

---

# Methods for the accurate estimation of confidence intervals on protein folding $\phi$ -values

---

INGO RUCZINSKI,<sup>1</sup> TOBIN R. SOSNICK,<sup>2</sup> AND KEVIN W. PLAXCO<sup>3</sup>

<sup>1</sup>Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA

<sup>2</sup>Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, USA

<sup>3</sup>Department of Chemistry and Biochemistry, and Interdepartmental Program in Biomolecular Science and Engineering, University of California, Santa Barbara, California 93106, USA

(RECEIVED March 21, 2006; FINAL REVISION July 21, 2006; ACCEPTED July 25, 2006)

## Abstract

$\phi$ -Values provide an important benchmark for the comparison of experimental protein folding studies to computer simulations and theories of the folding process. Despite the growing importance of  $\phi$  measurements, however, formulas to quantify the precision with which  $\phi$  is measured have seen little significant discussion. Moreover, a commonly employed method for the determination of standard errors on  $\phi$  estimates assumes that estimates of the changes in free energy of the transition and folded states are independent. Here we demonstrate that this assumption is usually incorrect and that this typically leads to the underestimation of  $\phi$  precision. We derive an analytical expression for the precision of  $\phi$  estimates (assuming linear chevron behavior) that explicitly takes this dependence into account. We also describe an alternative method that implicitly corrects for the effect. By simulating experimental chevron data, we show that both methods accurately estimate  $\phi$  confidence intervals. We also explore the effects of the commonly employed techniques of calculating  $\phi$  from kinetics estimated at non-zero denaturant concentrations and via the assumption of parallel chevron arms. We find that these approaches can produce significantly different estimates for  $\phi$  (again, even for truly linear chevron behavior), indicating that they are not equivalent, interchangeable measures of transition state structure. Lastly, we describe a Web-based implementation of the above algorithms for general use by the protein folding community.

**Keywords:** protein folding; transition state;  $\phi$ -value; free energy; estimate; standard error

**Supplemental material:** see [www.proteinscience.org](http://www.proteinscience.org)

Amino acid substitutions often affect the stability of both a protein's native state and the transition state that defines the rate with which it forms. A commonly employed experimental measure of these effects,  $\phi$ , represents the ratio of these variation-induced changes in relative free energies. As such,  $\phi$  is thought to indicate the extent to

which a given amino acid participates in native-like interactions in the folding transition state. Since its introduction more than fifteen years ago (Garvey and Matthews 1989; Goldenberg et al. 1989; Matouschek et al. 1989),  $\phi$ -value analysis has been applied with varying levels of completeness to more than two dozen proteins and has become a benchmark experimental method for characterizing folding transition states (Daggett and Fersht 2003). Recently, significant controversy has emerged over the precision that can be assigned to measures of this important experimental parameter (Sanchez and Kiefhaber 2003; Fersht and Sato 2004; Garcia-Mira et al. 2004; Settanni

---

Reprint requests to: Ingo Ruczinski, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA; e-mail: [iruczins@jhsph.edu](mailto:iruczins@jhsph.edu); fax: (410) 955-0958.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062230106>.

et al. 2005). For example, we have previously used blind, replicate studies performed in three independent laboratories to ascertain the precision with which  $\phi$  can be estimated, and found that the reproducibility of  $\phi$  under these circumstances is significantly poorer than is generally claimed in the literature (de los Rios et al. 2006).

As we have previously noted, the discrepancy between our observations and earlier reports of  $\phi$  precision may stem in part from widely differing and perhaps inappropriate methods for estimating confidence intervals for this parameter. Although many authors report “error bars” for their estimates of  $\phi$ , rarely, if ever, do they describe the error analysis in sufficient detail to allow for reproduction (see discussion in de los Rios et al. 2006). Moreover, even assuming that proper error propagation is employed (Zarrine-Afsar and Davidson 2004), typical propagation methods assume independent errors in the estimates of the change in the relative free energy of the folding transition state ( $\Delta\Delta G_{\ddagger}$ ) and the native state ( $\Delta\Delta G_U$ ) (for example, see Fersht 1999; Sanchez and Kiefhaber 2003), an assumption that has not previously been tested in the literature. Critically, the failure of this assumption could significantly affect the accuracy of confidence intervals estimated via this or other, similar approaches.

