BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

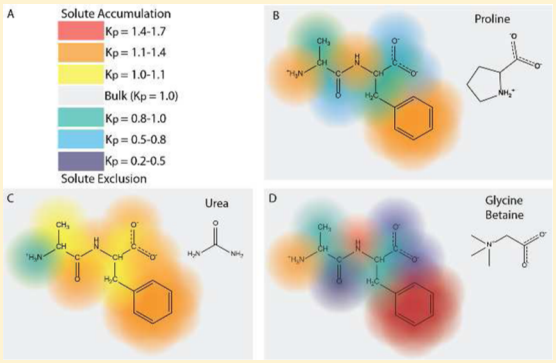
Quantifying Additive Interactions of the Osmolyte Proline with Individual Functional Groups of Proteins: Comparisons with Urea and Glycine Betaine, Interpretation of *m*-Values

Roger C. Diehl,[†,⊥] Emily J. Guinn,[‡,⊥,¶] Michael W. Capp,[†] Oleg V. Tsodikov,[§] and M. Thomas Record, Jr.*[,†,‡]

[†]Department of Biochemistry and [‡]Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, United States
[§]Department of Pharmaceutical Sciences, University of Kentucky, Lexington, Kentucky 40538, United States

Supporting Information

ABSTRACT: To quantify interactions of the osmolyte L-proline with protein functional groups and predict their effects on protein processes, we use vapor pressure osmometry to determine chemical potential derivatives $d\mu_2/dm_3 = \mu_{23}$, quantifying the preferential interactions of proline (component 3) with 21 solutes (component 2) selected to display different combinations of aliphatic or aromatic C, amide, carboxylate, phosphate or hydroxyl O, and amide or cationic N surface. Solubility data yield μ_{23} values for four less-soluble solutes. Values of μ_{23} are dissected using an ASA-based analysis to test the hypothesis of additivity and obtain α-values (proline interaction potentials) for these eight surface types and three inorganic ions. Values of μ_{23} predicted from these α-values agree with the experiment, demonstrating additivity. Molecular interpretation of α-values using the solute partitioning model yields partition coefficients (K_p) quantifying the local accumulation or exclusion of proline in the hydration water of each functional group. Interactions of proline with native protein surfaces and effects of proline on protein unfolding are predicted from α-values and ASA information and compared with experimental data, with results for glycine betaine and urea, and with predictions from transfer free energy analysis. We conclude that proline stabilizes proteins because of its unfavorable interactions with (exclusion from) amide oxygens and aliphatic hydrocarbon surfaces exposed in unfolding and that proline is an effective in vivo osmolyte because of the osmolality increase resulting from its unfavorable interactions with anionic (carboxylate and phosphate) and amide oxygens and aliphatic hydrocarbon groups on the surface of cytoplasmic proteins and nucleic acids.



INTRODUCTION

Osmotic stress is a common occurrence and an evolutionary challenge for bacteria, plants, and some types of animal cells. The α-imino acid L-proline is one of several "osmoprotectant" solutes (including the methylated α-amino acid glycine betaine [*N*,*N*,*N*-trimethylglycine; GB] and the β-imino acid ectoine) that are transported from the growth medium to high concentrations in the *E. coli* cytoplasm in response to osmotic stress.[1,2] Proline is also an important plant osmolyte.[3,4] Increased proline synthesis makes wheat more resistant to dehydration stress,[5] and proline accumulates in drought-stressed potatoes[6] and hot peppers.[7] Proline also increases the cold tolerance of maize.[8] Proline and other osmolytes have often been called "compatible" solutes because at molar concentrations they do not greatly affect enzyme activity and do not destabilize or denature globular proteins.[9,10] Nevertheless, these osmolytes are not inert; proline stabilizes globular proteins and destabilizes RNA secondary and tertiary structure.[11] Proline inhibits aggregation of the aggregation-prone P39A variant of the cellular retinoic acid-binding protein CRABP and of constructs containing a polyQ tract like that found in Huntington's disease.[12]

Previously, we quantified the preferential interactions of the solutes urea and GB with extensive sets of model compounds displaying the functional groups of biopolymers. We discovered that these interactions are quantitatively interpreted and predicted as additive contributions from interactions of urea or GB with the individual functional groups on the model compounds using an analysis motivated by the solute partitioning model[13−15] and based on water-accessible surface area.[15−21] We interpreted the observed accumulation or exclusion of urea or GB in the vicinity of these groups in terms of noncovalent interactions with these groups, relative to interactions with water. Because GB lacks a hydrogen bond donor (whereas water has two and urea has four), we interpreted unfavorable interactions of GB with amide and anionic O as an inability to compete with water for hydrogen

Received: May 29, 2013
Revised: July 29, 2013
Published: August 5, 2013

 © 2013 American Chemical Society dx.doi.org/10.1021/bi400683y | *Biochemistry* 2013, 52, 5997−6010

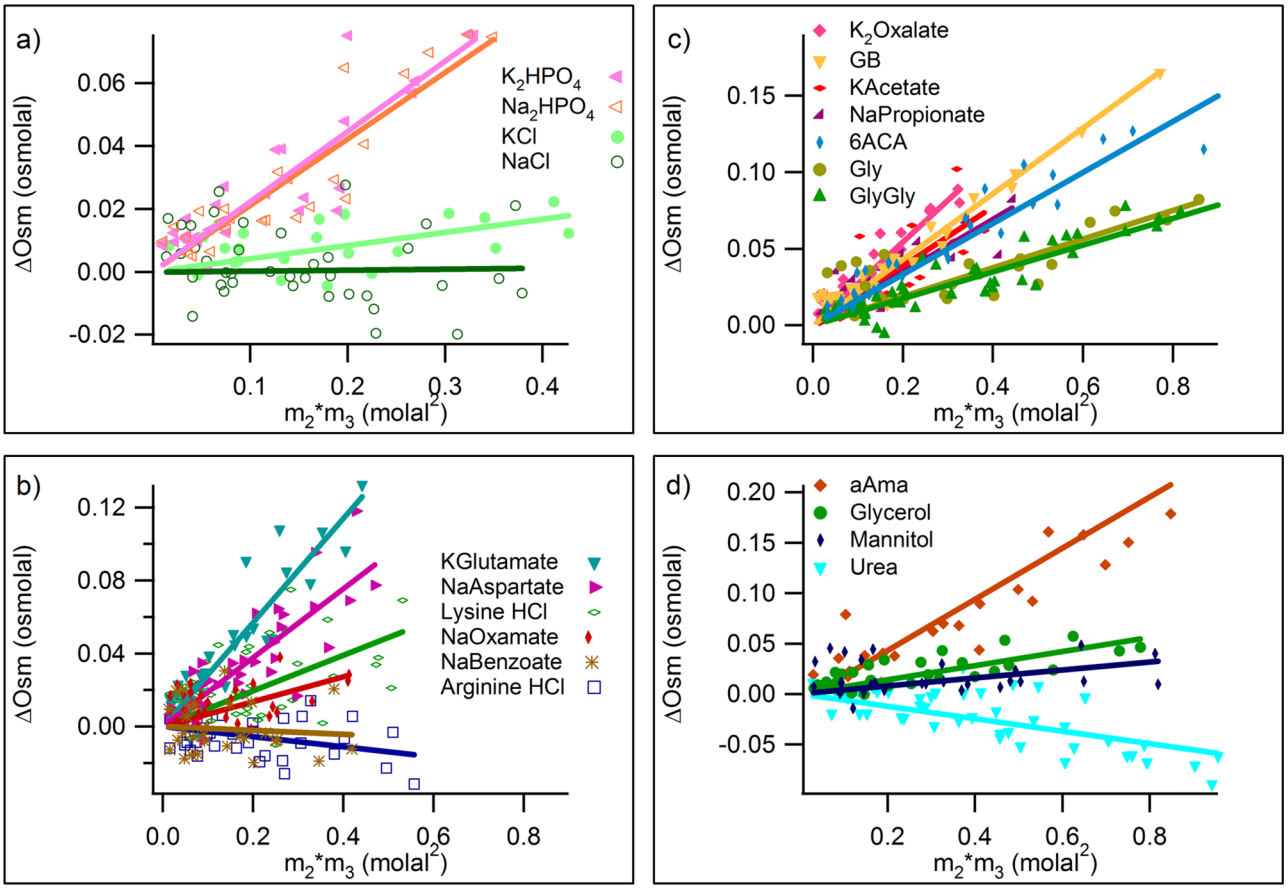


Figure 1. Quantifying interactions of proline with soluble model compounds by osmometry. Excess osmolalities ΔOsm of three component solutions of the model compound and proline, determined by VPO, are plotted as a function of m_2m_3, the product of the molal concentrations of model compound and proline. Groups of model compounds are (a) inorganic salts, (b) organic salts, (c) carboxylate salts and zwitterions, and (d) uncharged species. Linear least squares regression lines are shown for all compounds with slopes equal to μ_{23}/RT (eq 1).

bonding with these groups, whereas favorable interactions of urea with these groups are explained by urea having more ways to form hydrogen bonds or forming stronger hydrogen bonds.[17] Proline provides an interesting structural comparison because, like GB, proline is zwitterionic but, unlike GB, the proline imino group can function as a hydrogen bond donor, though not as effectively as the primary ammonium group of other amino acids or the primary amine groups of urea.

In the research reported here, quantitative thermodynamic information about the preferential interactions of proline with model compounds displaying the functional groups of proteins is obtained by vapor pressure osmometry (VPO, for sufficiently soluble model compounds) and solubility assays (for sparingly soluble model compounds). Both assays determine chemical potential derivatives $d\mu_2/dm_3 = \mu_{23}$, quantifying the effect of proline on the chemical potential of the model compound; μ_{23} also quantifies the preferential interaction of proline (component 3) with these model compounds (component 2) because of the dialysis equilibrium or preferential interaction coefficient $\Gamma_{23} = -\mu_{23}/\mu_{33}$, where $\mu_{33} = d\mu_3/dm_3$. These μ_{23} values are needed to interpret or predict effects of proline on biopolymer processes, for which the free energy derivative $d\Delta G^o_{obs}/dm_3$ = m-value = $\Delta\mu_{23}$. For soluble model compounds, values of μ_{23} from VPO measurements also provide the free energy of transfer of the model compound from water to 1 m proline solution, which cannot be accurately obtained from solubility data alone.

