Urea−Amide Preferential Interactions in Water: Quantitative Comparison of Model Compound Data with Biopolymer Results Using Water Accessible Surface Areas

Jonathan G. Cannon,[†] Charles F. Anderson,[‡] and M. Thomas Record, Jr.*[,‡,§]

Biophysics Program, Chemistry Department, and Biochemistry Department, University of Wisconsin−Madison, Madison, Wisconsin 53706

Received: March 13, 2007; In Final Form: April 23, 2007

Two fundamentally different thermodynamic approaches are in use to interpret or predict the effects of urea on biopolymer processes: one is a synthesis of transfer free energies obtained from measurements of the effects of urea on the solubilities of small, model compounds; the other is an analysis of preferential interactions of urea with a range of folded and unfolded biopolymer surfaces. Here, we compare the predictions of these two approaches for the contribution of urea−amide (peptide) interactions to destabilization of folded proteins by urea. For these comparisons, we develop independent thermodynamic analyses of osmometric and solubility data characterizing interactions of a model compound with urea (or any other solute) and apply them to all five model compounds (glycine, alanine, diglycine, glycylalanine, and triglycine) where both isopiestic distillation (ID) and solubility data in aqueous urea solutions are available. We use model-independent expressions to calculate μ_{23}^{ex}, the derivative of the "excess" chemical potential of solute "2" (either a model compound or a biopolymer) with respect to the molality of solute "3" (urea). Analyses of ID data for these systems reveal significant dependences of μ_{23}^{ex} on both m_2 and m_3, which must be taken into account in making comparisons with values of μ_{23}^{ex} obtained from solubility studies or from analyses of urea−biopolymer preferential interactions. Values of μ_{23}^{ex} calculated from model compound ID data at low m_2 and m_3 are directly proportional to the amount of polar amide (N, O) surface area, and not to any other type of surface. The proportionality constant in this limit, $\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}) = (1.0 \pm 0.1) \times 10^{-3}$ Å^{-2}, is very similar to that previously obtained by analysis of urea−biopolymer preferential interactions ($(1.4 \pm 0.3) \times 10^{-3}$ Å^{-2}). This level of agreement for amide surface in the low concentration limit, as well as the absence of any significant preferential interaction of urea with Gly and Ala, reinforces the conclusion that the primary preferential interaction of urea with protein surface is a favorable interaction (resulting in local accumulation of urea) at polar amide surface, located mostly on the peptide backbone. However, μ_{23}^{ex} for interactions of urea with these model amides is found from both ID and solubility data to be urea concentration-dependent, in contrast to the urea concentration independence of the analogous quantity for protein unfolding.

Introduction

Two thermodynamic strategies have been developed to analyze quantitatively the interactions of urea and other biochemical solutes with biopolymers. Both have been applied to predict and interpret the effects of changes in concentration of solutes on structural changes in biopolymer processes such as protein unfolding and protein−nucleic acid interactions. One approach is based on a synthesis of solubility data for model compounds (amino acids, small peptides).[1−6] These data are compared with the corresponding data for glycine (to account for contributions of $-NH_3^+$ and $-CO_2^-$ groups common to all amino acid monomers) in an analysis that yields "free energies of transfer" of the peptide backbone ($-CH_2CONH-$) and the various amino acid side chains from water to a solution containing urea or other solutes. Transfer free energies are weighted by the changes in water accessible surface area (ASA) of these groups in the process of interest and, by assuming additivity, used to predict the effects of solutes on protein unfolding or other protein processes.[2,4]

In this paper, we further develop an alternative approach to predict or interpret effects of urea and other solutes on biopolymer processes based on an analysis of thermodynamic data characterizing interactions of these solutes with biopolymer surfaces that differ widely in surface area and surface composition Thermodynamic inputs for this analysis include the preferential interaction coefficient[7−12] Γ_{μ_3} or, nearly equivalent, the excess chemical potential derivative $\mu_{23}^{ex} \equiv \partial \ln\gamma_2/\partial m_3$ (see eq 2 below). The model-independent thermodynamic results of this analysis are correlated with a coarse-grained description of biopolymer surface, where each "unified" atom (i.e., a C, N, or O atom with any covalently bonded hydrogen atoms) is classified as either cationic, anionic, polar amide, other polar, or nonpolar. An ASA-based analysis is used to quantify the preferential interaction of the solute with a unit area of each of these different types of biopolymer surface.[7−9] Preferential interactions normalized per unit ASA are interpreted quantitatively using a two-state (local, bulk) partitioning model[10,13,14] for the distribution of the solute between the local region defined

* Corresponding author. 433 Babcock Drive. Phone: (608) 262-5332. Fax: (608) 262-3453. E-mail: record@biochem.wisc.edu.
[†] Biophysics Program.
[‡] Chemistry Department.
[§] Biochemistry Department.

10.1021/jp072037c CCC: $37.00
Published on Web 07/21/2007

by the water of hydration of the biopolymer surface of interest and the remainder of the aqueous solution (the bulk phase).

Urea *m*-values (see eq 1 below) quantifying the effect of urea on unfolding of homologous series of globular proteins[15] and RNA[16] are observed to correlate with the total ASA exposed in unfolding. To interpret such correlations, we obtained and analyzed thermodynamic data characterizing preferential interactions of urea with a range of folded and unfolded protein and nucleic acid surfaces[9] and deduced that the only strong preferential interaction of urea is with polar amide surface of proteins and the corresponding functional groups of nucleic acid bases.[7,9−11] We also concluded that the local concentration of urea in the vicinity of polar amide surface exceeds its bulk concentration by approximately 70% (local-bulk partition coefficient, $K_p \equiv m_3^{loc}/m_3^{bulk} = 1.71 \pm 0.15$, calculated assuming that the hydration per unit ASA (b_1) of that surface is a monolayer: $b_1 = 0.11$ H_2O Å^{-2}).[7] No other correlations between biopolymer surface types and urea−biopolymer preferential interactions were observed; we therefore concluded that the local concentration of urea in the vicinity of other types of protein or native DNA surface (including charged surface, other (nonamide) polar surface, and nonpolar surface) is indistinguishable from its bulk concentration.

From the above analysis, we obtained a general relationship between the urea concentration dependence of the observed equilibrium constant K_{obs} (i.e., the equilibrium concentration quotient) of a biopolymer process and the change in water accessible polar amide surface area ($\Delta ASA_{polar\ amide}$) in the process:[7]

$$\left(\frac{\partial \ln K_{obs}}{\partial c_3}\right)_{m_2,T,P} = (1.4 \pm 0.3) \times 10^{-3}\, \Delta ASA_{polar\ amide} = \frac{\text{urea } m\text{-value}}{RT} \quad (1)$$

where the *m*-value is the derivative of the negative of the observed standard free energy change with respect to solute molarity (see eq 4 below). To test eq 1 by applying it to a structurally well-characterized protein−DNA interaction, the effect of low (nondenaturing) concentrations of urea on the binding of lac repressor to lac operator DNA was quantified.[7] Urea weakens the interaction: $\ln K_{obs}$ decreases linearly with increasing urea molarity. From the urea *m*-value of repressor−operator binding (-1.2 ± 0.1 kcal M^{-1} when corrected to constant salt activity), we deduced that this interaction buries 1500 ± 300 Å^2 of polar amide surface. This agrees quantitatively with structural analysis[7] which predicts that approximately 500 Å^2 of polar amide surface is buried in each of the three interfaces formed in complexation: the operator−DNA interface, the folding of the hinge region, and the interface between the assembled DNA binding domain and the core repressor. Urea was subsequently used as a quantitative probe of burial of polar amide surface in the steps of forming a transcriptionally competent promoter complex between *E. coli* RNA polymerase and promoter DNA.[17] The primary effect of urea was inhibition of the final step in open complex formation; from the urea dependence of K_{obs} and eq 1, we deduced that 100−150 residues of RNA polymerase must fold in this step to clamp the "downstream" region of promoter DNA.[17]

From analyses of the effects of urea on the solubility of amino acids, small oligopeptides, and related compounds,[1,5,18] Bolen and co-workers[4] concluded that the primary interaction of urea with the surface exposed in unfolding a globular protein is a favorable interaction with the peptide backbone. In addition, they concluded that a range of favorable and unfavorable interactions with various side chains are also significant. By contrast, no significant preferential interactions of urea with any type of surface other than polar amide surface have been detected in our previous analyses of urea−biopolymer interactions.[7,9]

Osmolality (proportional to the logarithm of water activity) measurements by isopiestic distillation (ID)[19−21] on two and three component solutions of urea and one of five amino acids (mono-, di-, and triglycine, alanine, and glycylalanine) were introduced as a less approximate thermodynamic alternative to solubility measurements. Free energies of transfer of these amino acids from water to urea calculated from their ID data by Schönert and co-workers[19,20] differed significantly under most conditions from the extant free energies of transfer from solubility data.[1]

While previous analyses[1,4,20] of the model compound thermodynamic data (both ID and solubility) have reported transfer free energies, our analyses of solute−biopolymer interactions and of *m*-values characterizing solute effects on protein folding and other biopolymer processes utilize another thermodynamic quantity which provides a simple, direct comparison among the different sources of data quantifying solute−biopolymer interactions. This thermodynamic quantity is the derivative of the excess part of the biopolymer or model compound chemical potential ($\mu_2^{ex} \equiv RT \ln\gamma_2$ where γ_2 is the activity coefficient defined on the mole fraction scale) with respect to the perturbing solute molality (m_3):

$$\mu_{23}^{ex} \equiv (\partial\mu_2^{ex}/\partial m_3)_{T,P,m_2} \equiv RT(\partial \ln \gamma_2/\partial m_3)_{T,P,m_2} \quad (2)$$

The quantity μ_{23}^{ex} is closely related to the preferential interaction coefficients $\Gamma_{\mu 3}$ and $\Gamma_{\mu 1,\mu 3}$:[12,14,22−27]

$$\Gamma_{\mu_3} = -\mu_{23}/\mu_{33} \cong \Gamma_{\mu_1,\mu_3} \quad (3)$$

where $\mu_{33} \equiv (\partial\mu_3/\partial m_3)_{m_2,P,T}$. [For eq 3, at the concentrations investigated experimentally ($m_2 \ll 0.01$ *m*), μ_{23}^{ex} is numerically indistinguishable from μ_{23} for biopolymers (but not for many of the solutes used as model compounds)].

