FOR THE RECORD

# BPPred: A Web-based computational tool for predicting biophysical parameters of proteins

CHRISTIAN D. GEIERHAAS,[1,2] ADRIAN A. NICKSON,[1,2] KRESTEN LINDORFF-LARSEN,[2] JANE CLARKE,[1,2] AND MICHELE VENDRUSCOLO[2]

[1]MRC Centre for Protein Engineering, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom
[2]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

(RECEIVED June 5, 2006; FINAL REVISION September 17, 2006; ACCEPTED September 18, 2006)

**Abstract**

We exploit the availability of recent experimental data on a variety of proteins to develop a Web-based prediction algorithm (BPPred) to calculate several biophysical parameters commonly used to describe the folding process. These parameters include the equilibrium *m*-values, the length of proteins, and the changes upon unfolding in the solvent-accessible surface area, in the heat capacity, and in the radius of gyration. We also show that the knowledge of any one of these quantities allows an estimate of the others to be obtained, and describe the confidence limits with which these estimations can be made. Furthermore, we discuss how the kinetic *m*-values, or the Beta Tanford values, may provide an estimate of the solvent-accessible surface area and the radius of gyration of the transition state for protein folding. Taken together, these results suggest that BPPred should represent a valuable tool for interpreting experimental measurements, as well as the results of molecular dynamics simulations.

**Keywords:** protein denaturation; urea; guanidine hydrochloride; guanidinium chloride; protein folding; *m*-values; SASA; radius of gyration; heat capacity; transition state; unfolded state; denatured state

The possibility of interpreting quantities readily measurable experimentally in terms of descriptors of protein strucure has contributed very significantly to our understanding of the folding process. In a seminal work, Myers et al. (1995) considered an earlier suggestion by Schellman (1978) and showed that the change in solvent-accessible surface area ($\Delta SASA$) upon unfolding is related linearly to the experimental $m_{\mathrm{D-N}}$-value (Pace 1986), which describes how the stability $\triangle G_{\mathrm{D-N}}$ of the native state of a protein decreases linearly with the concentration of denaturant (Tanford 1968, 1970):

$$\Delta G_{\mathrm{D-N}} = \Delta G^{\mathrm{water}}_{\mathrm{D-N}} - m_{\mathrm{D-N}}[\mathrm{denaturant}] \qquad (1)$$

The relationship between *m*-values and $\Delta SASA$ is extremely useful because it gives important insights into the determinants of protein stability and the equilibrium properties of proteins. In order to establish such a relationship, however, one needs an estimate of the value of *SASA* of the unfolded state. Despite recent advances (Mok et al. 2005), it is still very challenging to measure *SASA* directly in the denatured state. Myers et al. (1995) derived it from a tripeptide model (Shrake and Rupley 1973; Rose

Reprint requests to: Michele Vendruscolo, Department of Chemistry, University of Cambridge, Lensfield Rd., Cambridge CB2 1EW, UK; e-mail: mv245@cam.ac.uk; fax: 44-1223-763848; or Jane Clarke, MRC Centre for Protein Engineering, Department of Chemistry, University of Cambridge, Lensfield Rd., Cambridge CB2 1EW, UK; e-mail: jc162@cam.ac.uk; fax: 44-1223-336362.

*Abbreviations:* SASA, Solvent-accessible surface area; ΔSASA, change in solvent-accessible surface area upon unfolding; $C_p$, heat capacity; $\Delta C_p$, change in heat capacity upon unfolding; $R_g$, radius of gyration; $\Delta R_g$, change in radius of gyration upon unfolding; GdmCl, guanidinium chloride; SAXS, small angle X-Ray scattering; FRET, Förster resonance energy transfer; TS, transition state; *N*, number of residues; *R*, Pearson correlation coefficient; $\beta_T$, Beta Tanford.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.062383807.

et al. 1985; Miller et al. 1987; Lesser and Rose 1990); later it was shown that this procedure may overestimate it by ~25% (Creamer et al. 1995, 1997).

In this study, we establish an approximate relationship between the changes in the radius of gyration ($\Delta R_g$) upon protein folding and the *m*-values. Such a relationship is useful as $\Delta R_g$ is readily measurable experimentally by using small angle X-ray scattering (SAXS) (Lipman et al. 2003) or Förster resonance energy transfer (FRET) (Schuler et al. 2002). Moreover, since we can also show that a relationship exists between $\Delta R_g$ and $\Delta SASA$, we provide a method that in principle can provide an estimate of the latter independent from any model for the unfolded state.

Furthermore, we use the extensive body of experimental data that has become available since the original work of Myers et al. (1995) on the thermodynamics of protein unfolding to reanalyze a series of relationships between *m*-values, $\Delta SASA$, $\Delta C_p$, and the number *N* of amino acids in the protein; we also present new relationships between these quantities and $\Delta R_g$. We show that knowledge of any of these variables can be used to estimate all the others and present appropriate equations for making these estimates. A Web-based tool that allows users to use the equations that are presented here was developed, Bio-Physical PREDictions (BPPred, http://www-clarke.ch.cam.ac.uk). In addition to estimating several biophysical parameters for any given protein, the results are also visualized graphically, as is illustrated below. For example, if two properties are known, for example, the chain length *N* and the equilibrium *m*-value, their values can be entered through a Web-based interface and both the predicted and the actual properties are compared graphically against the values available for the other proteins present in the database. The BPPred Web server therefore enables users to verify whether any unusual behavior is displayed by a particular protein upon unfolding, for example, as a consequence of the presence of disulfide bridges.

A further motivation for the present study is given by the current challenge in protein folding to benchmark simulation against experiment (Dinner et al. 2000; Vendruscolo and Paci 2003; Sato et al. 2004). We have been particularly interested in using experimentally derived restraints to determine structures of transition states and other transient species on folding pathways. Transition state structures can be derived from restrained molecular dynamics using experimental Φ-values as restraints (Vendruscolo et al. 2001; Paci et al. 2002, 2003; Geierhaas et al. 2004), or extracted from molecular dynamics unfolding simulations at high temperatures (Daggett 2002; Gsponer and Caflisch 2002; Beck and Daggett 2004). In this case, S- or Φ-values calculated for the resulting structures are compared to experimental Φ-values to validate the results of the simulations (Daggett et al. 1996).

Another experimental parameter that reports on the properties of transition states is the Beta Tanford value ($\beta_T$), which is the ratio of equilibrium and kinetic *m*-values:

$$\beta_T = \frac{m_{D-TS}}{m_{D-N}} \text{ or } \beta_T = 1 - \frac{m_{TS-N}}{m_{D-N}} \quad (2)$$

$\beta_T$ is commonly used to infer the $\Delta SASA$ between the denatured state and the transition state (Vendruscolo and Dobson 2005). We show here that relationships between $\beta_T$, $\Delta SASA$, and $\Delta R_g$ can be used to verify the overall structural properties of transition state structures or, alternatively, $\beta_T$ might provide an additional experimental restraint for simulations.