Values of μ_{23} are interpreted at two levels, first using the hypothesis of additivity and a coarse-grained dissection of the model compound surface into the water-accessible surface area (ASA) of different functional groups. This first level of analysis yields intrinsic interaction potentials (called α-values) quantifying the interaction of proline with a unit area of each functional group. These α-values are sufficient to interpret or predict the interactions of proline with protein surfaces and the effects of proline on protein processes in terms of ASA information. At a second (molecular) level, these α-values are interpreted using the molecular thermodynamic analysis of the solute partitioning model (SPM) to obtain microscopic partition coefficients (K_p values) that quantify the accumulation or exclusion of proline in the water of hydration (b_1) of these groups. Comparison of K_p or α-values for interactions of proline, urea, and GB with those of the functional groups of proteins provides significant insight into the molecular interactions that lead to accumulation or exclusion of these solutes in the vicinity of each type of functional group. These K_p or α-values not only explain why proline is an effective osmolyte in vivo and a protein stabilizer but also allow a quantitative prediction of these effects from ASA information. We also compare α-values predictions of interactions of proline, glycine betaine, and urea with those of the peptide backbone and amino acid side chain with those obtained from amino acid solubility data using the group transfer free energy (GTFE) analysis.[22] These comparisons demonstrate the advantages of the α-value analysis to interpret

dx.doi.org/10.1021/bi400683y | Biochemistry 2013, 52, 5997−6010

or predict interactions of any solute (e.g., proline, urea, glycine betaine) with a protein and effects of these solutes on protein processes. In particular, the α-values determined here make proline very useful in conjunction with other solutes as a probe of the amount and type of surface buried in individual steps of protein processes and mechanisms.

MATERIALS AND METHODS

Details about methods and materials used (chemicals, osmometry/solubility procedures, etc.) and data analysis are provided in Supporting Information.

RESULTS

Selection of Model Compounds. Preferential interactions (μ_{23} values) of proline with 13 soluble salts (arginine hydrochloride, potassium acetate, KCl, potassium glutamate, K_2HPO_4, potassium oxalate, lysine hydrochloride, sodium aspartate, sodium benzoate, NaCl, sodium oxamate, sodium propionate, and Na_2HPO_4), 8 soluble nonelectrolytes (acetyl-Ala-methylamide (aAma), glycerol, glycine, GB, glycylglycine (glygly), mannitol, urea, and zwitterionic 6-aminocaproic acid (6ACA)), and low-solubility naphthalene were determined as part of this study. Preferential interactions of proline with three low-solubility cyclic dipeptides (cycloGG, cycloAA, cycloLA) were determined from literature data. Compounds investigated here were selected to provide at least two examples of each of the seven (coarse-grained) functional groups of proteins (aliphatic and aromatic C, carboxylate, amide and hydroxyl O, and amino and cationic N), as well as two examples of anionic phosphate O and of salts of the relevant inorganic ions, in order to vary the surface composition of these functional groups as widely as possible (e.g., urea, oxamate, aAma, and glygly to vary the ratio of amide O to amide N surface areas) and to allow interactions of proline with inorganic ions to also be determined. Preferential interactions of proline with low-solubility compounds (naphthalene, cycloGG, cycloAA, cycloLA) were determined from solubility data, and interactions with soluble compounds were determined by vapor pressure osmometry (VPO).

Determination of Interactions of Proline with Soluble Salts and Polar Compounds by Vapor Pressure Osmometry (VPO). To determine μ_{23} for interactions of proline with soluble model compounds, osmolality differences (ΔOsm) between proline model compound three-component solutions and the corresponding two-component solutions are determined as a function of the concentrations of model compound (m_2) and proline (m_3):

$$\Delta\mathrm{Osm} = \mathrm{Osm}(m_2, m_3) - \mathrm{Osm}(m_2, 0) - \mathrm{Osm}(0, m_3) \quad (1)$$

At low to moderate concentrations of the two solutes, multicomponent solution thermodynamic analysis[23−25] predicts that ΔOsm is proportional to the product of molal concentrations m_2m_3, with proportionality constant μ_{23}/RT for the proline−model compound interaction:

$$\Delta\mathrm{Osm} = (\mu_{23}/RT)m_2m_3 \quad (2)$$

The panels of Figure 1 plot ΔOsm as a function of m_2m_3 for all 21 model compounds investigated by VPO. In all cases, ΔOsm is found to be proportional to m_2m_3, showing that values of μ_{23} are independent of proline and model compound concentrations over the ranges investigated.

Determination of Preferential Interactions of Proline with Low-Solubility Model Compounds from Solubility Data. Preferential interactions of proline with four sparingly soluble model compounds (naphthalene and three cyclic dipeptides) are quantified from solubility data. An absorbance-based assay was used to determine the effect of proline on the solubility of naphthalene. Effects of proline on the solubility of three cyclic dipeptides (cycloAA, cycloGG, cycloLA) were obtained from the literature.[26] In Figure 2, logarithms of normalized solubilities are plotted vs proline molality m_3. (The quantity plotted is $-\ln(m_2^{ss}/m_{2,0}^{ss})$, where m_2^{ss} is the molal concentration of the model compound in the saturated solution and $m_{2,0}^{ss}$ is the corresponding solubility in the absence of proline.)

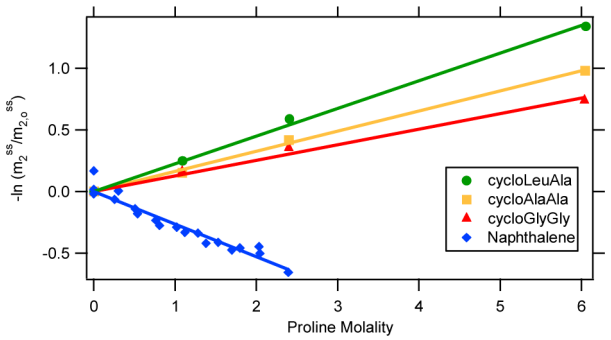


Figure 2. Quantifying interactions of proline with sparingly soluble compounds by solubility assays. The negative logarithm of the molal solubility of the model compound in proline solutions (m_2^{ss}), normalized by its solubility in water ($m_{2,o}^{ss}$), is plotted vs the molality of proline (m_3). Data for cyclic dipeptides are from ref 21. Slopes of these plots are equal to μ_{23}/RT (eq 3). In all cases, $m_{2,o}^{ss}$ is obtained by extrapolation of the full data set to $m_3 = 0$.

These plots are linear for all four solutes; from the thermodynamic analysis, slopes $-\mathrm{d}(\ln m_2^{ss})/\mathrm{d}m_3$ are approximately equal to μ_{23}/RT:

$$-\mathrm{d}(\ln m_2^{ss})/\mathrm{d}m_3 = \mu_{23}/[RT(1 + \varepsilon_2)] \approx \mu_{23}/RT \quad (3)$$

In eq 3, the self-interaction nonideality term, $\varepsilon_2 = (\mathrm{d}(\ln \gamma_2)/\mathrm{d}(\ln m_2))_{m_3} \approx 0$, provided that the saturated solution is very dilute in component 2. The linearity of the plotted data reveals that μ_{23} is independent of proline concentration over the range examined.

Experimentally determined μ_{23}/RT values (slopes) for all compounds shown in Figures 1 and 2 are presented in Table S1 in the Supporting Information. In the two sections that follow, trends in these μ_{23} values are discussed, after which they are quantitatively dissected into functional group contributions.

Qualitative Deductions from μ_{23} Values about Interactions of Proline with Functional Groups of Proteins. Inspection of μ_{23}/RT values (Figures 1 and 2 and Supporting Information Table S1) that quantify the preferential interactions of proline with this range of model compounds reveals the following:

(1) Proline increases the chemical potential of most model compounds studied ($\mu_{23} > 0$), indicating that interactions of proline with many of the functional groups of proteins must be unfavorable relative to interactions with water.
(2) Proline significantly reduces the chemical potential of naphthalene ($\mu_{23} < 0$). Therefore, interactions of proline

dx.doi.org/10.1021/bi400683y | Biochemistry 2013, 52, 5997−6010

Table 1. Solute-Functional Group Interaction Potentials (α-Values) and Local Solute Partition Coefficients (K_p) by Surface Type or Ion for Proline, Glycine Betaine, and Urea

	$10^4 \times \alpha_i$ (m^{-1} Å^{-2})			partition coefficient (K_p)[a]		
surface type (i)	proline	GB[39]	urea[19,17]	proline	GB	urea
aliphatic C	5.3 ± 1.3	3 ± 3	−1.1 ± 0.5	0.85 ± 0.04	0.92 ± 0.08	1.03 ± 0.02
aromatic C	−9.2 ± 0.9	−23 ± 4	−8.9 ± 0.5	1.26 ± 0.03	1.62 ± 0.11	1.28 ± 0.02
hydroxyl O	−0.7 ± 1.3	1 ± 2	−2.5 ± 0.6	1.02 ± 0.04	0.97 ± 0.06	1.08 ± 0.02
amide O	14.5 ± 4.5	28 ± 10	−8.5 ± 1.8	0.59 ± 0.13	0.24 ± 0.27	1.28 ± 0.06
amide N	−11.8 ± 3.2	−20 ± 7	−3.7 ± 1.6	1.33 ± 0.09	1.54 ± 0.19	1.10 ± 0.07
carboxylate O	16.6 ± 4.3	29 ± 2	−3.7 ± 1.6	0.53 ± 0.12	0.22 ± 0.06	1.13 ± 0.05
phosphate O	18.0 ± 5.2	49 ± 4	−5.8 ± 1.2	0.49 ± 0.15	0.00	1.18 ± 0.04
cationic N	−12.6 ± 4.3	−12 ± 4	1.6 ± 1.7	1.36 ± 0.12	1.32 ± 0.11	0.94 ± 0.05

	$10^2 \times \beta_{ion}$ (m^{-1})		
ionic species (ion)	proline	GB[39]	urea[19,17]
Na^+	−4 ± 4	0[b]	10 ± 1
K^+	−2 ± 4	5 ± 2	14 ± 2
Cl^-	4 ± 4	−4 ± 4	−16 ± 2

[a] K_p was calculated from the corresponding α-value using eq 5 and b_1 = 0.18 H_2O Å^{-2} for the local water of hydration of these functional groups, as determined previously,[28,29] except for phosphate, for which a minimum b_1 = 0.27 H_2O Å^{-2} was determined previously.[39] [b] β_{ion} for the Na–GB interaction was set to zero.[39]

with aromatic surfaces are favorable relative to interactions with water, causing an accumulation of proline (presumably a cation–π interaction).

(3) 6ACA, a zwitterionic ε-amino acid analog of glycine with four additional methylene groups, interacts much less favorably with proline than glycine does, as indicated by its more positive μ_{23} value. This indicates that the interaction of proline with an aliphatic surface is unfavorable relative to water, resulting in the exclusion of proline from aliphatic surfaces.

(4) Mannitol and glycerol have similar amounts of aliphatic surface area, but mannitol has much more hydroxyl surface area. Because the interaction of proline with mannitol is more favorable than its interaction with glycerol, the interaction of proline with hydroxyl surfaces must be favorable relative to interaction with water.