In this paper, we separately analyze solubility and ID data for the set of four amino acids and peptides for which both kinds of data are available in order to obtain quantitative information about μ_{23}^{ex} as a function of m_2 and m_3. Values of μ_{23}^{ex} evaluated as functions of m_2 and m_3 from ID measurements are compared with the corresponding quantities obtained from solubility measurements for saturated solutions of component 2 as a function of m_3. This comparison provides the first explicit, rigorous test of the quantitative accuracy of solubility for determining thermodynamic effects of urea interactions with amino acids in the low amino acid concentration limit relevant for comparison with biopolymer processes. In addition, we calculate values of μ_{23}^{ex}/ASA from ID studies of urea−amino acid interactions and of $\Delta\mu_{23}^{ex}/\Delta ASA$ for differences between pairs of the amino acids for comparisons with corresponding quantities obtained from analyses of biopolymer data.

Background and Analytical Methods. Chemical potential derivatives $\mu_{23}^{ex} \equiv RT(\partial \ln\gamma_2/\partial m_3)_{m_2}$ are the fundamental thermodynamic quantities that, for the purposes of this paper, are used to analyze or predict the effect of a solute (component 3) on a process involving a biopolymer. (Temperature and pressure are constant in these and all subsequent partial derivatives.) For the process of unfolding a protein (component 2) in excess urea (component 3, $m_3 \gg m_2$):

$$\text{urea } m\text{-value} \equiv -\left(\frac{\partial \Delta G^{\circ}_{\text{unfold}}}{\partial c_3}\right)_{m_2} \equiv RT\left(\frac{\partial \ln K^{\text{unfold}}_{\text{obs}}}{\partial c_3}\right)_{m_2} = RT\left(\frac{\partial \ln K^{\text{unfold}}_{\text{obs}}}{\partial m_3}\right)_{m_2}\left(\frac{\partial m_3}{\partial c_3}\right)_{m_2} \quad (4a)$$

$$= -RT\left(\frac{\partial \ln(\gamma_{2\text{unfolded}}/\gamma_{2\text{folded}})}{\partial m_3}\right)_{m_2}\left(\frac{\partial m_3}{\partial c_3}\right)_{m_2} \equiv -\Delta\mu^{\text{ex}}_{23}\left(\frac{\partial m_3}{\partial c_3}\right)_{m_2} \quad (4b)$$

where $K^{\text{unfold}}_{\text{obs}} \equiv ([\text{unfolded}]/[\text{folded}])_{\text{eq}}$. The designation component 2 is used for the protein in both folded and unfolded states. Details of the derivation of equations analogous to 4a,b have been given previously.[28] Protein activity coefficients, γ_2, are defined here on the mole fraction scale, though the ratio of protein activity coefficients for the unfolding process is the same on any concentration scale. Derivatives with respect to molarity (c_3) have typically been used in analyzing the effect of urea on protein stability, presumably because of the observed constancy of the m-value on this concentration scale over a wide range of urea molarity (see below). However, derivatives with respect to molality (m_3) avoid unnecessary complications (as well as some ambiguities) in the multicomponent thermodynamic analysis of solute−biopolymer and solute−model compound interactions.[29] Thus, we analyze the m-values using the derivatives μ^{ex}_{23} with respect to molality for comparisons with ID and solubility data.

Values of μ^{ex}_{23} and $\Delta\mu^{\text{ex}}_{23}$ can be interpreted using the solute partitioning model (SPM), a two-state description of the distribution of the solute between the biopolymer surface and the bulk solution.[11,13,14] In the SPM, developed to interpret the preferential interaction coefficient Γ_{μ_1,μ_3}, the quantities μ_{23} and $\Delta\mu_{23}$ are interpreted as:

$$\mu_{23} = -\mu_{33}\Gamma_{\mu_3} \cong -\mu_{33}\Gamma_{\mu_1,\mu_3} \cong -(K_{\text{P}} - 1)RTb_1\text{ASA}(1 + \epsilon_3)/m_1^{\bullet} \quad (5a)$$

$$\Delta\mu_{23} = -\mu_{33}\Delta\Gamma_{\mu_3} \cong -\mu_{33}\Delta\Gamma_{\mu_1,\mu_3} \cong -(K_{\text{P}} - 1)RTb_1\Delta\text{ASA}(1 + \epsilon_3)/m_1^{\bullet} \quad (5b)$$

where K_{P} is the solute partition coefficient ($K_{\text{P}} \equiv m_3^{\text{loc}}/m_3^{\text{bulk}}$), $b_1 \equiv B_1/\text{ASA}$ is the hydration (B_1 moles of water) per Å^2 of water accessible biopolymer surface, the nonideality correction $\epsilon_3 \equiv \partial \ln \gamma_3/\partial \ln m_3$, and $m_1^{\bullet}$ is the molality of water (55.5 mol/kg). In eq 5b, K_{P} and b_1 refer to that biopolymer surface which is exposed or buried in the process (and therefore contributes to ΔASA).[11] To obtain μ^{ex}_{23}, the derivative μ^{mix}_{23} of the ideal mixing expression ($\mu_2^{\text{mix}} \equiv \mu_2^{\circ} + RT \ln X_2$), is subtracted from μ_{23}:

$$\mu^{\text{ex}}_{23} = \mu_{23} - \mu^{\text{mix}}_{23} \quad (6)$$

where $\mu^{\text{mix}}_{23} = RT(\partial \ln X_2/\partial m_2) = -RT/(m_1^{\bullet} + m_2 + m_3)$. For biopolymers, μ^{mix}_{23} is typically insignificant in comparison to μ^{ex}_{23}, so $\mu^{\text{ex}}_{23} \cong \mu_{23}$. For small model compounds, μ^{mix}_{23} can be of the same magnitude as μ^{ex}_{23}.

The SPM predicts that μ^{ex}_{23}/ASA (eq 5a; or $\Delta\mu^{\text{ex}}_{23}/\Delta$ASA, eq 5b) is constant (independent of ASA) for interactions of urea or any other solute with a homologous series of biopolymers with the same surface composition and different surface areas. Studies of the dependence of μ^{ex}_{23}/ASA (or $\Delta\mu^{\text{ex}}_{23}/\Delta$ASA) on biopolymer (protein, nucleic acid) surface composition have been used to deduce the types of surface (e.g., anionic, polar amide, etc.) responsible for preferential interactions with a given solute.[7−9,11]

Evaluating μ^{ex}_{23} as a Function of m_2 and m_3 from Osmolality Data. For the three-component aqueous solutions of nonelectrolytes considered here, the osmolality Osm(m_2,m_3) is accurately represented by a truncated form of the following polynomial series in the two solute molalities, m_2 and m_3:

$$\text{Osm}(m_2,m_3) = \sum_{i=0}\sum_{j=0}\alpha_{ij}m_2^i m_3^j \quad (7)$$

Here, the coefficients α_{ij} depend only on T and P. [By comparison with the cluster expansion of the osmotic pressure (on the molar scale) for the case of two solutes, the nonideal parts of α_{11}, α_{20}, and α_{02} contain information about the interaction of two hydrated solute molecules (together with terms related to their partial molar volumes); likewise, the nonideal parts of α_{21}, α_{12}, α_{30}, and α_{03} contain information about interactions of the various combinations of three hydrated solute molecules (together with second-order virial coefficients and terms related to their partial molar volumes)] (C.F.A., unpublished). For the systems and solute concentration ranges analyzed here (see below), we find that eq 7 may be truncated at terms in which $i + j = 3$ (in some cases truncated at $i + j = 2$). From the generalized Gibbs−Duhem-based derivation (eq 10 of ref 29) in conjunction with this truncated form of eq 7, one obtains the following expression for μ_{23}:

$$\mu_{23} = RT(\alpha_{11} + \alpha_{21}m_2 + \alpha_{12}m_3) \quad (8a)$$

For these situations, eq 8a can be recast as a triple finite difference:

$$\mu_{23} = RT[\text{Osm}(m_2,m_3) - \text{Osm}(m_2) - \text{Osm}(m_3)]/m_2m_3 \equiv RT\Delta_{23}\text{Osm}/m_2m_3 \quad (8b)$$

Expressions algebraically equivalent to eq 8 for derivatives related to μ_{23} have been obtained previously.[20,30] (For electrolytes at low concentration ($\lesssim$0.1−0.5 m, depending on the electrolyte) eq 7 is inapplicable; μ^{ex}_{23} in this situation is directly obtainable from Osm,[29] although eq 8a,b is inapplicable in the low molality regime.)

The excess quantity μ^{ex}_{23} (cf. eq 2) is obtained from osmolality data by application of eqs 6 and 8a:

$$\frac{\mu^{\text{ex}}_{23}}{RT} = \alpha_{11} + \alpha_{21}m_2 + \alpha_{12}m_3 + (m_1^{\bullet} + m_2 + m_3)^{-1} = \left(\frac{\partial \ln \gamma_2}{\partial m_3}\right)_{m_2} \quad (9a)$$

For the limiting case $m_2 \rightarrow 0$, of general significance because virtually all biopolymer studies in vitro and in vivo are at low m_2 ($\ll 0.01$ m),

$$\left(\frac{\mu^{\text{ex}}_{23}}{RT}\right)_{m_2\rightarrow 0} = \alpha_{11} + \alpha_{12}m_3 + (m_1^{\bullet} + m_3)^{-1} = \left(\frac{\partial \ln \gamma_2}{\partial m_3}\right)_{m_2\rightarrow 0} \quad (9b)$$

For the double limit $m_2 \rightarrow 0$, $m_3 \rightarrow 0$,

$$\left(\frac{\mu_{23}^{ex}}{RT}\right)_{m_2\to 0,m_3\to 0} = \alpha_{11} + (m_1^{\bullet})^{-1} = \left(\frac{\partial \ln \gamma_2}{\partial m_3}\right)_{m_2\to 0,m_3\to 0} \quad (9c)$$

This double limit is used in the comparison of values of $\mu_{23}^{ex}/(RT\cdot\mathrm{ASA})$ obtained from model compound data with those obtained from biopolymer data.