## Results and Discussion

### *Database*

We compiled a database that includes experimental measurements of *m*-values in GdmCl and urea for a set of 30 disulfide-free, two-state proteins (see Table 1), which included several of the 22 proteins with similar characteristics considered by Myers et al. (1995). We used molecular dynamics (MD) simulations to calculate changes upon unfolding in structural properties ($\Delta R_g$ and $\Delta SASA$) of these proteins in order to assess their relationship to the corresponding experimental quantities. Simulations were carried out using the CHARMM 19 force field (Brooks et al. 1983) at 573 K and excluded volume interactions, following a procedure similar to that described by Dedmon et al. (2005) and Kristjansdottir et al. (2005), in which only the repulsive part of the Lennard-Jones potential was used in the molecular dynamics simulations. Each protein was simulated for 100 nsec, and usually the protein was already unfolded within the first few nanoseconds. The results of the MD simulations are, within the statistical errors, in agreement with the experimental values of $\Delta R_g$ (Millett et al. 2002; Kohn et al. 2004) in the cases in which the latter are available (Fig. 1A).

### *Reassessment of the relationships between equilibrium* m-*values,* $\Delta C_p$, ΔSASA, *and* N

Equilibrium *m*-values are related to the changes in heat capacity ($\Delta C_p$) upon unfolding (Livingstone et al. 1991; Spolar et al. 1992; Myers et al. 1995). Myers et al. (1995) also showed that the *m*-values are correlated linearly with $\Delta SASA$, $\Delta C_p$, and *N*. We have reassessed here these relationships by considering the new data that have become

**Table 1.** *Database*

| Protein | PDB | $N$ | $\Delta SASA$ (Å$^2$) | $\Delta R_g$ (Å) | GdmCl $m$ (cal mol$^{-1}$ M$^{-1}$) | Urea $m$ (cal mol$^{-1}$ M$^{-1}$) | $\Delta C_p$ (cal mol$^{-1}$ K$^{-1}$) |
|---|---|---|---|---|---|---|---|
| Engrailed homeodomain | 1ENH | 54 | 4400 ± 200 | 13.8 ± 4.0 | | 800[a] | |
| IgG binding domain of protein G | 1PGB | 56 | 4100 ± 200 | 13.4 ± 4.1 | 1800[b] | | 600[b] |
| SH3 domain of α-spectrin | 1SHG | 57 | 4500 ± 200 | 14.5 ± 4.1 | 1900[c] | 800[d] | 800[d] |
| Chymotrypsin inhibitor 2 | 2CI2 | 65 | 4800 ± 200 | 15.4 ± 4.7 | 1900[e] | | 700[f] |
| Cold shock protein B (*Bacillus subtilis*) | 1NMG | 67 | 4700 ± 200 | 15.4 ± 4.6 | 1800[g] | 800[g] | |
| Cold shock protein A (*E. coli*) | 3MEF | 69 | 4800 ± 200 | 15.6 ± 4.7 | 2200[h] | 700[h] | |
| Calbindin D9K | 1B1G | 75 | 5900 ± 200 | 17.6 ± 5.1 | | 1100[i] | |
| Ubiquitin | 1UBI | 76 | 6050 ± 200 | 17.7 ± 5.3 | 1800[j] | 1100[k] | 1400[l] |
| Rfa RBD | 1RFA | 78 | 6500 ± 300 | 18.4 ± 5.3 | | 1000[m] | |
| λ-Repressor | 1LMB | 80 | 6500 ± 300 | 18.3 ± 5.4 | 2400[m] | 1100[n] | |
| ADAh2 | 1O6X | 81 | 5500 ± 200 | 16.2 ± 5.5 | | 900[m] | |
| Acyl-coenzyme A binding protein | 1NTI | 86 | 6800 ± 200 | 18.3 ± 5.4 | 3500[o] | | |
| HPr (*B. subtilis*) | 2HPR | 87 | 7000 ± 300 | 19.6 ± 5.6 | 2200[p] | 1100[q] | 1200[r] |
| Barstar | 1BTA | 89 | 7500 ± 300 | 20.0 ± 5.5 | 2400[s] | 1300[s] | 1500[s] |
| TI I27 (human) | 1TIT | 89 | 6600 ± 300 | 19.0 ± 5.9 | 2500[t] | | 1400[u] |
| TNfn3 (human) | 1TEN | 89 | 7200 ± 300 | 19.2 ± 6.0 | 2600[v] | 1400[w] | 1200[w] |
| CTL9 | 1DIV | 92 | 7200 ± 300 | 18.4 ± 6.0 | | 1100[m] | |
| L23 | 1N88 | 96 | 7400 ± 300 | 19.3 ± 6.1 | | 800[m] | |
| U1A | 1URN | 96 | 8100 ± 300 | 21.3 ± 6.3 | 2100[m] | | |
| mAcP | 1APS | 98 | 8350 ± 300 | 21.0 ± 6.2 | | 1300[m] | |
| Src SH3 | 1SPR | 103 | 8800 ± 300 | 21.7 ± 6.3 | 1600[m] | | |
| FK binding protein (human) | 1FKD | 107 | 8900 ± 300 | 22.5 ± 6.7 | | 1400[x] | |
| Barnase | 1RNB | 109 | 9400 ± 300 | 22.7 ± 6.8 | 4400[y] | 1900[y] | 1700[z] |
| Che Y (*E. coli*) | 3CHY | 128 | 11300 ± 300 | 26.9 ± 7.9 | 2300[aa] | 1600[bb] | |
| Fatty acid binding protein (rat) | 1IFC | 131 | 11900 ± 300 | 27.2 ± 7.8 | 4500[cc] | 1800[cc] | |
| Interleukin 1-β | 5I1B | 151 | 13500 ± 400 | 30.1 ± 8.7 | 5500[dd] | | 1900[ee] |
| Ribonuclease H | 2RN2 | 155 | 13500 ± 400 | 29.4 ± 8.3 | 4500[ff] | 1900[ff] | |
| Dihydrofolate reductase (*E. coli*) | 4DFR | 159 | 13950 ± 400 | 30.7 ± 8.5 | | 2000[gg] | |
| T4 lysozyme | 1L63 | 162 | 14800 ± 400 | 29.8 ± 8.6 | 5500[hh] | 2000[hh] | 2600[ii] |
| Adenylate kinase (*Saccharomyces cerevisiae*) | 2AKY | 218 | 19650 ± 400 | 38.6 ± 10.6 | 7300[jj] | | 2100[jj] |

[a] Mayor et al. (2000); [b] O'Neil et al. (1995); [c] Viguera et al. (1994); [d] Viguera et al. (1994); [e] Itzhaki et al. (1995); [f] Jackson et al. (1993); [g] Perl et al. (1998); [h] Reid et al. (1998); [i] Akke and Forsen (1990); [j] Benitez-Cardoza et al. (2004); [k] Ibarra-Molero et al. (1999); [l] Wintrode et al. (1994); [m] Maxwell et al. (2005); [n] Huang and Oas (1995), Burton et al. (1996); [o] Kragelund et al. (1999); [p] Van Nuland et al. (1998); [q] Peterson et al. (1999); [r] Scholtz (1995); [s] Agashe and Udgaonkar (1995); [t] Wright et al. (2003); [u] Wright et al. (2004); [v] Rounsevell et al. (2005); [w] Clarke et al. (1997); [x] Main et al. (1999); [y] Clarke and Fersht (1993); [z] Griko and Privalov (1994); [aa] Filimonov et al. (1993); [bb] Munoz et al. (1994); [cc] Ropson et al. (1990); [dd] Covalt et al. (2001); [ee] Makhatadze et al. (1994); [ff] Dabora and Marqusee (1994); [gg] Ionescu et al. (2000); [hh] Zhang et al. (1993); [ii] Hu et al. (1992); [jj] Spuergin et al. (1995).