Here we explore the validity of the above-described independence assumption and the impact that this assumption has on  $\phi$ -value precision estimates. We do so via numerical simulations based on authentic experimental data collected from more than a dozen different laboratories. We derive an analytical solution to the estimation of  $\phi$ -value confidence intervals that explicitly corrects for correlations between the estimates of  $\Delta\Delta G_{\ddagger}$  and  $\Delta\Delta G_U$ , and also describe an alternative analysis approach that corrects for this effect implicitly. We find that both methods produce accurate estimates of the precision

when validated against numerical simulations. We also discuss the effects of the commonly employed techniques of calculating  $\phi$  from folding/unfolding rates estimated at non-zero denaturant concentrations, and via the assumption of parallel chevron arms. Lastly, we describe and publish a Web-based implementation of these algorithms for use by the general folding community.

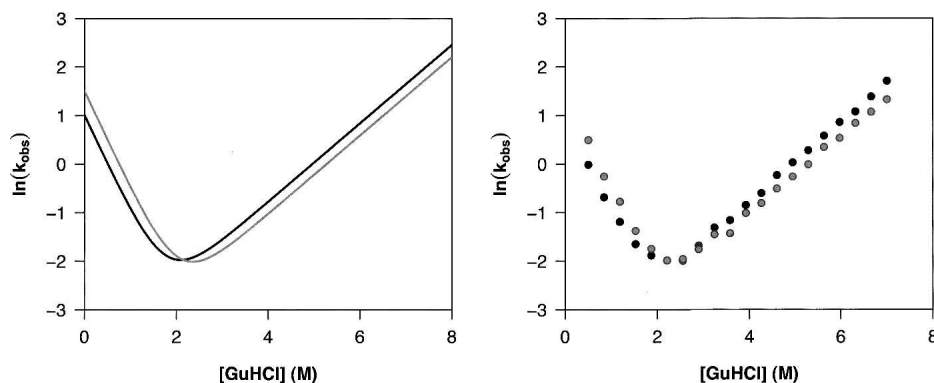
## Results

### *The independence assumption is false*

The parameter  $\phi$  is given by the ratio of the variation-induced change in the stability of the native and transition states:  $\phi = \Delta\Delta G_{\ddagger}/\Delta\Delta G_U$  (where the former is the free energy difference between the transition state and the unfolded state, and the latter is the free energy difference between the native and the unfolded state). Because both the numerator and denominator in the  $\phi$ -value expression are functions of common kinetic parameters [such as the log of the folding and unfolding rates,  $\ln(k_f)$  and  $\ln(k_u)$ ] estimated from the same two chevron plots (plots of experimentally observed log relaxation rates versus denaturant concentration; see Fig. 1), the estimates of these two changes in relative free energy are correlated; across a data set of 24 two-state chevron curves collected in three laboratories (de los Rios et al. 2006), we observe correlations as great as 0.83 (mean correlation 0.30), with all correlations being positive.

### *Effect of the correlation on confidence intervals*

To illustrate the point that the assumption of independence may lead to substantial differences in the confidence intervals, we have simulated data for a typical



**Figure 1.** In order to test our proposed methods for the determination of confidence intervals on  $\phi$ , we have generated a data set of 10,000 simulated experimental chevron curves by adding noise to points on two ideal curves. (*Left*) For these ideal curves, the log folding (unfolding) rates of the wild type and variant are 1 and 1.5 (−4 and −4.25), respectively, which yields  $\phi = 2/3$ . (*Right*) An example of the simulated experimental chevron curves obtained by adding random Gaussian noise of a magnitude typical of that reported in the literature (e.g., Maxwell et al. 2005).