Evaluation of μ_{23}^{ex} at Saturating m_2 from the Dependence of Solubility on m_3. Previous analyses (for example, ref 1) of effects of urea or other solutes on solubilities of amino acids and other model compounds reported transfer free energies. In this paper, we introduce an alternative analysis of these model compound solubility measurements which permits direct comparisons with ID and biopolymer data. Solubility studies measure the change in molality of the model compound (m_2) in a saturated solution (ss) as a function of the molality of a perturbing solute (dm_2^{ss}/dm_3). This derivative has been approximated experimentally by determining solubilities in the presence and absence of a perturbing solute or from differences in solubilities at more than one perturbing solute concentration:

$$\frac{dm_2^{ss}}{dm_3} = \left(\frac{\partial m_2}{\partial m_3}\right)_{\mu_2=\mu_2^{ss}} \quad (10a)$$

$$\frac{dm_2^{ss}}{dm_3} \cong \frac{\Delta m_2^{ss}}{\Delta m_3} = \frac{m_2^{ss} - m_2^{ss,0(3)}}{m_3} \quad (10b)$$

The superscript 0(3) denotes the absence of perturbing solute 3.

Tanford and Nozaki[1] determined m_2^{ss} at urea molalities (m_3) of 0, 2.26, 5.17, 8.97, and 14.04 *m*. We fit their data up to 5.17 *m* with a second-order polynomial (see Supporting Information) and used the derivative of the quadratic fit to determine $(\partial m_2/\partial m_3)_{\mu_2}$. Bolen and co-workers[4,18] determined m_2^{ss} only at $m_3 =$ 1.06 or 2.26 *m*, for which we therefore can only estimate the derivative dm_2^{ss}/dm_3 by eq 10b. The partial derivative $(\partial m_2/\partial m_3)_{\mu_2}$ is related exactly to the chemical potential derivative μ_{23} by:

$$\left(\frac{\partial m_2}{\partial m_3}\right)_{\mu_2} = -\frac{(\partial m_2/\partial m_3)_{m_2}}{(\partial m_2/\partial m_2)_{m_3}} = -\frac{\mu_{23}}{\mu_{22}} \quad (11)$$

Since the derivative $(\partial m_2/\partial m_3)_{\mu_2}$ is evaluated at $\mu_2 = \mu_2^{ss}$, the values of μ_{23} and μ_{22} in eq 11 are those applicable to the saturated solution. Then, μ_{23}^{ex} is calculated using eq 6. (The evaluation of μ_{22} is discussed in a following section: see comments on eq 12.)

Results and Discussion

Calculation of μ_{23}^{ex} from ID Data for Urea−Peptide (Amino Acid) Interactions. Osmolalities of three-component solutions of urea and each of the amino acids glycine (Gly), alanine (Ala), glycylalanine (GlyAla), diglycine (Gly_2), and triglycine (Gly_3) have been measured by ID at 25 °C by Schönert and co-workers[19,20] and Uedaira.[21] Schönert and co-workers analyzed these data using expressions algebraically similar to eqs 7 and 8 to obtain a free energy of transfer of the amino acid from water into urea solutions. To obtain the fundamental thermodynamic quantity μ_{23}^{ex} from their three-component data, we fit them together with relevant two-component ID data[19,20,31−36] to eq 7 (truncated at $i + j = 3$). Examination of the residuals of these fittings revealed systematic, rather than normal, distributions of the residuals in the Ala, Gly, and GlyAla data sets. Inclusion of higher order terms (up to $i + j = 5$) did not alleviate the problem. Omitting data points with osmolalities above 5.2 for Ala and 6 Osm for Gly and GlyAla removed these systematic trends (see Supporting Information), which appear to result from an offset in the high osmolality data and not from a defect in eq 7.[37] Below these osmolality cutoffs, osmolality data for each

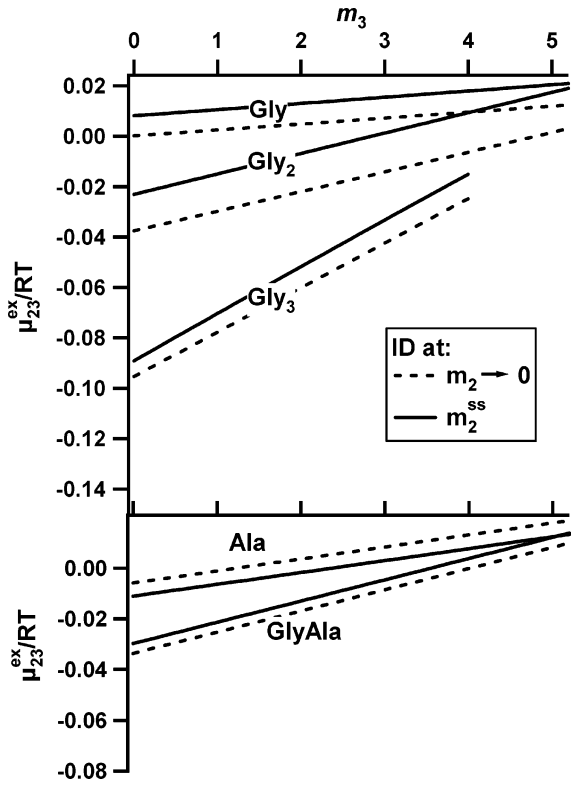

Figure 1. Urea−amino acid preferential interactions calculated from ID data. Values of μ_{23}^{ex}/RT (calculated with eq 9) for each amino acid are plotted as functions of urea molality (m_3) for two different limits of amino acid concentration (m_2): $m_2 \to 0$ (zero concentration limit; dashed lines); $m_2 \to m_2^{ss}$ (saturated solution limit; solid lines).

TABLE 1: Coefficients in Polynomial Expansion of $\mathrm{Osm}(m_2,m_3) = \sum_{i=0}\sum_{j=0}\alpha_{ij}m_2^i m_3^j$ ($i + j \le 3$) from ID Data for Ternary Solutions (Urea−Amino Acid−Water)[a]

	Gly	Gly_2	Gly_3	Ala	GlyAla
α_{10}	0.970 ± 0.004	0.971 ± 0.005	0.97 ± 0	1.003 ± 0.003	0.985 ± 0.004
α_{20}	−0.050 ± 0.004	−0.131 ± 0.009	−0.197 ± 0.055	−0.008 ± 0.005	−0.086 ± 0.009
α_{30}	0.007 ± 0.001	0.039 ± 0.004	−0.108 ± 0.185	0.0098 ± 0.002	0.047 ± 0.005
α_{01}	0.991 ± 0.001	0.991 ± 0.001	0.991 ± 0.001	0.991 ± 0.001	0.991 ± 0.001
α_{02}	−0.032 ± 0.001	−0.032 ± 0.001	−0.032 ± 0.001	−0.032 ± 0.001	−0.032 ± 0.001
α_{03}	0.002 ± 0.000	0.002 ± 0.000	0.002 ± 0.000	0.002 ± 0.000	0.002 ± 0.000
α_{11}	−0.018 ± 0.004	−0.056 ± 0.005	−0.114 ± 0.004	−0.024 ± 0.002	−0.052 ± 0.003
α_{21}	0.003 ± 0.001	0.009 ± 0.003	0.019 ± 0.009	−0.003 ± 0.001	0.003 ± 0.002
α_{12}	0.003 ± 0.001	0.008 ± 0.001	0.018 ± 0.001	0.005 ± 0.001	0.009 ± 0.001

[a] α_{00} defined as zero. α_{i0} and α_{0j} coefficients and errors were determined from third-order fits of two-component ID data; Ala,[20,36] Gly,[32] Gly_2,[32] Gly_3,[33] GlyAla,[32] and urea.[34,35] Two-component values for α_{i0} and α_{0j} were used in fits to determine α_{11}, α_{21}, and α_{12}. Errors on α_{ij} cross-coefficients are from least-squares fitting of three-component data; Gly and Ala (Table 2 of ref 20), Gly_2,[21] Gly_3, and GlyAla.[19]

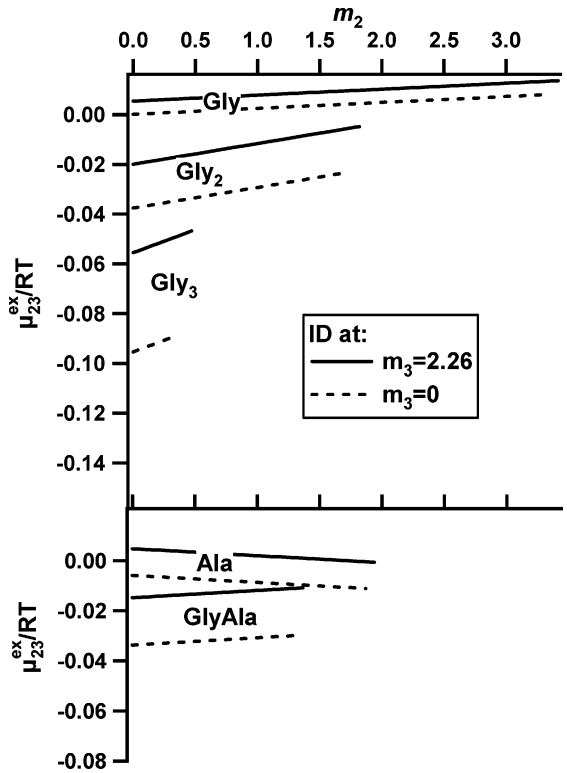

Figure 2. Urea−amino acid preferential interactions calculated from ID data. Values of μ_{23}^{ex}/RT (eq 9) for each amino acid are plotted as functions of amino acid molality (m_2) for two different urea concentrations (m_3): $m_3 \rightarrow 0$ (zero concentration limit; dashed lines); $m_3 = 2.26$ *m* urea (solid lines).

peptide are well-fit by polynomial series truncated at third power terms ($i + j = 3$) in solute molalities; so in this concentration range, eq 8a,b for μ_{23}^{ex} are effectively exact for the available data. We focused on the lower osmolality data because of their relevance for comparison with biopolymer data. Table 1 lists values of the fitting coefficients α_{ij} with uncertainties for the five amino acids investigated by ID. Values of μ_{23}^{ex}/RT obtained from eq 8 for those systems are plotted in Figures 1 and 2, and values of μ_{23}^{ex}/RT at representative urea and amino acid concentrations are listed in Table 2. The coefficient α_{11} dominates values of μ_{23}^{ex}/RT, while the m_2 and m_3 dependences of μ_{23}^{ex}/RT (i.e., the second derivatives μ_{232}^{ex} and μ_{233}^{ex}) arise from the coefficients α_{21} and α_{12}, respectively. (Values of α_{ij} were also calculated setting $\alpha_{10} = 1$ and $\alpha_{01} = 1$, the values expected[38] on the basis of the virial expansion corresponding to eq 7. The resulting changes in values of the cross coefficients α_{11}, α_{21}, and α_{12} were minimal, and the quality of the fit to the low concentration data was significantly reduced.)

