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[a,#], Emily J. Guinn[b,#,&], Michael W. Capp[a], Oleg V. Tsodikov[c], and

M. Thomas Record Jr. [a,b,*]

Departments of Biochemistry[a] and Chemistry[b], University of Wisconsin, Madison WI 53706 and Department of Pharmaceutical Sciences[c], University of Kentucky, Lexington KY 40538

[*] To whom correspondence should be addressed at Department of Biochemistry, 433 Babcock Dr., University of Wisconsin, Madison WI 53706, (608) 262-5332, mtrecord@wisc.edu

[#] These authors contributed equally to this research

[&] Present address: California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720

Supporting Information

SI Text:

Chemicals

Chemicals used in vapor pressure osmometry (VPO) and solubility assays were from Sigma-Aldrich, Fischer, Fluka, or Bachem and were all of the highest available grade (at least 98% purity). Potassium glutamate, GB, sodium aspartate, and potassium oxalate were obtained as monohydrates, while all other chemicals were obtained in anhydrous form. All samples were dissolved in water purified with a Barnstead E-pure system (Thermo Fisher Scientific) to no less than 15 MΩ (DI water).

Vapor Pressure Osmometry to Determine Proline-Model Compound Preferential Interactions

Samples were gravimetrically prepared and osmolalities were measured on a Wescor Vapro 5520 vapor pressure osmometer by a procedure adapted from Capp et al[1]. Osmolalities were determined for at least 6 series of 6 samples in which the model compound molality was held constant and the molality of proline was varied. For each series, the osmolality of the 2-component model compound-water solution was measured in addition to the series of 3-component model compound-proline-water samples. A series of 2 component samples of different proline concentrations was also prepared and measured separately. For all series, bracketing KCl standards were read with each sample and used to correct its osmolality using literature ID data for KCl[2].

Values of the chemical potential derivative $\mu_{23}=(d\mu_2/dm_3)_{m2}$ quantifying proline-model compound preferential interactions were obtained from osmolality data on three and two component solutions using Eq. 1 and 2 [1, 3, 4] in the text. In Eq. 1 for the excess osmolality ΔOsm, Osm(m_2,m_3) is the measured 3-component osmolality. Osm(m_2,0) is determined directly from

measurements on two component model compound-water solutions. $Osm(0,m_3)$ is determined by interpolation of a quadratic fit of osmolality vs. molality of the proline 2-component data. Due to high variability of osmometry readings at high osmolality, only results from solutions reading 2 Osm or less were included in data analysis. Some earlier assays used the technique described in Guinn et al[5] in which proline molality was held constant in series of experiments with variable model compound molality and $Osm(m_2,0)$ and $Osm(0,m_3)$ were both interpolated from fits of two component data. Results from these assays were the same within error as results using the technique described above.

Interactions of proline with native bovine serum albumin (BSA) were also determined by VPO. 2-component proline solutions and 3-component proline-BSA solutions were prepared gravimetrically and osmolalities were measured according to the methods used by Courtenay *et al*[6]. $Osm(0,m_3)$ was interpolated from a fit of two-component data, while $Osm(m_2,0)$ was determined from the osmolality of a two-component BSA solution, assuming the osmolality of a BSA solution is proportional to its molality in the small range of BSA concentrations studied (varied less than 10% in 2- and 3-component samples).

For all compounds measured by VPO, values of ΔOsm (eq. 1) were plotted against m_2m_3, the product of proline and model compound (or BSA) concentrations. These plots, shown for model compounds in Figure 1 and BSA in Figure 4, showed no evidence of systematic deviation from a linear relationship with a zero intercept. Linear least-squares slopes (μ_{23}/RT, Eq. 1) were calculated with intercepts fixed at the origin using the statistics program Igor Pro 5.05A.