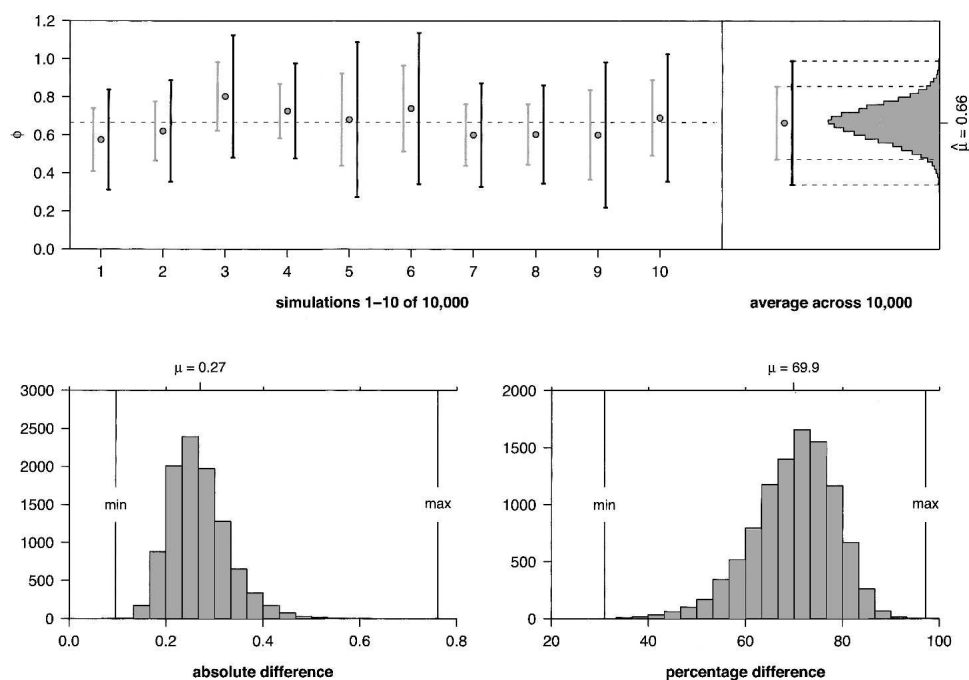
kinetic experiment. Using “synthetic” chevron curves for both wild-type (Fig. 1, black) and variant (Fig. 1, gray) proteins, we simulated data sets by adding random Gaussian noise with a mean of zero and a standard deviation of 0.05 (which is a typical root-mean-squared error observed in actual chevron experiments; see Figure 2 in de los Rios et al. 2006; data also available at <http://www.foldeomics.org>), at equally spaced denaturant concentrations. We simulated a total of 10,000 such independent data sets (see Fig. 1, for one realization of such a simulation) and used this large data set of “replicate” chevrons to calculate confidence intervals on  $\phi$  with and without taking the correlation between the free energy estimates into account. For each pair of chevrons (e.g., for the wild-type and variant protein), the kinetic parameters were fit (allowing for different  $m$ -values in the two chevrons), and estimates of  $\phi$  and a pair of confidence intervals were obtained. In this simulation we find that the confidence interval derived by taking the dependence of the estimates in free energy changes into account is *always smaller* than the one derived using the “traditional” error propagation method (the confidence inter-

vals for the first 10 simulations are shown in Figure 2; all confidence intervals in this manuscript are at the 95% level). For this synthetic data set, the average difference in confidence interval lengths is 0.27, which means that the incorrectly estimated confidence intervals are, on average, 70% larger than the true confidence intervals (Fig. 2).

#### An analytical solution

As the specific details of the analytical solution for the  $\phi$ -value confidence intervals is not, presumably, of great interest to the general practitioner, here we provide only a brief outline and state the key results. The complete details of the relevant derivation are provided in the Supplemental Materials.

Given the experimental error typically observed in the chevron plots, we infer that the parameter estimates obtained in the chevron curve fits follow (approximately) a multivariate normal distribution. Some standard linear model theory then immediately leads to the joint distribution of the estimates for the changes in free energy (for



**Figure 2.** The data from each of the 10,000 simulated experimental chevron curves were used to estimate  $\phi$  along with a pair of 95% confidence intervals, assuming that  $\Delta\Delta G_{\ddagger}$  and  $\Delta\Delta G_U$  are either correlated or uncorrelated. (Top) shown are the resulting pairs of confidence intervals for the first ten simulations. Accounting for the correlation between  $\Delta\Delta G_{\ddagger}$  and  $\Delta\Delta G_U$  estimates yielded shorter confidence intervals (light gray bar) than ignoring the correlation (dark gray bar). The estimated  $\phi$  is indicated by the dot between the respective confidence intervals. The average confidence intervals are shown on the right, centered around the mean estimated  $\phi$  ( $\mu = 0.66$ ). The distribution of the  $\phi$  estimates is shown as a histogram on the right side. Comparison between the confidence intervals derived with and without taking the correlation into account shows that the latter invariably produces a larger confidence interval for this data set (absolute differences between the confidence intervals derived without and with taking the correlation into account, bottom left; relative differences, bottom right). Indicated are the mean, minimum, and maximum differences between the interval pairs.

