where some of the denaturant concentrations were chosen on the folding and unfolding arms (design 4). In practice, the experimentalist can first derive the observed folding rates for an equally spaced set of denaturant concentrations relatively far apart, and then augment the design points at the extremes (near 0M and saturation) as well as the inferred inflection point. For these three designs, we analyzed data from 10,000 simulated synthetic chevrons using Gaussian distributions to generate the experimental error typically seen in laboratories. The power to detect the departure from linearity was 92% for design 1 and 90% for design 4, but only 29% for design 3. Thus, if the shape of the chevrons is not known and the assessment of linearity is of concern as well, it is important to also

choose some denaturant concentrations away from the extremes (i.e., the lowest and highest possible denaturant concentrations, as well as the inflection point).

Typical Φ error analysis is not optimal

In the previous paragraphs we have demonstrated that simple experimental design considerations can significantly improve the precision of estimates of the kinetic parameters of interest in the field. We also find, however, that the methods commonly employed to make inference based on these kinetic parameters (and functions thereof, such as the Φ -value) are not correct, and can lead to invalid conclusions.

Making valid inference

Methods for the determination of standard errors on Φ -value estimates often assume that the estimates of the changes in free energy of the transition and folded states are independent. It has previously been demonstrated, however, that this assumption is usually incorrect and that, in turn, this can severely affect estimates of Φ precision.⁹ The derivation of an analytical expression for the precision of Φ estimates (assuming linear chevron behavior) that explicitly takes this dependence into account (given in the Appendix) yields:

$$\text{se}(\hat{\Phi}) = |\Phi| \times \sqrt{\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)^2 - 2\rho_{\Delta}\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)\left(\frac{\sigma_U}{\Delta\Delta G_U}\right) + \left(\frac{\sigma_U}{\Delta\Delta G_U}\right)^2} \quad (5)$$

As the changes in free energy ($\Delta\Delta G_{\ddagger}$ and $\Delta\Delta G_U$) are usually non-negative and the correlation between their estimates from the chevron is typically positive, ignoring ρ_{Δ} typically leads to an over-estimation of the above variability (though this is not necessarily so: for example, $\Phi = 0$ implies $\Delta\Delta G_{\ddagger} = 0$, and the *estimate* of this change in free energy can easily be negative). The ratio of those two variances is bounded above by $1 - \rho_{\Delta}/(1 - \rho_{\Delta})$, which is achieved when the estimates of the changes of free energy relative to their standard errors, that is the ratios $\sigma_{\ddagger}/\Delta\Delta G_{\ddagger}$ and $\sigma_U/\Delta\Delta G_U$, are equal (see Appendix). For example, if $\rho_{\Delta} = 0.5$ (a value of 0.5 and larger is quite common in chevron experiments), the variability ignoring the correlation between the estimates of the free energy changes is twice as large as necessary if $\sigma_{\ddagger}/\Delta\Delta G_{\ddagger} = \sigma_U/\Delta\Delta G_U$, and therefore is a substantial waste of information.

Of note, a recently published manuscript¹¹ recognizes that the estimates of $\Delta\Delta G_{\ddagger}$ and $\Delta\Delta G_U$ are not independent variables because both are derived using the term $\ln(k'_f) - \ln(k_f)$, and that an assumption of independence would lead to incorrect error estimates on Φ . In an