Solubility Assays to Determine Naphthalene-Proline Preferential Interactions

Series of proline solutions from 0.25 m to 2.0 m were prepared gravimetrically, with several deionized water tubes prepared as controls. An excess (0.1g) of naphthalene was added to each tube, and after mixing the tubes were sealed with plastic film and placed in a shaking water bath at 25°C for no less than 7 days. Filtered 2 mL samples were collected from each tube and the absorbance at 275 nm was measured on a Cary UV-Vis spectrophotometer. The absorbance of a saturated solution was used as a measure of solubility, and was converted to the molal scale by Eqs. S1 and S2 below:

$$\ln\left(\frac{m_{2,ss}}{m_{2,ss}^0}\right) = \ln\left(\frac{A_{ss}\chi_{ss}}{A_{ss}^0\chi_{ss}^0}\right) \qquad \text{(S1)}$$

$$\chi = \left(\frac{m_3 M_3}{1000} + 1\right) \qquad \text{(S2)}$$

In Eqs. S1 and S2, $m_{2,ss}$ is the molality of naphthalene in a saturated solution where the molality of proline is m_3, M_3 is the molar mass of proline, A_{ss} is the absorbance of the solution, and the superscript "0" refers to the absence of proline.

Surface Area Calculations

Model compound structures were obtained from the online databases BMRB [7] or PubChem (http://pubchem.ncbi.nlm.nih.gov/) and converted to pdb files with the CACTUS SMILES translator (http://cactus.nci.nih.gov/translate/). Water-accessible surface areas were calculated using Surface Racer[8] and the Richards[9] set of radii as described in Guinn et al[5]. Accessible surface areas were divided into: aliphatic and aromatic carbon; anionic carboxylate, anionic phosphate, amide and hydroxyl oxygen; and amide and cationic nitrogen. Accessible surface areas for unfolded α-chymotrypsin (1YPH) and reduced carboxyamidated RNAse T1 (2BU4) were calculated using an extended model of the polypeptides (backbone ϕ=ψ=180°).

Justification for the use of this extended-chain model of the unfolded protein has recently been

obtained from analysis of denaturant m-values and heat capacity changes of protein

unfolding[10]. Extended peptide models were first built in Pymol (DeLano Scientific) and then

side chain rotamers were chosen in Coot [11] among preferred ones to minimize interactions and

avoid steric clashes with other side chains and the backbone. These ΔASA values and their ratio

are given in Table S3. The ASA of native BSA was calculated using Surface Racer[8] and the 1BM0

structure previously used in Capp *et al*[1], and these values are also given in Table S3. ASA data

for model compounds is given in Table S2.

Determination of proline α-values by global fit

α_i and β_{ion} values for interactions of proline with functional groups and inorganic ions

(Table 1) were determined by a global linear fit of the experimentally determined μ_{23}/RT values

from Table S1 and corresponding ASA values from Table S2, using Equation 4 and the statistics

program Igor 5.05A. Uncertainties reported for all α_i or β_{ion} values in Table 1 are fitting errors

reported by this program.

Transfer of amino acids from water to 1M proline, GB, or urea:

In the GTFE analysis of solute effects, a quantity called the free energy of transfer (Δg_{tr})

of an amino acid from water to a 1 M solution of the solute is determined from differences in

solubility of the amino acid in a 1 M solution of the solute ($m_2^{ss, 1M}$) and in water ($m_2^{ss, 0M}$) [12]:

$$\Delta g_{tr} = -RT d\ln m_2^{ss}/dm_3 \approx -RT\Delta\ln m_2^{ss}/\Delta m_3 = -RT\ln(m_2^{ss, 1M}/m_2^{ss, 0M}). \quad \text{Eq S1}$$

From Eq. 3, the Δg_{tr} defined in Eq S1 is related to μ_{23} by

$$\Delta g_{tr} = (\mu_{23}/(1 + \varepsilon_2)) \qquad\qquad \text{Eq S2}$$

To interpret solute m-values for protein unfolding determined for the usual experimental conditions where the concentration of the solute greatly exceeds that of the protein, the quantity of interest is μ_{23}.[13-16] Values of Δg_{tr} are only a good approximation to μ_{23} if the amino acid self-interaction nonideality term $\varepsilon_2 = (d\ln\gamma_2/d\ln m_2)_{m3}$ in Eq S2 negligibly small. For soluble amino acids, the ε_2 term is a significant correction, not available except by osmometry. For cases where ε_2 is significant, osmometry (Eqs. 1-2) is the preferred direct method of obtaining μ_{23}.