available since the work of Myers et al. (1995). In Figure 1B, we present the scatterplot between $\Delta SASA$ and $N$ for the database of 30 proteins that we considered. The existence of such a strong correlation enables us to estimate $\Delta SASA$ from $N$:

$$\Delta SASA = (-1520 \pm 40) + (98 \pm 1)N \quad R = 0.99 \pm 0.01 \quad (3)$$

The correlations for GdmCl and urea $m$-values with $\Delta SASA$ (Myers et al. 1995) are shown in Figure 2, A and B, respectively, for our database of 30 proteins. The linear fits are given by, respectively,

$$m_{GdmCl} = (10 \pm 50) + (0.36 \pm 0.01)\Delta SASA \quad R = 0.89 \pm 0.01 \quad (4)$$

$$m_{urea} = (270 \pm 20) + (0.13 \pm 0.01)\Delta SASA \quad R = 0.89 \pm 0.01 \quad (5)$$

The slope is ~2.7 times as large for GdmCl as for urea, in agreement with the fact that GdmCl is a stronger denaturant than urea (Myers et al. 1995).

Figure 2C shows the linear fit of the change in heat capacity upon unfolding with $\Delta SASA$:

$$\Delta C_p = (130 \pm 60) + (0.16 \pm 0.01)\Delta SASA \quad R = 0.94 \pm 0.03 \quad (6)$$

This equation can be used to estimate $\Delta C_p$ from the change in solvent-accessible surface area.

Therefore, all the relationships that Myers et al. (1995) observed for $\Delta SASA$ estimated using the glycine-tripeptide model (Shrake and Rupley 1973; Rose et al. 1985; Miller et al. 1987; Lesser and Rose 1990) are also valid if the change in solvent-accessible surface area is calculated from the present molecular dynamics simulations. These equations can be used to quantify, for example, $\Delta SASA$ from

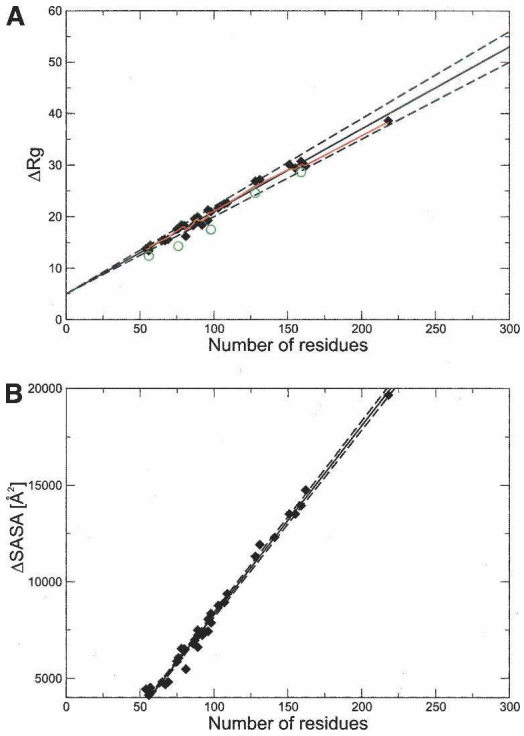


**Figure 1.** (*A*) Correlation between the change in radius of gyration upon unfolding ($\Delta R_g$) and *N* for the 32 proteins in Table 1. The red line represents the nonlinear fit of $\Delta R_g$ to *N* (Flory 1988). For comparison, experimental values of $\Delta R_g$ are indicated with green circles; data for protein G, ubiquitin, Che Y, mACP, and DHFR (Millett et al. 2002; Kohn et al. 2004). (*B*) Correlation between the change in solvent-accessible surface area ($\Delta SASA$) and the number of residues *N* (Myers et al. 1995). Solid lines are the linear fits, and dashed lines represent the statistical deviation, estimated using the jackknife method (Miller 1974).

*m*-values, and thus determine the SASA of the denatured state of a protein, since the SASA of the native state can be determined from the native structure.

### *Additional relationships between* N, m-*value, and* $\Delta C_p$

In addition to the correlations determined by Myers et al. (1995), we also analyzed the correlations between *N*, *m*-values, and $\Delta C_p$. We thus defined the following relationships (Fig. 3A, Equation 7; Fig. 3B, Equation 8; Fig. 3C, Equation 9):

$$m_{\mathrm{GdmCl}} = (-580 \pm 50) + (36 \pm 1)N \quad R = 0.88 \pm 0.06 \quad (7)$$

$$m_{\mathrm{urea}} = (70 \pm 20) + (13 \pm 1)N \quad R = 0.88 \pm 0.05 \quad (8)$$

$$\Delta C_p = (-130 \pm 70) + (16 \pm 1)N \quad R = 0.94 \pm 0.03 \quad (9)$$

### *Relationships involving* $\Delta R_g$

We investigated whether the correlations provided by Myers et al. (1995) for $\Delta SASA$ are also valid for $\Delta R_g$. The correlation for the proteins in Table 1 between GdmCl and urea *m*-values and the change of $R_g$ is shown in Figure 4, A and B:

$$m_{\mathrm{GdmCl}} = (-1720 \pm 100) + (225 \pm 5)\Delta R_g \quad R = 0.86 \pm 0.08 \quad (10)$$

$$m_{\mathrm{urea}} = (-260 \pm 30) + (76 \pm 1)\Delta R_g \quad R = 0.86 \pm 0.08 \quad (11)$$

The knowledge of *m*-values (GdmCl or urea) can thus be used to estimate the values of $\Delta R_g$ by inverting these relationships. The slope of the linear regression is about three times larger for GdmCl than for urea (225 ± 5 cal mol$^{-1}$ M$^{-1}$ Å$^{-1}$ compared to 76 ± 1 cal mol$^{-1}$ M$^{-1}$ Å$^{-1}$), again reflecting the fact that GdmCl is a stronger denaturant than urea (Myers et al. 1995).