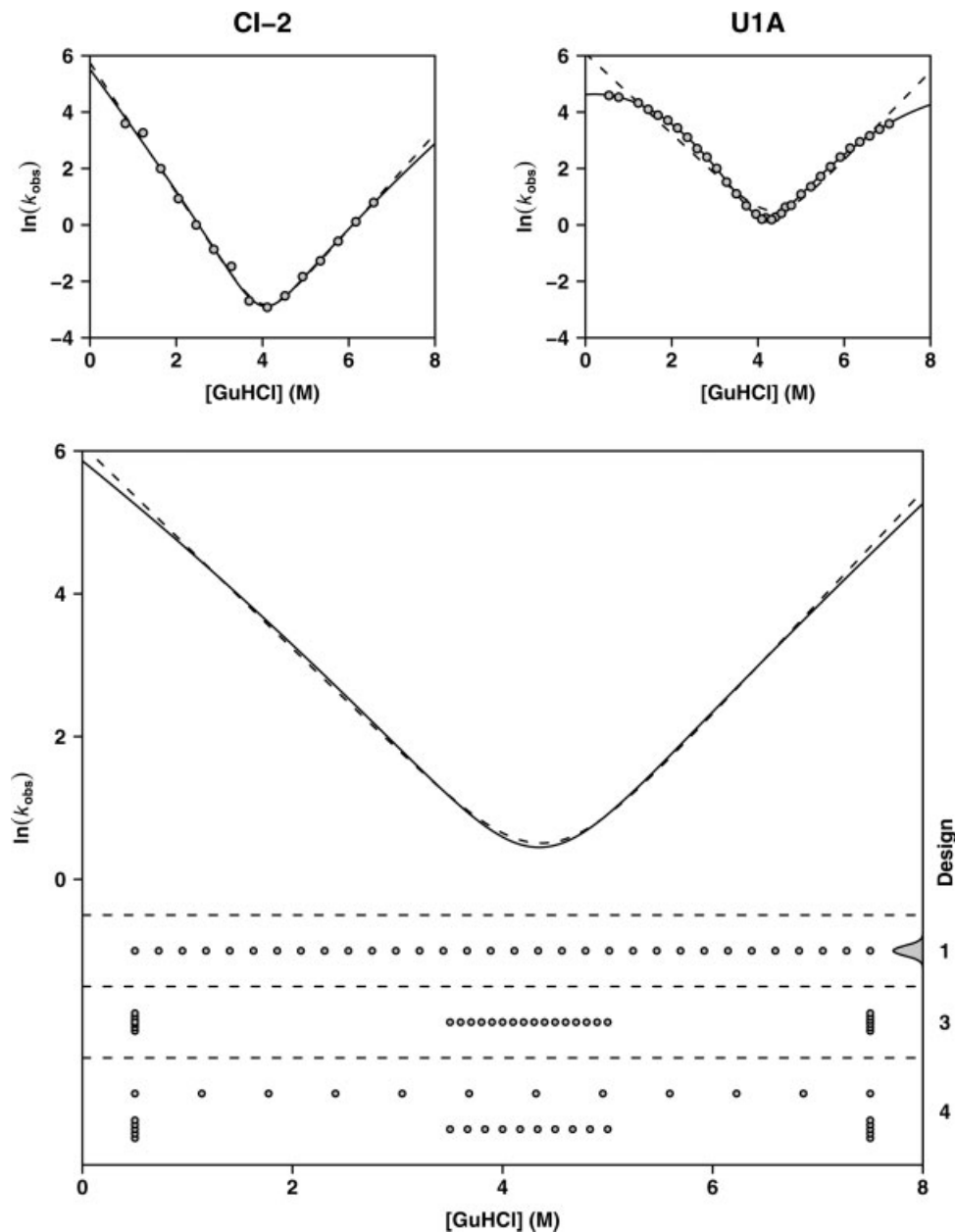


Figure 3

Upper panels: the observed kinetic data and fitted chevron curves for the proteins CI-2 and U1A (www.foldeomics.org). The linear chevron fits are indicated by dashed lines, the quadratic fits are indicated by solid lines. Lower panel: the setup for the simulation to investigate the statistical power to detect the departure from linearity. The assumed quadratic chevron, simulated as a mixture between the two curves in the upper right U1A panel, is shown as a solid line, the linear chevron is shown as a dashed line. Three experimental setups defined by the locations of the 32 denaturant concentrations are considered. Design 1 refers to the traditional design of equally spaced denaturant concentrations, design 3 refers to the previously described choice of denaturant concentrations (in this case, 8 measurements each at the extrema, and another 16 measurements are taken around the inflection point). The denaturant concentrations in design 4 are chosen in two stages: first a set of equally spaced denaturant concentrations are chosen (12 in this case), and then this set is augmented by design points at the extremes (near 0M and saturation, here 5 points each) as well as the inferred inflection point (10 points). The variability of the assumed measurement errors was chosen as a standard Gaussian distribution with mean zero and standard deviation typically seen in laboratories, indicated by the bell shaped curve next to the label for design 1.

attempt to alleviate this problem and derive an accurate estimation of the Φ error, the authors rewrote Φ as the ratio $1/(1 - b/a)$, where $a = \ln(k'_f) - \ln(k_f)$, and $b = \ln(k'_u) - \ln(k_u)$. Unfortunately, however, a and b are

still dependent as they are both obtained from estimates that are derived using the same, single set of data (each chevron yields an estimate for the folding and for the unfolding rate, and the former contributes to the term a ,

and the latter contributes to the term b). In fact, linear model theory yields that the covariance between a and b is equal to the sum of the covariances between the estimates of the folding and unfolding rates from the respective chevrons (see Appendix). The fact that this covariance was neglected in Ref. 11, unfortunately, renders their reported algorithm for error analysis incorrect.

Sato et al.¹¹ further argue that the estimates for Φ at the extremes ($\Phi = 0$ and $\Phi = 1$) should be the most precise and accurate. The standard error of the Φ estimate (which defines the precision) is, however, a function in several parameters [see Eq. (5)]. Therefore comparisons such as the above can only make sense (if at all) if all but the parameter of interest (here, Φ) are considered constant. The above equation can be re-written (see Appendix) as

$$\text{se}(\hat{\Phi}) = \left| \frac{\sigma_U}{\Delta\Delta G_U} \right| \times \sqrt{\left(\Phi - \rho_\Delta \frac{\sigma_\ddagger}{\sigma_U} \right)^2 + (1 - \rho_\Delta^2) \left(\frac{\sigma_\ddagger}{\sigma_U} \right)^2} \quad (6)$$

which contains the term $\Delta\Delta G_\ddagger$ implicitly (in Φ). Assuming that the parameters σ_\ddagger , σ_U , $\Delta\Delta G_U$, ρ_Δ are constant, the above as a function of Φ is minimized if $\Phi = \rho_\Delta \times \sigma_\ddagger/\sigma_U$, which is not equal to zero or one in general. Moreover, if the correlation between the estimates of the changes in free energy ρ_Δ is ignored (or is zero in truth), the above simplifies to

$$\text{se}(\hat{\Phi}) = \left| \frac{\sigma_U}{\Delta\Delta G_U} \right| \times \sqrt{\Phi^2 + \left(\frac{\sigma_\ddagger}{\sigma_U} \right)^2} \quad (7)$$

This term is indeed minimized if $\Phi = 0$, however *maximized* if $\Phi = 1$. Moreover, the considerations of the variabilities in the parameter estimates (and the “propagated errors”) alone do not allow for any conclusions about the accuracy (i.e. the potential bias of the estimator). In fact, in many estimation procedures there is a bias/variance trade-off: higher precision in the estimates can often be obtained at the price of incurring a bias (see for example the following chapter on estimation at non-zero denaturant concentrations).

