Chemistry & Biology, Vol. 10, 1205–1213, December, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.11.017

A New H/D Exchange- and Mass Spectrometry-Based Method for Thermodynamic Analysis of Protein-DNA Interactions

Liyuan Ma and Michael C. Fitzgerald*
Department of Chemistry
Duke University
Durham, North Carolina 27708

Summary

The application of SUPREX (stability of unpurified proteins from rates of H/D exchange) to the thermodynamic analysis of protein-DNA complexes is described. A series of five model protein-DNA complexes involving two known DNA binding proteins, Arc repressor and CopG, were analyzed in order to determine the accuracy, precision, and generality of the SUPREX technique for quantifying the strength of protein-DNA interactions. For protein-DNA complexes that reversibly unfold in a two-state manner, we demonstrate that reasonably precise K_d values in agreement with those determined by conventional techniques can be determined by SUPREX. In the case of protein-DNA complexes that are not well modeled by a two-state unfolding mechanism, we find that relative binding affinities can be determined in the SUPREX experiment.

Introduction

Many important biological processes are mediated by proteins that interact with unique DNA sequences. The thermodynamic analysis of protein-DNA interactions is important for understanding such biological processes. Conventional approaches for characterizing the strength of protein-DNA interactions have involved the use of gel mobility shift assays, fluorescence anisotropy, analytical ultracentrifugation, intercalating dyes, and filter binding assays [1–3]. While these methods have been widely used in a number of studies of protein-DNA binding interactions, there are several disadvantages and experimental limitations to their application. Fluorescence anisotropy [2] and gel mobility shift assays [3], the two most frequently used methods for measuring dissociation constants (K_d values) of protein-DNA complexes, have the disadvantage that they require selective labeling of either the protein or the DNA prior to analysis. An additional disadvantage to the gel mobility shift assay is that the protein-DNA complex is analyzed by moving through a gel matrix and not analyzed directly in solution under true equilibrium conditions. All of the above approaches are also limited to the analysis of highly purified protein-DNA samples, and they are not readily amenable to automation and high-throughput analyses.

Recently, we reported on a new strategy for studying the thermodynamic properties of protein-ligand binding interactions, including those that involve small molecules, nucleic acids, peptides, and other proteins [4]. The strategy involves the use of an H/D exchange- and mass spectrometry-based approach, termed SUPREX (stability of unpurified proteins from rates of H/D exchange), to measure the change in a protein's thermodynamic stability upon ligand binding [5–7]. Ultimately, the measured change in stability (i.e., the binding free energy) is used to calculate the dissociation constant (K_d value) of the protein-ligand complex. Significantly, the SUPREX technique is not subject to the same limitations that are currently associated with conventional approaches for the quantitative analysis of protein-DNA interactions. SUPREX does not require selective labeling of the protein or the DNA with a specific fluorophore or radioactive probe. SUPREX affords the unique ability to perform analyses on unpurified protein samples, such as those found in complex cell lysates [5, 6] or in vivo [8], and the SUPREX protocol is also amenable to automation and high-throughput analyses.

Our SUPREX studies to date have only included the analysis of only one protein-DNA complex, the ternary complex formed between Trp repressor, l-tryptophan, and a 25 bp oligonucleotide [4]. In the case of this complex, SUPREX provided a reasonably accurate and precise quantitative measure of the DNA binding affinity. Here, we report on the detailed application of SUPREX to the thermodynamic study of protein-DNA interactions. We have studied the DNA binding properties of two additional model DNA binding proteins, Arc repressor (ArcR) and CopG. ArcR is a homodimeric DNA binding protein, comprised of 53 amino acid subunits, that represses transcription of the P_{ant} gene during the lytic growth of the bacteriophage P22 [9–11]. ArcR binds specific DNA sequences (i.e., cognate DNA) with high affinity (K_d values on the order of 10^{-9} M) and binds random DNA sequences (i.e., noncognate DNA) with significantly decreased affinities (K_d values in the range 10^{-6} to 10^{-7} M) [12, 13]. CopG is a homodimeric protein with a 45 amino acid polypeptide chain that regulates its own synthesis and that of RepB by binding to the CopG-RepB promoter region [14]. Studies have shown that the K_d value for CopG binding its cognate DNA sequence is on the order of 10^{-9} M [15]. Our experiments on the five protein-DNA complexes in this work serve to define the accuracy, precision, and generality of SUPREX for the thermodynamic analysis of protein-DNA interactions.

Results

The binding affinities of ArcR and of CopG to the DNA fragments outlined in Figure 1 were characterized by SUPREX. The *O1*, *L*, and *R* DNA fragments used in our experiments with ArcR were identical to those used by Sauer and coworkers in earlier studies of ArcR's DNA binding properties [12]. The sequence of the *NS* fragment corresponds to the operator sequence of another transcription factor, the tryptophan repressor, and is unrelated to cognate sequences of ArcR. The *O2* fragment is identical to that used in an earlier structural study of CopG binding to its operator sequence.

*Correspondence: michael.c.fitzgerald@duke.edu

O1: 5'-ATCGATGATAGAAGCACTCTACTATCG-3'

3'-TAGCTACTATCTTCGTGAGATGATAGC-5'

L : 5'-TGACATGATAGAAGCACGT-3'
3'-ACTGTACTATCTTCGTGCA-5'

R: 5'-AGCACTCTACTATATTC-3'
3'-TCGTGAGATGATATAAG-5'

NS: 5'-ATCGAACTAGTTAACTAGTACGCAA-3'
3'-TAGCTTGATCAATTGATCATGCGTT-5'

O2 : 5'-GATTGCATTGAGTGCACGG-3'
3'-CTAACGTAACTCACGTGCC-5'

Figure 1. Nucleotide Sequences of the Double-Stranded DNA Fragments Used in the Binding Assays of This Study

The underlined sequences in each fragment correspond to the nucleotides that directly interact with the protein.

Samples of ArcR and ArcR bound to the *O1*, *L*, *R*, and *NS* DNA fragments as well as samples of CopG and CopG bound to the *O2* DNA fragment were subjected to SUPREX analyses as described in the Experimental Procedures. Shown in Figure 2 is a MALDI mass spectrum typical of those used to generate the SUPREX curves in this work. The major ion signals detected in the spectrum were from the ArcR protein and the protein calibrants. The acidic nature of the MALDI matrix denatures the protein-DNA complex during MALDI sample preparation. Therefore, it is not surprising that an ion signal corresponding to the intact protein-DNA complex was not detected. However, it is important to note that denaturation of the protein-DNA complexes during the MALDI analysis does not interfere with our ability to readout their thermodynamic properties in the solution phase. This is because the H/D exchange rates measured by MALDI in the SUPREX experiment are defined by the solution phase properties of each protein and protein-DNA complex.