example, see Seber 1977). A Taylor expansion around the true value of  $\phi$  then gives an approximation for the distribution of the estimate of  $\phi$ . The latter is valid when the estimate of the change in free energy in the folded state is large compared to its standard error (a good rule of thumb is about three times larger; for example, see Marsaglia 1965; Hinkley 1969). The standard error for the estimate of  $\phi$  is then given by

$$\text{se}(\hat{\phi}) = |\Phi| \times \sqrt{\left(\frac{\sigma_f}{\Delta\Delta G_f}\right)^2 - 2\rho_{\Delta\Delta G} \left(\frac{\sigma_f}{\Delta\Delta G_f}\right) \left(\frac{\sigma_u}{\Delta\Delta G_u}\right) + \left(\frac{\sigma_u}{\Delta\Delta G_u}\right)^2} \quad (1)$$

where  $\sigma_f$  and  $\sigma_u$  denote the respective standard errors of the estimates of  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$ , and  $\rho_{\Delta\Delta G}$  denotes the correlation between the estimates of  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$ . The above is, in principle, standard error propagation for a function in two terms. In the relevant published literature, however, the cross term that accounts for the dependence between the arguments is often omitted (for example, see Fersht 1999). Relying on such an approach for error propagation leads to a formulation for the standard error that is lacking the middle term in the square root above (for example, see Sanchez and Kiefhaber 2003). As seen above, however, omitting this term can (but does not necessarily have to) lead to significantly inflated confidence intervals. The correlation between  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$  is given by

$$\begin{aligned} \rho_{\Delta\Delta G} &= \text{corr}(\Delta\Delta G_f, \Delta\Delta G_u) \\ &= \frac{(\sigma_f)^2 + (\sigma_u^*)^2 - \rho\sigma_f\sigma_u - \rho^*\sigma_f^*\sigma_u^*}{\sigma_f\sigma_u} \quad (2) \end{aligned}$$

where  $\sigma_f$  and  $\sigma_u$  denote the standard errors of the estimates of the folding and unfolding rates of the wild type,  $\rho$  denotes the correlation between the estimates of the folding and unfolding rate in the wild type, and the asterisks refer to the values associated with the variant protein. It is noteworthy that there are two types of correlations in Equation 2 that have traditionally been ignored in  $\phi$  error analysis: the correlations between the estimates of the folding and unfolding rates [ $\ln(k_f)$  and  $\ln(k_u)$ ] in the two chevron curves, and the correlation between the changes in the free energies  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$  (with the former correlation contributing to the latter correlation). For example, in the data set of 24 two-state chevron curves collected in three laboratories (de los Rios et al. 2006), the observed correlations between  $\ln(k_f)$  and  $\ln(k_u)$  range from 0.30 to 0.56 (mean 0.38). A positive correlation between the estimates of  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$

means that errors in the numerator and the denominator tend to cancel one another, thus leading to a decrease in the size of the confidence intervals on  $\phi$ . It appears worth noting that, in principle, the correlation between  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$  in Equation 2 could be negative and thus, in some circumstances, ignoring this term could instead lead to *underestimated* confidence intervals. In particular, this possibility could result from longer extrapolations required in order to estimate  $\ln(k_u)$  at 0 M denaturant, or if many more data points are available on the folding arms compared to the unfolding arms, leading to lower standard errors for the folding rates. The latter possibility is not relevant to the data sets we have employed in this study. Further discussion of the effect of ignoring  $\rho_{\Delta\Delta G}$  on the absolute and relative differences between the confidence intervals is given in the Supplemental Materials.

The analytical solution described above correctly estimates  $\phi$ -confidence intervals: when this approach is applied to the 10,000 independently generated data sets described above, we find that, as expected, 95.8% of the calculated “95% confidence intervals” encompass the true  $\phi$ . In contrast, ignoring the correlation between  $\Delta\Delta G_f$  and  $\Delta\Delta G_u$  generates inappropriately large confidence intervals: under these circumstances 99.8% of the putative “95%” confidence intervals encompass the true  $\phi$ . While this discrepancy may appear small, it is very significant; when we equivalently estimate  $\pm 1$  standard error intervals (which should correspond to  $\sim 67\%$  confidence intervals) we find that, whereas 68% of the confidence intervals cover the true  $\phi$  when we take the correlation into account, 90% of the intervals encompass  $\phi$  when we ignore this correlation (i.e., the putative 67% confidence interval is, in fact, a 90% confidence interval).