Thus, in short, the commonly asserted notion that estimates for Φ at the extrema are the most accurate and precise is significantly misplaced.

Estimating kinetic parameters at non-zero concentrations

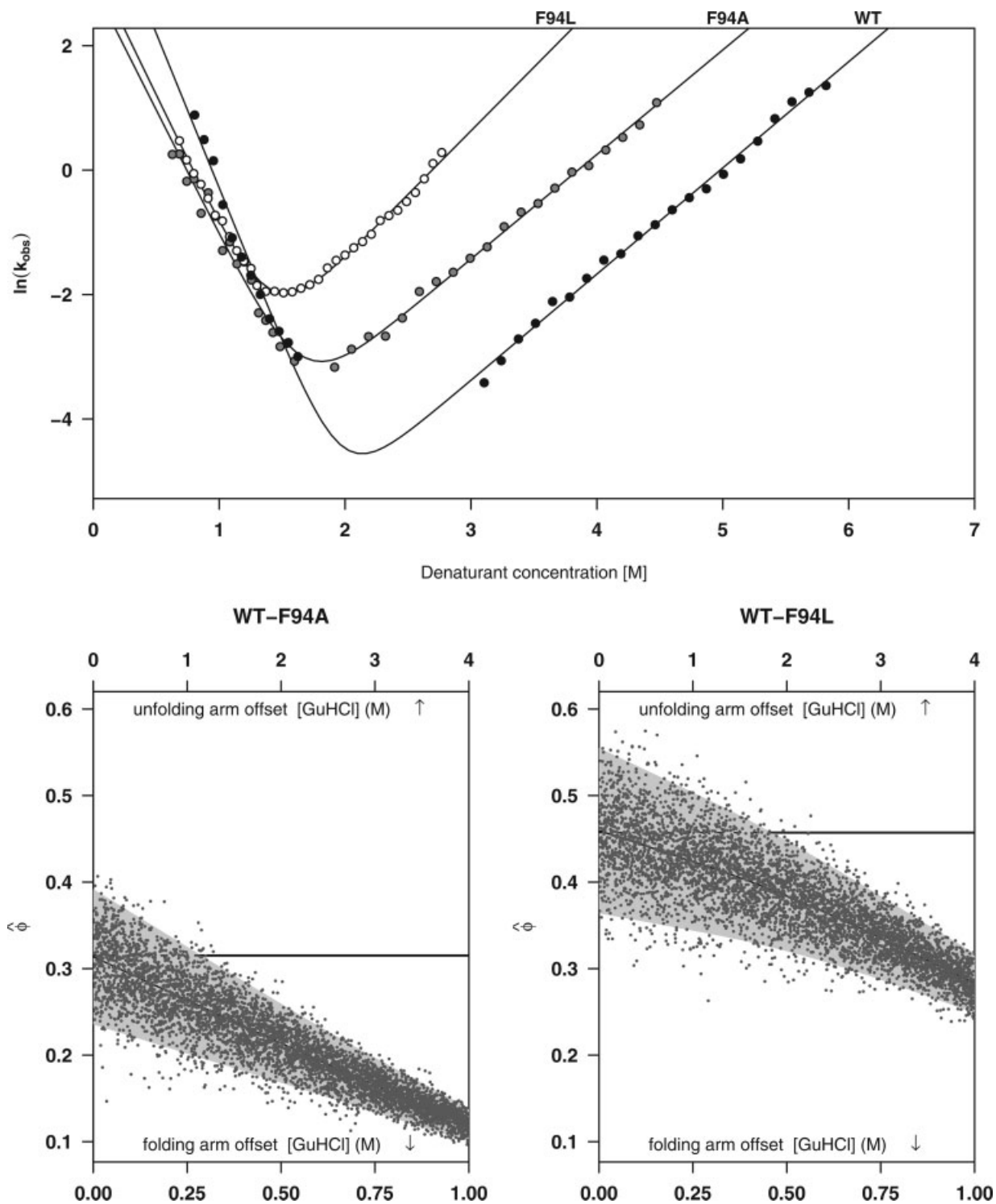
To decrease the standard error in the estimate of Φ , some scientists estimate the kinetic parameters at non-zero concentrations; that is, they do not extrapolate the chevron arms to $0M$ denaturant concentration. As was previously pointed out,⁹ this is only valid if the chevron