Behavior of μ_{23}^{ex} in the Double Limit ($m_2 \rightarrow 0$, $m_3 \rightarrow 0$). From eq 9c in the double limit of low amino acid and urea concentrations ($m_2 \rightarrow 0$, $m_3 \rightarrow 0$), the solute-specific parts of μ_{23}^{ex}/RT and of the solute preferential interaction coefficient $\Gamma_{\mu 3}$ are determined by α_{11}, the coefficient of the m_2m_3 term in the osmolality expansion in eq 7. Figure 1 shows that values of μ_{23}^{ex}/RT for the amino acids Gly and Ala are near zero in this double limit (0.000 for Gly, −0.005 for Ala), indicating from eq 9c that the preferential interactions of urea with Gly and Ala (relative to interactions with water) are essentially zero. For the series Gly, Gly$_2$, and Gly$_3$, limiting values of μ_{23}^{ex}/RT (i.e., α_{11}) become increasingly negative with an increasing degree of oligomerization, from 0.00 m^{-1} (Gly) to −0.04 m^{-1} (Gly$_2$), to −0.09 m^{-1} (Gly$_3$; see Table 2, column 1), indicating the existence of a favorable preferential interaction of the di- and tripeptides with urea, which increases with the number of amide groups and is absent in glycine and alanine.

Dependence of μ_{23}^{ex} on m_3 ($m_2 \rightarrow 0$). From eq 9a, the solute-specific part of the m_3 dependence of μ_{23}^{ex} is determined by α_{12}. Values of α_{12}, given in Table 1, are positive for all five amino acids, increasing from values near zero for the two monomeric amino acids (Gly, 0.003; Ala, 0.005) to larger values for the dipeptides (Gly$_2$, 0.008; GlyAla, 0.009) and the tripeptide (Gly$_3$, 0.018). Figure 1 shows that for each peptide μ_{23}^{ex}/RT increases significantly with increasing urea molality. Differences in

TABLE 2: Excess Chemical Potential Derivatives (μ_{23}^{ex}/RT) × 10^3 (m^{-1}) for Urea−Amino Acid Interactions from Osmolality or Solubility Data[a]

	$(\mu_{23}^{ex}/RT) \times 10^3$ (m^{-1}) evaluated at $m_2 = 0$ by ID, urea molality				$(\mu_{23}^{ex}/RT) \times 10^3$ (m^{-1}) evaluated at m_2^{ss} by ID, urea molality				$(\mu_{23}^{ex}/RT) \times 10^3$ (m^{-1}) evaluated at m_2^{ss} by solubility, urea molality				
solute	0.0	1.06	2.26	5.17	0.0	1.06	2.26	5.17	0.0	1.06	2.26	5.17	ref
glycine	0 ± 4[b]	3 ± 4	5 ± 4	12 ± 5	8 ± 6	11 ± 6	14 ± 6	21 ± 7	8 ± 1	8 ± 1	8 ± 1	7 ± 1	*c*
									7 ± 1	7 ± 1			*d*
Gly$_2$	−38 ± 5	−29 ± 5	−20 ± 6	3 ± 7	−21 ± 8	−12 ± 8	−3 ± 8		−4 ± 10	−3 ± 11	1 ± 12	7 ± 20	*c*
									−15 ± 2	−15 ± 2			*d*
Gly$_3$	−95 ± 4	−77 ± 5	−56 ± 5	−4 ± 7	−89 ± 5	−69 ± 6	−47 ± 7		−140 ± 18	−90 ± 8	−51 ± 15	0 ± 12	*c*
									−110 ± 12	−86 ± 14			*d*
alanine	−6 ± 2	−1 ± 2	5 ± 2	18 ± 2	−11 ± 2	−6 ± 2	−1 ± 2	13 ± 3	−5 ± 4	−1 ± 4	3 ± 5	13 ± 7	*c*
									−2 ± 7		−2 ± 7		*c*
GlyAla	−34 ± 3	−25 ± 3	−15 ± 3	9 ± 4	−30 ± 4	−21 ± 4	−11 ± 4	13 ± 4					
Gly$_2$ − Gly	−38 ± 6	−32 ± 6	−25 ± 7	−9 ± 9	−29 ± 10	−23 ± 10	−17 ± 10		−12 ± 10	−11 ± 11	−8 ± 12	−2 ± 20	*c*
									−22 ± 3	−22 ± 3			*d*
Gly$_3$ − Gly$_2$	−57 ± 6	−48 ± 7	−36 ± 8	−7 ± 10	−68 ± 9	−57 ± 10	−44 ± 11		−136 ± 20	−87 ± 14	−52 ± 19	−7 ± 23	*c*
									−95 ± 13	−71 ± 15			*d*
(Gly$_3$ − Gly)/2	−48 ± 3	−40 ± 3	−30 ± 3	−8 ± 4	−48 ± 4	−40 ± 4	31 ± 5		−71 ± 9	−49 ± 4	−30 ± 7	−4 ± 6	*c*
									−59 ± 6	−47 ± 7			*d*
Ala − Gly	−6 ± 4	−4 ± 4	0 ± 4	6 ± 5	−19 ± 6	−17 ± 6	−15 ± 6	−8 ± 8	−13 ± 4	−9 ± 4	−5 ± 5	4 ± 7	*c*
GlyAla − Ala	−28 ± 4	−24 ± 4	−20 ± 4	−9 ± 4	−19 ± 4	−15 ± 4	−10 ± 4	0 ± 5					

[a] Errors were propagated by standard methods. Error in μ_{22} was propagated from 1 standard deviation (SD) in fit coefficients from least-squares fitting of two-component ID data. [b] Errors were propagated from 1 SD in fit coefficients from least-squares fitting of ID data. Covariance is unaccounted for. Covariance is zero at the 0 *m* limits and would, if significant, decrease the propagated error at the solubility limit. [c] Calculated with data from refs 1, 5, and 18 using quadratic fits and the derivative in eq 10a. Errors were propagated from 1 SD in fit coefficients from least-squares quadratic fitting of solubility data. [d] Calculated with data from refs 5 and 18 using the difference relationship in eq 10b. Uncertainty in solubility measurements was assumed to be ±0.1 g of solute 2 in determination of the solubility limit. Error in $\partial m_2/\partial m_3$ is from error weighted linear fits of solubility data.

μ_{23}^{ex}/RT between Gly_2 and Gly or GlyAla and Ala provide information on contributions to μ_{23}^{ex}/RT from the peptide group and polar amide atoms (see below). Figure 1 shows that these differences are also functions of urea concentration (see below) and are reduced at higher m_3. Though values of μ_{23}^{ex}/RT at m_2^{ss} differ from those at $m_2 \rightarrow 0$ (Table 2, columns 4−8), Figure 1 shows that effects of increasing m_3 on μ_{23}^{ex}/RT at m_2^{ss} parallel those observed at $m_2 \rightarrow 0$. Analysis of solubility data acquired at multiple urea concentrations yields trends in μ_{23}^{ex}/RT with m_3 that are qualitatively similar to those observed by analyzing ID.[39]

Dependence of μ_{23}^{ex} on m_2 ($m_3 \rightarrow 0$ and as a Function of m_3). The solute-specific part of the m_2 dependence of μ_{23}^{ex} is determined by α_{21} (eq 9a). Values of α_{21} given in Table 1 are positive for all amino acids except alanine and increase in magnitude from near zero for the monomeric amino acids (Ala, −0.003; Gly, 0.003) to larger positive values for dipeptides (GlyAla, 0.003; Gly_2, 0.005) and the tripeptide (Gly_3, 0.019). Both magnitudes and trends of α_{21} are very similar to those observed for α_{12} (see above). The resulting m_2 dependences of μ_{23}^{ex} are displayed in Figure 2 and Table 2. The left four columns of Table 2 give values of μ_{23}^{ex}/RT from ID measurements at the $m_2 = 0$ limit, and the center columns give values from ID at the m_2^{ss} limit. Only ID (but not solubility) data can be used to investigate the m_2 dependence of μ_{23}^{ex}/RT.

In vitro studies of effects of urea on protein folding and other protein processes are almost invariably performed at low (picomolar to micromolar) protein concentrations. VPO studies of interactions of urea with BSA used protein concentrations in the range of 150−250 mg/mL (2−5 m*m*), at the high end of the accessible concentration range. Even at the highest protein concentrations, the concentration of urea was in excess in comparison to the concentration of protein or of each type of amino acid in the protein. Essentially all in vitro protein concentrations therefore correspond to the $m_2 \rightarrow 0$ limit of model compound data.