We must also stress that the rule of thumb “95% confidence interval is estimate plus minus two standard errors” can be overly optimistic. If the number of observed rates obtained in the experiments is small, the use of a t-quantile (which adjusts the “rule of thumb” value of 2 by the number of degrees of freedom available) is more appropriate than the factor of 2 usually employed (which itself is the rounded version of the normal quantile 1.96). We find that, for our simulations, nominal coverage is obtained by setting the degrees of freedom equal to  $n - 10$ , where  $n$  is the total number of data points in the fit (from both chevron curves), and 10 is the total number of parameters fitted (four kinetic parameters plus the variance for the error in each of the two chevron plots).

#### *An alternative analysis method*

The analytical approach described above (in which the two chevron curves are fit separately and the confidence

interval is calculated from the relevant parameter estimates) explicitly accounts for the correlation between  $\Delta\Delta G_{\ddagger}$  and  $\Delta\Delta G_U$ . This approach, however, presents two potential drawbacks to the practitioner. First, the relevant formulas are somewhat complicated and their use can be tedious. Second, many commonly employed commercial software packages do not return the requisite parameters. In particular, an estimate for the correlation between the folding and the unfolding rates, which is necessary to obtain valid standard errors for the estimate of  $\phi$ , is often omitted (which might also be an explanation why formulas that incorrectly lack this term are apparently widely employed).

The inconvenience of the above implementation can be avoided by the parameterization of a single model where both chevron curves are fit simultaneously, with  $\phi$  being one of the fit parameters. To fit the data from wild type and variant simultaneously, one has to create a Boolean indicator variable  $I_{VP}$ , equal to zero or one if an observation is from the wild-type or variant protein, respectively. Then one can fit the following model (which includes the denaturant concentration and the indicator variable as predictors) to the raw chevron data:

$$\begin{aligned} \ln(k_{obs}) &= \ln \left[ \exp \left( \ln(k_f) + \Delta_1 I_{VP} + (m_f + \Delta_2 I_{VP}) \frac{[denaturant]}{RT} \right) \right. \\ &\quad \left. + \exp \left( \ln(k_u) + \Delta_3 I_{VP} + (m_u + \Delta_4 I_{VP}) \frac{[denaturant]}{RT} \right) \right] \end{aligned} \quad (3)$$

In the above parameterization  $\Delta_1$  and  $\Delta_3$  denote the differences in the logarithms of the folding and unfolding rates respectively between the wild-type and variant proteins. Equivalently,  $\Delta_2$  and  $\Delta_4$  denote, respectively, the differences in the logarithms of the folding and unfolding  $m$ -values between the wild-type and variant proteins. The fact that  $\phi = \Delta_1/(\Delta_1 - \Delta_3)$ , i.e., that  $\Delta_3 = -\Delta_1(1 - \phi)/\phi$ , leads to the model

$$\begin{aligned} \ln(k_{obs}) &= \ln \left[ \exp \left( \ln(k_f) + \Delta_1 I_{VP} + (m_f + \Delta_2 I_{VP}) \frac{[denaturant]}{RT} \right) \right. \\ &\quad \left. + \exp \left( \ln(k_u) - \Delta_1 I_{VP} \frac{1 - \phi}{\phi} + (m_u + \Delta_4 I_{VP}) \frac{[denaturant]}{RT} \right) \right] \end{aligned} \quad (4)$$

In case  $\phi$  is close to zero, however, the above parameterization might lead to numerical instability. The following, equivalent model can be used in all instances when  $\phi$  is not close to one:

$$\begin{aligned} \ln(k_{obs}) &= \ln \left[ \exp \left( \ln(k_f) - \Delta_3 I_{VP} \frac{\phi}{1 - \phi} + (m_f + \Delta_2 I_{VP}) \frac{[denaturant]}{RT} \right) \right. \\ &\quad \left. + \exp \left( \ln(k_u) + \Delta_3 I_{VP} + (m_u + \Delta_4 I_{VP}) \frac{[denaturant]}{RT} \right) \right] \end{aligned} \quad (5)$$

The above models can generally be fit with any statistical software package that can also fit chevron curves, and the output should return the standard error for  $\phi$ . This standard error is usually derived numerically: the optimization procedure finds the global optimum of the function implemented, i.e., the highest or lowest peak in the “landscape” on the parameter space, and the standard errors are derived by investigating the curvature of said landscape around the optimum.

The underlying assumptions differ slightly from the first described approach. For example, fitting two separate chevron curves allows for experimental errors (spread of the data around the chevron curve) of different magnitude (i.e., standard deviation), while the latter approach fits only one parameter for the magnitude. Given, however, that these data are usually collected by a single practitioner on a single device, this assumption may be reasonable (and, if the assumption is actually met, a degree of freedom for the error estimation is gained).