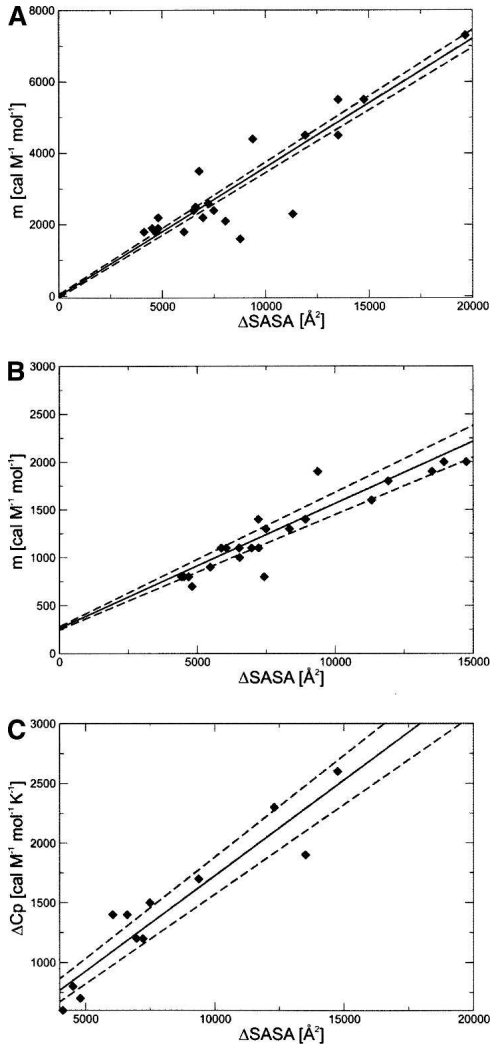


**Figure 2.** Correlation between $\Delta SASA$ and (*A*) GdmCl equilibrium *m*-values, (*B*) urea equilibrium *m*-values, and (*C*) $\Delta C_p$. The solid line is the linear fit, and the dashed lines represent the statistical deviation, estimated using the jackknife method (Miller 1974).

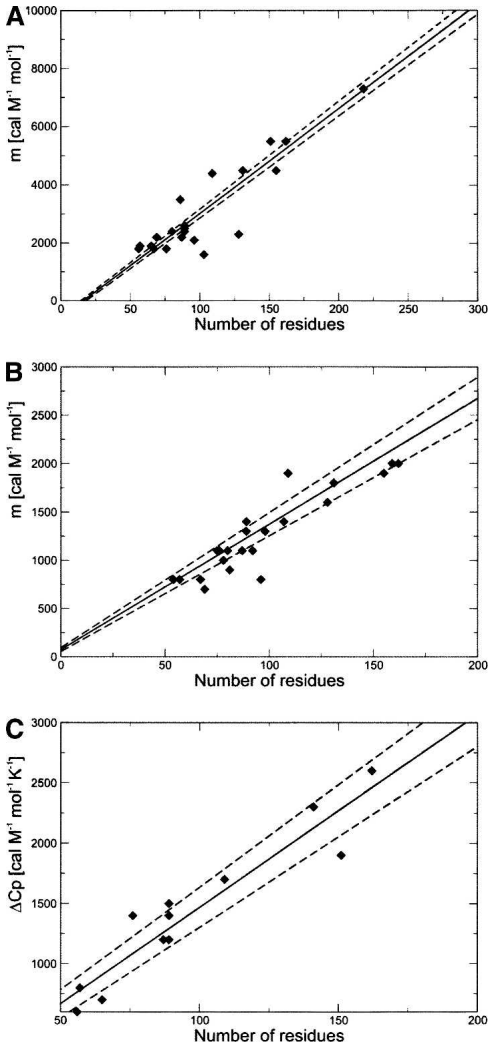


**Figure 3.** Correlation between the chain length $N$ and (*A*) GdmCl equilibrium $m$-values, (*B*) urea equilibrium $m$-values, and (*C*) $\Delta C_p$. These relations can be used to estimate $m$-values and $\Delta C_p$ from $N$. The solid line is the linear fit, and the dashed lines represent the statistical deviation, estimated using the jackknife method (Miller 1974).

Although the correlation between $m$-values and $\Delta R_g$ is statistically significant, there are considerable deviations for proteins with a similar number of residues or a similar $\Delta R_g$. The major outliers for GdmCl are src SH3 and CheY (*Escherichia coli*), and for urea, barnase. These deviations can be attributed to the presence of residual structure in the denatured state. For example, barnase has been shown to exhibit residual structure in the denatured state (Bond et al. 1997). The value of the total charge of the protein can also affect the value of $\Delta R_g$. A highly charged protein will be more expanded in the unfolded state, compared to a neutral protein because of electrostatic repulsions. These effects are a likely reason for the observed pH dependence of the $m$-values for RNase A, RNase T1, and barnase (Pace et al. 1990, 1992). Electrostatic effects can be influenced by GdmCl, which is dissociated in solution (Monera et al. 1994). The comparison of $m$-values and $\Delta R_g$ values represents therefore a useful tool for detecting persistent residual structure in the denatured state of proteins.

Furthermore, we find a linear correlation between $\Delta R_g$ and $\Delta C_p$ (Fig. 4C):

$$\Delta C_p = (-630 \pm 110) + (100 \pm 6)\Delta R_g \quad R = 0.94 \pm 0.04 \quad (12)$$

Therefore, the change in heat capacity and the $m$-values does not only correlate to the amount of buried surface

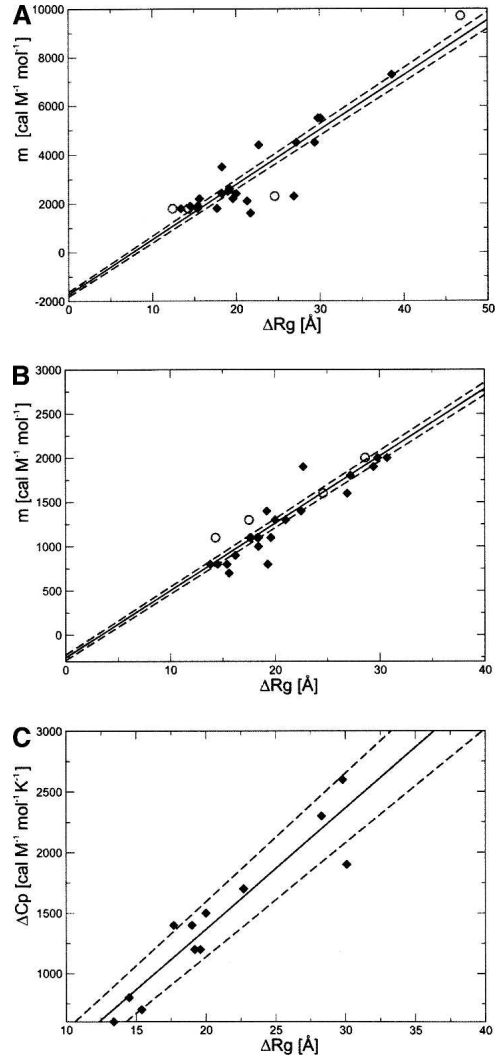


**Figure 4.** Correlation between $\Delta R_g$ and (*A*) GdmCl equilibrium $m$-values, (*B*) urea equilibrium $m$-values, and (*C*) $\Delta C_p$. The solid line is the linear fit, and the dashed lines represent the statistical deviation, estimated using the jackknife method (Miller 1974). For comparison, we have also included the experimental values of $\Delta R_g$ for protein G, ubiquitin, Che Y, mACP, DHFR, and yPGK (open circles in *A* and *B*) (data from Millett et al. 2002).

that is exposed during unfolding but also to the expansion of the molecule.