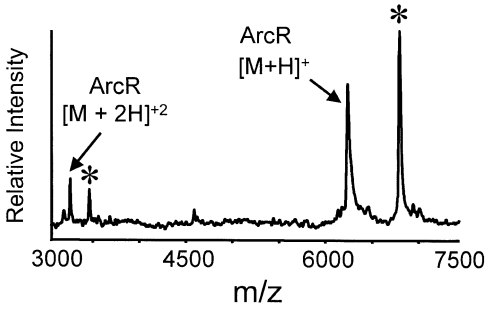

Figure 2. MALDI Mass Spectrum Acquired in the SUPREX Analysis of ArcR in the Presence of Excess *L*

The spectrum represents the data used to generate the 2.1 M urea point in Figure 4A. The ion signal detected for ArcR at m/z 6243.6 indicates a mass gain of 15.3 over the fully protonated molecule. Two peaks labeled with an "*" are attributed to the two calibrants. The resolution, full width half maximum (FWHM), for the ArcR and two calibrant peaks was 222, 342, and 319, respectively.

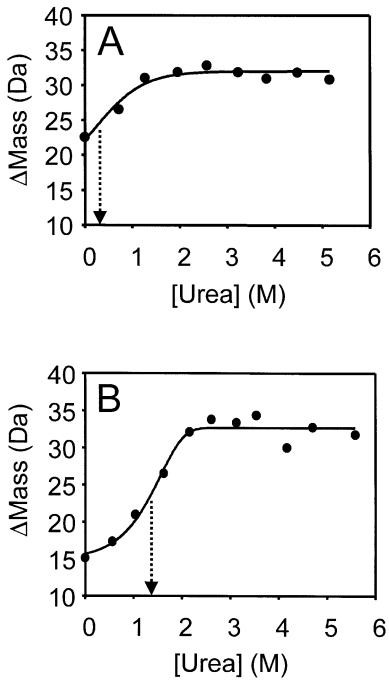

Figure 3. SUPREX Curves Recorded for ArcR and CopG in the Absence of DNA

(A) The ArcR data was acquired with t = 300 s and [P] = 160 μM. (B) The CopG data was acquired with t = 2400 s and [P] = 150 μM. The vertical dashed arrows denote the transition midpoint (i.e., $C^{1/2}_{SUPREX}$ value) of each SUPREX curve.

Shown in Figure 3 are representative SUPREX curves obtained for ArcR and CopG in the absence of DNA, and shown in Figure 4 are representative SUPREX curves obtained for ArcR and CopG complexed with DNA. The $C^{1/2}_{SUPREX}$ values extracted from the SUPREX curves in Figures 3 and 4 are summarized in Table 1. Also included in Table 1 are the $C^{1/2}_{SUPREX}$ values obtained from additional SUPREX curves recorded for each protein and protein-DNA complex (data not shown). The $C^{1/2}_{SUPREX}$ values recorded for each protein and protein-DNA complex varied with exchange time, t, and protein concentration, P. Therefore, in order to directly compare the $C^{1/2}_{SUPREX}$ values obtained for the proteins and protein-DNA complexes in this study, the $C^{1/2}_{SUPREX}$ values in Table 1 were normalized to the same protein concentration (8 μM) and to the same exchange time (3600 s) (see Experimental Procedures). The normalized $C^{1/2}_{SUPREX}$ values we obtained are summarized in Table 1.

Our analysis of CopG, CopG-*O2*, ArcR-*L*, and ArcR-*O1* included the acquisition of SUPREX curves using both different exchange times and different protein concentrations. As predicted by Equation 2 in the Experimental Procedures, the $C^{1/2}_{SUPREX}$ value of each SUPREX curve shifted to a lower [urea] when longer exchange times were used and for a given exchange time the $C^{1/2}_{SUPREX}$ value was shifted to a higher [urea] with increasing protein concentration (see Table 1). Ultimately, the $C^{1/2}_{SUPREX}$ values obtained in these experiments were plotted as a function of t and P, according to Equation 2. Plots were generated in which the oligomeric state of

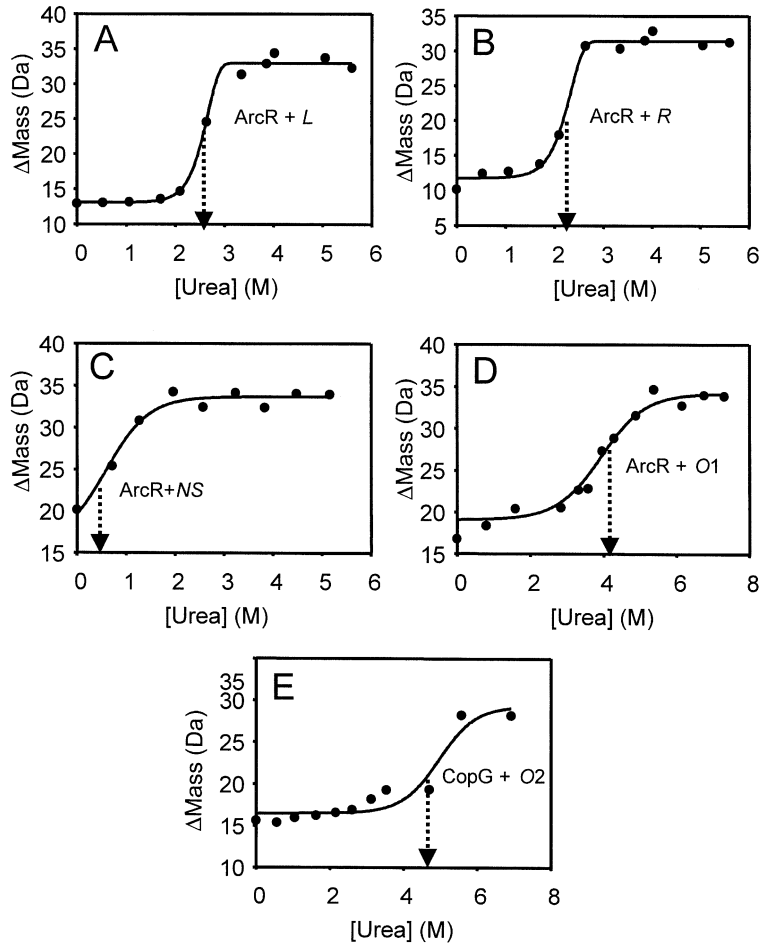

Figure 4. SUPREX Curves Recorded for ArcR and CopG in the Presence of Excess DNA

(A) ArcR-*L* data was acquired with t = 3600 s and [*P*] = 10 μM.