Evaluation of μ_{23}^{ex} at Saturating m_2 from Solubility Studies as a Function of Urea Concentration. Solubilities of glycine, diglycine, and triglycine in urea solutions have been measured by Nozaki and Tanford[1] and Auton and Bolen.[5] We analyzed these data to calculate values of μ_{23}^{ex}/RT using eqs 6, 10a, and 11. Since the solubility of diglycine in water differs significantly in these two sets and the data cover different urea concentration ranges, the data from these sources were analyzed both separately and together. (The data of Talukdar et al.[40] for urea effects on the solubility of glycine, diglycine, and triglycine exhibit substantially more scatter and were therefore omitted from our analysis.) Values of the derivative dm_2^{ss}/dm_3 at different urea concentrations were determined by analysis of the difference (eq 10a) for the data of Bolen and co-workers[5,18] and by differentiation of quadratic fits of the data of Nozaki and Tanford[1] combined with those of Bolen and co-workers.[5,18]

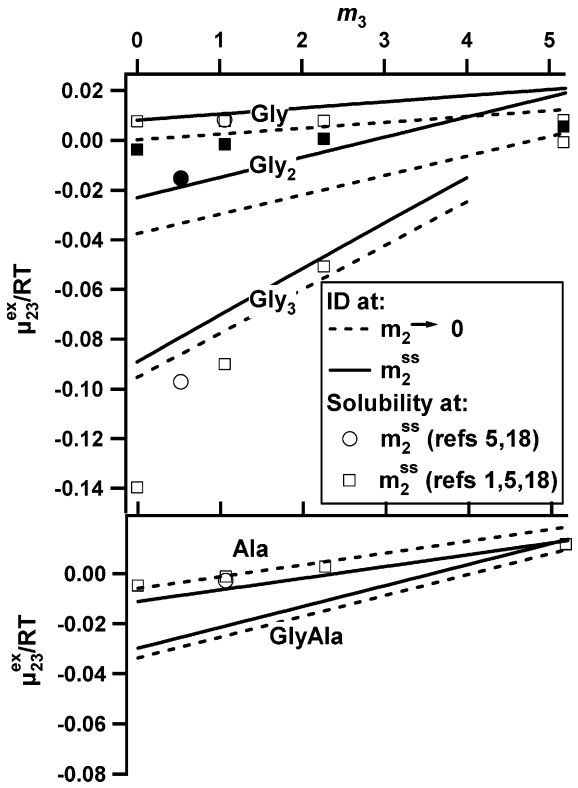

Figure 3. Urea−amino acid interactions calculated from ID data are represented by lines as in Figure 1. Values of μ_{23}^{ex}/RT (eqs 6, 10a, and 11) at saturated amino acid concentrations (m_2^{ss}) from solubility data are indicated by symbols: (●), from refs 5 and 18, calculated using the difference relationship in eq 10a; and (■), from refs 1, 5 and 18 , calculated using the derivative relationship in eq 10a. Values for Gly, Gly_3, and Ala are indicated by open symbols; values for Gly_2 are indicated by filled symbols. No solubility data are available for GlyAla.

To determine $\mu_{23}(m_2^{ss})$ from these data using eq 11, corresponding values of μ_{22} are required; these were determined from the Gibbs−Duhem relationship:

$$\mu_{22} = \frac{-m_1^{\bullet}\mu_{12} - m_3\mu_{23}}{m_2} \cong \frac{-m_1^{\bullet}\mu_{12}}{m_2} \qquad (12)$$

where $\mu_{12} = -(RT/m_1^{\bullet})(\partial \mathrm{Osm}/m_2)_{m_3,T,P}$ which is determined independently by analyzing corresponding osmometric data. The ideal mixing approximation to μ_{22} ($\mu_{22}^{ideal} = RT/m_2$) is not sufficiently accurate to analyze solubility data for moderately soluble species; eq 12 is necessary for an accurate determination of μ_{22}. In general, μ_{12} should be evaluated from three-component osmometric data. For the systems analyzed here, below 6 *m* urea, values of μ_{12} for urea−water mixtures differ by less than 2% from those for urea−amino acid−water mixtures, and values

TABLE 3: Water Accessible Surface Areas (ASA) and Values of $(-\mu_{23}^{ex}/(RT \cdot ASA_{total}))^a$ for Amino Acids and Peptides

			surface types							
			polar amide (amide N, O)		anionic (carboxylate O)		cationic (amino N)		nonpolar[b] (C)	
solute	total ASA (Å^2)	$-\mu_{23}^{ex}/(RT \cdot ASA_{total})$ (at m_2, $m_3 \rightarrow 0$)	ASA (Å^2)	fraction	ASA (Å^2)	fraction	ASA (Å^2)	fraction	ASA (Å^2)	fraction
Gly	209	0.0×10^{-4}	0	0	91	0.43	54	0.26	65	0.31
Ala	225	0.0×10^{-4}	0	0	90	0.4	55	0.25	80	0.35
Gly_2	297	1.3×10^{-4}	42	0.14	91	0.31	54	0.18	110	0.37
GlyAla	317	1.1×10^{-4}	34	0.11	86	0.27	54	0.17	143	0.45
Gly_3	385	2.5×10^{-4}	84	0.22	91	0.24	54	0.14	155	0.39

[a] μ_{23}^{ex} (eq 2) obtained from analysis of ID data (see Table 2). [b] Including amide carbon (~5 ± 3 Å^2).

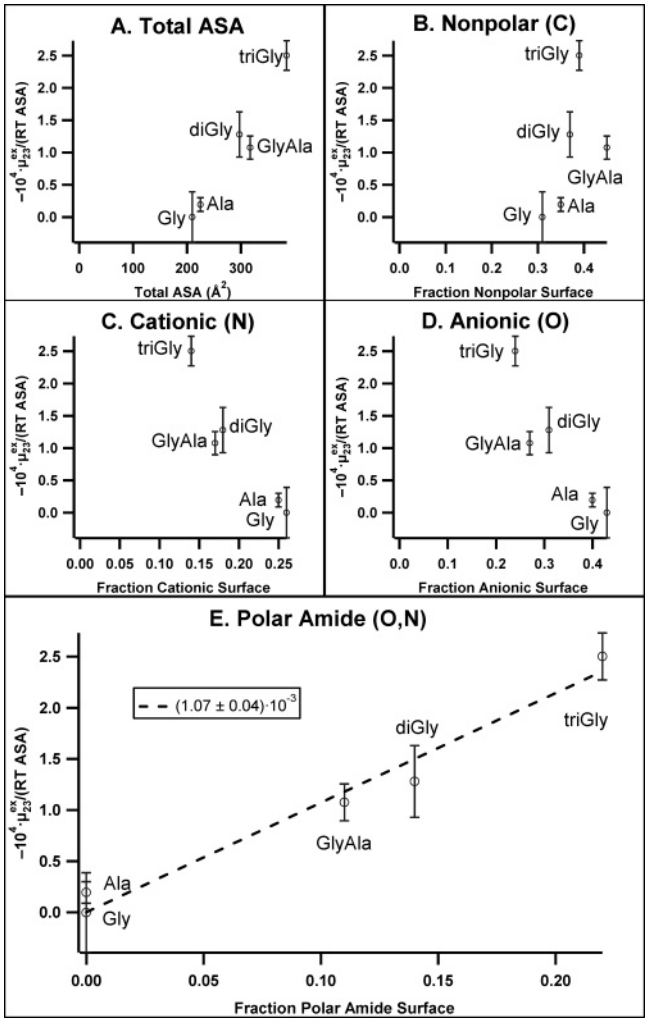

Figure 4. Tests of correlations between urea−amino acid preferential interactions normalized by total amino acid ASA ($-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA})$) and different types of water accessible surface area: (A) nonpolar ASA, (B) anionic ASA, (C) cationic ASA, and (D) polar amide ASA. In panel E, the dashed line represents an error weighted, least-squares fit through the origin to the data and yields $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{polar\ amide}}) = (-1.07 \pm 0.04) \times 10^{-3}$.

of μ_{23}^{ex}/RT (Table 2) are the same (within uncertainty) when calculated from three-component or two-component data.[39]

Table 2 includes values of μ_{23}^{ex}/RT from solubility data compared with μ_{23}^{ex}/RT from ID data at the same molalities as well as some additional molalities. These values of μ_{23}^{ex}/RT from solubility are also plotted in Figure 3. Figure 3 shows that values of μ_{23}^{ex}/RT obtained at the solubility limit of the amino acid (m_2^{ss}) and at various m_3 from solubility studies are of the same magnitude as values of this derivative obtained for this concentration regime from polynomial analysis of ID data. Quantitative discrepancies are most notable for triglycine at low urea concentration, where solubility values of μ_{23}^{ex}/RT are substantially more negative than expected from the osmolality data. An accurate calculation of μ_{23}^{ex} from solubility data requires a value of μ_{22} at the m_2 and m_3 of the solubility experiments, which cannot be evaluated using only solubility data but is readily obtained by ID. In addition, the ID data indicate significant, systematic differences between values of μ_{23}^{ex} at m_2^{ss} and in the limit $m_2 \rightarrow 0$ appropriate for analysis of biopolymer experiments. For polar model compounds with moderate to high solubility, only ID provides unambiguous

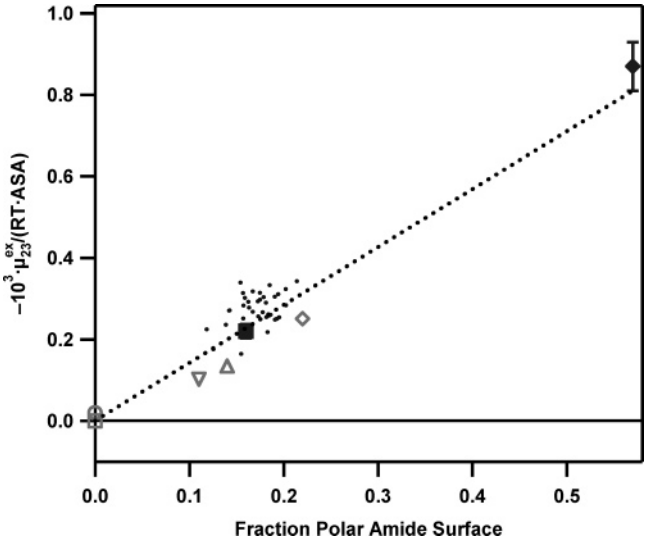

Figure 5. Comparison of ASA correlations of urea preferential interaction data for model compounds and biopolymers. Values of $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA})$ for preferential interactions of urea per unit area of small solutes (open symbols: glycine □; diglycine △; triglycine ◇; alanine ○; glycylalanine ▽) and values of $-\Delta\mu_{23}^{ex}/(RT\cdot\Delta\mathrm{ASA})$ for preferential interactions of urea per unit area of biopolymer surface exposed in unfolding (filled symbols: α-helix unfolding[9,52] ◆; lacHP unfolding[41] ■; other protein unfolding data analyzed in ref 7 •) are plotted versus the fraction of total ASA or ΔASA which is polar amide surface. The dotted line is the least-squares fit to the biopolymer data. (The quantity reported in previous papers, $\Gamma_{\mu 3}/m_3$, approximately equals $-\mu_{23}^{ex}$ for biopolymers in the $m_3 \rightarrow 0$ limit.)

information about μ_{23}^{ex} in the limit $m_2 \rightarrow 0$ (and $m_3 \rightarrow 0$) applicable to analyses and interpretations of urea effects on biopolymer processes.