We find that $N$ also correlates with $\Delta R_g$ (Fig. 1A). Therefore, an estimate for $\Delta R_g$ of a given protein with $N$ residues can be obtained from

$$\Delta R_g = (5.0 \pm 0.1) + (0.16 \pm 0.01)N \quad R = 0.99 \pm 0.01 \quad (13)$$

Kohn et al. (2004) and Millett et al. (2002) recently presented an extensive analysis of chemically and thermally unfolded proteins. They showed that the $R_g$ of unfolded proteins can be related to $N$ by using a random-coil model (Flory 1988):

$$R_g = R_0 N^{\nu} \quad (14)$$

where $R_0$ is a constant related to the persistence length of the protein and $\nu$ is an exponent that describes how $R_g$ scales with the length $N$ of the protein. By fitting data from a database of 28 chemically denatured, disulfide-free proteins, Kohn et al. (2004) obtained $R_0 = 2.08 \pm 0.19$ Å and $\nu = 0.598 \pm 0.029$, in agreement with both the theoretical values of $\nu = 0.6$ (Flory 1988) and of $\nu = 0.588$ calculated for excluded-volume polymers in a good solvent (LeGuillou and Zinn-Justin 1977). In our case, a similar type of fitting provided values of $R_0 = 2.07$ and $\nu = 0.61$, both compatible with the results of Kohn et al. The apparent existence of a linear relationship between $\Delta R_g$ and $N$ (Equation 13) is surprising as $R_g$ grows with $N^{0.6}$ in the unfolded state and with $N^{1/3}$ in the folded state. However, this result can be explained by the small range of values for $N$ considered in our database, for which $\Delta R_g$ can be approximated by a linear relationship (Fig. 1A).

It has been shown that the $R_g$ values of proteins represented by using a repulsive hard-sphere potential should obey the random-coil model (Creamer et al. 1995, 1997; Goldenberg 2003). These results, however, do not necessarily imply the complete absence of residual structure. Fitzkee and Rose (2004) showed that a ''rigid-segment model,'' in which known protein structures are partitioned alternately into rigid segments linked by individual flexible residues, gives values for $R_0$ and $\nu$ similar to those of Kohn et al. (2004) and Millett et al. (2002), despite the presence of native-like structural elements. The apparent discrepancy between the SAXS experimental results, which are consistent with a random-coil behavior, and other studies that indicate the presence of residual structure can be explained if the proportion of conformers exhibiting residual structure in the unfolded state is rather low (Lindorff-Larsen et al. 2004). These conformers may not affect the result of ensemble-averaged measurements, as the SAXS technique is most sensitive to large values of $R_g$ (Kohn et al. 2004). With these considerations in mind, we decided to use molecular dynamics simulations to estimate $\Delta R_g$ in the cases in which experimental measurements are not yet available. The relations between the number of residues $N$ and either $\Delta R_g$ or $\Delta SASA$ (Equations 13 and 3, respectively) suggest a correlation between the change in radius of gyration and the change in surface-accessible area upon unfolding. Indeed, there is a significant linear correlation between $\Delta R_g$ and $\Delta SASA$ (Fig. 5). As already mentioned, it is very difficult to measure $\Delta SASA$ of a protein directly; in contrast, $\Delta R_g$ can be obtained from experiments (Millett et al. 2002; Kohn et al. 2004). We suggest that $\Delta SASA$ can be estimated from $R_g$ measurements according to the relationship

$$\Delta SASA = (-4500 \pm 70) + (610 \pm 3)\Delta R_g \quad R = 0.99 \pm 0.01 \quad (15)$$

### *Estimation of* $R_g$ *and* SASA *of the transition state for folding*

One of the motivations for this work was to establish convenient criteria to benchmark the properties of transition states for protein folding obtained from molecular dynamics simulations. A kinetic $m$-value analysis not only provides information about the $R_g$ and $SASA$ of the unfolded state, but also about the $R_g$ and $SASA$ of the transition state for folding. The knowledge of equilibrium and kinetic $m$-values allows the value of $R_g^{TS}$ of the transition state of the protein folding reaction to be estimated

$$\beta_T = \frac{a_r + b_r\left(R_g^D - R_g^{TS}\right)}{a_r + b_r\left(R_g^D - R_g^N\right)} \quad (16)$$

where $R_g^D$ is the radius of gyration in the denatured state, $a_r$ and $b_r$ are the constants of the linear correlation between $m$-values and $R_g$ (Equations 10 and 11), where we assumed

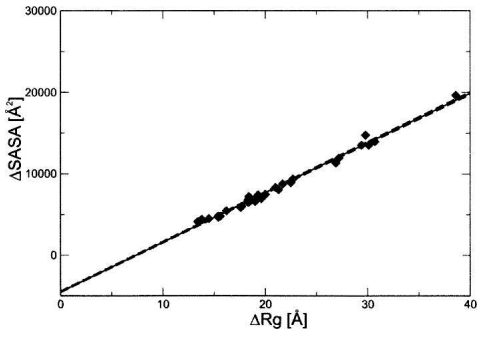


**Figure 5.** Correlation between $\Delta SASA$ and $\Delta R_g$. The solid line is the linear fit, and the dashed lines represent the statistical deviation, estimated using the jackknife method (Miller 1974).

that the same relationships hold for the transition state, and $\beta_T$ is the ratio of equilibrium and kinetic *m*-values (Equation 2). Given Equation 16, the radius of gyration of the transition state can be estimated from experimentally derived quantities, that is, if $\beta_T$ is known and $R_g^D$ is estimated from Equation 10 or Equation 11. Alternatively, $R_g^D$ can be estimated from molecular dynamics simulations.

The statistical error on the estimate for $R_g$ of the transition state can be expressed as

$$\delta R_g^{TS} = \left|\left(-\frac{a_r}{b_r} + R_g^N - R_g^D\right)\delta\beta_T\right| + \left|\beta_T \delta R_g^N\right| + \left|(1-\beta_T)\delta R_g^D\right| + \left|\left(\frac{1-\beta_T}{b_r}\right)\delta a_r\right| + \left|\left(\frac{-a_r + a_r\beta_T}{b_r^2}\right)\delta b_r\right| \quad (17)$$

where $\delta\beta_T$ is the experimental error on $\beta_T$, $\delta R_g^N$ and $\delta R_g^D$ are the statistical deviations of the radius of gyration of the native state and of the denatured state, respectively, and $\delta a_r$ and $\delta b_r$ are the deviations of the constants in Equation 10 or 11.