(B) ArcR-*R* curve was acquired with t = 3600 s and [*P*] = 10 μM.

(C) ArcR-*NS* data curve was acquired with t = 300 s and [*P*] = 8 μM.

(D) ArcR-*O1* data curve was acquired with t = 3600 s and [*P*] = 8 μM.

(E) The CopG-*O2* data curve was acquired with t = 3600 s and [*P*] = 16 μM.

The vertical dashed arrows denote the transition midpoint (i.e., $C^{1/2}_{SUPREX}$ value) of each SUPREX curve.

the folded protein, *n*, was set to 1, 2, 3, and 4 (see Figures 5 and 6). The data for CopG and ArcR-*L* were linear (correlation coefficient, >0.9) when *n* was set to 2. The data for the ArcR-*O1* complex and the CopG-*O2* complex were not linear (correlation coefficient, <0.9) for any *n* value we set (1, 2, 3, or 4).

Discussion

Ideal Protein Systems

Ideal protein systems for SUPREX analyses are those systems that exhibit so-called EX2 H/D exchange behavior (i.e., the protein's refolding rate is much faster than the intrinsic exchange rate of an unprotected amide proton) and that are well modeled by two-state, reversible unfolding processes (i.e., no intermediate states are populated). In most cases, the right experimental conditions (i.e., buffer pH and temperature) can be chosen to ensure that the protein under study exhibits EX2 exchange behavior. Under the conditions of the SUPREX experiments performed here, all of the proteins and protein-DNA complexes appeared to exhibit EX2 exchange behavior as judged by the fact that only one population of deuterated protein molecules was detected in the mass spectrometry readout of our SUPREX experiments. If a protein's refolding rate is slower than the intrinsic exchange rate of an unprotected amide proton (so-called EX1 exchange behavior), then two distinct populations of deuterated protein molecules would be detected (one in which all of the globally protected H's were exchanged and one in which none of the globally protected were exchanged).

Of the proteins and protein complexes employed in our study, only ArcR, ArcR-*L*, ArcR-*R*, and ArcR-*NS* have been previously shown to exhibit two-state unfolding behavior [12]. In the case of these complexes, accurate ΔG_f° values can be determined from the $C^{1/2}_{SUPREX}$ values we determined in our SUPREX experiments using Equation 2 [6, 7]. Indeed, the ΔG_f° value of 8.8 ± 0.2 kcal/mol that we determined by SUPREX for ArcR is in reasonable agreement with the literature value of 9.5 ± 0.2 kcal/mol [16]. Comparable ΔG_f° values to those summarized in Table 2 for the ArcR-*R*, ArcR-*L*, and ArcR-*NS* complexes have not been reported in the literature. However, we note that the SUPREX-derived K_d values we determined for these ArcR-DNA complexes using the ΔG_f° values in Table 2 were all within 5-fold of previously reported values obtained using conventional techniques [12]. We note that the literature K_d values reported in Table 2 were determined in gel-shift assays using a binding buffer (10 mM Tris, pH 7.5, 3 mM $MgCl_2$, 0.1 mM EDTA, 100 mM KCl, 100 μg/mL bovine serum albumin,

Table 1. Summary of SUPREX Data Collected in This Study

Proteins	*t* (s)	*P* (μM)[a]	$C^{1/2}_{SUPREX}$ (M)[b]	Normalized $C^{1/2}_{SUPREX}$ (M)[c]	Average $C^{1/2}_{SUPREX}$ (M)[d]	$\Delta C^{1/2}_{SUPREX}$ (M)[e]
ArcR	300	160	0.3	−2.6	−2.6 ± 0.1	—
	300	160	0.3	−2.6		
	300	160	0.4	−2.5		
ArcR + *L*	960	1	1.9	2.3	2.4 ± 0.1	5.0 ± 0.1
	1800	1	1.5	2.3		
	3600	10	2.6	2.4		
	27,060	10	1.5	2.5		
	41,400	10	1.4	2.6		
ArcR + *R*	3600	10	2.2	2.1	2.1 ± 0.1	4.7 ± 0.1
	3600	10	2.3	2.2		
	15,240	10	1.3	2.0		
	41,400	10	1.0	2.2		
ArcR + *NS*	300	8	0.3	−1.0	−0.9 ± 0.1	1.7 ± 0.1
	300	8	0.6	−0.7		
	300	8	0.4	−1.0		
ArcR + *O1*	3900	0.8	4.0	5.3	4.2 ± 0.5	6.8 ± 0.5
	7200	0.8	2.7	4.3		
	18,000	0.8	1.7	3.8		
	19,800	0.8	2.6	4.8		
	45,000	0.8	1.6	4.2		
	3600	8	3.9	3.9		
	17,400	8	3.3	4.2		
	46,560	8	2.7	4.1		
	54,600	8	2.0	3.5		
CopG	300	40	2.2	−1.6	−1.6 ± 0.1	—
	600	40	1.4	−1.7		
	1200	40	1.0	−1.5		
	600	46	1.4	−1.8		
	1800	150	1.8	−1.6		
	2400	150	1.3	−1.7		
CopG + *O2*	1800	1.6	3.5	4.4	4.2 ± 0.4	5.8 ± 0.4
	5400	1.6	3.1	4.9		
	3600	16	5.0	4.3		
	5400	16	4.2	3.9		
	12,600	16	3.8	4.3		
	14,400	16	3.2	3.8		
	19,800	16	2.9	3.8		

[a] Expressed in monomer.
[b] $C^{1/2}_{SUPREX}$ values determined by SUPREX.
[c] Normalized $C^{1/2}_{SUPREX}$ values (i.e., values in column 4 were normalized to same *t* and same [*P*]).
[d] Averaged $C^{1/2}_{SUPREX}$ values in column 5 with standard deviation.
[e] Transition midpoint shift upon DNA binding.

and 0.02% Nonidet NP-40), a loading buffer (50% glycerol in 10 mM Tris, pH 8.0, 1 mM EDTA), and a runing buffer (0.5 × TBE) that was slightly different in composition from the binding buffer and the SUPREX exchange buffers (see Experimental Procedures) used in this work.