Contributions to μ_{23}^{ex} from Preferential Interactions of Urea with Different Types of Model Compound Surface: Dominance of Interactions with Polar Amide Surface. To interpret values of μ_{23}^{ex} for urea−model compound interactions in terms of interactions of urea with the different types of model compound surface, the starting point (motivated by the SPM for biopolymers, eq 5a) is to express these values per unit of model compound ASA. Values of $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{total}})$, listed in Table 3, range from near zero for Gly and Ala to near 1×10^{-4} m^{-1} Å^{-2} for the dipeptides and 2.5×10^{-4} m^{-1} Å^{-2} for triglycine. This trend in $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{total}})$ indicates that urea interacts more strongly with functional groups present in the di- and tripeptides than with the monomeric amino acids and clearly correlates with the number of amide groups in these model compounds. To establish a quantitative relationship, we introduce the same coarse-grained decomposition of the water accessible surface of model compounds as introduced previously for biopolymers.[7−9] Four chemically distinct types of surface of the five amino acids and oligopeptides are considered: polar amide (N, O of amide groups), anionic (O of $-CO_2^-$ groups), cationic ($-NH_3^+$), and nonpolar (C). Table 3 lists the amounts of each type of surface and the fraction that each contributes to total ASA of each model compound. Figure 4 plots values of $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{total}})$ versus each of these coarse-grained compositional variables and versus total ASA. Panel E of Figure 4 shows $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{total}})$ is directly proportional to the fraction of polar amide surface in these model compounds; values of $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{total}})$ are not proportional to any other type of surface. The proportionality constant relating $-\mu_{23}^{ex}/(RT\cdot\mathrm{ASA}_{\mathrm{total}})$ to the fraction of polar amide surface quantifies the average intrinsic strength of the urea−amide

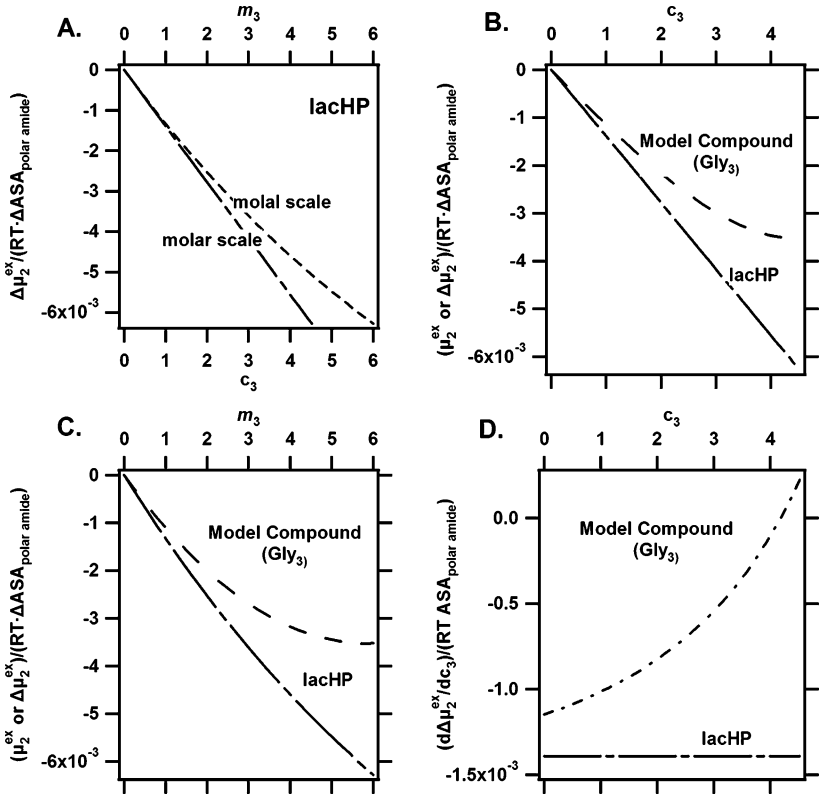

Figure 6. Comparison of predictions of model compound- and biopolymer-based analyses of the contribution of urea−amide preferential interactions to the effect of urea concentration on the unfolding of the lac repressor DNA-binding domain (lacHP). (A) $\Delta\mu_2^{ex}/(RT{\cdot}\Delta ASA_{polar\ amide})$ for unfolding lacHP as a function of urea concentration: molal scale (short dashes), molar scale (long−short dashes). (B,C) Comparison of $\Delta\mu_2^{ex}/(RT{\cdot}\Delta ASA_{polar\ amide})$ for lacHP (from A) with $\mu_2^{ex}/(RT{\cdot}ASA_{polar\ amide})$ for triglycine as a function of urea concentration on molar (B) and molal (C) scales. (D) Comparison of predicted contributions of polar amide surface to urea *m*-values: molar scale derivatives $(\partial\mu_2^{ex}/\partial c_3)/(RT{\cdot}ASA_{polar\ amide})$ for triglycine and $(\partial\Delta\mu_2^{ex}/\partial c_3)/(RT{\cdot}\Delta ASA_{polar\ amide})$ for lacHP are plotted as functions of urea molarity.

preferential interaction for these model compounds: $-\mu_{23}^{ex}/(RT{\cdot}ASA_{polar\ amide}) = (1.07 \pm 0.04) \times 10^{-3}$ Å^{-2}. Individual values of $-\mu_{23}^{ex}/(RT{\cdot}ASA_{polar\ amide})$ for all three amide-containing compounds fall in the range $(1.0 \pm 0.1) \times 10^{-3}\ m^{-1}$ Å^{-2}. Gly and Ala have no amide surface, and (as expected) their values of μ_{23}^{ex}/RT are very close to zero.

Comparison of Model Compound and Biopolymer Data for Urea−Amide Interactions. A comparison of the thermodynamics of urea−model compound interactions (μ_{23}^{ex}) in the limit $m_2 \rightarrow 0$, $m_3 \rightarrow 0$ with analogous thermodynamic data for urea−biopolymer interactions and urea effects on biopolymer processes in the same double limit is shown in Figure 5. For both model compounds and biopolymers, values of $-\mu_{23}^{ex}/(RT{\cdot}ASA_{total})$ for interactions with urea (and of $-\Delta\mu_{23}^{ex}/(RT{\cdot}\Delta ASA_{total})$ for effects of urea on processes) are plotted versus the fraction of ASA (or of ΔASA) which is polar amide surface. Model compound data are from Figure 4E. Biopolymer data plotted in Figure 5 are from our previous compilation for the surfaces exposed in unfolding α-helical, alanine based peptides[7] (where 57% of the total ASA exposed in unfolding is polar amide surface and for which $-\Delta\mu_{23}^{ex}/(RT{\cdot}\Delta ASA_{total}) = (8.7 \pm 0.6) \times 10^{-4}\ m^{-1}$ Å^{-2}), lacHP[9] (16% polar amide surface; $-\Delta\mu_{23}^{ex}/(RT{\cdot}\Delta ASA_{total}) = (2.0 \pm 0.1) \times 10^{-4}\ m^{-1}$ Å^{-2}), and the large set of protein unfolding data (where 12−21% of the total ASA exposed in unfolding is polar amide surface and for which $-\Delta\mu_{23}^{ex}/(RT{\cdot}\Delta ASA_{total})$ ranges from approximately $1.6 \times 10^{-4}\ m^{-1}$ Å^{-2} to $3.4 \times 10^{-4}\ m^{-1}$ Å^{-2}).[7]

A least-squares, linear fit through all of the biopolymer data of Figure 5 has an intercept of zero within the uncertainty ($(1.7 \pm 2.0) \times 10^{-5}$), indicating that $-\mu_{23}^{ex}/(RT{\cdot}ASA_{total})$ ($-\Delta\mu_{23}^{ex}/(RT{\cdot}\Delta ASA_{total})$ for processes) is proportional to the fraction of polar amide surface. From the slope, one obtains $-\mu_{23}^{ex}/(RT{\cdot}ASA_{polar\ amide}) = (1.4 \pm 0.1) \times 10^{-3}\ m^{-1}$ Å^{-2}. From similar analyses (based on $\Gamma_{\mu 3}$ rather than μ_{23}^{ex}), we previously concluded that urea interacts (preferentially relative to water) to a significant extent only with polar amide surface of biopolymers.[7] When the uncertainties in individual members of the protein unfolding data set are taken into account, the result is $-\Delta\mu_{23}^{ex}/(RT{\cdot}\Delta ASA_{polar\ amide}) = (1.4 \pm 0.3) \times 10^{-3}\ m^{-1}$ Å^{-2}.[7]

Figure 5 indicates that the strength of interaction of urea with a unit area of polar amide surface of model compounds ($(1.07 \pm 0.04) \times 10^{-3}\ m^{-1}$ Å^{-2} in the limit of $m_2 \rightarrow 0$, $m_3 \rightarrow 0$) shows a slight, systematic difference from the strength of urea interactions with biopolymers ($(1.4 \pm 0.3) \times 10^{-3}\ m^{-1}$ Å^{-2}).[7] With increasing urea concentration, larger differences between these two data sets are observed, as discussed next.