Using the aforementioned data (de los Rios et al. 2006), we find that the estimates for  $\phi$  and their respective standard errors, as expected, were virtually identical to those obtained using our analytical solution (with correlations  $>0.999$ ). We note, however, that if the experimental error on the two chevron curves differs, the standard errors estimated by this approach can significantly deviate from the true precision. This approach also suffers from a potential inconvenience: the estimate of the precision, for example, on a specific kinetic parameter (such as the wild-type  $m_f$ ) can vary depending on which variant is employed in the above equation. Further, this implementation does not explicitly estimate standard errors for many parameters that may be of interest to the practitioner, such as  $\Delta\Delta G_U$ , without re-parameterization. These issues can be avoided if one employs the explicit approach described above (and, as described below, available online).

#### *Error estimation at non-zero denaturant concentrations or with parallel chevron arms*

The above implementations estimate  $\phi$  using the estimated folding and unfolding rates expected in the absence of denaturant. Because these values represent long extrapolations from the observed data (which are necessarily collected at non-zero denaturant concentrations)

the errors associated with them can be rather large. In an effort to reduce the impact of such errors on  $\phi$  precision, many authors employ folding and unfolding rates estimated at non-zero denaturant concentrations. This referencing is readily implemented in the above algorithms by substituting the term “([denaturant] –  $c$ )” for the term “[denaturant]” in the respective formulas, where  $c$  represents the denaturant concentration at which the folding or unfolding rates are defined. When we employ this approach we find that, indeed, confidence intervals can be reduced significantly. For example, using experimental data collected for several variations of the protein common-type acylphosphatase (ctAcP), we find that the breadth of the confidence intervals can be more than three times greater if extrapolations to zero denaturant are employed, compared to calculations employing folding and unfolding rates estimated at 1 M and 4 M denaturant (Fig. 3).

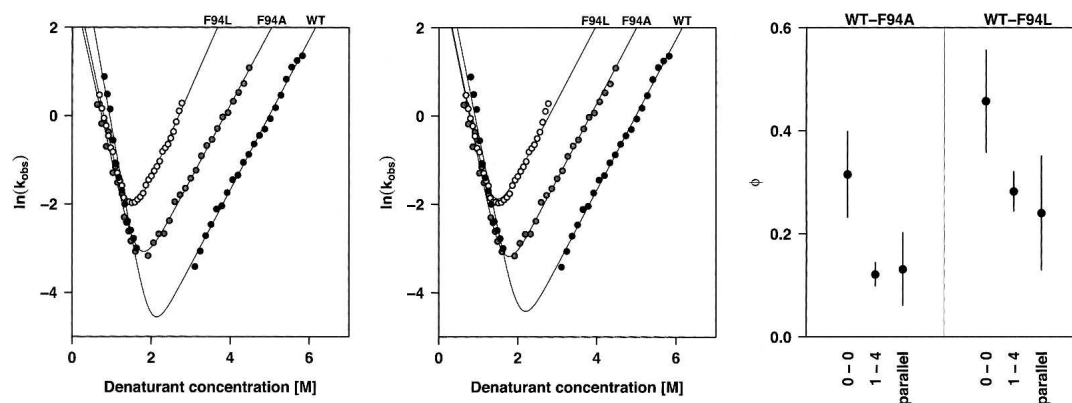
An important cautionary note, however, is in order with regard to this approach. Namely,  $\phi$ -values estimated at non-zero denaturant values are not equivalent to  $\phi$ -values obtained in the absence of denaturant. For example, the  $\phi$ -values we derive for ctAcP using folding and unfolding rates obtained at 1 M and 4 M denaturant respectively are statistically significantly different than those derived via extrapolation to zero denaturant (Fig. 3, right). Thus, while this analysis method may improve the confidence

intervals on  $\phi$ , it does so at the expense of arbitrarily introducing a potentially significant bias into  $\phi$  analysis.

Alternatively, the precision on  $\phi$  estimates can be improved by assuming that the  $m$ -values of the wild-type and variant proteins are identical (i.e., that the relevant chevron arms are parallel). This identity is achieved with the above algorithms simply by fitting a single parameter for the slope of the folding arms and a second, single parameter for the slope of the unfolding arms. Under these constraints, the same value of  $\phi$  will be returned at any denaturant concentration. Unfortunately, however, the assumption that  $m$  is invariant upon variation may be unwarranted. Indeed, for several variants in the protein ctAcP we find that the estimate of  $\phi$  produced via this assumption of parallel chevron arms differs significantly from the estimate produced in the absence of such a constraint and thus this approach can also introduce potentially significant bias into  $\phi$  analysis (Fig. 3). It also appears that in this example the true chevron arms are not parallel, and as a result, the precision is not improved.