A potential complication associated with using SUPREX to measure the thermodynamic stabilities of proteins and protein-DNA complexes is that the required use of deuterium perturbs the equilibrium under study. In particular, deuterons are less effective than protons at forming hydrogen bonds. However, the magnitude of this effect is small and not likely to effect our SUPREX measurements. In fact, the results of studies on several protein-ligand binding reactions have revealed that such isotope effects on ΔG determinations are not measurable by conventional techniques [17]. It has also been noted that even large proteins with 50–100 hydrogen bonds only produce isotope effects on the order <0.5 kcal/mol [18]. Considering the small number of hydrogen bonds that are present at protein-DNA interfaces, it is unlikely that the small decrease in stability expected for deuteron-containing hydrogen bonds will impact the K_d determinations in this work, especially with the precision of the SUPREX-derived $\Delta\Delta G_f^\circ$ values in Table 2, ±0.4 kcal/mol.

One goal of this work was to determine if the SUPREX-based strategy described here for studying the strength of protein-DNA interactions could be used to differentiate between specific and nonspecific DNA binding interactions. Our results with the ArcR-*R*, ArcR-*L*, and ArcR-*NS* complexes in this work indicate that the accuracy and precision of the SUPREX technique are more than adequate to differentiate such binding affinities that typically differ by 2–5 kcal/mol.

The thermodynamic parameters reported in Table 2 were obtained from SUPREX experiments in which urea was used as the chemical denaturant. Other denaturants, such as GdmCl, can also be used for SUPREX analyses. However, in this work we found that the use of urea, as opposed to GdmCl, was critical to the suc-

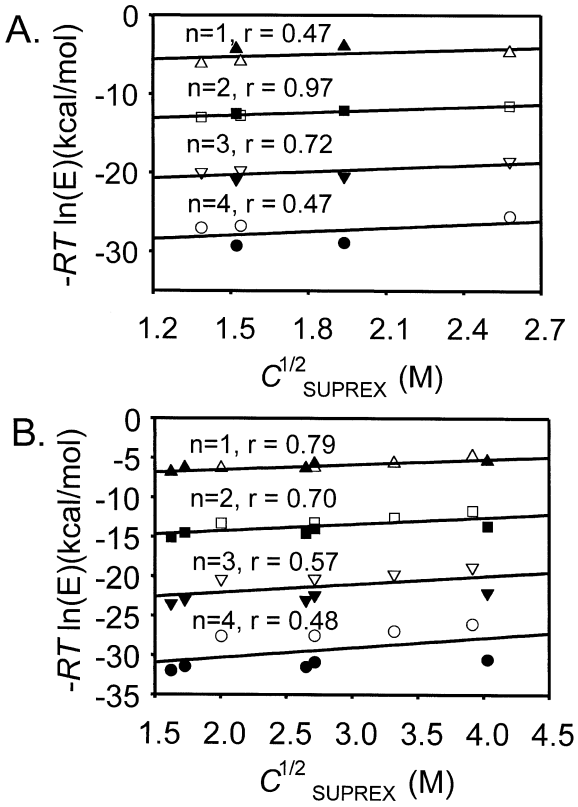

Figure 5. Experimental Determinations of ArcR *n* Values

Plots of $-RT$ ln(E) versus $C^{1/2}{}_{SUPREX}$, where E represents $[((\langle k_{int}\rangle t/0.693) - 1)/(n^n [P]^{n-1}/2^{n-1})]$. Data are shown for (A) ArcR-*L* and (B) ArcR-*O1*. In each case, the solid lines are the results of linear least squares analyses of the data. The upward triangles, squares, downward triangles, and circles represent *n* values of 1, 2, 3, and 4 (respectively). The data points from different concentrations of protein are indicated by filled symbols (lower concentration) and open symbols (higher concentration), respectively. The correlation coefficients for the data points on each line are indicated in the plot.

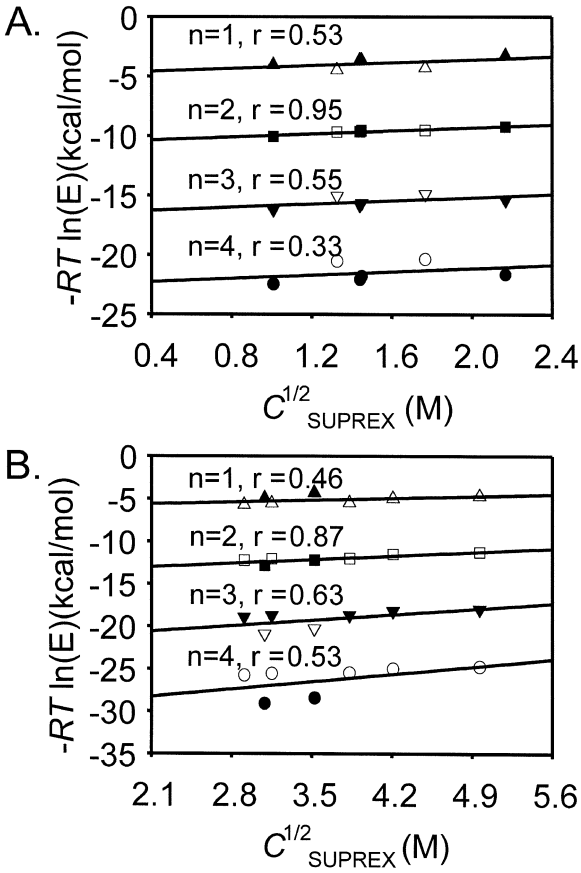

Figure 6. Experimental Determinations of CopG *n* Values

Plots of $-RT$ ln(E) versus $C^{1/2}{}_{SUPREX}$, where E represents $[((\langle k_{int}\rangle t/0.693) - 1)/(n^n [P]^{n-1}/2^{n-1})]$. Data are shown for (A) CopG and (B) CopG-*O2*. In each case, the solid lines are the results of linear least squares analyses of the data. The upward triangles, squares, downward triangles, and circles represent *n* values of 1, 2, 3, and 4 (respectively). The data points from different concentrations of protein are indicated by filled symbols (lower concentration) and open symbols (higher concentration), respectively. The correlation coefficients for the data points on each line are indicated in the plot.