Different Urea Concentration Dependences of μ_2^{ex} (and μ_{23}^{ex}) for Interactions of Urea with the Water Accessible Amide Surface of Model Compounds and Biopolymers. Plots of ΔG_{obs}° or ln K_{obs} for protein and α helix unfolding as a function of urea molarity are almost invariably linear over very wide ranges of urea concentrations. The urea *m*-value is therefore independent of urea molarity, which is explained in the context of the solute partitioning model (eq 5) by the lack of significant nonideality of urea solutions (on the molar scale) and the lack of significant urea concentration dependences of the hydration b_1 and the urea partition coefficient K_P.[11,14] In situations where the *m*-value is independent of urea concentration, the *m*-value is simply related to the difference in excess

chemical potentials between unfolded and folded states, $\Delta\mu_2^{ex}$, by the expression:

$$m\text{-value} = \frac{\Delta\mu_2^{ex}}{c_3} = \frac{RT\Delta \ln \gamma_2}{c_3} \quad (13)$$

and a conventional *m*-value plot is the same as a plot of $\Delta\mu_2^{ex}$ versus urea molarity. Figure 6A summarizes this behavior for the marginally stable DNA binding domain of lac repressor (lac headpiece; lacHP[41]). For comparison with model compound data, plotted values of $\Delta\mu_2^{ex}$ are normalized by $RT\cdot\Delta ASA_{polar\ amide}$ where $\Delta ASA_{polar\ amide}$ is the positive amount of polar amide surface exposed to water in unfolding this protein (560 Å^2).[9] Figure 6A also shows that $\Delta\mu_2^{ex}$ is a nonlinear function of urea molality, which results entirely from the derivative $(\partial m_3/\partial c_3)$ in eq 4 and the nonlinear relationship between molality and molarity ($m_3 = c_3/(1 - c_3\bar{V}_3)$, where $\bar{V}_3$ is the partial molar volume of solute 3).

For the model amides investigated by ID and analyzed here, the quantity corresponding to the *m*-value $(\Delta\mu_2^{ex}/c_3)$ is not independent of urea concentration. Integration of eq 9b (for the case of $m_2 \rightarrow 0$) for the derivative of μ_2^{ex} of the amide with respect to urea molality m_3 yields

$$\frac{\mu_2^{ex}}{m_3} = RT\left[\alpha_{11} + 0.5\alpha_{12}m_3 - \frac{1}{m_3}\ln(m_1^{\bullet} + m_3)\right] \quad (14)$$

Therefore, for amide−urea interactions, μ_2^{ex}/m_3 depends significantly on m_3; conversion to the molar scale somewhat reduces but does not eliminate this dependence. Figure 6B,C compares values of $\mu_2^{ex}/(RT\cdot ASA_{polar\ amide})$ for triglycine with values of $\Delta\mu_2^{ex}/(RT\cdot\Delta ASA_{polar\ amide})$ for lacHP on both molar (Figure 6B) and molal (Figure 6C) scales. Very similar behavior is obtained for the other model amides investigated. At the level of *m*-values (i.e., derivatives of the curves in Figure 6B), Figure 6D plots the ASA normalized *m*-value (i.e., $(\partial\Delta\mu_2^{ex}/\partial c_3)/\Delta ASA_{polar\ amide}$) for lacHP unfolding (representative of the protein data set[7]) and the corresponding quantity for triglycine ($(\partial\mu_2^{ex}/\partial c_3)/ASA_{polar\ amide}$) versus urea concentration. At low urea concentration, these intrinsic urea interaction strengths are in close agreement, as discussed above, but deviate by more than 30% at 2 M urea as a result of the urea concentration dependence of its preferential interaction with amide surface of triglycine or other model compounds.

Clearly, use of the model compound ID data to model the contribution of urea−amide interactions to *m*-values for protein unfolding introduces a much larger urea concentration dependence of these *m*-values (or of $\Delta\mu_{23}^{ex}$) than what is experimentally observed. The large difference in concentration dependences of μ_{23}^{ex} for interactions of urea with amides and biopolymers and the small difference in values of $\mu_{23}^{ex}/(RT\cdot ASA_{polar\ amide})$ in the zero urea concentration limit remain unexplained. These discrepancies between μ_{23}^{ex} for model compounds and biopolymers do not affect the use of eq 1 to interpret *m*-values and urea concentration dependences of equilibrium constants K_{obs} for biopolymer processes, especially in the limit of low urea concentration.

Concluding Discussion

Solubility and other thermodynamic data characterizing the interactions of small model compounds with water and cosolvents (denaturants, osmolytes, other solutes, and salts) have been analyzed in various ways to interpret heat capacity changes and solute *m*-values of protein folding and other biopolymer processes. Here, we obtain the chemical potential derivative μ_{23}^{ex} by analyzing model compound solubility and osmolality data. This derivative provides a direct measure of how changing the molality of a perturbing solute 3 affects the "excess" chemical potential of solute 2. Unlike other quantities (ΔG_{tr}°, *m*-values) frequently used to analyze solute effects, values of μ_{23}^{ex} from different experimental methods and for different model compounds, biopolymers, and biopolymer processes are readily and directly compared. We apply a recently derived analysis[29] that uses a single and testable approximation (a polynomial fit of osmolality data versus molality, including terms of the type shown in eq 7) to obtain values of μ_{23}^{ex} for urea−amino acid interactions from ID data. In addition, the current study applies a model-independent thermodynamic analysis to solubility data to obtain μ_{23}^{ex} and thereby compare solubility and three-component ID data.

These analyses reveal systematic differences between ID data (extrapolated to the solubility limit) and solubility data for interactions of urea with triglycine, diglycine, and glycine. Values of μ_{23}^{ex} from ID data depend significantly on both m_2 and m_3 for the amide-containing amino acids, with small dependences detected for Gly and Ala. Values of μ_{23}^{ex} for triglycine obtained from solubility data also show a strong dependence on m_3, but for glycine, diglycine, and alanine, this derivative does not depend on m_3 outside of reported errors. Values of μ_{23}^{ex} from ID appear to be sufficiently accurate to extrapolate to the low m_2 concentration limit of the model compound appropriate for comparison with biopolymer studies and to the solubility limit to compare with μ_{23}^{ex} for solubility data. By comparison with ID results (Figure 3), solubility data provide less accurate measures of the m_3 dependence of μ_{23}^{ex} for amino acid−urea interactions and cannot be used to examine any m_2 dependence.

Interactions of urea with amino acid side chains, quantified by solubility by Nozaki and Tanford,[1] make contributions to the *m*-value (or to μ_{23}^{ex}) which are relatively independent of urea concentration. Some examples of urea concentration dependent contributions of urea−amide interactions to the *m*-value (or μ_{23}^{ex}) were observed by these authors but discounted as probably artifactual. Values of μ_{23}^{ex} from ID data for preferential interactions of urea with the amide-containing amino acids (Figure 3) exhibit even larger urea concentration dependences. In addition to these studies with urea, the effect of concentration of the perturbing solute sarcosine (*N*-methylglycine) on the solubility of amino acids has also been systematically examined.[42] Calculations of μ_{23}^{ex} (not shown) from these solubility data over the range 0−6 M sarcosine reveal moderate dependences of μ_{23}^{ex} on sarcosine concentration for many amino acids.

To obtain side chain transfer free energies using model compound solubility data, the transfer free energy of glycine at its solubility limit (~3.3 *m*) is subtracted from that of other amino acids with very different solubilities[4] (ranging from ~0.02 to 15 *m*), thereby potentially introducing systematic errors. For urea−amino acid interactions, our analyses based on ID data indicate that these systematic errors are small, in part because μ_{23}^{ex} for urea−glycine preferential interactions is near zero at all glycine and urea concentrations. For studies with other perturbing solutes such as sarcosine, glycine betaine, or trimethylamine-*N*-oxide (TMAO), where values of μ_{23}^{ex} for interactions with glycine are more significant, the m_2 and m_3 dependences of μ_{23}^{ex}

are expected to introduce more significant systematic errors in values for side chain and backbone contributions calculated from solubility data. Our "coarse-grained" structural analysis of the thermodynamic data reveals that urea interacts primarily with polar amide ASA on both model compounds and biopolymers and reveals a somewhat smaller magnitude of urea interactions with polar amide surface of Gly_2, Gly_3, and GlyAla compared with biopolymer surface. The analysis of Auton and Bolen[4] based on solubility data also deduces that the primary interaction of urea is with amide (backbone) surface but finds that interactions of urea with many side chains (in addition to those with amide groups) are significant.

Molecular dynamics (MD) simulations with native and denatured proteins in aqueous urea solutions have been performed by several laboratories to investigate urea−protein interactions and the molecular basis of unfolding by urea.[43−47] While thermodynamic quantities like μ_{23}^{ex} have not been computed, aspects of the results of these simulations can be semiquantitatively compared with thermodynamic analyses. Bennion and Daggett observed residence times for urea in the vicinity of polar groups which substantially exceeded those for urea at nonpolar groups or for water at polar and nonpolar surface.[44] These authors also reported numbers of urea (44) and water (157) molecules at the surface of an unfolded protein (chymotrypsin inhibitor 2) in 8 M (13 *m*) urea.[43] This simulation yields an average local urea molality of 15.6 *m* in the first layer of water, or a composite $K_P = 1.2$, similar to that predicted by application of the SPM to unfolded protein surface, which is ∼16% polar amide and consequently is predicted to exhibit an average $K_P = 1.12$ (calculated assuming $b_1 = 0.11$ $H_2O/Å$,[2] i.e. one layer of water for comparison with the MD data[43]).[11] Work in our laboratory suggests that $b_1 = 0.19 \pm 0.03$ $H_2O/Å^2$ or almost two water layers;[8,9,48] however, the information provided by Bennion and Daggett[43,44] does not allow for comparison with this larger hydration layer. Smith[46] proposed a quantitative comparison of MD simulations with thermodynamic data, which should provide a means of comparing predicted and observed values of μ_{23}^{ex} and of testing the SPM.