Table S4 compares values of Δg_{tr} for transfer of amino acids from water to 1 M solutions of proline, GB and urea obtained entirely from solubility data with those obtained for sufficiently soluble amino acids from osmometric data and with those predicted from α-values (Table 1), which are derived from analysis of combined data sets using osmometric data for soluble solutes and solubility data for relatively insoluble solutes. Values of Δg_{tr} obtained solubility data give the free energy change for the process of transferring an amino acid from a *saturated* solution of this amino acid in the absence of the solute to a *saturated* solution of this amino acid in the presence of 1 M solute). Values of Δg_{tr} obtained from VPO experiments or predicted from α-values ($\Delta g_{tr} = \mu_{23}$) give the free energy change for the process of transferring an amino acid from a *dilute* solution of this amino acid in the absence of the solute to a *dilute* solution of the amino acid in the presence of 1 M solute. α-Value predictions of Δg_{tr} of amino acids from water to 1 M solutions of proline, GB and urea are compared in the bar graphs of Fig. S1.

VPO determinations of Δg_{tr} in Table S4 are part of the training sets used to obtain α-values for these three solutes, and agreement between VPO results and α-value predictions of

Δg_{tr} is similar to that obtained in comparisons with the full training sets in Fig 3 (no systematic

differences; average magnitude of differences ~20% except for cases where $\Delta g_{tr} = \mu_{23} \approx 0$).   In

general, for all three solutes, larger and more systematic differences are observed between

values of Δg_{tr} obtained from solubility data and those obtained from VPO data or predicted

from α-values. For transfer of most amino acids to 1 M proline, 1 M GB and 1 M urea,

solubility-derived values of Δg_{tr} are more positive than those determined by VPO or predicted

from α-values; in about half these cases the differences are significantly larger than the

differences between VPO-derived values of Δg_{tr} and α-value predictions. These systematic

differences are observed for both soluble and relatively insoluble amino acids, and so cannot be

explained only by the ε_2 self-interaction term.
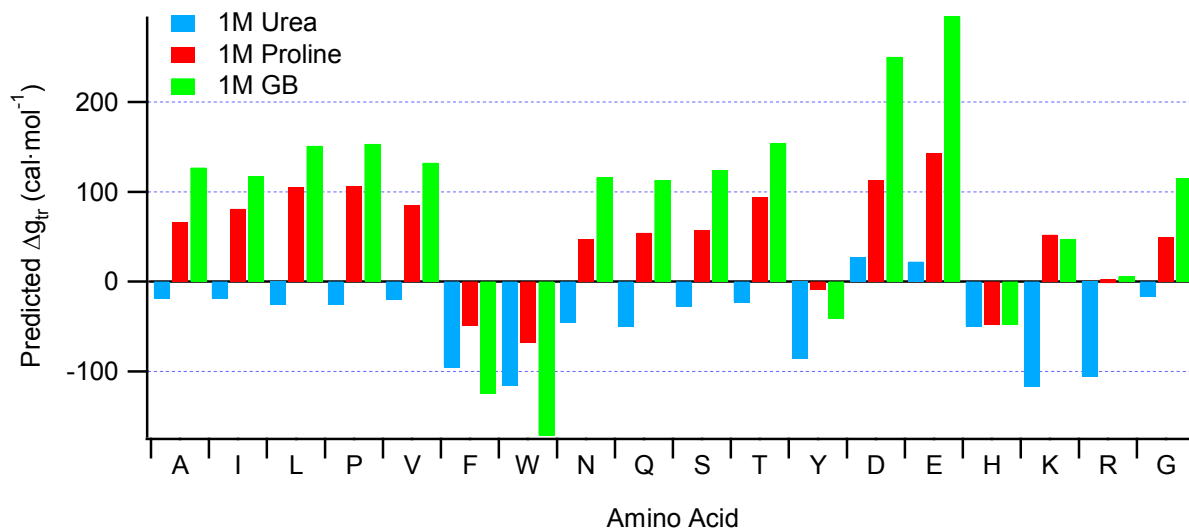
FIGURE:

*Figure S1: Predicted transfer free energies for the transfer of amino acids to 1M proline, GB, and urea
solution, as calculated using α_i and β_{ion} values in Table 1. Calculations for aspartate and glutamate
include a sodium counterion, while calculations for lysine and arginine include a chloride counterion.*

Table S1: Values of μ_{23}/RT for Interactions of Proline with Model Compounds (25 ^{o}C)

Model Compound	Method	Experimental μ_{23}/RT[b] (m^{-1})	Predicted μ_{23}/RT (m^{-1})
aAma	VPO	0.21±0.01	0.22±0.05
Arginine HCl	VPO	-0.028±0.007	0.002±0.097
cAA[a]	solubility	0.16±0.01	0.16±0.05
cGG[a]	solubility	0.13±0.01	0.14±0.05
cLA[a]	solubility	0.22±0.01	0.22±0.05
Glycerol	VPO	0.070±0.006	0.052±0.023
Glycine	VPO	0.094±0.007	0.076±0.050
Glycine Betaine	VPO	0.22±0.01	0.24±0.04
GlyGly	VPO	0.087±0.004	0.13±0.05
KAcetate	VPO	0.19±0.02	0.20±0.06
KCl	VPO	0.042±0.007	0.028±0.058
KGlutamate	VPO	0.29±0.01	0.24±0.09
K2HPO4	VPO	0.22±0.01	0.24±0.11
K2Oxalate	VPO	0.27±0.01	0.29±0.11
Lysine HCl	VPO	0.097±0.012	0.078±0.079
Mannitol	VPO	0.039±0.009	0.050±0.035
NaAspartate	VPO	0.19±0.01	0.17±0.08
NaBenzoate	VPO	-0.011±0.014	-0.031±0.057
NaCl	VPO	0.003±0.011	0.007±0.057
NaOxamate	VPO	0.068±0.011	0.12±0.06
NaPropionate	VPO	0.17±0.01	0.19±0.06
Naphthalene	solubility	-0.27±0.01	-0.25±0.03
Na2HPO4	VPO	0.21±0.01	0.20±0.11
Urea	VPO	-0.062±0.006	-0.077±0.047
6-ACA (zwitterionic)	VPO	0.17±0.01	0.16±0.06

a- Calculated from solubility data from ref 16. Solubility data for cAG and cVV from ref 16 were not included in the analysis because of large deviations between experimental and predicted μ_{23}/RT values.

b-Experimental error is assigned as 5% or the fitting error given by IGOR, whichever is larger.

Table S2: Accessible Surface Area Contributions by Surface Type for Model Compounds