cess of our thermodynamic studies. Accurate ΔG_f° values (i.e., values comparable to those obtained in conventional equilibrium unfolding experiments) could be obtained from SUPREX analyses on each DNA binding protein in the absence of DNA using either urea or GdmCl [6]. However, the DNA-induced stability changes were only detected in our SUPREX experiments when urea was used as the chemical denaturant. The presence of GdmCl in our SUPREX buffers appeared to interfere with the binding reaction between each protein and its cognate DNA. In contrast, the presence of urea in our SUPREX buffers did not appear to have a measurable effect on the binding reaction between each protein and its cognate DNA. This is evidenced by the good agreement between our SUPREX-derived K_d values for the protein-DNA complexes in this study and the K_d values previously reported in the literature for these complexes (see Table 2). We also note that no intrinsic urea dependence to the protein-DNA binding interaction was observed, as the *m* values we determined for the ArcR-DNA complexes in this work were comparable to the *m* value previously reported for ArcR alone (see below).

As part of this work we also found that both the *m* and *n* values in Equation 2 can be determined experimentally from SUPREX data collected on ideal protein systems. Our analysis of the ArcR-*L* complex included the acquisition of a series of different SUPREX curves in which both the exchange time and the protein concentration were varied. A plot of the resulting $C^{1/2}{}_{SUPREX}$ values versus $-RT\ln[((\langle k_{int}\rangle t/0.693) - 1)/(n^n [P]^{n-1}/2^{n-1})]$ was only linear in the case where *n* was assumed to be 2, indicating that the protein was a dimer in its folded state (see Figure 5). These results are consistent with the findings of other biophysical studies on the ArcR-*L* complex in which ArcR has been shown to bind as a dimer to the *L* fragment.

The *m* value of 1.1 kcal/(mol × M), determined from the slope of the $n = 2$ plot in Figure 5A, was also comparable to the *m* value of 1.39 kcal/(mol × M), determined from urea-induced equilibrium unfolding experiment on ArcR [16]. We also note that a $C^{1/2}{}_{SUPREX}$ value versus $-RT\ln[((\langle k_{int}\rangle t/0.693) - 1)/(n^n [P]^{n-1}/2^{n-1})]$ plot of the

Table 2. Thermodynamic Parameters Determined for the Ideal Protein Systems in This Study

Protein Systems	ΔG_f°(kcal/mol)[a]	$\Delta\Delta G_f^\circ$(kcal/mol)[b]	SUPREX K_d[c]	Literature K_d [d]
ArcR	−8.8 ± 0.2	—	—	—
ArcR + *L*	−14.4 ± 0.3	−5.6 ± 0.4	0.3 ± 0.2 nM	0.3 ± 0.2 nM
ArcR + *R*	−14.0 ± 0.3	−5.2 ± 0.4	1.0 ± 0.6 nM	1.5 ± 0.9 nM
ArcR + *NS*	−10.7 ± 0.2	−1.9 ± 0.3	0.3 ± 0.2 μM	1.3 ± 0.4 μM
CopG	−10.6 ± 0.2	—	—	—

[a] Calculated using Equation 2 reported with standard error.
[b] Folding free energy change upon DNA binding and reported with standard error.
[c] Calculated using Equation 3 reported with standard error.
[d] From [12].

ArcR-*R* data in Table 1 yielded a straight line ($r = 0.98$) with a slope identical to that determined for the ArcR-*L* complex. Unfortunately, it was not possible to collect SUPREX curves as a function of protein concentration and exchange time on ArcR alone, as such a high protein concentration (160 μM) and a short exchange time (300 s) was required to produce a measurable $C^{1/2}_{SUPREX}$ value (i.e., a $C^{1/2}_{SUPREX}$ value greater than 0 M urea). SUPREX analyses at increased exchange times and/or at decreased protein concentrations yielded $C^{1/2}_{SUPREX}$ values that were not experimentally accessible (i.e., $C^{1/2}_{SUPREX} < 0$ M urea) due to the relatively low thermodynamic stability of ArcR.

CopG is more stable than ArcR, and it was possible to generate SUPREX curves and determine $C^{1/2}_{SUPREX}$ values for CopG using different exchange times and protein concentrations. Our plot of the resulting $C^{1/2}_{SUPREX}$ values versus $-RT\ln[((<k_{int}>t/0.693) - 1)/(n^n\ [P]^{n-1}/2^{n-1})]$ was also only linear in the case where *n* was assumed to be 2. This is consistent with the findings of other structural studies on CopG in which the folded protein was found to be a dimer [19]. The good fit of our CopG data to Equation 2 using $n = 2$ also suggests that the SUPREX transition is well modeled by a two-state unfolding mechanism involving folded dimer and unfolded monomer. However, additional urea-induced equilibrium unfolding experiments using other spectroscopic probes are clearly needed to confirm the two-state folding properties of CopG.

Nonideal Protein Systems

Earlier biophysical studies on the ArcR-*O1* complex have established that the folding and assembly reaction of this complex is not two state [12]. These earlier biophysical studies concluded that ArcR binds to the *O1* DNA fragment as a tetramer and that the folding and assembly reaction involves partially folded intermediate states in which one ArcR dimer is bound to individual half-sites of the operator sequence (i.e., either the left or right half-site). Despite the multistate folding properties of the ArcR-*O1* complex, we were able to record SUPREX curves for this complex (see Figure 4D). However, the SUPREX data we collected was different in two respects from that collected on the ideal protein systems described above. First, the transition regions of the SUPREX curves were very broad. They spanned almost 2 M urea as compared to about 1 M urea for the ArcR-*R* and ArcR-*L* complexes, for example (see Figure 4), and second, the protein concentration-dependent data were not well fit to Equation 2 using *n* values of 1, 2, 3, or 4 (see Figure 5B). It is also noteworthy that the SUPREX data we obtained for the CopG-*O2* complex is very similar to that obtained for the ArcR-*O1* complex (i.e., the SUPREX transition is broad and the data are not well fit to Equation 2 using *n* values of 1, 2, 3, or 4). These results on ArcR-*O1* and CopG-*O2* suggest that such broad SUPREX transitions and the lack of a good fit of the data to Equation 2 using whole integers for *n* are indications of nonideal behavior in the SUPREX experiment.