A parallel analysis to that developed here for urea has been applied to interpret thermodynamic data for preferential interactions of glycine betaine with biopolymers[7,17] and model compounds (J.G.C. et al., in preparation). The analysis is being extended to preferential interactions of other neutral solutes and salts with biopolymers and model compounds. These studies will test the osmophobic hypothesis[5,49−51] that osmolytes stabilize proteins primarily through unfavorable preferential interactions with regions of the peptide backbone exposed in unfolding and will extend the test to solute effects on processes involving folded protein and nucleic acid surfaces. For glycine betaine, analyses of values of μ_{23}^{ex} for interactions with a variety of biopolymer (and model compound) surfaces by the method of Figure 5 have revealed a strong correlation with anionic surface (carboxylate and phosphate oxygen) and possibly a weak correlation with polar amide surface.[7,8] The long-term goal of these studies is to obtain relationships between solute−biopolymer preferential interactions and water accessible biopolymer surface (ASA) and to use this information to predict and interpret the effects of solutes on biopolymer processes.

Acknowledgment. The authors wish to thank the reviewers for their detailed and helpful comments on this manuscript. The support of this research by the NIH (GM 47022) is gratefully acknowledged. J.G.C. was supported in part by an NIH Molecular Biophysics Traineeship.

Appendix

Calculation of μ_{22} from Two-Component ID Data for Urea and Amino Acids. Two-component ID data for urea and the amino acids are from several different sources (Ala,[20,36] Gly,[32] Gly_2,[32] Gly_3,[33] GlyAla,[32] urea[34,35]). ID data for urea extend to ∼20 *m*, well beyond the molalities of the three-component data analyzed. ID data for the amino acids extend to molalities within 5% of the solubility limits. In fitting urea, glycine, diglycine, alanine, and glycylalanine data, systematic trends in the residuals resulting from second-order polynomial fits to osmolality data are eliminated by third-order polynomial fits. F tests indicate that fourth-order fits do not significantly improve fitting accuracy. Triglycine data are well-fit by a second-order polynomial, and using a third-order polynomial does not significantly change values of μ_{23}^{ex}/RT. Values of the coefficients α_{i0}, and α_{0j} are tabulated in Table 1. This section addresses only two-component solutions, so osmolality coefficients (α_i) are assigned a single subscript. Solute self-interactions are largely determined by the coefficient α_2; however, there is great uncertainty in determining solute self-interactions as $m_2 \rightarrow 0$. For an uncharged solute as its molality approaches zero, osmolality must approach this solute molality. From ID data, which cannot be obtained below ∼0.2 *m*, α_1 is not precisely 1, and the differences from 1 are of a magnitude similar to α_2. This means that solute self-interactions are not precisely determined; however, certain trends in self-interactions are unaffected by constraining α_1 to 1 or by allowing it to float in fitting two-component ID data. Values of μ_{22}^{ex}/RT normalized by polar amide ASA are of the same order of magnitude as $\mu_{23}^{ex}/(RT \cdot ASA_{polar\ amide})$ (i.e., 10^{-3} m^{-1} $Å^{-2}$). For urea, Gly, Gly_2, and GlyAla, μ_{22}^{ex}/RT becomes less negative, and for Ala, μ_{22}^{ex}/RT becomes more positive, with increasing m_2 because of positive values of α_3. For Gly_3, α_3 is not significantly different from zero.

Supporting Information Available: Brief derivation of eq 8a from eq 10 of ref 29. Solubility data used in linear and quadratic fits for calculating values found in Table 2 and Figure 3. Residuals of polynomial fits to complete three-component isopiestic distillation data sets and residuals of polynomial fits found in Table 1 to truncated three-component ID data. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1963**, *238*, 4074.
(2) Tanford, C. *J. Am. Chem. Soc.* **1964**, *86*, 2050.
(3) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1970**, *245*, 1648.
(4) Auton, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15065.
(5) Auton, M.; Bolen, D. W. *Biochemistry* **2004**, *43*, 1329.
(6) Robinson, D. R.; Jencks, W. P. *J. Am. Chem. Soc.* **1965**, *87*, 2462.
(7) Hong, J.; Capp, M. W.; Saecker, R. M.; Record, M. T. *Biochemistry* **2005**, *44*, 16896.
(8) Felitsky, D. J.; Cannon, J. G.; Capp, M. W.; Hong, J.; Van Wynsberghe, A. W.; Anderson, C. F.; Record, M. T., Jr. *Biochemistry* **2004**, *43*, 14732.
(9) Hong, J.; Capp, M. W.; Anderson, C. F.; Saecker, R. M.; Felitsky, D. J.; Anderson, M. W.; Record, M. T., Jr. *Biochemistry* **2004**, *43*, 14744.
(10) Courtenay, E. S.; Capp, M. W.; Record, M. T. *Protein Sci.* **2001**, *10*, 2485.
(11) Courtenay, E. S.; Capp, M. W.; Saecker, R. M.; Record, M. T. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 72.
(12) Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record, M. T. *Biochemistry* **2000**, *39*, 4455.
(13) Record, M. T.; Anderson, C. F. *Biophys. J.* **1995**, *68*, 786.
(14) Felitsky, D. J.; Record, M. T., Jr. *Biochemistry* **2004**, *43*, 9276.
(15) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138.

(16) Shelton, V. M.; Sosnick, T. R.; Pan, T. *Biochemistry* **1999**, *38*, 16831.
(17) Kontur, W. S.; Saecker, R. M.; Davis, C. A.; Capp, M. W.; Record, M. T. *Biochemistry* **2006**, *45*, 2161.
(18) Wang, A. J.; Bolen, D. W. *Biochemistry* **1997**, *36*, 9101.
(19) Schonert, H.; Stroth, L. *Biopolymers* **1981**, *20*, 817.
(20) Rafflenbeul, L.; Pang, W. M.; Schonert, H.; Haberle, K. *Z. Naturforsch., C: Biosci.* **1973**, *28*, 533.
(21) Uedaira, H. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3068.
(22) Record, M. T.; Zhang, W. T.; Anderson, C. F. *Adv. Protein Chem.* **1998**, *51*, 281.
(23) Timasheff, S. N.; Xie, G. *Biophys. Chem.* **2003**, *105*, 421.
(24) Timasheff, S. N. *Adv. Protein Chem.* **1998**, *51*, 355.
(25) Hong, J.; Capp, M. W.; Anderson, C. F.; Record, M. T. *Biophys. Chem.* **2003**, *105*, 517.
(26) Anderson, C. F.; Felitsky, D. J.; Hong, J.; Record, M. T. *Biophys. Chem.* **2002**, *101*−*102*, 497.
(27) Anderson, C. F.; Courtenay, E. S.; Record, M. T. *J. Phys. Chem. B* **2002**, *106*, 418.
(28) Anderson, C. F.; Record, M. T. *J. Phys. Chem.* **1993**, *97*, 7116.
(29) Anderson, C. F.; Record, M. T. *Biophys. Chem.* **2004**, *112*, 165.
(30) Robinson, R. A.; Stokes, R. H. *J. Phys. Chem.* **1961**, *65*, 1954.
(31) Smith, E. R. B.; Smith, P. K. *J. Biol. Chem.* **1937**, *117*, 209.
(32) Ellerton, H. D.; Reinfelds, G.; Mulcahy, D. E.; Dunlop, P. J. *J. Phys. Chem.* **1964**, *68*, 398.
(33) Smith, E. R. B.; Smith, P. K. *J. Biol. Chem.* **1940**, *135*, 273.
(34) Bower, V. E.; Robinson, R. A. *J. Phys. Chem.* **1963**, *67*, 1524.
(35) Scatchard, G.; Hamer, W. J.; Wood, S. E. *J. Am. Chem. Soc.* **1938**, *60*, 3061.
(36) Smith, P. K.; Smith, E. R. B. *J. Biol. Chem.* **1937**, *121*, 607.
(37) The offset in the high osmolality data may arise if the ID reference solute was changed from NaCl at lower osmolalities to another solute at higher osmolalities. Rard, J. A.; Platford, R. F. Experimental Methods: Isopiestic. In *Activity Coefficients in Electrolyte Solutions*; 2nd ed.; Pitzer, K. S., Ed.; CRC Press: Boca Raton, FL, 1991; p 209.
(38) Stigter, D.; Hill, T. L. *J. Phys. Chem.* **1959**, *63*, 551.
(39) Cannon, J. G. *Developing Osmolytes as Nano-Probes of Biopolymer Surfaces, Interfaces, and Conformational Changes*; University of Wisconsin: Madison, WI, 2007.
(40) Talukdar, H.; Rudra, S.; Kundu, K. K. *Can. J. Chem.* **1988**, *66*, 461.
(41) Felitsky, D. J.; Record, M. T., Jr. *Biochemistry* **2003**, *42*, 2202.
(42) Liu, Y. F.; Bolen, D. W. *Biochemistry* **1995**, *34*, 12884.
(43) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6433.
(44) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142.
(45) Smith, L. J.; Jones, R. M.; van Gunsteren, W. F. *Proteins: Struct., Funct., Bioinfo.* **2005**, *58*, 439.
(46) Smith, P. E. *J. Phys. Chem. B* **2004**, *108*, 18716.
(47) Baynes, B. M.; Trout, B. L. *J. Phys. Chem. B* **2003**, *107*, 14058.
(48) Pegram, L. M.; Record, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14278.
(49) Street, T. O.; Bolen, D. W.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13997.
(50) Bolen, D. W. *Methods* **2004**, *34*, 312.
(51) Auton, M. T.; Bolen, D. W. *Biophys. J.* **2003**, *84*, 164A.
(52) Scholtz, J. M.; Barrick, D.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 185.