An analogous relationship can be given for $\Delta SASA$:

$$\beta_T = \frac{a_s + b_s(SASA_D - SASA_{TS})}{a_s + b_s(SASA_D - SASA_N)} \quad (18)$$

where $SASA_D$ is the solvent-accessible surface area in the denatured state, $a_s$ and $b_s$ are the constants of the linear correlation (Equations 4 and 5), and $SASA_{TS}$ is the solvent-accessible surface area of the transition state for folding. The statistical error for the solvent-accessible surface area of the transition state is

$$\delta SASA_{TS} = \left|\left(-\frac{a_s}{b_s} + SASA_N - SASA_D\right)\delta\beta_T\right| + |\beta_T \delta SASA_N| + |(1-\beta_T)\delta SASA_D| + \left|\left(\frac{1-\beta_T}{b_r}\right)\delta a_s\right| + \left|\left(\frac{-a_s + a_s\beta_T}{b_s^2}\right)\delta b_s\right| \quad (19)$$

where $\Delta SASA_N$ and $\Delta SASA_D$ are the statistical deviations of *SASA* of the native state and of the denatured state, respectively, and $\delta a_s$ and $\delta a_s$ are the deviations of the constants in Equation 4 or 5.

We have previously determined the structures of the transition states of several proteins using experimental Φ-values as restraints in molecular dynamics simulations (Vendruscolo et al. 2001; Paci et al. 2002, 2003; Geierhaas et al. 2004). We applied the relationships to estimate *SASA* and $R_g$ in the transition states of four proteins, CI2, TNfn3, mACP, and TI I27 (see Table 2). These proteins have been chosen because their $\beta_T$ values range from 0.6 to 0.95. Thus, the transition state structures range from being fairly heterogeneous to being very native-like. It is important to note that, although the values for *a* and *b* are different in Equations 16 and 18 depending on the choice of GdmCl or urea as denaturant to measure the *m*-values, the results for $R_g$ and *SASA* of the transition state are the same within the statistical error (Table 2). We compared the estimated values of $R_g$ and *SASA* of the transition state to those obtained from restrained simulations, finding a good agreement (Table 2) despite the fairly large statistical errors and the potential presence of residual structure in the denatured state, which was not accounted for in our simulations.

**Table 2.** *Estimated values for* $\Delta R_g$ *and* SASA *of the transition state for folding*

| Protein | $R_g$ simulation[a] (Å) | $R_g$ GdmCl[b] (Å) | $R_g$ urea[c] (Å) | *SASA* simulation[a] (Å) | *SASA* GdmCl[d] (Å$^2$) | *SASA* urea[e] (Å$^2$) |
|---|---|---|---|---|---|---|
| TI I27[f] | 12.9 ± 0.2 | 13.5 ± 1.1 | 13.7 ± 1.3 | 5500 ± 200 | 5500 ± 500 | 5600 ± 600 |
| TNfn3[g] | 14.9 ± 1.4 | 15.3 ± 2.9 | 16.1 ± 3.0 | 7100 ± 600 | 6600 ± 800 | 7000 ± 1000 |
| CI2[h] | 14.4 ± 2.3 | 14.5 ± 3.7 | 16.1 ± 3.5 | 6100 ± 750 | 6400 ± 750 | 7100 ± 900 |
| mAcP | 14.0 ± 0.3 | 15.8 ± 3.0 | 16.7 ± 3.1 | 7200 ± 400 | 6900 ± 1000 | 7300 ± 1100 |

[a] Obtained by ensemble-averaged molecular dynamics simulations restrained by Φ-values (Best and Vendruscolo 2006).
[b] Obtained by inverting Equation 16 and using $R_g$ of the denatured state from molecular dynamics simulations with excluded volume interactions, values of $a_r$ and $b_r$ from Equation 10 (GdmCl).
[c] Obtained by inverting Equation 16 and using $R_g$ of the denatured state from molecular dynamics simulations with excluded volume interactions, values of $a_r$ and $b_r$ from Equation 11 (urea).
[d] Obtained by inverting Equation 18 and using *SASA* of the denatured state from molecular dynamics simulations with excluded volume interactions, values of $a_s$ and $b_s$ from Equation 4 (GdmCl).
[e] Obtained by inverting Equation 18 and using *SASA* of the denatured state from molecular dynamics simulations with excluded volume interactions, values of $a_s$ and $b_s$ from Equation 5 (urea).
[f] Transition state of TI I27 dominant at low concentrations of denaturant and at moderate temperatures, native-like transition state; $\beta_T = 0.95 \pm 0.05$ (Wright et al. 2003).
[g] Transition state of TNfn3; $\beta_T = 0.7 \pm 0.07$ (Hamill et al. 2000).
[h] Transition state of CI2; $\beta_T = 0.6 \pm 0.06$ (Itzhaki et al. 1995).
[i] Transition state of mACP; $\beta_T = 0.8 \pm 0.08$ (Chiti et al. 1999).

Deviations from a spherical form might also influence the predictions of $R_g^{TS}$ (Geierhaas et al. 2004).

We also investigated whether it is possible to reduce the fairly large statistical error in the estimate of $R_g$ from $\beta_T$ by a simultaneous fitting of *m*-values, *N*, and $\Delta SASA$ or $\Delta R_g$:

$$m = \alpha_r + \beta_r \Delta R_g + \chi_r N \quad (20)$$

$$m = \alpha_s + \beta_s \Delta SASA + \chi_s N \quad (21)$$

The statistical errors are, however, too large in this case to allow reliable predictions (data not shown). A more extensive database than the one we used (Table 1), which should become available through future experimental studies, should make this type of prediction possible.

**Table 3.** *Pairwise linear relationships between* m-*values,* N, $\Delta R_g$, $\Delta$SASA, *and* $\Delta C_p$

| | *m*-value (cal $M^{-1}$ $mol^{-1}$) | *N* | $\Delta SASA$ (Å$^2$) | $\Delta R_g$ (Å) | $\Delta C_p$ (cal $M^{-1}$ $mol^{-1}$ $K^{-1}$) |
|---|---|---|---|---|---|
| *m*-value (cal $M^{-1}$ $mol^{-1}$) | — | 7, 8 | 4, 5 | 10, 11 | 22, 23[a] |
| *N* | | — | 4 | 13 | 9 |
| $\Delta SASA$ (Å$^2$) | | | — | 15 | 6 |
| $\Delta R_g$ (Å) | | | | — | 12 |
| $\Delta C_p$ (cal $M^{-1}$ $mol^{-1}$ $K^{-1}$) | | | | | — |

The numbers in the table indicate the equations that provide the relationships between pairs of quantities.
[a](Equation 22) $\Delta C_p = -336 + 0.66 m_{GdmCl}$ $R = 0.87$; (Equation 23) $\Delta C_p = 117 + 1.1 m_{urea}$ $R = 0.88$. Equations from Myers et al. (1995), Figure 4.

## Conclusions

We have presented BPPred, a Web-based tool to predict *m*-values and several other descriptors of the folding process. These predictions exploit the good accuracy provided by linear relationships between the change in radius of gyration upon unfolding and the *m*-values from GdmCl and from urea unfolding. Such relationships are analogous to those determined by Myers et al. (1995) between the change in the surface-accessible area upon unfolding and the *m*-values. The latter result requires a model for the unfolded state, from which the corresponding surface area can be calculated, as its experimental measurement is challenging. Instead, $R_g$ can be readily measured experimentally and therefore the new relationships in principle do not depend on any assumption about unfolded states. However, as there is only a very limited number of proteins for which both $\Delta R_g$ and *m*-values are known, we used in this study estimates obtained from molecular dynamics simulations. When the comparison is possible, the obtained values for $\Delta R_g$ are compatible with experimental data within the statistical error. This relationship could be refitted and made completely independent from any particular model used to describe the unfolded state, when more systematic measurements of the $R_g$ of unfolded states become available. In addition, we have also reported a linear correlation between the change in radius of gyration and the change in solvent-accessible surface area upon unfolding. This relationship provides an estimate of $\Delta SASA$ from the values of $\Delta R_g$ that can be obtained from experiments.