Unfortunately, accurate ΔG_f° values can not be extracted from SUPREX curves on nonideal protein systems using Equation 2. However, our data suggests that the $\Delta C^{1/2}_{SUPREX}$ values obtained on the nonideal protein systems in this study can serve as good qualitative measures of DNA binding affinity. For protein-ligand systems of similar size (i.e., with similar *m* values) the magnitude of a $\Delta C^{1/2}_{SUPREX}$ can be directly related to binding affinity. Note that the $\Delta C^{1/2}_{SUPREX}$ value obtained for ArcR-*O1* is larger than the $\Delta C^{1/2}_{SUPREX}$ value measured for the ArcR-*L* and ArcR-*R* complexes. This is consistent with an earlier finding that cooperative interactions between ArcR dimers contribute ~5 kcal/mol to the stabilization of the ArcR tetramer on the *O1* DNA fragment. We also note that the $\Delta C^{1/2}_{SUPREX}$ value determined for ArcR-*O1* is slightly larger than that obtained for CopG-*O2*. This suggests that the binding affinity of ArcR to the *O1* fragment is greater than the binding affinity of CopG to the *O2* fragment. This conclusion is consistent with the results of conventional biophysical studies in which the K_d value of ArcR-*O1* has been calculated to be 0.2 nM, whereas the K_d value of the CopG-*O2* complex has been estimated to be on the order of 1 nM [15].

Significance

We have described a new analytical technique for the thermodynamic analysis of protein-DNA interactions in solution. The technique involves the application of a H/D exchange and MALDI mass spectrometry method, termed SUPREX, to measure the change in a protein's thermodynamic stability upon DNA binding. For protein-DNA complexes that reversibly unfold in a two-state fashion (i.e., partially folded intermediate states are not populated), the stability changes measured by SUPREX can be used to calculate reasonably accurate dissociation constants (i.e., K_d values). For protein-DNA complexes that unfold in a multistate fashion (i.e.,

partially folded intermediate states are populated), the SUPREX method enables the evaluation of relative binding affinities of different protein-DNA complexes.

Important experimental advantages of the described SUPREX-based method over conventional methods for characterizing the thermodynamic properties of protein-DNA complexes include (1) the ability to analyze samples in complex mixtures, (2) the ability to perform measurements in a high-throughput and automated fashion, and (3) the ability to analyze protein-DNA complexes directly in solution without the attachment of covalent labels to the protein or DNA.

Experimental Procedures

Reagents

Deuterium oxide (99.9 atom %D), sodium deuteroxide, and deuterium chloride were purchased from Aldrich (Milwaukee, WI). Deuterated phosphoric acid was obtained from Cambridge Isotope Laboratories (Andover, MA), and urea was obtained from Mallinckrodt Baker, Inc. (Paris, KY). Sinapinic acid (SA) was either from Acros Organics (Pittsburgh, PA) or from Aldrich. Trifluoroacetic acid (TFA) was obtained from Halocarbon (River Edge, NJ) and acetonitrile (MeCN) was from Fisher (Pittsburgh, PA). The protein mass standards (bovine ubiquitin and insulin) were purchased from Sigma (St. Louis, MO). The protein mass standard 4-oxalocrotonate tautomerase (4-OT) was chemically synthesized in house.

General Methods and Instrumentation

MALDI mass spectra were acquired on a Voyager DE Biospectrometry Workstation (PerSeptive Biosystems, Inc, Framingham, MA) in the linear mode using a nitrogen laser (337 nm). SUPREX samples were prepared for MALDI analysis as described below. SA was used as the matrix in all of our MALDI analyses. Spectra for all proteins were collected in the positive ion mode, and all experiments utilized an acceleration voltage of 25 kV, a grid voltage between 23 and 24 kV, a guide wire voltage of 75 V, and a delay time of 225 ns. Each spectrum obtained was the sum of 32 laser shots. The raw intensity versus time data in each mass spectrum was with an in-house Microsoft Excel macro that performed the following operations: a smoothing of the data using a 19-point floating average, a two-point mass calibration using the ion signals from the protein calibrants, and center of mass determination for the $[M + H]^+$ ion signal of the analyte.

UV/Vis absorbance measurements were recorded on a Hewlett Packard 8452A Diode Array UV/Vis Spectrophotometer. Urea concentrations were determined by refractive index measurements using a Bausch & Lomb (Rochester, NY) refractometer [20]. All pH measurements were recorded using a Jenco 6072 pH meter (San Diego, CA) equipped with a Futura calomel pH electrode from Beckman Instruments (Fullerton, CA). Deuterated exchange buffer pD values were determined from pH measurements by adding 0.4 to the measured pH [21].

Protein-DNA Samples

The 53 and 45 amino acid polypeptide chains of ArcR and CopG (respectively) were obtained by total chemical synthesis using highly optimized SPSS protocols for Boc-chemistry as described elsewhere [22, 23]. The crude synthetic product from each synthesis was purified by RP-HPLC using a C18 Vydac column (2.2 × 12.0 cm, 300 Å) under the following conditions: 3 mL/min flow rate and 30%–50% linear gradient of buffer B in buffer A (buffer A = 0.1% TFA in water, and buffer B = 90% acetonitrile in water containing 0.09% TFA). Pure fractions (as judged by electrospray ionization mass spectrometry) were pooled, frozen, and lyophilized to a dry white solid. The pure, lyophilized protein product from each synthesis was folded by dissolution of each sample (~5 mg/ml) in a folding buffer that contained 50 mM TrisHCl, 100 mM KCl, and 0.2 mM EDTA (pH 7.4). While the protein samples in this work were prepared by total chemical synthesis, we note that proteins obtained by other methods (e.g., by recombinant DNA techniques) are also amenable to SUPREX analyses.

The oligonucleotides used in these experiments were purchased from Integrated DNA Technologies, Inc. and used without further purification. Duplex DNAs were formed by mixing equimolar amounts of the appropriate complementary single-stranded oligonucleotides in an annealing buffer (10 mM TrisHCl, pH 7.5, 100 mM KCl, 3 mM $MgCl_2$, and 0.1 mM EDTA), incubating at an elevated temperature (~80°C) for 5 min, and then cooling down the solution to room temperature.