#### A Web-based implementation

We have implemented the above calculations in scripts that run from a Web site, accessible through the software section of the first author's home page (<http://biostat.jhsph.edu/~iruczins/software/phi/>). Chevron data can be uploaded



**Figure 3.** Because estimates of the folding and unfolding rate in the absence of denaturant require long extrapolations from the observed data (necessarily collected at non-zero denaturant concentrations), the errors associated with such estimates are often rather large. In an effort to reduce the impact of such errors on  $\phi$  precision, many authors estimate  $\phi$  using the estimated folding and unfolding rates at non-zero denaturant concentrations. This approach can significantly reduce the confidence intervals on  $\phi$ . For example, using data from the positions in the protein ctAcP (A. Pandit and T.R. Sosnick, unpubl.), we find that the 95% confidence intervals and standard errors on  $\phi$  are about three times larger when estimated in the absence of denaturant than when estimated using folding and unfolding rates estimated at 1 M and 4 M denaturant, respectively. However, this comes at a price: the  $\phi$ -value estimates so produced differ significantly from the estimate produced in the absence of these constraints. (Left) The kinetic data obtained on the wild-type ctAcP and two variants (F94L, F94A), and the individual chevron fits. (Center) The kinetic data and the chevron fits, assuming parallel chevron arms. (Right) Estimates and confidence intervals for the  $\phi$ -values estimated in the absence of denaturant (0–0), using folding and unfolding rates estimated at 1 M and 4 M denaturant (1–4), and assuming parallel chevron arms (parallel). Thus while the latter approaches can improve the precision with which  $\phi$  is reported, they do so at the cost of introducing significant and difficult to interpret bias into  $\phi$ -value analysis.

as a spreadsheet or as tab-delimited text file, and the simultaneous analysis of any number of variants is possible. Several user options are available, including the possibility of fitting parallel chevron arms, and the option to measure the folding and unfolding rates at non-zero denaturant concentrations. Once executed, the script creates a Web page with tabulations of the relevant kinetic parameters and their standard errors, as well as estimates of the changes in free energy between variants, and the estimate for  $\phi$  (including its standard error and confidence interval). These tables can also be downloaded in spreadsheet format. In order to ensure confidentiality, the uploaded data are automatically deleted after the calculations are completed.

## Discussion

The large majority of published  $\phi$ -value papers lack detailed descriptions of how the reported confidence intervals on  $\phi$  were derived, or fail to report any measure of uncertainty at all. In addition, when this information is given, it appears that often the methods employed were suboptimal, as they do not take into account important correlations between the changes in the free energies of the native and transition states. In this manuscript, we presented two rigorous statistical approaches to obtaining accurate standard errors and confidence intervals for  $\phi$  derived using kinetic data obtained from fitted chevron curves. Assuming that the proteins exhibit truly linear chevron curves, our approaches explicitly or implicitly correct for the relevant correlations between the various fitted parameters. As these calculations are somewhat tedious, we have also implemented a Web-based server to perform these analysis methods for the general community.

The results and discussion presented here highlight a number of important issues that have seen relatively little discussion in the literature. For example, it does not appear to be widely appreciated that the precision of  $\phi$  (Equation 1) depends on the magnitude of the free energy changes of the transition and folded states relative to their standard errors. As previously noted (de los Rios et al. 2006), this emphasizes that claims of  $\Delta\Delta G_U$  cutoffs above which  $\phi$  can be estimated precisely are misplaced; while the absolute value of  $\Delta\Delta G_U$  factors into the estimate for the standard error  $\phi$ , other factors are also critical. Conversely, it is possible (in theory) to obtain any desired precision for the estimate of  $\phi$ , even if  $\Delta\Delta G_U$  is arbitrarily low (but not zero, i.e., no change in free energy between the folded structures): the standard errors of the changes in free energy can be made arbitrarily small by increasing the number of measurements on the chevron curves, and therefore, the standard error as given (Equation 1) can be made arbitrarily small (de los Rios et al. 2006). Whether this is feasible in practice however, appears less likely

due to limitations in time and resources as well as other constraints.