(5) Proline has a nearly neutral ($\mu_{23} \sim 0$) interaction with NaCl and KCl, so interactions with K^+ and Na^+ must be similar in magnitude and opposite in sign to the interaction with the Cl^- anion. The simplest interpretation of these results, verified by the quantitative analysis described below, is that interactions of proline with K^+, Na^+, and Cl^- are small in magnitude.

(6) Preferential interactions of proline with phosphate salts (K_2HPO_4, Na_2HPO_4) and small carboxylate salts (potassium acetate, sodium propionate, potassium oxalate) are among the most unfavorable ($\mu_{23} > 0$). This indicates that proline is excluded from the vicinity of anionic carboxylate and phosphate oxygens, because its interaction with Na^+ and K^+ is small in magnitude.

(7) Lysine hydrochloride and arginine hydrochloride both have considerable carboxylate and aliphatic surface area, yet the interaction of proline with lysine hydrochloride is only slightly unfavorable and the interaction of proline with arginine hydrochloride is slightly favorable. Therefore, the interaction of proline with ammonium and guanidinium nitrogens must be sufficiently favorable to compensate for the unfavorable carboxylate interaction deduced above, leading to accumulation. Moreover, proline interacts more favorably with 6ACA than with GB, indicating a favorable interaction of proline with the ammonium group because 6ACA and GB display similar amounts of aliphatic and carboxylate surface.

Quantifying Preferential Interactions of Proline with Functional Groups of Proteins and Phosphate. Previous research to quantify and interpret preferential interactions of GB[16] and urea[17] with a similar model compound data set revealed that μ_{23} values for these solutes were well-interpreted as additive contributions from the interactions of these solutes with the individual functional groups on the model compounds. For model compounds that are salts, additive contributions to μ_{23} from interactions with the inorganic ions that make up the electroneutral component must be included. The contribution to μ_{23} from the interaction of the solute (urea, GB, proline) with a particular functional group on the model compound is represented as the product of an intrinsic interaction potential per unit area (designated α_i) and the ASA of that group:

$$\frac{\mu_{23}}{RT} = \sum_i \alpha_i (\mathrm{ASA})_i + \sum_{ion} \nu_{ion} \beta_{ion} \quad (4)$$

In applying eq 4 to salts with one or more inorganic ions, β_{ion} is the contribution to μ_{23}/RT from the interaction of the solute with an inorganic ion, and ν_{ion} is the number of such ions per formula unit of the salt.

For interactions of GB[16] and urea[17] with model compounds displaying one or more of the functional groups of proteins, α-values obtained from eq 4 reproduce μ_{23}/RT values to within ±20% for most compounds in the training set. Here, values of μ_{23}/RT (Supporting Information Table S1) quantifying the interactions of proline with 25 model compounds were globally fit to eq 4, using ASA values in Supporting Information Table S2, to obtain eight interaction potentials (α-values) quantifying interactions of proline with a unit area of each of the principal functional groups of proteins and anionic phosphate O and three interaction potentials β_{ion} quantifying interactions of proline with each inorganic ion. Fitted values of α and β_{ion} are listed in Table 1 along with the corresponding α- and β_{ion}-values previously obtained for urea and GB.[17]

To test this additivity analysis (eq 4), values of μ_{23}/RT for interactions of proline with the 25 model compounds in the training set were calculated from the α_i- and β_{ion}-values given in

dx.doi.org/10.1021/bi400683y | Biochemistry 2013, 52, 5997–6010

Table 1 and the ASA values given in Supporting Information Table S2. Figure 3 compares predicted and observed values of

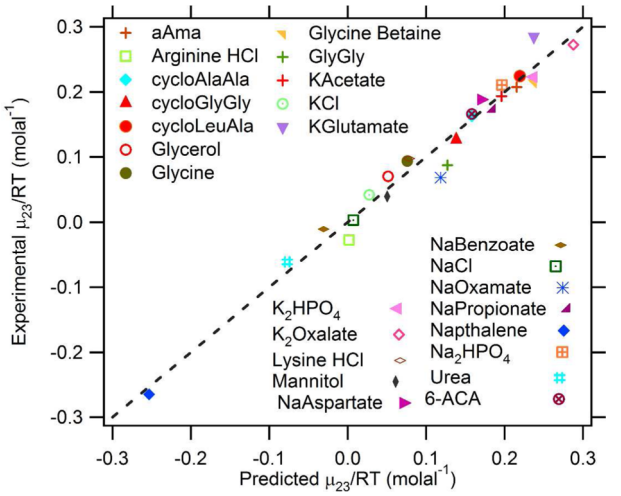


Figure 3. Comparison of observed and predicted proline–model compound interactions. Experimentally determined μ_{23}/RT for interactions of proline with the training set of 25 model compounds are compared with μ_{23}/RT predicted from 7 functional group α-values and 4 β_{ion}-values (Table 1) and ASA information (Supporting Information Table S2). The dashed line represents equality of predicted and experimental values.

μ_{23}/RT with the dashed diagonal line representing perfect correspondence. Very good agreement is obtained; the average deviation is ±17% where μ_{23}/RT is greater in magnitude than 0.03 m^{-1} and is ±0.03 m^{-1} for compounds where $\mu_{23}/RT \approx 0$ (NaCl, arginine hydrochloride, sodium benzoate).

To test the redundancy and robustness of the fit, we asked which compounds could be eliminated from the data set and be predicted from α-values obtained from analysis of the remaining 24 compounds. Fourteen compounds (cycloLA, cycloAA, cycloGG, 6-aminocaproic acid, sodium oxamate, potassium oxalate, sodium propionate, potassium glutamate, potassium acetate, lysine HCl, arginine HCl, glycine, sodium aspartate, and glycylglycine) can be removed with only small effects (±1 SE) on proline–functional group α-values, proline–ion (β_{ion}) interaction potentials, and predicted μ_{23}/RT values, including the prediction for the omitted compound. Indeed, at least one combination of 5 compounds (cycloAA, 6ACA, diglycine, NaOxamate, NaAspartate, 6ACA) is sufficiently redundant with the other 20 that it can be eliminated without significant effect on α-values, proline–ion (β_{ion}) interaction potentials, and predicted μ_{23}/RT values. However, none of the 11 compounds urea, naphthalene, KCl, aAma, NaCl, Na benzoate, mannitol, glycerol, Na_2HPO_4, K_2HPO_4, and GB can be eliminated without reducing the quality of the fit and of the prediction of their μ_{23}/RT values. Na benzoate and naphthalene, mannitol and glycerol, and Na_2HPO_4 and K_2HPO_4 are needed to determine α-values for aromatic, hydroxyl O, and phosphate O, respectively. NaCl and KCl are needed to obtain accurate β-values, and the other three compounds represent unique ratios of surface types needed to decouple otherwise correlated interactions.

Interpreting α-Values Using the SPM. According to the solute partitioning model (SPM), preferential interactions are manifested as the partitioning of the solute between the local water of hydration of a particular functional group and the bulk solution.[27] If the preferential interaction between the solute and a functional group is favorable ($\alpha_i < 0$), the solute competes effectively with the water of hydration and accumulates in the vicinity of that group, resulting in a local solute concentration that exceeds its bulk concentration. If the solute–functional group preferential interaction is unfavorable ($\alpha_i > 0$), the solute does not compete effectively with the water of hydration and is excluded from the vicinity of that group, resulting in a local solute concentration that is smaller than bulk. Accumulation has a thermodynamic effect analogous to weak binding of the solute and means that addition of the solute favors the direction of a process that exposes that functional group, whereas exclusion favors the direction that buries that functional group.

In the context of the SPM, the accumulation or exclusion of proline in the vicinity of a functional group on a model compound (or protein) is quantified by a partition coefficient $K_{p,i} = m_{3,local}/m_{3,bulk}$, where $m_{3,local}$ and $m_{3,bulk}$ are local and bulk concentrations of proline. Except at very high concentrations of the model compound (component 2), $m_{3,bulk} = m_3$. As previously described,[16,17] $K_{p,i}$ is related to α_i by eq 5, where $b_{1,i}$ is the surface density of water in the hydration layer.

$$\alpha_i = -\frac{(K_{p,i} - 1)b_{1,i}(1 + \varepsilon)}{55.5} \quad (5)$$

Analysis of μ_{23} or α-values for highly excluded solutes and salts provide a lower bound value $b_1 = 0.18$ Å^{-2} for both hydrocarbon and oxygen functional groups.[16,28,29] Table 1 displays K_p values for proline and compares them to values obtained previously for urea and GB.[17] Application of K_p values from Table 1 to predict surface maps of accumulation or exclusion of proline in the vicinity of a representative peptide (AlaPhe), comparisons of K_p values for proline with those of urea and GB, and chemical explanations of these K_p values are given in the Discussion.

Interaction of Proline with Bovine Serum Albumin (BSA). In Figure 4, excess osmolalities (ΔOsm; eq 1) of

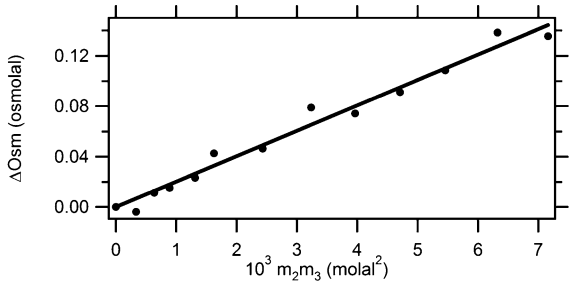


Figure 4. Quantifying interactions of proline with bovine serum albumin (BSA) by osmometry. Excess osmolalities ΔOsm of three component solutions of BSA and proline, determined by VPO, are plotted as a function of m_2m_3, the product of the molal concentrations of BSA and proline. A linear least-squares regression line is shown; from the slope $\mu_{23}/RT = 20.2 \pm 0.6$ m^{-1} (eq 2).

proline–BSA solutions as a function of the concentrations of these solutes are plotted as a function of the product of proline and BSA concentrations (m_2m_3) to determine μ_{23}/RT by eq 2. The linearity of this plot shows that μ_{23} is independent of proline and BSA concentration; from the slope, $\mu_{23}/RT = 20.2 \pm 0.6$ m^{-1} or $\mu_{23} = 11.9 \pm 0.3$ kcal mol^{-1} m^{-1} (Table 2). When normalized by the ASA of native BSA (2.8×10^4 Å^2),[20,30] these values are similar to those obtained for interactions of proline with GB and the other small solutes from which proline is

dx.doi.org/10.1021/bi400683y | Biochemistry 2013, 52, 5997–6010

highly excluded (Table 1), indicating that the interaction of proline with the BSA component is very unfavorable. Strong exclusion of proline from the surfaces of native proteins (and presumably also nucleic acids because the proline–phosphate interaction is highly unfavorable; Table 1) is a major reason why proline is an effective osmolyte in the concentrated intracellular biopolymer milieu.