The linear correlations between *m*-values, the chain length *N*, and the change in $R_g$, *SASA*, and $C_p$ upon unfolding that we discussed are summarized in Table 3. Such relationships can be used to estimate any one of these quantities from the knowledge of the other one. These relationships should represent valuable tools for interpreting experimental measurements in terms of the structural properties of the unfolded and of the transition states, as well as providing new ways of enhancing the powerful synergy between experiment and theory in protein folding that has been developed over the last several years.

## Acknowledgments

C.D.G. and A.A.N. hold a Wellcome Trust Prize Studentship. J.C. is a Wellcome Trust Senior Research Fellow. M.V. is supported by the Royal Society and the Leverhulme Trust.

## References

Agashe, V.R. and Udgaonkar, J.B. 1995. Thermodynamics of denaturation of barstar: Evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* **34:** 3286–3299.

Akke, M. and Forsen, S. 1990. Protein stability and electrostatic interactions between solvent exposed charged side chains. *Proteins* **8:** 23–29.

Beck, D.A. and Daggett, V. 2004. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* **34:** 112–120.

Benitez-Cardoza, C.G., Stott, K., Hirshberg, M., Went, H.M., Woolfson, D.N., and Jackson, S.E. 2004. Exploring sequence/folding space: Folding studies on multiple hydrophobic core mutants of ubiquitin. *Biochemistry* **43:** 5195–5203.

Best, R.B. and Vendruscolo, M. 2006. Structural interpretation of hydrogen exchange protection factors in proteins: Characterization of the native state fluctuations of CI2. *Structure* **14:** 97–106.

Bond, C.J., Wong, K.B., Clarke, J., Fersht, A.R., and Daggett, V. 1997. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proc. Natl. Acad. Sci.* **94:** 13409–13413.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. 1983. CHARMM—A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4:** 187–217.

Burton, R.E., Huang, G.S., Daugherty, M.A., Fullbright, P.W., and Oas, T.G. 1996. Microsecond protein folding through a compact transition state. *J. Mol. Biol.* **263:** 311–322.

Chiti, F., Taddei, N., White, P.M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C.M. 1999. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* **6:** 1005–1009.

Clarke, J. and Fersht, A.R. 1993. Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32:** 4322–4329.

Clarke, J., Hamill, S.J., and Johnson, C.M. 1997. Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **270:** 771–778.

Covalt, Jr., J.C., Roy, M., and Jennings, P.A. 2001. Core and surface mutations affect folding kinetics, stability and cooperativity in IL-1 β: Does alteration in buried water play a role? *J. Mol. Biol.* **307:** 657–669.

Creamer, T.P., Srinivasan, R., and Rose, G.D. 1995. Modelling unfolded states of peptides and proteins. *Biochemistry* **34:** 16245–16250.

Creamer, T.P., Srinivasan, R., and Rose, G.D. 1997. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* **36:** 2832–2835.

Dabora, J.M. and Marqusee, S. 1994. Equilibrium unfolding of *Escherichia coli* ribonuclease H: Characterization of a partially folded state. *Protein Sci.* **3:** 1401–1408.

Daggett, V. 2002. Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc. Chem. Res.* **35:** 422–429.

Daggett, V., Li, A., Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **257:** 430–440.

Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C.M. 2005. Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **127:** 476–477.

Dinner, A.R., Sali, A., Smith, L.J., Dobson, C.M., and Karplus, M. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25:** 331–339.

Filimonov, V.V., Prieto, J., Martinez, J.C., Bruix, M., Mateo, P.L., and Serrano, L. 1993. Thermodynamic analysis of the chemotactic protein from *Escherichia coli*, CheY. *Biochemistry* **32:** 12906–12921.

Fitzkee, N.C. and Rose, G.D. 2004. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci.* **101:** 12497–12502.

Flory, P.J. 1988. *Statistical mechanics of chain molecules*. Hanser/Gardner Publications, Inc., Cincinnati, Munich.

Geierhaas, C.D., Paci, E., Vendruscolo, M., and Clarke, J. 2004. Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J. Mol. Biol.* **343:** 1111–1123.

Goldenberg, D.P. 2003. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **326:** 1615–1633.

Griko, Y.V. and Privalov, P.L. 1994. Thermodynamic puzzle of apomyoglobin unfolding. *J. Mol. Biol.* **235:** 1318–1325.

Gsponer, J. and Caflisch, A. 2002. Molecular dynamics simulations of protein folding from the transition state. *Proc. Natl. Acad. Sci.* **99:** 6719–6724.

Hamill, S.J., Steward, A., and Clarke, J. 2000. The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297:** 165–178.

Hu, C.Q., Kitamura, S., Tanaka, A., and Sturtevant, J.M. 1992. Differential scanning calorimetric study of the thermal unfolding of mutant forms of phage T4 lysozyme. *Biochemistry* **31:** 1643–1647.

Huang, G.S. and Oas, T.G. 1995. Structure and stability of monomeric λ repressor: NMR evidence for two-state folding. *Biochemistry* **34:** 3884–3892.

Ibarra-Molero, B., Loladze, V.V., Makhatadze, G.I., and Sanchez-Ruiz, J.M. 1999. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge–charge interactions to protein stability. *Biochemistry* **38:** 8138–8149.

Ionescu, R.M., Smith, V.F., O'Neill, Jr., J.C., and Matthews, C.R. 2000. Multistate equilibrium unfolding of *Escherichia coli* dihydrofolate reductase: Thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. *Biochemistry* **39:** 9540–9550.

Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254:** 260–288.

Jackson, S.E., Moracci, M., elMasry, N., Johnson, C.M., and Fersht, A.R. 1993. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32:** 11259–11269.

Kohn, J.E., Millett, I.S., Jacob, J., Zagrovic, B., Dillon, T.M., Cingel, N., Dothager, R.S., Seifert, S., Thiyagarajan, P., Sosnick, T.R., et al. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci.* **101:** 12491–12496.

Kragelund, B.B., Osmark, P., Neergaard, T.B., Schiodt, J., Kristiansen, K., Knudsen, J., and Poulsen, F.M. 1999. The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nat. Struct. Biol.* **6:** 594–601.

Kristjansdottir, A.G., Lindorff-Larsen, K., Fieber, W., Dobson, C.M., Vendruscolo, M., and Poulsen, F.M. 2005. Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J. Mol. Biol.* **347:** 1053–1062.

LeGuillou, J.C. and Zinn-Justin, J. 1977. Critical exponents for the *n*-vector model in three dimensions from field theory. *Phys. Rev. Lett.* **39:** 95–98.