Stock solutions of the protein-DNA complexes were prepared by mixing each folded protein sample with the desired ratio of duplex DNA, diluting the sample in folding buffer, and then equilibrating the resulting solution for at least 4 hr. The concentration of protein (in *n*-mer equivalents) in these stock solutions was typically between 8 to 160 μM, and the concentration of DNA was typically between 30 to 200 μM. The DNA concentration in these stock solutions was also in excess of the protein concentration. The buffer composition was essentially that of the folding buffer. Ultimately, the protein-DNA samples in these stock solutions were subjected to SUPREX analyses as described below.

SUPREX Data Collection and Analysis

The protocol we used for normal SUPREX analyses was similar to that described previously [6]. In brief, H/D exchange reactions were initiated by 10-fold dilution of the above stock solutions of protein or protein-DNA samples into a series of deuterated exchange buffers. The series of deuterated exchange buffers used in these experiments contained 20 mM phosphate (pD 6.0), 100 mM KCl, and concentrations of urea that varied between 0 and 8 M. After a specific exchange time, a 1 ml aliquot of each protein-containing exchange buffer was combined with 9 ml of an ice-cold MALDI matrix solution that included the proteins used as internal mass standards. The matrix solution consisted of a saturated, aqueous solution of sinapinic acid (SA) containing 45% MeCN and 0.1% TFA (pH ~3.0). The low temperature and low pH of the matrix solution effectively quenched the H/D exchange reaction. Subsequently, 1 ml of the exchange solution was spotted on a stainless steel MALDI plate and the solvent was evaporated under a gentle flow of air. A total of ten replicate spectra were collected from various regions of the crystalline MALDI sample spot to determine the mass change (relative to a fully protonated sample) at each concentration of urea. The mass change was determined by averaging results from the ten spectra.

Ultimately, the mass change relative to a fully protonated sample was plotted as a function of [urea], and the data were fit to the following four-parameter sigmoidal equation using a nonlinear regression routine in Sigma Plot™ in order to extract a $C^{1/2}_{SUPREX}$ value.

$$\Delta Mass = \Delta M_0 + \frac{a}{1 + e^{-\frac{[Denaturant] - C^{1/2}_{SUPREX}}{b}}} \quad (1)$$

In Equation 1, ΔM_o is the change in mass measured before the globally protected hydrogens in the protein exchanged with deuterons (i.e., the pretransition baseline), *a* is the amplitude of the curve in Da, [*Denaturant*] is the denaturant concentration, $C^{1/2}_{SUPREX}$ is the [*Denaturant*] at the transition midpoint of the curve, and *b* is a parameter that describes the steepness of the transition. In fitting our Δ*Mass* versus [*Denaturant*] data sets to Equation 1, all of the parameters in the equation were typically allowed to "float."

Equation 2 was used to calculate ΔG_f° values from the $C^{1/2}_{SUPREX}$ values generated in this work.

$$-RT\left[\ln\frac{\left(\frac{\langle k_{int}\rangle t}{0.693} - 1\right)}{\left(\frac{n^n}{2^{n-1}}[P]^{n-1}\right)}\right] = mC^{1/2}_{SUPREX} + \Delta G_f^\circ \quad (2)$$

In Equation 2, *m* is defined as $\delta\Delta G_f^\circ/\delta$[urea] and can be estimated based on the number of residues in a protein [24] or determined experimentally as described below, *R* is the gas constant, *T* is the temperature in Kelvin, $\langle k_{int}\rangle$ is the average intrinsic exchange rate of an unprotected amide proton that can be estimated by using the

SPHERE program [25, 26], n is the number of protein oligomers involved in the folding reaction, t is the H/D exchange time, and [P] is the protein concentration expressed in n-mer equivalents.

In all of our ΔG_f° value calculation, values for R, T, [P], and t were predefined, and $<k_{int}>$ values were estimated using SPHERE [25, 26]. In the case of ArcR-*L* and CopG, both the m and n values used in our calculations were experimentally determined by collecting a series of SUPREX curves using different H/D exchange times (t) and different protein concentrations ([P]), by extracting $C^{1/2}{}_{SUPREX}$ values from each curve, and by generating plots of $C^{1/2}{}_{SUPREX}$ versus $-RT\ln[((<k_{int}>t/0.693) - 1)/(n^n\,[P]^{n-1}/2^{n-1})]$. Ultimately, a linear-least squares analysis of the data in these plots yielded the equation of a line in which the y intercept corresponded to ΔG_f° and the slope corresponded to the m value (see Equation 2). In the case of ArcR-*R*, only the m value was experimentally determined, and an n value of 2 was assigned to the complex based on the experimental results obtained for ArcR-*L*. Experimental determinations of m and n values were not possible in the case of ArcR and ArcR-*NS* (see text); therefore, the experimentally determined values obtained for ArcR-*L* were used in our ΔG_f° values calculations for these two complexes.

Equation 2 was also used to normalize the $C^{1/2}{}_{SUPREX}$ values we obtained for the protein and protein-DNA complexes in this work. This was accomplished in the following manner. The $C^{1/2}{}_{SUPREX}$ value we experimentally determined at each protein concentration and exchange time was used in Equation 2 to calculate a ΔG_f° value. This ΔG_f° value was then used in Equation 2 to calculate a $C^{1/2}{}_{SUPREX}$ value when the exchange time and protein concentration were set to 3600 s and 8 μM, respectively. For comparative purposes, the normalized $C^{1/2}{}_{SUPREX}$ values reported in Table 2 were all calculated using the same n value ($n = 2$). We note that the use of different n values (i.e., 1, 3, 4, etc.) in our normalization procedure only changed the magnitude of individual $C^{1/2}{}_{SUPREX}$ values; the relative magnitudes of the resulting $C^{1/2}{}_{SUPREX}$ values were unchanged as long as the same n value was used for each complex.