DISCUSSION

Prediction of Interactions of Proline with Peptide and Protein Surfaces; Comparison with Urea and Glycine Betaine Interactions. The K_p values of Table 1 allow one to predict the local accumulation or exclusion of proline in the vicinity of the functional groups of any compound displaying these groups, including any peptide or protein surface. These predictions not only provide molecular insight into the thermodynamics of these interactions but also will provide a first level of comparison with results of MD simulations or spectroscopic measurements. As one representative example, Figure 5 shows the predicted pattern of accumulation and exclusion of proline at the surface of a model dipeptide, AlaPhe, and compares this predicted pattern with those predicted for urea and glycine betaine (GB). Red, orange, and yellow indicate strong to weak accumulation; green, blue, and violet indicate weak to strong exclusion. Figure 5 shows that the pattern of accumulation and exclusion of proline in the vicinity of this peptide is more similar to that of GB than to that of urea. Although urea is slightly to moderately accumulated near most of the functional groups of proteins, both proline and GB exhibit a wide range of favorable and unfavorable interactions. Both proline and GB are strongly excluded from amide and anionic oxygens and moderately (proline) or weakly (GB) excluded from aliphatic hydrocarbons. Furthermore, both are accumulated at aromatic hydrocarbon surfaces as strongly (proline) or more strongly (GB) than urea is. Both proline and GB are also strongly accumulated at amide and cationic nitrogens, whereas urea is modestly accumulated at amide N and weakly excluded from cationic N. The contrasting interactions of these three solutes with aliphatic C and amide O and N surfaces should make them collectively an excellent set of quantitative probes of the amount and composition of the surface buried in processes involving protein folding in which ~90% of the surface buried is hydrocarbon or amide.

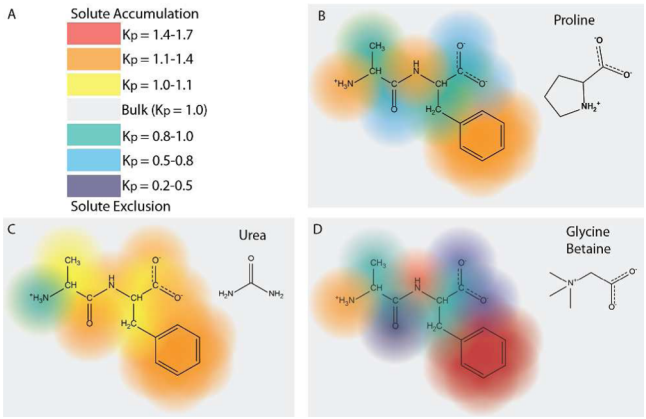


Figure 5. Predicted accumulation or exclusion of proline from the functional groups of the AlaPhe dipeptide; comparison with urea and glycine betaine predictions. Panel A gives a color scale of local partition coefficients K_p (see eq 5) used to quantify the accumulation (K_p > 1; red to yellow) or exclusion (K_p < 1; green to violet) of a solute like proline in the vicinity of functional groups of a model compound or protein. Panels B–D illustrate the predictions using K_p values from Table 1 for the accumulation or exclusion of (B) proline, (C) urea, and (D) GB in the vicinity of the individual functional groups of the AlaPhe dipeptide. Note the relatively uniform accumulation predicted for urea (except at the ammonium group) in contrast to the large differences in accumulation and exclusion of proline and GB at the different functional groups of AlaPhe.

Molecular Interpretation of Proline Interactions; Comparison with GB and Urea Interactions. Here, we propose molecular interpretations of K_p values for interactions of proline with the functional groups of proteins (cf. Table 1 and Figure 5). These molecular interpretations are supported by comparisons of K_p values for proline with those of interactions of urea and GB with these functional groups. Because these K_p values come from a "one-way" dissection of the solute–model compound μ_{23} data, they (like the α-values they are derived from) in general are interpreted as ASA-weighted averages of interactions of all the functional groups of the solute (proline, urea, or GB) with the specified model compound functional group.

The following characteristics of the molecular surfaces of proline, GB, and urea appear significant for a molecular interpretation of their interactions. Zwitterionic proline, like GB, is expected to be a good hydrogen bond acceptor, with ~80 Å^2 of anionic carboxylate oxygen ASA (~30% of total ASA), but the capabilities of these zwitterions as hydrogen bond donors are absent (GB) or limited (proline) as compared to those of urea or of amino acids with a cationic ammonium group. Cationic nitrogen ASA of proline is only 38 Å^2 (14% of total ASA), compared with 77 Å^2 for glycine and 0 Å^2 for GB. A majority of the ASA of both proline and GB is aliphatic carbon, which accounts for 151 Å^2 (56%) of total proline ASA and 197 Å^2 (71%) of total GB ASA, though the methyl groups bonded to the cationic N of GB may not behave like methyl groups bonded to neutral C atoms. The surface of polar urea is almost entirely amide N (129 Å^2, 69%) and amide O (49 Å^2, 26%) with a small exposure of the central carbon (8 Å^2, 4%).

From this surface comparison, urea is expected to interact better than proline (and much better than GB) with amide or anionic oxygens that require a hydrogen bond donor (NH→O) to compete with water and form a hydrogen bond to these groups. However, urea is also expected to interact less well with amide or cationic nitrogens than proline and GB because the larger ASA and anionic charge of the carboxylate oxygens of proline and GB should make them better hydrogen bond acceptors (O←HN model compound) than the amide oxygen of urea. (The amide nitrogens of urea are not expected to be effective hydrogen bond acceptors because of the delocalization of their lone pair electrons.) The contributions of the aliphatic surfaces of proline and GB (but not urea) to the interactions of these solutes with hydrocarbon (or polar) functional groups on model compounds are difficult to predict at present. "Two-way" dissections of sets of solute–model compound μ_{23} values into additive contributions from interactions of individual functional groups on each species, in progress, are needed to elucidate these contributions and confirm the deductions reported below.

Amide, Carboxylate, and Phosphate O. Interactions of proline with amide O (K_p = 0.6 ± 0.1), carboxylate O (K_p = 0.5 ± 0.1), and phosphate O (K_p = 0.5 ± 0.2) are decidedly unfavorable and of very similar magnitude. In all three cases, the local concentration of proline in the vicinity of these

dx.doi.org/10.1021/bi400683y | *Biochemistry* 2013, 52, 5997–6010

oxygens is predicted to be about half (50−60%) of the bulk proline concentration. GB is much more strongly excluded from phosphate and carboxylate O than is proline. ($K_p \approx 0$ for GB−phosphate O and $K_p \approx 0.2$ for GB−carboxylate O; predicted local concentrations of GB near anionic O groups are therefore ≤20% of the bulk GB concentration). Urea, by contrast, is moderately accumulated near amide O ($K_p = 1.28 \pm 0.06$), carboxylate O ($K_p = 1.13 \pm 0.05$), and phosphate O ($K_p = 1.18 \pm 0.04$).

For all three solutes, the close correspondence between K_p values for these chemically similar but Coulombically different oxygens indicates that the origins of these effects are not primarily Coulombic attraction or repulsion but rather are chemical (hydrogen bonding) interactions. Indeed, the interactions of all three solutes with amide and anionic O are well-explained in terms of the competition between these solutes and water to interact with these oxygens by hydrogen bonding. Water is the hydrogen bond donor in its interactions with both amide and anionic oxygens, so to interact favorably with these oxygens in water, the solute must be a better hydrogen bond donor (in quality or quantity of hydrogen bonds) than water.

Urea can form an NH→O hydrogen bond in four ways, as compared to the two ways to form an OH→O hydrogen bond involving water. GB lacks a hydrogen bond donor and so cannot effectively compete with water to interact with amide or anionic O. Proline falls between these extremes, having two ways (like water) to form an NH→O hydrogen bond. Water wins this competition and proline is excluded ($K_p < 1$) from these O groups, probably because interactions of the remainder of the proline molecule with amide or anionic O are unfavorable and not because the NH→O hydrogen bond is less favorable than the OH→O hydrogen bond.

Amide and Cationic N. Interactions of proline with amide N ($K_p = 1.33 \pm 0.09$) and cationic (ammonium, guanidinium) N ($K_p = 1.36 \pm 0.12$) are moderately and equally favorable; the local proline concentration in the vicinity of these groups is predicted to exceed its bulk concentration by about 33%. The interaction of GB with amide N is at least as favorable ($K_p = 1.5 \pm 0.2$) as it is for proline, whereas the interaction of GB with cationic N ($K_p = 1.32 \pm 0.11$) is comparably favorable to that of proline. By contrast, the interaction of urea with amide N ($K_p = 1.10 \pm 0.07$) is only modestly favorable, whereas the interaction of urea with cationic N is modestly unfavorable ($K_p = 0.94 \pm 0.05$).

Comparison of the K_p values for these chemically similar but Coulombically different nitrogens indicates that the origins of these effects are not primarily Coulombic attraction or repulsion but rather are chemical (hydrogen bonding) interactions. Indeed, the interactions of all three solutes with amide and cationic N are well explained in terms of the competition between these solutes and water to interact with these nitrogens by hydrogen bonding. In their interactions with water, these nitrogens primarily function as hydrogen bond donors. To interact favorably with these nitrogens in water, the solute must be a better hydrogen bond acceptor (in quality or quantity of hydrogen bonds) than water.

Favorable interactions of proline and GB with amide and cationic N presumably arise from the formation of strong NH→O hydrogen bonds from these N to the carboxylates of proline and GB. Because the observed effects of all these interactions are the same within uncertainty, clearly the charge state of the N (polar or cationic) is not a major determinant of the interaction strength. Interactions of other functional groups on proline and GB with amide and cationic N, including NH→N hydrogen bonds[17] involving the ammonium group of proline, are expected to be unfavorable. Interactions of urea with amide and cationic N are much less favorable than those of proline and GB with these groups, presumably because the carboxylate groups on proline and GB are better hydrogen bond acceptors than the amide oxygen on urea and because NH→N hydrogen bonds involving the amide N of urea and cationic or amide N of the model compound are unfavorable relative to hydrogen bonds with water.

Hydroxyl O. The strength of interaction of proline and OH groups is very similar to interactions with water ($K_p = 1.02 \pm 0.04$) and not very different from the marginally unfavorable interaction between GB and OH ($K_p = 0.97 \pm 0.06$) and the favorable urea OH interaction ($K_p = 1.08 \pm 0.02$). These K_p values could be near unity because the interactions of the OH functional group with all functional groups on these solutes are very similar to interactions with water or due to compensations between stronger and weaker interactions.