A second concern regards the effects of employing different fitting procedures for the estimation of  $\phi$ . As we have shown, the three approaches widely employed in the literature (including the determination of the kinetic parameters at denaturant concentrations other than zero, and the assumption of parallel chevron arms) sometimes produce wildly differing estimates (Fig. 3). Which of these nonequivalent estimates is the most appropriate for comparison to theory or simulation is unclear.

Finally, a distinction needs to be made between precision and accuracy. Whereas precision means low variability, accuracy refers to a lack of bias. The calculations presented in this manuscript only address the issue of precision; while the standard error of the estimate of  $\phi$  can be made arbitrarily small, this statistic is derived only from the experimental error around the given chevron curves. In order to obtain estimates that are not only precise but also accurate, utmost care must be employed to avoid systematic experimental biases (although they are always present, if one looks closely enough). Due to inevitably differing experimental conditions, however, different laboratories might obtain different estimates. This behavior has been observed in many settings and fields, including physics (Youden 1972), biology (Irizarry et al. 2005), and even in  $\phi$ -value analysis (de los Rios et al. 2006). Hopefully, though, this variance component is small (at least, in comparison to the variability due to the unavoidable measurement error). It is important to keep in mind that the magnitude of this variance component cannot be established from a single experiment, not even (necessarily) from replicate experiments in the same laboratory. Reproducibility, by definition, means that another scientist can repeat the experiment and arrive at the same conclusions.

## Acknowledgments

This work is supported in part by grants from the NIH (T.R.S. GM55694, K.W.P. GM62868-01A2). I.R. was supported by a faculty innovation award from the Johns Hopkins University.

## References

- Daggett, V. and Fersht, A. 2003. The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* **4**: 497–502.
- de los Rios, M.A., Muralidhara, B.K., Wildes, D., Sosnick, T.R., Marqusee, S., Wittung-Stafshede, P., Plaxco, K.W., and Ruczinski, I. 2006. On the precision of experimentally determined protein folding rates and  $\phi$ -values. *Protein Sci.* **15**: 553–563.
- Fersht, A. 1999. *Structure and mechanism in protein science*. p. 211. W.H. Freeman, New York.
- Fersht, A. and Sato, S. 2004.  $\phi$ -Value analysis and the nature of protein folding transition states. *Proc. Natl. Acad. Sci.* **101**: 7976–7981.
- Garcia-Mira, M.M., Bohringer, D., and Schmid, F.X. 2004. The folding transition of the cold shock protein is strongly polarized. *J. Mol. Biol.* **339**: 555–569.

- Garvey, E.P. and Matthews, C.R. 1989. Effects of multiple replacements at a single position on the folding and stability of dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **28**: 2083–2093.
- Goldenberg, D.P., Frieden, R.W., Haack, J.A., and Morrison, T.B. 1989. Mutational analysis of a protein-folding pathway. *Nature* **338**: 127–132.
- Hinkley, D.V. 1969. On the ratio of two correlated normal random variables. *Biometrika* **56**: 635–639.
- Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G., et al. 2005. Multiple-laboratory comparison of microarray platforms. *Nat. Methods* **2**: 329–330.
- Marsaglia, G. 1965. Ratios of normal variables and ratios of sums of uniform variables. *J. Am. Stat. Assoc.* **60**: 193–204.
- Matouschek, A., Kellis, J.T., Serrano, L., and Fersht, A. 1989. Mapping the transition-state and pathway of protein folding by protein engineering. *Nature* **340**: 122–126.
- Maxwell, K.L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M.A., Brown, A.G., Friel, C.T., Hedberg, L., Horng, J.C., Bona, D., Miller, E.J., et al. 2005. Protein folding: Defining a standard set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**: 602–616.
- Sanchez, I.E. and Kiefhaber, T. 2003. Origin of unusual  $\phi$ -values in protein folding: Evidence against specific nucleation sites. *J. Mol. Biol.* **334**: 1077–1085.
- Seber, G.A. 1977. *Linear regression analysis*. Wiley, New York.
- Settanni, G., Rao, F., and Caflish, A. 2005.  $\phi$ -Value analysis by molecular dynamics simulations of reversible folding. *Proc. Natl. Acad. Sci.* **102**: 628–633.
- Youden, W. 1972. Enduring values. *Technometrics* **14**: 1–11.
- Zarrine-Afsar, A. and Davidson, A.R. 2004. The analysis of protein folding kinetic data produced in protein engineering experiments. *Methods Enzymol.* **34**: 41–50.