arms of the wild-type and variant proteins are parallel, and the data are fitted accordingly. Otherwise, this procedure results in a bias in the estimates of the kinetic parameters, whose magnitude obviously depends (among others) on the differences in the slopes of the folding and unfolding arms respectively, and on the denaturant concentration where the estimates are determined. To demonstrate this effect more clearly, we augmented the results obtained in the analysis of the kinetic data from the protein common-type acylphosphatase (ctAcP) with a simulation. Previously, the kinetic parameters were estimated at $0M/0M$ as well as $1M/4M$ denaturant concentrations for the folding/unfolding arms of the wildtype, and the F94A and F94L variant chevrons. These two approaches yielded wildly different results, in particular for the estimates for Φ .⁹ To show the trade-off between bias and variability, we simulated 5001 synthetic chevrons for the wildtype and the two variants, and derived the Φ -values at 5001 equally spaced denaturant concentration pairs between $0/0$ and $1/4 M$ (one set of chevrons was analyzed for one pair of denaturant concentration to obtain independent results). As expected, the further away from $0M$ denaturant we estimate Φ , the smaller the variability around the estimates, but also, the larger the incurred bias (Fig. 4). Thus, although the common practice of estimating kinetic parameters away from $0M$ denaturant typically results in improved precision, it does so at the cost of introducing biases, some of which can significantly degrade accuracy (see discussion later).

DISCUSSION

In this manuscript we have described simple alterations to chevron analysis, a mainstay of experimental protein folding studies, that lead to much improved precision in the key parameters of interest. Moreover, as they only involve the selection of the denaturant concentrations chosen by the experimentalist no additional resources are required and the suggested changes are readily implemented.

Our conclusions are consistent with previous studies of the effect of design point choice on the precision of the parameter estimates in linear models, in which closed form solutions for the standard errors (as functions of the predictors) have been described. One of the key results of linear model theory is the observation that placing 50% of the design points at the each of the extremes of a region of linear behavior yields, for example the best precision on estimates of the slope of the regression line. Because the two arms of a chevron plot are typically linear, the conclusions drawn from linear model theory obviously appear to hold for some aspects of chevron analysis as well. Also similar to the case of chevron curves, departure from linearity cannot be detected if the design points are placed at the extremes,

**Figure 4**

Upper panel: The kinetic data obtained on the wild type of common-type acylphosphatase protein (ctAcP) and the two variants F94L and F94A, published in Ruczinski et al.⁹ (Lower panels) Simulated Φ value estimates for the F94A (left) and F94L (right) variants of ctAcP. The Φ -values were estimated at 5001 equally spaced denaturant concentration pairs ranging from 0M for both the folding and unfolding arm to 1M for the folding and 4M for the unfolding arm. One set of chevrons was analyzed for one pair of denaturant concentration to obtain independent results. The horizontal lines indicate the values of Φ when estimated at 0M denaturant for both folding and unfolding arm. The gray areas and the lines decreasing in denaturant concentration (i.e. folding and unfolding arms “offset”) are non-parametric estimates of the mean Φ and its standard deviation, indicating the decrease in the variability of the Φ estimate and the increase in bias.

and alternative strategies have to be explored if the nature of the relationship between predictors and response is to be investigated. In this manuscript, we have proposed such a strategy for chevrons that allows for detection of nonlinear features, and introduced a statistical test based on the likelihoods of the linear and the quadratic model. The power to detect a departure from linearity obviously depends on the magnitude of the curvature in the chevron, the number of measurements taken, the experimental error, and the location of the denaturant concentrations chosen. In our analysis and simulation (based on typical laboratory settings), slight changes in the selection of the design points yielded very powerful tests to detect very small departures from linearity.

In addition to the experimental design, we also discussed several statistical issues with regard to the analysis of kinetic data. For example, various estimates derived from kinetic experiments, such as the estimates of folding and unfolding rates derived from the same chevron, the estimates of the changes of free energy comparing native and transition states, are dependent quantities, and ignoring these dependencies can yield incorrect estimates of standard errors and confidence intervals on these important parameters. We also showed that the dependencies of the estimates of the changes in free energy persisted after variable transformations, despite claims in the literature that this particular approach yields independent quantities (and subsequently assumed in the error analysis for Φ). In fact, no variable transformation for this purpose is necessary, as relatively straightforward algebra leads to a closed-form solution for the standard error of Φ (as described in detail in Ref. 9). For convenience, we have developed a web server (<http://biostat.jhsph.edu/~iruczins/software/phi/>) that accepts kinetic data as a spreadsheet and lets the user calculate rigorous confidence intervals on Φ . It is emphasized again that the variability in the estimate of Φ foremost depends on the magnitude of $\Delta\Delta G_U$ relative to its standard error, not on the absolute value of $\Delta\Delta G_U$. The debate in the literature on how large $\Delta\Delta G_U$ has to be to derive a reasonably precise estimate of Φ has been extensive,^{8,12–14} but largely missing the point (in particular, by simply collecting more kinetic data and thus increasing the number of data points, the estimates of all parameters, including Φ , can be made arbitrarily small). We have also shown that the notion of “estimates of Φ are most accurate and precise at the extreme values”¹¹ is unfounded, or even incorrect.