Aliphatic C. Proline, like GB, is only modestly excluded from aliphatic C surfaces (proline $K_p = 0.85 \pm 0.04$; GB $K_p = 0.92 \pm 0.08$); the average local concentrations of these solutes in the vicinity of aliphatic C surfaces are only 8−15% smaller than their bulk concentrations. Anionic carboxylate oxygens on these amino acids, like those of sulfate and phosphate groups, are expected to be almost completely excluded from aliphatic C (i.e., $K_p \approx 0$).[17,29] The cationic group of proline may also be excluded from aliphatic C on the basis of the behavior of the ammonium ion;[14] interactions of alkylated ammonium cations with aliphatic C have not been quantified. Favorable aliphatic C−aliphatic C interactions must largely compensate for these unfavorable interactions.

Unlike proline and GB, urea is slightly accumulated ($K_p = 1.03 \pm 0.02$) at aliphatic C. This favorable but small interaction is an important part of the reason that urea is a protein denaturant, because aliphatic C accounts for approximately two-thirds of the surface that gets exposed in unfolding a globular protein. It seems unlikely that the C surface of urea (4% of total) accounts for this net favorable interaction. Although the interactions of amide N and O with aliphatic C are not expected to be intrinsically favorable, they must be less unfavorable than the interactions of water with aliphatic C.

Aromatic C. Proline interacts quite favorably with aromatic C ($K_p = 1.26 \pm 0.03$). By comparison, the GB−aromatic C interaction is significantly more favorable ($K_p = 1.62 \pm 0.11$), and the urea aromatic C interaction is comparable in strength ($K_p = 1.28 \pm 0.02$). Because the interaction of the carboxylate oxygens of proline and GB with aromatic C are expected to be highly unfavorable,[17] clearly the interaction of the substituted ammonium cationic groups and of aliphatic C with aromatic C must be quite favorable. For GB, this combined interaction is interpreted as a cation−π interaction.[16] Although proline also has a formal positive charge on its ammonium group, it is expected to be a less effective cation−π donor because its ammonium hydrogens form interactions with water that must be broken to form a cation−π interaction. This conclusion is supported by enzyme inhibition studies showing that more highly substituted ammonium groups are better cation−π donors.[31] Possible interpretations of the favorable interaction of urea with aromatic C include hydrogen bonding to and stacking of the urea π system on the aromatic ring. [17]

dx.doi.org/10.1021/bi400683y | *Biochemistry* 2013, 52, 5997−6010

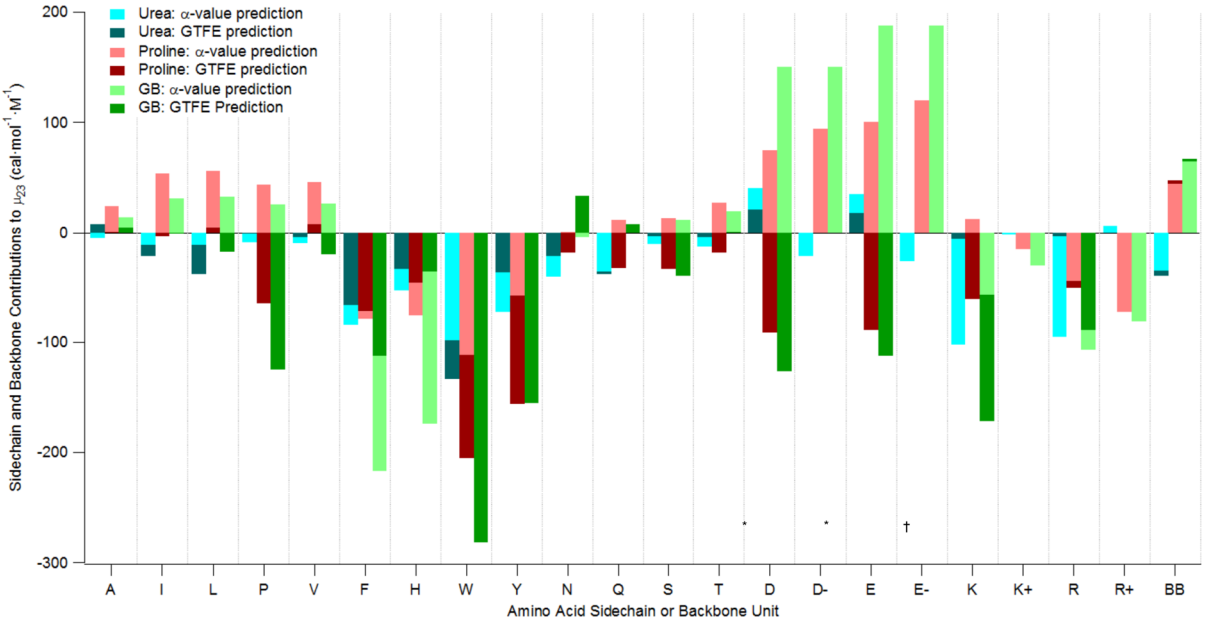


Figure 6. Comparison of α-value and group transfer free energy (GTFE) predictions of interactions of proline, urea, and GB with amino acid side chains and the peptide backbone (BB). α-Value predictions use functional group α-values from Table 1 (converted from molal to molar scale at 1 M) and ASA information from Supporting Information Table S2. GTFE predictions are from ref 29; values for W and Y side chains were not given. α-Value predictions for charged side chains are given both without counterion (specified as D^-, E^-, K^+, R^+) and with counterion (D^*, E^*, $K^\dagger$, $R^\dagger$; where * indicates Na^+ and † indicates Cl^-). For charged side chains, GTFE predictions are only available with the counterion (i.e., D^*, E^*, $K^\dagger$, $R^\dagger$). Values for histidine are for the uncharged side chain. For both α-value and GTFE predictions, the backbone is modeled as half the surface area of cyclic glycylglycine (cGG/2). Values plotted are equivalent to free energies of transfer from water to 1 M solute.

Prediction of Proline Interactions with Amino Acid Side Chains and Peptide Backbone of Proteins; Comparison with Those of Urea and GB and with the GTFE Approach. Interactions of proline, urea, and GB with most amino acid side chains and the peptide backbone of proteins are straightforward to predict from solute–functional group α-values (Table 1) and the amount and composition of their ASA (Supporting Information Table S2). This allows a direct comparison between results of the α-value (μ_{23}-based) analysis used here and the group transfer free energy (GTFE) approach to analyze the effects of urea and osmolytes on protein processes, which interprets amino acid and dipeptide solubility data to obtain free energies (ΔG_{tr}) for transfer of amino acid side chains and the peptide backbone from water to a 1 M solution of the solute in question.[22,32−35] At least for sparingly soluble amino acids, values of ΔG_{tr} for amino acids are closely related to values of μ_{23} for the solute–amino acid interaction on the molar scale (see Supporting Information) and are compared in Supporting Information Table S4.

Figure 6 shows that α-value and GTFE predictions for interactions of proline, urea, and GB with the peptide backbone, modeled as cGG/2, are in quantitative agreement. Interactions of proline and GB with this backbone model are moderately unfavorable, whereas interaction of urea is moderately favorable. Interpreted using the α-value analysis, the unfavorable interactions of proline and GB with the peptide backbone result from their unfavorable interactions with amide O and aliphatic C, which are dominant over their favorable interactions with amide N, whereas the favorable interaction of urea results from its favorable interactions with all backbone groups, especially the amide O.

Figure 6 also shows the very significant differences between α-value and GTFE predictions for interactions of proline, GB, and urea with most amino acid side chains. The α-value analysis predicts that the largest unfavorable interactions of proline (and also GB) are with anionic (D^-, E^-) side chains; interactions of proline (also GB) with both the aliphatic C and carboxylate O of these side chains are unfavorable. α-Value predictions for the electroneutral combination of anionic side chain and Na^+ counterion are not very different because the interactions of proline and GB with Na^+ are small. By contrast, GTFE predictions for interactions of proline and GB with these D and E electroneutral side chains are quite favorable; these predictions are difficult to explain at the molecular level. The unfavorable interaction of proline and GB with anionic side chains predicted by α-value analysis is consistent with the function of these solutes as osmolytes in the concentrated cytoplasmic protein solution, with a high concentration of surface-exposed carboxylates.

For urea, the α-value analysis predicts favorable interactions with anionic (D^-, E^-) side chains because of the favorable interactions of urea with both carboxylate O and aliphatic C but predicts that urea interacts unfavorably with the electroneutral side chain–cation combination because of the very unfavorable interaction of urea with Na^+. GTFE analysis predicts weaker unfavorable interactions of urea with the electroneutral D and E side chains.

The α-value analysis predicts that proline (also GB) interacts moderately unfavorably with aliphatic hydrocarbon side chains (A, I, L, P, V) because the proline–aliphatic C α-value (also GB) is unfavorable. On the other hand, GTFE analysis predicts that interactions of proline and GB with these aliphatic side chains are weakly either unfavorable or favorable. The unfavorable interaction of proline (also GB) with aliphatic side chains predicted by α-value analysis is consistent with the function of these solutes as in vivo osmolytes and as protein stabilizers because the surface of globular proteins is 50% hydrocarbon and the surface exposed in unfolding is 70%

dx.doi.org/10.1021/bi400683y | *Biochemistry* 2013, 52, 5997−6010

Table 2. Comparison of Predicted and Observed Proline *m*-Values for Protein Unfolding and μ_{23} Values for Interaction of Proline with Native BSA Surface

	proline			GB			urea		
protein process	experimental	α-value prediction	GTFE prediction[37]	experimental	α-value prediction	GTFE prediction[37]	experimental	α-value prediction	GTFE prediction[37]
ribonuclease T1 unfolding *m*-value (kcal mol^{-1} M^{-1})	0.8 ± 0.1 [22]	1.1 ± 0.5[a]	−0.03 ± 0.6	0.6 ± 0.1 [22]	0.7 ± 1.1[a]	−0.2 ± 0.7	−1.2 ± 0.1[22]	−2.0 ± 0.2[a]	−1.2 ± 0.6
α-chymo-trypsin unfolding *m*-value (kcal mol^{-1} M^{-1})	6.0 ± 0.4[36]	4.3 ± 1.4[a]	1.8 ± 2.2	5.6 ± 0.3[36]	3.9 ± 3.1[a]	1.4 ± 3.1	−3.0 ± 0.4[36,c]	−4.5 ± 0.5[a]	−4.4 ± 2.0
Native BSA μ_{23} value (kcal mol^{-1} m^{-1})	11.9 ± 0.3	7.7 ± 2.0[b]	−6.6[d]	21.7[17]	10.8 ± 3.3[b]	−12.1[d]	−4.5[17]	−3.6 ± 0.8[b]	−2.3[d]

[a]α-value predictions use an extended model of the unfolded state (see text). [b]α-value predictions use the ASA calculated from PDB 1BM0[18] and neglect the interactions of the solute with the Na^+ ions of the electroneutral BSA component. [c]Calculated as the initial slope of a quadratic fit of ΔG_{obs} vs urea molarity.[36] [d]GTFE predictions use side chain and backbone values from ref 22 and the ASA calculated from PDB 1BM0;[18] these predictions implicitly include counterions of charged amino acid side chains.[37]

hydrocarbon. Interactions of urea with these aliphatic side chains are predicted by α-value analysis to be favorable but much smaller in magnitude than for proline and GB. On the other hand, GTFE analysis predicts a mixture of favorable and unfavorable interactions of urea with these chemically similar aliphatic side chains, mostly larger in magnitude than those predicted by α-value analysis.