The ΔG_f° value we determined for each protein and protein-DNA complex were used to derive a $\Delta\Delta G_f^\circ$ value for each protein-DNA binding reaction. These $\Delta\Delta G_f^\circ$ values were used in Equation 3 to calculate a K_d value for each protein-DNA complex [27, 28].

$$K_d = [L]/(e^{-\Delta\Delta G_f^0/NRT} - 1) \quad (3)$$

In Equation 3, [L] is the concentration of free DNA, N is the number of independent binding sites, and $\Delta\Delta G_f^\circ$ is the change in folding free energy upon DNA binding. For the model protein systems in this study, $N = 1$ (i.e., there is one DNA binding site in each protein-DNA complex studied here). In our experiments, [L] was calculated using Equation 4 [29].

$$[L] = L_{total} - \frac{P_{total} + L_{total} + K_d - \sqrt{(P_{total} + L_{total} + K_d)^2 - 4P_{total}L_{total}}}{2} \quad (4)$$

In Equation 4, L_{total} is the total concentration of DNA, P_{total} is the total concentration of protein, and K_d is the dissociation constant for the complex. In our experiments on the protein-DNA complex in this work, the [L] was always in the range of 3–6 μM.

Acknowledgments

Financial support for this work was from a PECASE Award (NSF-CHE-00-94224) and an ASMS Research Award (to M.C.F.). The authors thank Kendall D. Powell for helpful discussion throughout the course of this work. We are also grateful to Thomas E. Wales who prepared the ArcR and CopG constructs used in this work.

Received: July 18, 2003
Revised: September 5, 2003
Accepted: September 23, 2003
Published: December 19, 2003

References

1. Winzor, D.J., and Sawyer, W.H. (1995). Quantitative Characterization of Ligand Binding (New York: Wiley-Liss).
2. Hill, J.J., and Royer, C.A. (1997). Fluorescence approaches to study of protein-nucleic acid complexation. Methods Enzymol. *278*, 390–416.
3. Carey, J. (1991). Gel retardation. Methods Enzymol. *208*, 103–117.
4. Powell, K.D., Ghaemmaghami, S., Wang, M.Z., Ma, L., Oas, T.G., and Fitzgerald, M.C. (2002). A general mass spectrometry-based assay for the quantitation of protein-ligand binding interactions in solution. J. Am. Chem. Soc. *124*, 10256–10257.
5. Ghaemmaghami, S., Fitzgerald, M.C., and Oas, T.G. (2000). A quantitative, high-throughput screen for protein stability. Proc. Natl. Acad. Sci. USA *97*, 8296–8301.
6. Powell, K.D., Wales, T.E., and Fitzgerald, M.C. (2002). Thermodynamic stability measurements on multimeric proteins using a new H/D exchange- and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry-based method. Protein Sci. *11*, 841–851.
7. Powell, K.D., and Fitzgerald, M.C. (2003). Accuracy and precision of a new H/D exchange- and mass spectrometry-based technique for measuring the thermodynamic properties of protein-peptide complexes. Biochemistry *42*, 4962–4970.
8. Ghaemmaghami, S., and Oas, T.G. (2001). Quantitative protein stability measurement in vivo. Nat. Struct. Biol. *8*, 879–882.
9. Breg, J.N., Boelens, R., George, A.V.E., and Kaptein, R. (1989). Sequence-specific ^{1}H NMR assignment and secondary structure of the Arc repressor of bateriophage P22, as determined by two-dimensional ^{1}H NMR spectrometry. Biochemistry *28*, 9826–9833.
10. Raumann, B.E., Rould, M.A., Pabo, C.O., and Sauer, R.T. (1994). DNA recognition by β-sheets in the Arc repressor-operator crystal structure. Nature *367*, 754–757.
11. Vershon, A.K., Youderian, P., Susskind, M.M., and Sauer, R.T. (1985). The bacteriophage P22 arc and mnt repressors: overproduction, purification, and properties. J. Biol. Chem. *260*, 12124–12129.
12. Brown, B.M., and Sauer, R.T. (1993). Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers. Biochemistry *32*, 1354–1363.
13. Brown, B.M., Bowie, J.U., and Sauer, R.T. (1990). Arc repressor is tetrameric when bound to operator DNA. Biochemistry *29*, 11189–11195.
14. Del Solar, G.H., Perez-Martin, J., and Espinosa, M. (1990). Plasmid pLSI-encoded RepA protein regulates transcription from *repAB* promoter by binding to a DNA sequence containing a 13-base pair symmetric element. J. Biol. Chem. *265*, 12569–12575.
15. Del Solar, G., Albericio, F., Eritja, R., and Espinosa, M. (1994). Chemical synthesis of a fully active transcriptional repressor protein. Proc. Natl. Acad. Sci. USA *91*, 5178–5182.
16. Milla, M.E., and Sauer, R.T. (1994). Arc repressor: folding kinetics of a single-domain, dimeric protein. Biochemistry *33*, 1125–1133.
17. Oas, T.G., and Toone, E.J. (1997). Thermodynamic solvent isotope effects and molecular hydrophobicity. Adv. Biophys. Chem. *6*, 1–52.
18. Krantz, B.A., Moran, L.B., Kentsis, A., and Sosnick, T.R. (2000). D/H amide kinetic isopote effects reveal when hydrogen bonds form during protein folding. Nat. Struct. Biol. *7*, 62–71.
19. Gomis-Ruth, F.X., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., Gonzalez, A., Espinosa, M., Del Solar, G., and Coll, M. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. EMBO J. *17*, 7404–7415.
20. Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods Enzymol. *131*, 266–280.
21. Glasoe, P.K., and Long, F.A. (1960). Use of glass electrodes to measure acidities in deuterium oxide. J. Phys. Chem. *64*, 188–190.
22. Wales, T.E., and Fitzgerald, M.C. (2001). The energetic contribution of backbone-backbone hydrogen bonds to the thermody-

namic stability of a hyperstable P22 arc repressor mutant. J. Am. Chem. Soc. *123*, 7709–7710.
23. Schnolzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S.B.H. (1992). Highly optimized solid phase peptide synthesis protocols for Boc-chemistry. Int. J. Pept. Protein Res. *40*, 180–193.
24. 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.
25. Zhang, Y.-Z. (1995). Structural biology and molecular biophysics, thesis, University of Pennsylvania, Philadelphia, Pennsylvania.
26. Bai, Y., Milne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins *17*, 75–76.
27. Schellman, J. (1975). Macromolecular binding. Biopolymers *14*, 999–1018.
28. Pace, C.N., and McGrath, T. (1980). Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. J. Biol. Chem. *255*, 3862–3865.
29. Segel, I.H. (1975). Enzyme Kinetics. (New York: John Wiley & Sons).