Finally, we have shown that although the common practice of estimating kinetic parameters away from 0M denaturant typically results in improved precision, it does so at the cost of introducing biases. Critically, at best these biases lead to difficulty in interpreting Φ and at worst they significantly degrade its accuracy. For example, if the bias arises due to the failure of the assumption of parallel folding arms the calculated Φ value will differ

from the Φ value that would be observed at zero denaturant. This, perhaps, may not be critical (if the structure of the transition state at non-zero denaturant is of interest), but it will complicate interpretation of Φ if the goal is to understand the folding transition state in the absence of perturbative co-solvents. A perhaps more damaging bias arises, however, when the assumption of parallel unfolding arms is violated. Under these circumstances the denominator in the Φ -value relationship ($\Delta\Delta G_U$) does not accurately represent the equilibrium free energy of the protein at the denaturant concentration at which the numerator ($\Delta\Delta G_{\ddagger}$) is defined, leading to a calculated Φ value that differs not only from the Φ value at zero denaturant, but that also differ from the true Φ value at the selected, non-zero denaturant concentration. Thus, if the assumption of parallel chevron arms breaks down, the common practice of estimating kinetic parameters away from 0M denaturant will lead to improved precision, but at the cost of reduced accuracy.

In summary, we hope that we have provided some useful experimental guidelines for the practitioner to improve his or her kinetic analysis, and that we clarified some misconceptions, particularly about Φ -value analysis, that exist in the literature. We also hope that this manuscript might provide a useful introduction for quantitatively oriented researchers who develop an interest in protein folding kinetics.

REFERENCES

- Levinthal C. How to fold graciously. *Proceedings: Mossbauer Spectroscopy in Biological Systems, University of Illinois Bulletin* 1969; 67:22–24.
- Gillespie B, Plaxco KW. Using protein folding rates to test protein folding theories. *Annu Rev Biochem* 2004;73(0066-4154 (Print)): 837–859.
- Plaxco KW, Dobson CM. Time-resolved biophysical methods in the study of protein folding. *Curr Opin Struct Biol* 1996;6:630–663.
- Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rios MA, Brown AG, Friel CT, Hedberg L, Horng JC, Bona D, Miller EJ, Vallee-Belisle A, Main ERG, Bemporad F, Qiu L, Teilum K, Vu ND, Edwards AM, Ruczinski I, Poulsen FM, Kragelund BB, Michnick SW, Chiti F, Bai Y, Hagen SJ, Serrano L, Oliveberg M, Raleigh DP, Wittung-Stafshede P, Radford SE, Jackson SE, Sosnick TR, Marqusee S, Davidson AR, Plaxco KW. Protein folding: defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 2005;14:602–616.
- Raleigh DP, Plaxco KW. The protein folding transition state: what are phi-values really telling us? *Protein Pept Lett* 2005;12:117–122.
- Fulton KE, Bate MA, Faux NG, Mahmood K, Betts C, Buckle AM. Protein folding database (pfd 2.0): an online environment for the international foldomics consortium. *Nucleic Acids Res* 2007; 35(Database issue):D304–D307.
- de los Rios MA, Muralidhara BK, Wildes D, Sosnick TR, Marqusee S, Wittung-Stafshede P, Plaxco KW, Ruczinski I. On the precision of experimentally determined protein folding rates and phi-values. *Protein Sci* 2006;15:553–563.
- Sanchez IE, Kiefhaber T. Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J Mol Biol* 2003; 334(5):1077–1085.

9. Ruczinski I, Sosnick TR, Plaxco KW. Methods for the accurate estimation of confidence intervals on protein folding phi-values. *Protein Sci* 2006;15:2257–2264.
10. Sanchez IE, Kiefhaber T. Non-linear rate-equilibrium free energy relationships and hammond behavior in protein folding. *Biophys Chem* 2003;100:397–407.
11. Sato S, Religa TL, Fersht AR. Phi-analysis of the folding of the b domain of protein a using multiple optical probes. *J Mol Biol* 2006;360(4):850–864.
12. Fersht AR, Sato S. Phi-Value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci USA* 2004;101:7976–7981.
13. Garcia-Mira MM, Boehringer D, Schmid FX. The folding transition state of the cold shock protein is strongly polarized. *J Mol Biol* 2004;339:555–569.
14. Settanni G, Rao F, Caffisch A. Phi-value analysis by molecular dynamics simulations of reversible folding. *Proc Natl Acad Sci USA* 2005;102(3):628–633.
15. Geyer CJ. Le cam made simple: asymptotics of maximum likelihood without the llr or clt or sample size going to infinity. Technical Report 643 (revised), School of Statistics, University of Minnesota; 2005.
16. Seber GA. Linear regression analysis. New York: Wiley; 1977.
17. Marsaglia G. Ratios of normal variables and ratios of sums of uniform variables. *J Am Statistical Assoc* 1965;60:193–204.
18. Hinkley DV. On the ratio of two correlated normal random variables. *Biometrika* 1969;56:635–639.