Not all proline−side chain interactions are unfavorable. The α-value analysis predicts favorable interactions of proline (also GB) with aromatic (F, H, W, Y) and cationic (K^+, R^+) amino acid side chains because interactions of proline and GB with aromatic C are favorable, and interactions of these solutes with cationic N are sufficiently favorable to outweigh their unfavorable interaction with aliphatic C. Urea is also predicted to interact favorably with aromatic side chains, whereas its unfavorable interaction with cationic N causes it to interact unfavorably with the R^+ side chain and neutrally with K^+. For most of these situations, the GTFE analysis predicts the same sign but a very different magnitude than the α-value analysis, and in some cases predicts the opposite sign, as for the interaction of proline with the electroneutral K side chain.

Predictions of the α-value analysis for the interaction of proline (also GB and urea) with the partially polar side chains of N, Q, S, and T are small in magnitude. These are neutral to unfavorable for proline (also GB) because unfavorable interactions with aliphatic C and amide O outweigh the favorable interactions of these osmolytes with amide N and hydroxyl O. Urea is predicted by the α-value analysis to interact favorably, though weakly, with all the functional groups of all four of these side chains. GTFE analysis also predicts modest interactions of these solutes with these four side chains, differing, in some cases, in sign from the α-value predictions.

Predictions of α-Value Analysis for the Contributions of the Peptide Backbone and Individual Side Chains to Proline *m*-Values for Protein Unfolding; Comparison with α-Value Predictions for Urea and GB, GTFE Predictions, and Experimental Data. Proline unfolding *m*-values have been reported for α-chymotrypsin and RCAM (reduced, carboxymethylated) RNaseT1 and compared with urea and GB *m*-values.[22,36] Addition of proline stabilizes both proteins against unfolding, with *m*-values of 0.8 ± 0.1 kcal mol^{-1} M^{-1} for RNase T1(25 °C) and 4.3 ± 1.4 kcal mol^{-1} M^{-1} for α-chymotrypsin (60 °C). GB is found to be as stabilizing as proline, whereas urea is of course destabilizing. Urea *m*-values are similar in magnitude but opposite in sign to those of proline and GB. Table 2 compares these experimental *m*-values with those predicted from α-values for these solutes (Table 1; 25 °C) and the amount and composition of the ΔASA using an extended model for the unfolded state (see Materials and Methods) and the equation

$$m\text{-value} = \Delta\mu_{23} = RT\sum \alpha_i \Delta \mathrm{ASA}_i \quad (6)$$

where $\Delta \mathrm{ASA}_i$ is the change in ASA of functional group i in unfolding.

For RNase T1, the α-value analysis (eq 6) predicts a stabilizing proline unfolding *m*-value of 1.1 ± 0.5 kcal mol^{-1} M^{-1}, in agreement with the observed proline *m*-value. GTFE analysis for RNase T1 predicts a small destabilizing proline *m*-value of −0.03 ± 0.60 kcal mol^{-1} M^{-1} at 25 °C[22] using an average of extended and more structured models for the unfolded state.[37] (The uncertainties in these RNase T1 predictions are large enough that they may not be significantly different.) Also for RNase T1, Table 2 shows that the α-value prediction for the stabilizing GB *m*-value agrees with the experimental value, whereas the α-value prediction for the destabilizing urea *m*-value agrees only semiquantitatively with the experimental value. The GTFE prediction for urea is in quantitative agreement with the experimental urea *m*-value, but GTFE predicts a small destabilizing GB *m*-value.

For α-chymotrypsin, Table 2 shows that *m*-values predicted at 25 °C from α-value analysis, using the structure of monomeric α-chymotrypsin (1YPH) and an extended chain model of the unfolded state, also agree reasonably well with experimental *m*-values, but because the temperature coefficients of the α-values are unknown, this comparison cannot be made quantitatively. For proline and GB, GTFE predictions at 25 °C differ from both the α-value predictions and experimental *m*-values; for urea, GTFE and α-value predictions agree with one another (though using different models of the unfolded state) and with the experimental *m*-value.

The composition of the protein surface exposed in unfolding α-chymotrypsin is quite representative of the average for globular proteins. Hence, we use α-chymotrypsin unfolding and the α-values of Table 1 to predict the contributions of the peptide backbone and most of the different amino acid side chains to the proline unfolding *m*-value and compare the effects of proline with those of urea and GB at this level. Similar graphs for GTFE predictions for these solutes are shown in Figure 3 of ref 17. Figure 7 illustrates these results on bar graphs, using the ASA-normalized contributions of each side chain and the

dx.doi.org/10.1021/bi400683y | Biochemistry 2013, 52, 5997−6010

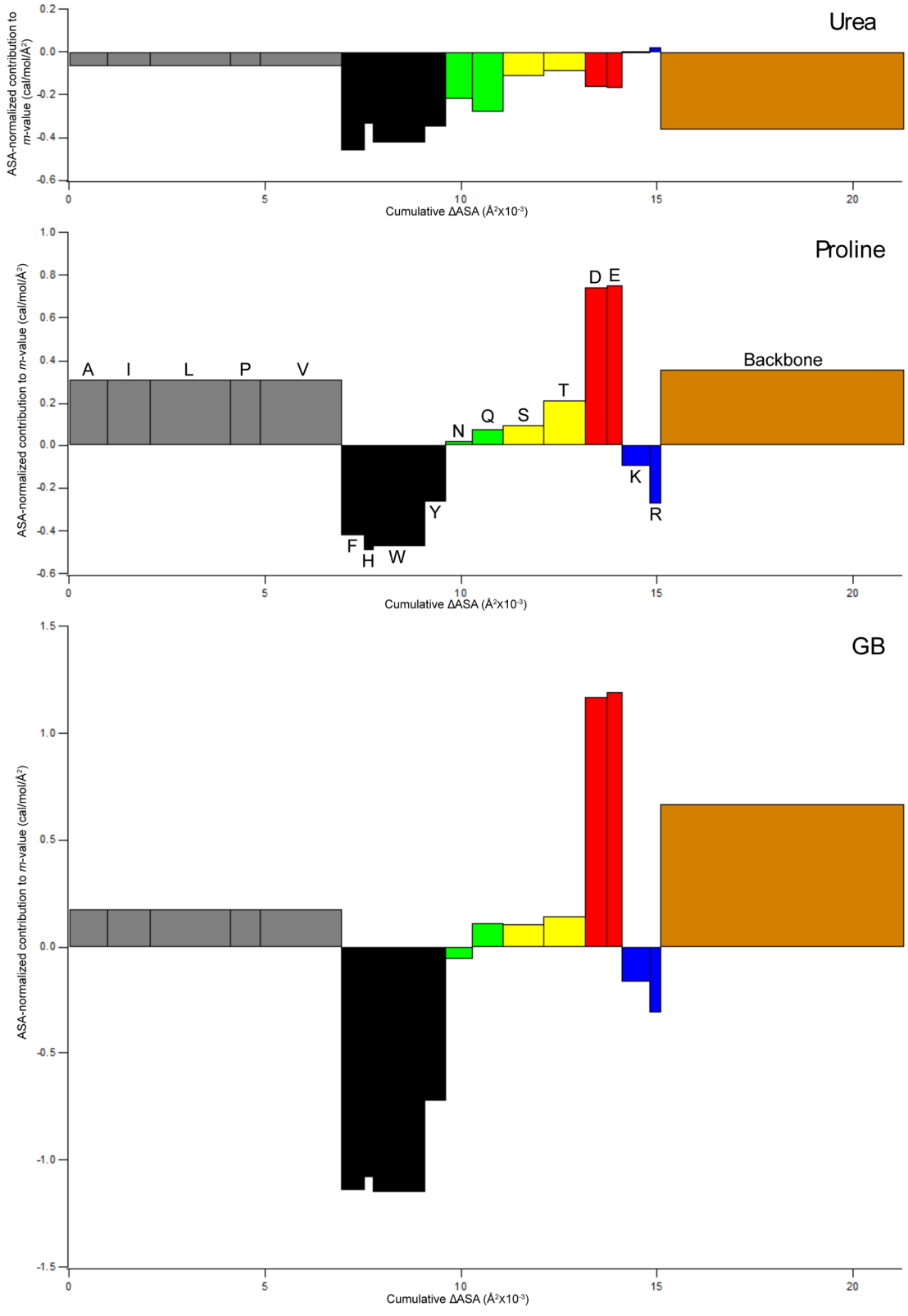


Figure 7. Comparison of α-value predictions of backbone and side chain contributions to proline (panel B), urea (panel A), and GB (panel C) *m*-values for unfolding chymotrypsin (PDB code 1YPH). For each side chain or the peptide backbone, the area of the rectangle is the contribution to the *m*-value and is the same as the contribution to μ_{23} (or $\Delta\mu_{23}$) predicted from α-values and ASA information in Figure 6. The width of the rectangle is the total amount of accessible surface area of that side chain (or backbone) exposed in unfolding chymotrypsin, and the height of the rectangle is the contribution per Å^2 of that side chain or backbone to the *m*-value. The same scale is used for all three solutes. Bar color denotes the type of amino acid (aliphatic amino acids are gray, aromatics are black, amides are green, hydroxyls are yellow, anionic are red, cationic are blue, and the backbone is brown). Cysteine and methionine are not shown.

dx.doi.org/10.1021/bi400683y | *Biochemistry* 2013, 52, 5997−6010

Table 3. Comparison between α-Value and GTFE-Based Analyses of Solute Effects on Protein Processes

	α-value analysis	GTFE analysis[22,37]
experimental method for high-solubility model compounds:	vapor pressure osmometry	solubility
experimental method for low-solubility model compounds:	solubility or micelle formation[17]	solubility
building blocks of proteins used to quantify solute interactions:	functional groups (aliphatic and aromatic C; carboxylate, amide and hydroxyl O)	peptide backbone, amino acid side chains
number of model compounds:	variable (25 for proline, 27 for GB,[39] 46 for urea[17])	21 (20 amino acids + cGG)
number of interaction values (α or GTFE/ASA) obtained from analysis:	7 (7 functional groups)	20 (19 side chains + backbone)
redundancy of model compound set:	yes	no
use of hypothesis of additivity of contributions:	yes, by unified atom or functional group (eq 4)	yes, by side chain or backbone
test of hypothesis of additivity:	yes, at solute−functional group level	no
treatment of counterions of charged amino acids:	separated from side chain, quantified individually	treated implicitly with side chain; not separable
unfolded state model:	all-trans extended (but other models can be used)	average of extended and structured (but other models can be used)
uncertainty values:	based on error propagation from α-value uncertainty	not reported

backbone to the solute *m*-value as the *y* axis and the ΔASA of unfolding as the *x* axis. The same scale is used for the *y* axis for all three solutes to allow easy visual comparison of the predicted effects. Side chains are grouped, with aliphatic and aromatic side chains on the left and polar and charged side chains as well as the polar backbone on the right. Areas of each portion of the bar graph represent the contributions of individual side chains and the peptide backbone to the predicted solute *m*-value for unfolding α-chymotrypsin at 25 °C.