Compound	Source	Contribution to ASA (Å^2)							
		Aliphatic C	Aromatic C	Hydroxyl O	Amide O	Amide N	Carboxylate O	Phosphate O	Cationic N
aAma	PubChem	261	0	0	70	19	0	0	0
Alanine	BMRB	91	0	0	0	0	86	0	77
Arginine HCl	BMRB	104	0	0	0	0	83	0	184
Asparagine	BMRB	48	0	0	37	60	85	0	41
cAA [17]	Cactus	181	0	0	83	48	0	0	0
cGG [17]	PubChem	115	0	0	93	48	0	0	0
cLA [17]	Cactus	263	0	0	83	33	0	0	0
Glutamine	BMRB	75	0	0	42	61	80	0	61
Glycerol	BMRB	116	0	140	0	0	0	0	0
Glycine	BMRB	56	0	0	0	0	86	0	77
Glycine Betaine	BMRB	197	0	0	0	0	81	0	0
GlyGly	PubChem	99	0	0	31	12	86	0	77
Histidine	BMRB	47	66	0	0	61	76	0	69
Isoleucine	BMRB	174	0	0	0	0	71	0	66
KAcetate	BMRB	99	0	0	0	0	96	0	0
KGlutamate	BMRB	82	0	0	0	0	172	0	60
K2HPO4	Cactus	0	0	56	0	0	0	150	0
K2Oxalate	BMRB	21	0	0	0	0	185	0	0
Leucine	BMRB	178	0	0	0	0	83	0	54
Lysine HCl	BMRB	135	0	0	0	0	78	0	130
Mannitol	BMRB	126	0	241	0	0	0	0	0
NaAspartate	BMRB	65	0	0	0	0	150	0	61
NaBenzoate	BMRB	10	164	0	0	0	91	0	0
NaOxamate	Cactus	18	0	0	44	61	92	0	0
NaPropionate	BMRB	132	0	0	0	0	91	0	0
Naphthalene	PubChem	0	275	0	0	0	0	0	0
Na2HPO4	Cactus	0	0	56	0	0	0	150	0
Phenylalanine	BMRB	43	148	0	0	0	77	0	71
Proline	BMRB	151	0	0	0	0	80	0	38
Serine	BMRB	65	0	48	0	0	83	0	65
Threonine	BMRB	109	0	44	0	0	84	0	38
Tryptophan	BMRB	46	160	0	0	35	80	0	58
Tyrosine	BMRB	51	109	57	0	0	81	0	55
Urea	PubChem	8	0	0	49	129	0	0	0
Valine	BMRB	148	0	0	0	0	81	0	63
6-ACA (zwitterionic)	Cactus	180	0	0	0	0	101	0	82

Table S3: ΔASA Information for Protein Unfolding and ASA Information for Native BSA (Table 2)

	Contribution to ASA or ΔASA (Å^2)						
Protein Process	**Aliphatic C**	**Aromatic C**	**Hydroxyl O**	**Amide O**	**Amide N**	**Carboxyl O**	**Cationic N**
Ribonuclease T1 Unfolding (11,14-16)[a]	3313	1215	447	1555	1068	334	243
α-chymotrypsin Unfolding[18a]	12723	1843	1015	3075	1666	697	398
Native BSA (1BM0)[1]	14667	417	642	2888	1014	4540	3747

a- Assumes an extended denatured state

Table S4: Transfer Free Energies for Amino Acids from Water to 1 M[d] proline, GB, or urea solution

		$\Delta G_{transfer}$ (cal·mol^{-1})								
		Transfer to 1M Proline			Transfer to 1M Glycine Betaine			Transfer to 1M Urea		
Amino Acid	Solubility in water (m)[19]	Solubility Data[20,21]	VPO Data[a]	α-value prediction[b]	Solubility Data[21]	VPO Data[a,1]	α-value prediction[b]	Solubility Data[19,21]	VPO Data[a,5]	α-value prediction[b]
Pro	15.8	58		106	53	141	153	10	-37	-25
NaAsp	5.02	31	121	113	62	288	250	31		28
Ser	4.08	89		57	137		124	7		-25
ArgHCl	4.07	62	-18	1	69	-15	6	7	-93	-109
LysHCl	3.95	62	63	52	7	96	47	5	-106	-115
NaGlu	3.69	33	*184*	143	66	*312*	295	28	*35*	22
Gly	3.34	122	60	49	178	90	115	10^c	-13	-16
Ala	1.86	122		66	183		127	18^c	-20	-19
Thr	0.817	104		94	179		154	6		-30
Val	0.489	130		85	159		132	6		-22
Gln	0.287	90		54	186		116	-27		-53
His	0.277	77		-48	143		-48	-23		-44
Ile	0.255	119		81	177		117	-11		-21
Asn	0.201	104		47	212		113	-11		-50
Phe	0.170	51		-49	66		-124	-55		-95
Leu	0.164	127		105	161		151	-27		-26
Trp	0.067	-83		-68	N/A		-171	-123		-112
Tyr	0.003[22]	-34		-9	N/A		-41	-26		-86