APPENDIX

Distributions of the Kinetic Parameters

Using least squares or maximum likelihood methods, we infer that the estimates of the kinetic parameters follow approximately a multivariate normal distribution.¹⁵ As the chevron curves for the two amino acid sequences are fit separately, kinetic parameter estimates from the same chevron fit will be correlated, while kinetic parameter estimates from different chevron fits will be independent. Mathematically, this means that

$$\mathbf{Y} = \begin{bmatrix} \ln(\widehat{k}_f) \\ \ln(\widehat{k}_u) \\ \ln(\widehat{k}'_f) \\ \ln(\widehat{k}'_u) \end{bmatrix} \sim N \left(\begin{bmatrix} \ln(k_f) \\ \ln(k_u) \\ \ln(k'_f) \\ \ln(k'_u) \end{bmatrix}, \begin{bmatrix} (\sigma_f)^2 & \rho\sigma_f\sigma_u & 0 & 0 \\ \rho\sigma_f\sigma_u & (\sigma_u)^2 & 0 & 0 \\ 0 & 0 & (\sigma'_f)^2 & \rho'\sigma'_f\sigma'_u \\ 0 & 0 & \rho'\sigma'_f\sigma'_u & (\sigma'_u)^2 \end{bmatrix} \right) \quad (\text{A1})$$

Here, $\ln(k_f)$ denotes the logarithm of the folding rate of the wild type, $\ln(k_u)$ denotes the logarithm of the unfolding rate of the wild type, and $\ln(k'_f)$ and $\ln(k'_u)$ denote the respective parameters for the mutant. The symbol σ is used for the respective standard errors, and ρ is used for the respective correlations.

An elementary result of linear model theory¹⁶ states that if \mathbf{Y} is a random vector of length n following a mul-

tivariate normal distribution, that is $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and \mathbf{C} is a matrix with n columns and p rows, then $\mathbf{CY} \sim N_p(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)$. We use this theorem to derive the joint distribution of $\ln(\widehat{k}_f) - \ln(\widehat{k}'_f)$ and $\ln(\widehat{k}_u) - \ln(\widehat{k}'_u)$, and the joint distribution of the estimates in the changes in free energy, as all of these are linear functions of the kinetic parameters ($\widehat{\Delta\Delta G}_\ddagger = RT \times [\ln(\widehat{k}_f) - \ln(\widehat{k}'_f)]$, and $\widehat{\Delta\Delta G}_U = RT \times [\ln(\widehat{k}_f) - \ln(\widehat{k}'_f) - \ln(\widehat{k}_u) + \ln(\widehat{k}'_u)]$). Choosing

$$\mathbf{C} = \begin{bmatrix} +1 & 0 & -1 & 0 \\ 0 & +1 & 0 & -1 \end{bmatrix}, \quad (\text{A2})$$

yields

$$\begin{aligned} \mathbf{CY} &= \begin{bmatrix} +1 & 0 & -1 & 0 \\ 0 & +1 & 0 & -1 \end{bmatrix} \times \begin{bmatrix} \ln(\widehat{k}_f) \\ \ln(\widehat{k}_u) \\ \ln(\widehat{k}'_f) \\ \ln(\widehat{k}'_u) \end{bmatrix} \\ &= \begin{bmatrix} \ln(\widehat{k}_f) & -\ln(\widehat{k}'_f) \\ \ln(\widehat{k}_u) & -\ln(\widehat{k}'_u) \end{bmatrix} \sim N \left(\begin{bmatrix} \ln(k_f) & -\ln(k'_f) \\ \ln(k_u) & -\ln(k'_u) \end{bmatrix}, \begin{bmatrix} (\sigma_f)^2 + (\sigma'_f)^2 & \rho\sigma_f\sigma_u + \rho'\sigma'_f\sigma'_u \\ \rho\sigma_f\sigma_u + \rho'\sigma'_f\sigma'_u & (\sigma_u)^2 + (\sigma'_u)^2 \end{bmatrix} \right) \quad (\text{A3}) \end{aligned}$$

Hence the covariance between $\ln(\widehat{k}_f) - \ln(\widehat{k}'_f)$ and $\ln(\widehat{k}_u) - \ln(\widehat{k}'_u)$ is equal to the sum of the covariances between the estimates of the folding and unfolding rates from the respective chevrons. In particular, they are not independent, as stated in Ref. 11.

The Standard Error of $\widehat{\Phi}$

To derive the distribution of $\widehat{\Delta\Delta G}_\ddagger$ and $\widehat{\Delta\Delta G}_U$, we choose

$$\mathbf{C} = RT \times \begin{bmatrix} +1 & 0 & -1 & 0 \\ +1 & -1 & -1 & +1 \end{bmatrix}, \quad (\text{A4})$$

where R is the gas constant, and T is the absolute temperature. Therefore

$$\begin{aligned} \mathbf{CY} &= RT \times \begin{bmatrix} +1 & 0 & -1 & 0 \\ +1 & -1 & -1 & +1 \end{bmatrix} \times \begin{bmatrix} \ln(\widehat{k}_f) \\ \ln(\widehat{k}_u) \\ \ln(\widehat{k}'_f) \\ \ln(\widehat{k}'_u) \end{bmatrix} \\ &= RT \times \begin{bmatrix} \ln(\widehat{k}_f) - \ln(\widehat{k}'_f) \\ \ln(\widehat{k}_f) - \ln(\widehat{k}'_f) - \ln(\widehat{k}_u) + \ln(\widehat{k}'_u) \end{bmatrix} \\ &= \begin{bmatrix} \widehat{\Delta\Delta G}_\ddagger \\ \widehat{\Delta\Delta G}_U \end{bmatrix} \sim N \left(\begin{bmatrix} \Delta\Delta G_\ddagger \\ \Delta\Delta G_U \end{bmatrix}, \begin{bmatrix} \sigma_\ddagger^2 & \rho_\Delta\sigma_\ddagger\sigma_U \\ \rho_\Delta\sigma_\ddagger\sigma_U & \sigma_U^2 \end{bmatrix} \right), \quad (\text{A5}) \end{aligned}$$