Figure 7 predicts that proline−backbone and net proline−side chain contributions to the unfolding *m*-value are positive (unfavorable) and of similar magnitude. The group of five aliphatic (A, I, L, P, V) side chains and the peptide backbone, whose intrinsic interactions with proline are predicted to be moderately unfavorable (~0.3 cal/(mol Å^2); see Figure 7), are predicted to make the largest positive contributions to the *m*-value because they make up the majority of the ΔASA (~30% each). Intrinsic interactions of proline with anionic (D, E) side chains are more strongly unfavorable (~0.7 cal/(mol Å^2)) but make a smaller contribution to the *m*-value because their contribution to the ΔASA is small. Intrinsic interactions of proline with aromatic (F, H, W, Y) and cationic (K, R) side chains are predicted to be favorable: these are relatively strong (~0.3−0.5 cal/(mol Å^2)) for aromatic side chains and moderately strong (~0.1−0.3 cal/(mol Å^2)) for cationic side chains. Intrinsic interactions of proline with polar (N, Q, S, T) side chains are predicted to be unfavorable with small to moderate magnitudes (~0−0.2 cal/(mol Å^2)). Of these four types of side chains (anionic, aromatic, cationic, polar), only for aromatic side chains is the product of intrinsic interaction strength and contribution to the ΔASA of unfolding (given by the area in Figure 7) sufficiently large to make a major contribution (in this case favorable) to the predicted proline *m*-value.

Comparison of predictions of the α-value analysis for proline with those for GB and urea in Figure 7 shows that, with the exception of the asparagine side chain, contributions to the proline and GB *m*-values for the different side chains and the peptide backbone exhibit the same sign but different magnitudes. Because proline interacts less favorably than GB with uncharged and cationic side chains and more favorably than GB with anionic side chains and the backbone, the predicted *m*-values for these two osmolytes are very similar. For proline, contributions to the predicted *m*-value from unfavorable interactions with the peptide backbone and with side chains are similarly large. For GB, on the other hand, the large favorable interaction with aromatic side chains almost completely counterbalances the smaller unfavorable interactions with aliphatic, polar, and anionic side chains, resulting in a small net contribution of side chain interactions to protein stability; the stabilizing effect of GB is largely the result of its unfavorable interactions with the backbone. The destabilizing effect of urea, as discussed previously, is distributed relatively equally between favorable interactions with the backbone and with all side chains except R, especially with aromatic side chains and the amide-containing side chains of N and Q.

Interactions of Proline with Native Bovine Serum Albumin; Comparison with Interactions of GB and Urea. In the previous section, α-value analysis was used to predict the interactions of proline, urea, and GB with the functional groups exposed in unfolding a globular protein and thereby interpret the unfolding *m*-value. Also important are the interactions of proline and the other solutes with the surface of native proteins and nucleic acids that determine their ability to function efficiently as osmolytes in the concentrated biopolymer milieu of the cytoplasm. If present in the growth medium, proline and GB are imported into the *E. coli* cytoplasm under conditions of osmotic stress. Accumulation of proline allows the cell to maintain a larger fraction of the amount of cytoplasmic water present before the osmotic stress than the mixture of osmolytes (principally glutamate and trehalose) that is synthesized in minimal media.[38] GB is even more effective in this regard, is imported in preference to proline, and allows the cell to retain almost all the original amount of water (and growth rate) over a wide range of osmolality.[2] Urea, which is membrane permeable, is not used as an *E. coli* osmolyte.[1]

Table 2 shows predicted contributions to μ_{23} for the interactions of proline, GB, and urea with the native BSA surface. This surface (55% hydrocarbon, 20% amide, 16% carboxylate, 13% cationic) may be a useful model for the surface composition of *E. coli* cytoplasmic proteins. VPO measurements of the interaction of proline with BSA (Supporting Information Figure S1) yield μ_{23} = 11.9 ± 0.3 kcal mol^{-1} m^{-1} at 25 °C (Table 2). This is a significantly unfavorable preferential interaction, corresponding to the modest exclusion of proline from the vicinity of BSA (the average local proline concentration is ~20−25% less than

dx.doi.org/10.1021/bi400683y | *Biochemistry* 2013, 52, 5997−6010

bulk). From α-values (Table 1) and the surface area and composition of native BSA (Supporting Information Table S3), eq 4 predicts the value at 25 °C (neglecting the small contribution of Na^+ counterions of the electroneutral BSA component) is μ_{23} = 7.7 ± 2.0 kcal mol^{-1} m^{-1} for proline, in semiquantitative agreement with the experimental value. GB is a more effective in vivo osmolyte than proline.[38] Consistent with this, the previously reported preferential interaction of GB with BSA (observed $\mu_{23} \approx$ 21.7 kcal mol^{-1} m^{-1}; predicted μ_{23} = 10.8 ± 3.3 kcal mol^{-1} m^{-1}) [39] is more unfavorable than for proline, though agreement with the μ_{23} value predicted from α-values, ASA and composition information (μ_{23} = 10.8 ± 3.3 kcal mol^{-1} m^{-1}) is not as good as for proline. Urea is much less effective than either proline or GB at boosting osmolality in concentrated biopolymer solutions: urea–BSA preferential interactions are favorable (observed $\mu_{23} \approx -4.5$ kcal mol^{-1} m^{-1}; predicted $\mu_{23} = -3.6 \pm 0.8$ kcal mol^{-1} m^{-1}). Normalized by surface area, the interactions of all three solutes with the native protein surface of BSA are more unfavorable (or, for urea, less favorable) than their interactions with the surface exposed in unfolding globular proteins (see, for example, Table 2). The unfavorable interactions of GB and proline with native protein surfaces and with nucleic acid phosphates[2] make them very good choices as osmolytes in the concentrated biopolymer environment and relatively high surface:volume ratio of the cytoplasm.[2]

Comparisons of α-Value and Group Transfer Free Energy (GTFE) Analyses. Figure 6 shows that the α-value and GTFE analyses yield significantly different predictions for the interactions of all three solutes with many amino acid side chains and thus, in general, give different m-value predictions (Figure 7). Table 3 presents a summary comparison of these two analyses. A first important difference is the input data for these two analyses. The GTFE analysis is based entirely on solubility data for amino acids and dipeptides; as discussed previously[17,25] (also see Supporting Information), these solubility data provide a simply interpretable transfer free energy (i.e., a direct measure of μ_{23}) only for solutes where the approximation in eq 3 is adequate. GTFE analysis uses glycine as a reference to dissect these amino acid transfer free energies and obtain side chain contributions, which are not dissected further. (This GTFE dissection assumes that the contributions of the α-carbon, ammonium nitrogen, and carboxylate oxygens of each amino acid to its transfer free energy are the same as for glycine, even though the various side chains reduce the water-accessible surface area of these groups from their values for glycine (see Supporting Information Table S2).) By contrast, α-value analysis is based on VPO measurements for soluble model compounds displaying functional groups of proteins and solubility data for low-solubility model compounds displaying these functional groups.

Compared to the GTFE analyses, the α-value analyses of interactions of proline, GB, and urea with protein functional groups are based on more extensive model compound data sets and obtain fewer parameters (α-values) quantifying interactions of the solute with the groups on the model compounds. Because the number of model compounds used in the α-value analyses greatly exceed the number of fitted α- and β-values, these analyses provide some redundancy and in each case test and validate (see for example Figure 3) the hypothesis of additivity (eq 4). Additivity at the level of solute–side chain and solute–backbone interactions is assumed but not tested in the GTFE analysis. Interactions of proline, GB, and urea with the sets of model compounds used in the α-value analysis are indeed well-described as the sum of interactions of these solutes with the individual functional groups of the model compounds (eq 4). The α-values for interactions of these three solutes with the functional groups of proteins exhibit reasonable trends and make good chemical sense when interpreted in terms of noncovalent interactions (hydrogen bonding, cation–π, etc.).

At a more detailed level, α-value analysis separates interactions of the solute with charged side chains from interactions with their counterion, which is not possible with the GTFE analysis. This is predicted to have only a minor effect for proline and GB, which interact only weakly with these counterions, but should have a larger effect for urea.

Finally, and most significantly, only α-value analysis can be used to obtain a quantitative interpretation of solute thermodynamic and kinetic m-values in terms of the amount and composition of the biopolymer surface buried in the steps of a protein process like folding or binding. A recent example is the use of urea and GuHCl kinetic m-values together with activation heat capacity data for folding and unfolding rate constants for 13 proteins to quantify the amount of amide and hydrocarbon surface buried in folding to and from the high free energy transition state (TS).[40] Solute m-value data have also been obtained for protein–nucleic acid interactions to probe for coupled folding and formation of new interfaces in the steps of transcription initiation by RNA polymerase[41−43] as well as in the binding of lac repressor to lac operator DNA.[44] For a full interpretation of these data, α-values for interactions of these solutes with nucleic acid surfaces are needed; such data are now available only for urea.[19]

■ ASSOCIATED CONTENT

Supporting Information

Details about methods and materials used (chemicals, osmometry and solubility procedures, etc.); data analysis; tabulated experimental data and ASA values; further comparisons to the GTFE model. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*M.T.R. E-mail: mtrecord@wisc.edu. Tel: (608) 262-5332.

Present Address

¶E.J.G.: California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720, United States.

Author Contributions

⊥R.C.D and E.J.G. contributed equally to this research.

Funding

This research was supported by NIH grant GM47022 (M.T.R.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the reviewers for their helpful comments.

■ ABBREVIATIONS:

GB: glycine betaine (*N*,*N*,*N*-trimethylglycine); GTFE: group transfer free energy; ASA: accessible surface area; VPO: vapor pressure osmometry; SPM: solute partitioning model; TS: transition state

REFERENCES

(1) Cayley, D. S., Guttman, H. J., and Record, M. T. (2000) Biophysical Characterization of Changes in Amounts and Activity of Escherichia coli Cell and Compartment Water and Turgor Pressure in Response to Osmotic Stress. *Biophys. J. 78*, 1748−1764.