a- VPO determinations of transfer free energies (μ_{23}) are from Table S1 for proline, from Guinn *et. al.* 2011[5] for urea, and from Capp *et. al.* 2009[1] for glycine betaine. Values in italics are from KGlu data. To convert values of μ_{23}/RT from these sources from m^{-1} to M^{-1} units, the molality of a 1M solution of urea (1.046 m), proline (1.093 m), or GB (1.109 m) was used as a conversion factor.

b- α-value predictions are from Table S1, converted to a molar scale as in footnote a. Predictions for for NaAsp, ArgHCl, LysHCl, and NaGlu include the inorganic ion.

c- Corrected for amino acid self-interactions to obtain μ_{23} (see Eq S2).

N/A- not available

REFERENCES:

1. Capp, M. W., Pegram, L. M., Saecker, R. M., Kratz, M., Riccardi, D., Wendorff, T., and Record, J. G. C. a. 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.

2. Archer, D. G. (1999) Thermodynamic Properties of the KCl + H2O System, *J.Phys.Chem.Ref.Data 28*, 1-17.

3. 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.

4. 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.

5. 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. 108*, 16932-16937.

6. Courtenay, E. S., Capp, M. W., Anderson, C. F., and Record, M. T. (2000) Vapor Pressure Osmometry Studies of Osmolyte-Protein Interactions: Implications for the Action of Osmoprotectants in Vivo and for the Interpretation of "Osmotic Stress" Experiments in Vitro, *Biochemistry 39*, 4455-4471.

7. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucl. Acid Res. 28*, 235-242.

8. Tsodikov, O. V., Record, M. T., and Sergeev, Y. V. (2002) Novel Computer Program for Fast Exact Calculation of Accessible and Molecular Surface Areas and Average Surface Curvature, *J.Comp.Chem. 23*, 600-609.

9. Lee, B., and Richards, F. M. (1971) The Interpretation of Protein Structures: Estimation of Static Accessibility, *Journal of Molecular Biology 55*, 379-400.

10. Guinn, E. J., Kontur, W. S., Tsodikov, O. V., Shkel, I., and Record, M. T. (2013) Probing the Protein Folding Mechanism Using Solutes and Heat Capacity Changes, *Proc. Nat. Acad. Sci. USA In press*.

11. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D- Biological Crystallography 60*, 2126-2132.

12. 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, *Biophysical Chemistry 159*, 90-99.

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

14. Courtenay, E. S., Capp, M. W., Saecker, R. M., and Record, M. T. (2000) Thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding: Interpretation of urea and guanidinium chloride m-values and their correlation with changes

in accessible surface area (ASA) using preferential interaction coefficients and the local-bulk domain model, *Proteins: Struct., Funct., and Genetics 41*, 72-85.

15.   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. Nat. Acad. Sci. U.S.A. 108*, 16932.

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.   Venkatesu, P., Lee, M.-J., and Lin, H.-m. (2008) Effect of osmolyte or GdnHCl on volumetric properties of aqueous solutions containing cyclic dipeptides, *Biochem. Eng. Journal 38*, 326-340.

18.   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.

19.   Auton, M., and Bolen, D. W. (2005) Prediction the energetics of osmolyte-induced protein folding/unfolding, *Proc. Natl. Acad. Sci. 102*, 15065-15068.

20.   Qu, Y., Bolen, C. L., and Bolen, D. W. (1998) Osmolyte-driven contraction of a random coil protein, *Proc. Natl. Acad. Sci. 95*, 9268-9273.

21.   Auton, M., Bolen, D. W., and Rosgen, J. (2008) Structural thermodynamics of protein preferential solvation: Osmolyte solvation of proteins, aminoacids, and peptides, *Proteins 73*, 802-813.

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