where

$$\begin{aligned}\sigma_{\ddagger}^2 &= \text{var}(\widehat{\Delta\Delta G_{\ddagger}}) \\ &= (RT)^2 \times [(\sigma_f)^2 + (\sigma'_f)^2], \sigma_U^2 = \text{var}(\widehat{\Delta\Delta G_U}) \\ &= (RT)^2 \times [(\sigma_f)^2 + (\sigma'_f)^2 \\ &\quad + (\sigma_u)^2 + (\sigma'_u)^2 - 2\rho\sigma_f\sigma_u - 2\rho\sigma'_f\sigma'_u], \\ \rho_{\Delta}\sigma_{\ddagger}\sigma_U &= \text{cov}(\widehat{\Delta\Delta G_{\ddagger}}, \widehat{\Delta\Delta G_U}) \\ &= \text{corr}(\widehat{\Delta\Delta G_{\ddagger}}, \widehat{\Delta\Delta G_U}) \times \text{se}(\widehat{\Delta\Delta G_{\ddagger}}) \times \text{se}(\widehat{\Delta\Delta G_U}) \\ &= (RT)^2 \times [(\sigma_f)^2 + (\sigma'_f)^2 - \rho\sigma_f\sigma_u - \rho\sigma'_f\sigma'_u]. \quad (\text{A6})\end{aligned}$$

Here, $\text{var}()$ denotes the variance, $\text{cov}()$ the covariance, $\text{corr}()$ the correlation, and $\text{se}()$ the standard error of the respective arguments. As the Φ -value estimate is not a linear function of the estimates of the changes in free energy, we use standard error propagation (based on a Taylor-series expansion) to derive an approximate distribution: in essence, if we have an estimator $\hat{\theta}$ for a parameter θ , following a multivariate normal distribution with mean θ and variance covariance matrix Σ , then for a non-constant and differentiable function f ,

$$f(\hat{\theta}) \approx N(f(\theta), \mathbf{V}). \quad (\text{A7})$$

where

$$\mathbf{V} = \left(\frac{\delta f}{\delta \theta}\right)^T \Sigma \left(\frac{\delta f}{\delta \theta}\right) \Big|_{\theta=\hat{\theta}}. \quad (\text{A8})$$

In this context, the above approximation holds if the absolute values of the estimates of the changes in free energy are large compared to their standard errors (see for example the discussions in Refs. 17 and 18). To use this result to derive a distribution for $\hat{\Phi}$, consider $f(\theta) = f(\Delta\Delta G_{\ddagger}, \Delta\Delta G_U) = \Delta\Delta G_{\ddagger}/\Delta\Delta G_U$. Since

$$\frac{\delta f}{\delta \Delta\Delta G_{\ddagger}} = \frac{1}{\Delta\Delta G_U} \quad \text{and} \quad \frac{\delta f}{\delta \Delta\Delta G_U} = -\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U^2}, \quad (\text{A9})$$

it follows that

$$\hat{\Phi} = \frac{\widehat{\Delta\Delta G_{\ddagger}}}{\Delta\Delta G_U} \approx N(\Phi, \text{var}(\hat{\Phi})), \quad (\text{A10})$$

and the variance of the estimate of Φ is given by

$$\begin{aligned}\text{var}(\hat{\Phi}) &= \frac{1}{(\Delta\Delta G_U)^2} [\Delta\Delta G_U, -\Delta\Delta G_{\ddagger}] \begin{bmatrix} \sigma_{\ddagger}^2 & \rho_{\Delta}\sigma_{\ddagger}\sigma_U \\ \rho_{\Delta}\sigma_{\ddagger}\sigma_U & \sigma_U^2 \end{bmatrix} \begin{bmatrix} \Delta\Delta G_U \\ -\Delta\Delta G_{\ddagger} \end{bmatrix} \frac{1}{(\Delta\Delta G_U)^2} \\ &= \frac{1}{(\Delta\Delta G_U)^4} (\sigma_{\ddagger}^2 \times (\Delta\Delta G_U)^2 - 2\rho_{\Delta}\sigma_{\ddagger}\sigma_U \times \Delta\Delta G_{\ddagger}\Delta\Delta G_U + \sigma_U^2 \times (\Delta\Delta G_{\ddagger})^2) \\ &= \left(\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U}\right)^2 \times \left[\frac{\sigma_{\ddagger}^2}{(\Delta\Delta G_{\ddagger})^2} - 2\rho_{\Delta} \frac{\sigma_{\ddagger}\sigma_U}{\Delta\Delta G_{\ddagger}\Delta\Delta G_U} + \frac{\sigma_U^2}{(\Delta\Delta G_U)^2} \right] \\ &= \Phi^2 \times \left[\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)^2 - 2\rho_{\Delta} \left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right) \left(\frac{\sigma_U}{\Delta\Delta G_U}\right) + \left(\frac{\sigma_U}{\Delta\Delta G_U}\right)^2 \right]. \quad (\text{A11})\end{aligned}$$

Effects of Ignoring the Correlation Between the Free Energy Estimates

When ρ_{Δ} is ignored (i.e. assumed to be zero), the estimate for the variability in $\hat{\Phi}$ is simply

$$\text{var}_{\rho_{\Delta}=0}(\hat{\Phi}) = \Phi^2 \times \left[\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)^2 + \left(\frac{\sigma_U}{\Delta\Delta G_U}\right)^2 \right]. \quad (\text{A12})$$