(2) Cayley, S., and Record, M. T. (2003) Roles of Cytoplasmic Osmolytes, Water, and Crowding in the Response of Escherichia coli to Osmotic Stress: Biophysical Basis of Osmoprotection by Glycine Betaine. *Biochemistry 42*, 12596−12609.

(3) Hare, P. D., Cress, W. A., and Van Staden, J. (1998) Dissecting the roles of osmolyte accumulation during stress. *Plant, Cell Environ. 21*, 535−553.

(4) Aubert, S., Hennion, F., Bouchereau, A., Gout, E., Bligny, R., and Dorne, A. J. (1999) Subcellular compartmentation of proline in the leaves of the subantarctic Kerguelen cabbage Pringlea antiscorbutica R. Br. In vivo 13C-NMR study. *Plant, Cell Environ. 22*, 255−259.

(5) Sairam, R. K., Rao, K. V., and Srivastava, G. C. (2002) Differential response of wheat genotypes to long term salinity stress in relation to oxidative stress, antioxidant activity and osmolyte concentration. *Plant Science 163*, 1037−1046.

(6) Knipp, G., and Honermeier, B. (2006) Effect of water stress on proline accumulation of genetically modified potatoes (*Solanum tuberosum* L.) generating fructans. *J. Plant Physiol. 163*, 392−397.

(7) Anjum, S. A., Farooq, M., Xie, X.-y., Liu, X.-j., and Ijaz, M. F. (2012) Antioxidant defense system and proline accumulation enables hot pepper to perform better under drought. *Sci. Hort.(Amsterdam, Neth.) 140*, 66−73.

(8) Farooq, M., Aziz, T., Wahid, A., Lee, D.-J., and Siddique, K. H. M. (2009) Chilling tolerance in maize: agronomic and physiological approaches. *Crop Pasture Sci. 60*, 501−516.

(9) Yancey, P. H. (1994) in *Cellular and Molecular Physiology of Cell Volume Regulation* (Strange, K., Ed.), pp 81−109, CRC Press, Boca Raton, FL.

(10) Yancey, P. H., Clark, M. E., Hand, S. C., Bowlus, B. D., and Somero, G. N. (1982) Living with water stress: evolution of osmolyte systems. *Science 217*, 1214−1222.

(11) Lambert, D., and Draper, D. E. (2007) Effets of Osmolytes on RNA Secondary and Tertiary Structure Stabilities and RNA-Mg^{2+} Interactions. *J. Mol. Biol. 370*, 993−1005.

(12) Ignatova, Z., and Gierasch, L. M. (2006) Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc. Natl. Acad. Sci. U. S. A. 103*, 13357−13361.

(13) Record, M. T., and Anderson, C. F. (1995) Interpretation of Preferential Interaction Coefficients of Nonelectrolytes and of Electrolyte Ions in Terms of a Two-Domain Model. *Biophys. J. 68*, 786−794.

(14) Pegram, L. M., and Record, M. T. (2008) Thermodynamic Origin of Hofmeister Ion Effects. *J. Phys. Chem. B 112*, 9428−9436.

(15) Pegram, L. M., and Record, M. T. (2006) Partitioning of atmospherically relevent ions between bulk water and the water/vapor interface. *Proc. Nat. Acad. Sci. U. S. A. 103*, 14278−14281.

(16) Capp, M. W., Pegram, L. M., Saecker, R. M., Kratz, M., Riccardi, D., Wendorff, T., Cannon, J. G., and Record, M. T. (2009) Interactions of the Osmolyte Glycine Betaine with Molecular Surfaces in Water: Thermodynamics, Structural Interpretation, and Prediction of m-values. *Biochemistry 48*, 10372−10379.

(17) Guinn, E. J., Pegram, L. M., Capp, M. W., Pollock, M. N., and Record, M. T. (2011) Quantifying why urea is a protein denaturant whereas glycine betaine is a protein stabilizer. *Proc. Natl. Acad. Sci.U. S. A. 108*, 16932−16937.

(18) Record, M. T., Guinn, E. J., Pegram, L. M., and Capp, M. W. (2013) Introductory Lecture: Interpreting and predicting Hofmeister salt ion and solute effects on biopolymer and model processes using the solute partitioning model. *Faraday Discuss. 160*, 9−44.

(19) Guinn, E. J., Schwinefus, J. J., Cha, H. K., McDevitt, J. L., Merker, W. E., Ritzer, R., Muth, G. W., Engelsgjerd, S. W., Mangold, K. E., Thompson, P. J., Kerins, M. J., and Record, M. T. (2013) Quantifying functional group interactions that determine urea effects on nucleic acid helix formation. *J. Am. Chem. Soc. 135*, 5828−5838.

(20) Courtenay, E. S., Capp, M. W., and Record, M. T. (2001) Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects on protein processes and changes in water-accessible surface area. *Protein Sci. 10*, 2485−2497.

(21) 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.

(22) Auton, M., Rosgen, J., Mikhail, S., Holthauzen, L. M. F., and Bolen, D. W. (2011) Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophys. Chem. 159*, 90−99.

(23) Robinson, R. A., and Stokes, R. H. (1961) Activity coefficients in aqueous solutions of sucrose, mannitol, and their mixtures at 25 degrees C. *J. Phys. Chem. 65*, 1954−1958.

(24) Anderson, C. F., and Record, M. T. (2004) Gibbs-Duhem-based relationships among derivatives expressing the concentration dependences of selected chemical potentials for a multicomponent system. *Biophys. Chem. 112*, 165−175.

(25) Cannon, J. G., Anderson, C. F., and Record, M. T. (2007) Urea-Amide Preferential Interactions in Water: Quantitative Comparison of Model Compound Data with Biopolymer Results Using Water Accessible Surface Areas. *J. Phys. Chem. B 111*, 9675−9685.

(26) Venkatesu, P., Lee, M.-J., and Lin, H.-m. (2007) Thermodynamic Characterization of the Osmolyte Effect on Protein Stability and the Effect of GdnHCl on the Protein Denatured State. *J. Phys. Chem. B 111*, 9045−9056.

(27) Felitsky, D. J., and Record, M. T. (2004) Application of the Local-Bulk Partitioning and Competitive Binding Models to Interpret Preferential Interactions of Glycine Betaine and Urea with Protein Surface. *Biochemistry 43*, 9276−9288.

(28) Pegram, L. M., and Record, M. T. (2007) Hofmeister Salt Effects on Surface Tension Arise from Partitioning of Anions and Cations between Bulk Water and the Air-Water Interface. *J. Phys. Chem. B 111*, 5411−5417.

(29) Pegram, L. M., and Record, M. T. (2008) Thermodynamic Origin of Hofmeister Ion Effects. *J. Phys. Chem. B 112*, 9428−9436.

(30) Felitsky, D. J., Cannon, J. G., Capp, M. W., Hong, J., Wynsberghe, A. W. V., Anderson, C. F., and Record, M. T. (2004) The Exclusion of Glycine Betaine from Anionic Biopolymer Surface: Why Glycine Betaine is an Effective Osmoprotectant but also a Compatible Solute. *Biochemistry 43*, 14732−14743.

(31) Salonen, L. M., Bucher, C., Banner, D. W., Haap, W., Mary, J.-L., Benz, J., Kuster, O., Seiler, P., Shweizer, W. B., and Diederich, F. (2009) Cation-π Interactions at the Active Site of Factor Xa: Dramatic Enhancement upon Stepwise N-Alkylation of Ammonium Ions. *Angew. Chem., Int. Ed. 48*, 811−814.

(32) Auton, M., and Bolen, D. W. (2004) Additive Transfer Free Energies of the Peptide Backbone Unit That are Independent of the Model Compound and the Choice of Concentration Scale. *Biochemistry 43*, 1329−1342.

(33) Auton, M., and Bolen, D. W. (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. U. S. A. 102*, 15065−15068.

(34) Auton, M., Holthauzen, L. M. F., and Bolen, D. W. (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. U. S. A. 104*, 15317−15322.

(35) Nozaki, Y., and Tanford, C. (1963) The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions. *J. Biol. Chem. 238*, 4074−4081.

(36) Attri, P., Venkatesu, P., and Lee, M.-J. (2010) Influence of osmolytes and denaturants on the structure and enzyme activity of alpha-chymotrypsin. *J. Phys. Chem. B 114*, 1471−1478.

(37) Auton, M., and Bolen, D. W. (2007) Application of the transfer model to understand how naturally occuring osmolytes affect protein stability. *Methods Enzymol. 428*, 397−418.

(38) Cayley, S., Lewis, B. A., and Record, M. T. (1992) Origins of the Osmoprotective Properties of Betaine and Proline in *Escherichia coli* K-12. *Biophys. J. 174*, 1586−1595.

(39) Capp, M. W., Pegram, L. M., Saecker, R. M., Kratz, M., Riccardi, D., Wendorff, T., Cannon, J. G., and Record, M. T. (2009) Interactions of the Osmolyte Glycine Betaine with Molecular Surfaces in Water: Thermodynamics, Structural Interpretation, and Prediction of m-values. *Biochemistry 48*, 10372−10379.
(40) Guinn, E. J., Kontur, W. S., Tsodikov, O. V., Shkel, I., and Record, M. T. (2013) Probing the Protein Folding Mechanism Using Denaturant and Temperature Effects on Rate Constants. *Proc. Natl. Acad. Sci. U. S. A.*, in press.
(41) Kontur, W. S., Saecker, R. M., Davis, C. A., Capp, M. W., and Record, M. T. (2006) Solute Probes of Conformational Changes in Open Complex (RPo) Formation by Escherichia coli RNA Polymerase that the lambdaPr Promoter: Evidence for Unmasking of the Active Site in the Isomerization Step and for Large-Scale Coupled Folding in the Subsequent Conversion to RPo. *Biochemistry 45*, 2161−2177.
(42) Kontur, W. S., Capp, M. W., Gries, T. J., Saecker, R. M., and Record, M. T. (2010) Probing DNA binding, DNA opening, and assembly of a downstream clamp/jaw in Escherichia Coli RNA polymerase-lambdaPR promotor complexes using salt and the physiological anion glutamate. *Biochemistry 49*, 4361−4373.
(43) Kontur, W. S., Saecker, R. M., Capp, M. W., and Record, M. T. (2008) Late Steps in the Formation of E. coli RNA Polymerase - LambdaPR Promotor Open Complexes: Characterization of Conformational Changes by Rapid [Perturbant] Upshift Experiments. *J. Mol. Biol. 376*, 1034−1047.
(44) Hong, J., Capp, M. W., Saecker, R. M., and Record, M. T. (2005) Use of Urea and Glycine Betaine to Quantify Coupled Folding and Probe the Burial of DNA Phosphates in Lac Repressor-Lac Operator Binding. *Biochemistry 44*, 16896−16911.