Lesser, G.J. and Rose, G.D. 1990. Hydrophobicity of amino acid subgroups in proteins. *Proteins* **8:** 6–13.

Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C.M., Poulsen, F.M., and Vendruscolo, M. 2004. Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. *J. Am. Chem. Soc.* **126:** 3291–3299.

Lipman, E.A., Schuler, B., Bakajin, O., and Eaton, W.A. 2003. Single-molecule measurement of protein folding kinetics. *Science* **301:** 1233–1235.

Livingstone, J.R., Spolar, R.S., and Record, M.T. 1991. Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface-area. *Biochemistry* **30:** 4237–4244.

Main, E.R.G., Fulton, K.F., and Jackson, S.E. 1999. Folding pathway of FKBP12 and characterisation of the transition state. *J. Mol. Biol.* **291:** 429–444.

Makhatadze, G.I., Clore, G.M., Gronenborn, A.M., and Privalov, P.L. 1994. Thermodynamics of unfolding of the all β-sheet protein interleukin-1 β. *Biochemistry* **33:** 9327–9332.

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.

Mayor, U., Johnson, C.M., Daggett, V., and Fersht, A.R. 2000. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci.* **97:** 13518–13522.

Miller, R.G. 1974. The jackknife—A review. *Biometrica* **61:** 1–15.

Miller, S., Janin, J., Lesk, A.M., and Chothia, C. 1987. Interior and surface of monomeric proteins. *J. Mol. Biol.* **196:** 641–656.

Millett, I.S., Doniach, S., and Plaxco, K.W. 2002. Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv. Protein Chem.* **62:** 241–262.

Mok, K.H., Nagashima, T., Day, L.J., Hore, P.J., and Dobson, C.M. 2005. Multiple subsets of side-chain packing in partially folded states of α-lactalbumins. *Proc. Natl. Acad. Sci.* **102:** 8899–8904.

Monera, O.D., Kay, C.M., and Hodges, R.S. 1994. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci.* **3:** 1984–1991.

Munoz, V., Lopez, E.M., Jager, M., and Serrano, L. 1994. Kinetic characterization of the chemotactic protein from *Escherichia coli*, CheY. Kinetic analysis of the inverse hydrophobic effect. *Biochemistry* **33:** 5858–5866.

Myers, J.K., Pace, C.N., and Scholtz, J.M. 1995. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4:** 2138–2148.

O'Neil, K.T., Hoess, R.H., Raleigh, D.P., and DeGrado, W.F. 1995. Thermodynamic genetics of the folding of the B1 immunoglobulin-binding domain from streptococcal protein G. *Proteins* **21:** 11–21.

Pace, C.N. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131:** 266–280.

Pace, C.N., Laurents, D.V., and Thomson, J.A. 1990. pH dependance of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry* **29:** 2564–2572.

Pace, C.N., Laurents, D.V., and Erickson, R.E. 1992. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochemistry* **31:** 2728–2734.

Paci, E., Vendruscolo, M., Dobson, C.M., and Karplus, M. 2002. Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* **324:** 151–163.

Paci, E., Clarke, J., Steward, A., Vendruscolo, M., and Karplus, M. 2003. Self-consistent determination of the transition state for protein folding: Application to a fibronectin type III domain. *Proc. Natl. Acad. Sci.* **100:** 394–399.

Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M.A., Jaenicke, R., and Schmid, F.X. 1998. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat. Struct. Biol.* **5:** 229–235.

Peterson, R.W., Nicholson, E.M., Thapar, R., Klevit, R.E., and Scholtz, J.M. 1999. Increased helix and protein stability through the introduction of a new tertiary hydrogen bond. *J. Mol. Biol.* **286:** 1609–1619.

Reid, K.L., Rodriguez, H.M., Hillier, B.J., and Gregoret, L.M. 1998. Stability and folding properties of a model β-sheet protein, *Escherichia coli* CspA. *Protein Sci.* **7:** 470–479.

Ropson, I.J., Gordon, J.I., and Frieden, C. 1990. Folding of a predominantly β-structure protein: Rat intestinal fatty acid binding protein. *Biochemistry* **29:** 9591–9599.

Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H., and Zehfus, M.H. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science* **229:** 834–838.

Rounsevell, R.W., Steward, A., and Clarke, J. 2005. Biophysical investigations of engineered polyproteins: Implications for force data. *Biophys J.* **88:** 2022–2029.

Sato, S., Religa, T.L., Daggett, V., and Fersht, A.R. 2004. Testing protein-folding simulations by experiment: B domain of protein A. *Proc. Natl. Acad. Sci.* **101:** 6952–6956.

Schellman, J.A. 1978. Solvent denaturation. *Biopolymers* **17:** 1305–1322.

Scholtz, J.M. 1995. Conformational stability of HPr: The histidine-containing phosphocarrier protein from *Bacillus subtilis*. *Protein Sci.* **4:** 35–43.

Schuler, B., Lipman, E.A., and Eaton, W.A. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419:** 743–747.

Shrake, A. and Rupley, J.A. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79:** 351–371.

Spolar, R.S., Livingstone, J.R., and Record, M.T. 1992. Use of liquid-hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **31:** 3947–3955.

Spuergin, P., Abele, U., and Schulz, G.E. 1995. Stability, activity and structure of adenylate kinase mutants. *Eur. J. Biochem.* **231:** 405–413.

Tanford, C. 1968. Protein denaturation. *Adv. Protein Chem.* **23:** 121–282.

Tanford, C. 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24:** 1–95.

Van Nuland, N.A., Meijberg, W., Warner, J., Forge, V., Scheek, R.M., Robillard, G.T., and Dobson, C.M. 1998. Slow cooperative folding of a small globular protein HPr. *Biochemistry* **37:** 622–637.

Vendruscolo, M. and Dobson, C.M. 2005. Towards complete descriptions of the free-energy landscapes of proteins. *Philos. Transact. A Math. Phys. Eng. Sci.* **363:** 433–450.

Vendruscolo, M. and Paci, E. 2003. Protein folding: bringing theory and experiment closer together. *Curr. Opin. Struct. Biol.* **13:** 82–87.

Vendruscolo, M., Paci, E., Dobson, C.M., and Karplus, M. 2001. Three key residues form a critical contact network in a protein folding transition state. *Nature* **409:** 641–645.

Viguera, A.R., Martinez, J.C., Filimonov, V.V., Mateo, P.L., and Serrano, L. 1994. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry* **33:** 2142–2150.

Wintrode, P.L., Makhatadze, G.I., and Privalov, P.L. 1994. Thermodynamics of ubiquitin unfolding. *Proteins* **18:** 246–253.

Wright, C.F., Lindorff-Larsen, K., Randles, L.G., and Clarke, J. 2003. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Biol.* **10:** 658–662.

Wright, C.F., Steward, A., and Clarke, J. 2004. Thermodynamic characterisation of two transition states along parallel protein folding pathways. *J. Mol. Biol.* **338:** 445–451.

Zhang, T., Bertelsen, E., Benvegnu, D., and Alber, T. 1993. Circular permutation of T4 lysozyme. *Biochemistry* **32:** 12311–12318.