The absolute difference in those variabilities is

$$\begin{aligned}\text{var}_{\rho_{\Delta}=0}(\hat{\Phi}) - \text{var}(\hat{\Phi}) &= \Phi^2 \times 2\rho_{\Delta} \left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right) \left(\frac{\sigma_U}{\Delta\Delta G_U}\right) \\ &= 2\Phi \times \frac{\rho_{\Delta}\sigma_{\ddagger}\sigma_U}{(\Delta\Delta G_U)^2}. \quad (\text{A13})\end{aligned}$$

This formula states that the smaller ρ_{Δ} , σ_{\ddagger} , σ_U , and Φ (for a given $\Delta\Delta G_U$), the smaller the absolute difference in the estimated variabilities. Also, the larger $\Delta\Delta G_U$ (for a given Φ), the smaller that difference. To investigate the relative increase in the estimate of the variability of $\hat{\Phi}$ when ρ_{Δ} is ignored, we simplify the notation by writing

$$\begin{aligned}V_1 &= \Phi^2 (a^2 - 2\rho_{\Delta}ab + b^2), \\ V_2 &= \Phi^2 (a^2 + b^2),\end{aligned} \quad (\text{A14})$$

where $a = \sigma_{\ddagger}/\Delta\Delta G_{\ddagger}$ and $b = \sigma_U/\Delta\Delta G_U$. Using $b = c \times a$, we define

$$\frac{V_1}{V_2} = 1 - \frac{2\rho_{\Delta}ab}{a^2 + b^2} = 1 - \frac{2\rho_{\Delta}ca^2}{a^2 + c^2a^2} = 1 - \frac{2\rho_{\Delta}c}{1 + c^2} =: f(c) \quad (\text{A15})$$

The first derivative of this function f is

$$f'(c) = -\frac{2\rho_{\Delta}(1-c)}{(1+c^2)^2}, \quad (\text{A16})$$

which is equal to zero iff $c = 1$. The function f attains a global minimum at $c = 1$, that is when $\sigma_{\ddagger}/\Delta\Delta G_{\ddagger} = \sigma_U/\Delta\Delta G_U$. Hence V_1/V_2 is bounded below by $f(1) = 1 - \rho_{\Delta}$, and converges to 1 (i.e. towards equal variance estimates) as c moves away from this point in either direction. Equivalently, V_2/V_1 is bounded above by $1 -$

$\rho_{\Delta}/(1 - \rho_{\Delta})$. The lengths of the confidence intervals scales with the standard errors of the free energy estimates, and therefore the ratio of confidence intervals is bounded by $\sqrt{V_2}/\sqrt{V_1} = \sqrt{V_2/V_1} = \sqrt{1 - \rho_{\Delta}/(1 - \rho_{\Delta})}$.

Precision of the Estimate of Φ in the Extremes

To investigate the precision of the estimate of Φ as a function of Φ (and, of course, other kinetic parameters), we re-write Eq. (18) as

$$\begin{aligned} \text{se}(\hat{\Phi}) &= |\Phi| \times \sqrt{\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)^2 - 2\rho_{\Delta}\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)\left(\frac{\sigma_U}{\Delta\Delta G_U}\right) + \left(\frac{\sigma_U}{\Delta\Delta G_U}\right)^2} \\ &= \sqrt{\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)^2 \left(\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U}\right)^2 - 2\rho_{\Delta}\left(\frac{\sigma_{\ddagger}}{\Delta\Delta G_{\ddagger}}\right)\left(\frac{\sigma_U}{\Delta\Delta G_U}\right)\left(\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U}\right) + \left(\frac{\sigma_U}{\Delta\Delta G_U}\right)^2 \left(\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U}\right)^2} \\ &= \left|\frac{\sigma_U}{\Delta\Delta G_U}\right| \times \sqrt{\left(\frac{\sigma_{\ddagger}}{\sigma_U}\right)^2 - 2\rho_{\Delta}\frac{\sigma_{\ddagger}}{\sigma_U}\left(\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U}\right) + \left(\frac{\Delta\Delta G_{\ddagger}}{\Delta\Delta G_U}\right)^2} \\ &= \left|\frac{\sigma_U}{\Delta\Delta G_U}\right| \times \sqrt{\Phi^2 - 2\rho_{\Delta}\frac{\sigma_{\ddagger}}{\sigma_U}\Phi + \rho_{\Delta}^2\left(\frac{\sigma_{\ddagger}}{\sigma_U}\right)^2 - \rho_{\Delta}^2\left(\frac{\sigma_{\ddagger}}{\sigma_U}\right)^2 + \left(\frac{\sigma_{\ddagger}}{\sigma_U}\right)^2} \\ &= \left|\frac{\sigma_U}{\Delta\Delta G_U}\right| \times \sqrt{\left(\Phi - \rho_{\Delta}\frac{\sigma_{\ddagger}}{\sigma_U}\right)^2 + (1 - \rho_{\Delta}^2)\left(\frac{\sigma_{\ddagger}}{\sigma_U}\right)^2} \quad (\text{A17}) \end{aligned}$$