PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Entropy Reduction in Unfolded Peptides (and Proteins) due to Conformational Preferences of Amino Acid Residues

Reinhard Schweitzer-Stenner and Siobhan E. Toal*

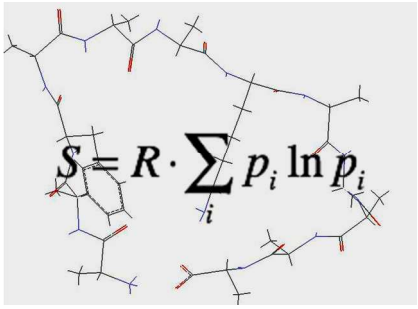
Department of Chemistry, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA

To be submitted to: PhysicalChemistryChemicalPhysics

*Corresponding author, Phone: 215-895-2268, Fax: 215-895-1265, email: rschweitzer-stenner@drexel.edu

Table of Contents Entry

The conformational entropy of amino acid residues in unfolded peptides and proteins is estimated from conformational backbone distributions and compared with experimental and computational values.



Abstract

As established by several groups over the last 20 years, amino acid residues in unfolded peptides and proteins do not exhibit the unspecific random distribution as assumed by the classical random coil model. Individual amino acid residues in small peptides were found to exhibit different conformational preferences. Here, we utilize recently obtained conformational distributions of guest amino acid residues in GxG peptides to estimate their conformational entropy, which we find to be significantly lower than the entropy of an assumed random coil like distribution. Only at high temperature do backbone entropies approach random coil like values. We utilized the obtained backbone entropies of the investigated amino acid residues to estimate the loss of conformational entropy caused by a coil→helix transition and identified two subsets of amino acid residues for which the thus calculated entropy losses correlate well with the respective Gibbs energy of helix formation obtained for alanine based host-guest systems. Calculated and experimentally derived entropic losses were found to be in good agreement. For most of the amino acid residues investigated entropic losses derived from our GxG distributions correlate very well with corresponding values recently obtained from MD simulations biased by conformational propensities derived from truncated coil libraries. Both, conformational entropy and the entropy of solvation exhibit a strong, residue specific temperature dependence, which can be expected to substantially affect the stability of unfolded states. Altogether, our results provide strong evidence for the notion that conformational preferences of amino acid residues matter with regard to the thermodynamics of peptide and protein folding.

Introduction

The general view of the unfolded state of peptides and proteins that is still dominant in textbooks invokes the random coil concept from polymer physics. It considers water as a good solvent for the polypeptide chains, thus maximizing backbone accessibility to water. With respect to the backbone coordinates φ and ψ individual residues are assumed to sample the entire sterically allowed space as visualized in the Ramachandran plot (Figure 1).[1] With the exception of glycine and proline, conformational manifolds of individual residues are considered to be very similar and independent of their respective neighbors in a polypeptide chain (isolated pair hypothesis, IPH).[2] Thus, the unfolded state is described as a highly disordered and entropic entity, the basic properties of which do not depend significantly on the amino acid composition. Prediction of the random coil model with regard to parameters like the radius of gyration and the hydrodynamic radius of unfolded proteins were verified by numerous experiments,[3] which cemented the trust of the scientific community in this description of the unfolded state.

Over the last 10-15 years, however, multiple evidence has emerged suggesting that on the local level (i.e. a few amino acid residues) the random coil model offers a too simplistic view of the unfolded state of peptides and proteins alike. The analysis of restricted coil libraries, for instance, show different structural preferences of amino acid residues in segments of proteins that are not incorporated in regular secondary structures such α-helices, β-sheets and turns.4-6 This is particularly true for alanine, which was found to exhibit a preference for conformations generally found in polyproline II helices (pPII).[7-14] This conformation is located in the upper left quadrant of Ramachandran space with canonical backbone dihedral angles φ=-75° and ψ=150°.[15] Aromatic and aliphatic side chains such as phenylalanine and valine were found to exhibit a modest preference for more extended β-strand-like conformations.[16] The analysis of coil library

distributions further revealed the conformational propensities depend on the nature of the nearest neighbors, which is at variance with IPH.[6, 17, 18] In order to analyze conformational preferences of amino acid residues in solution, a variety of spectroscopic techniques (NMR, IR, Raman, Raman optical activity, vibrational circular dichroism (VCD)) have been applied to short peptides in (mostly aqueous) solution.[8, 10, 19, 20] In principle, these studies confirm the existence of conformational preference, thus corroborating the analysis of coil libraries. However, differences between coil library and solution distributions of amino acid residues are noteworthy.[16]

Protein folding is a transition from a disordered to an ordered state. It is generally thought to involve a massive reduction of conformational entropy, which is compensated by a gain in enthalpy, and for hydrophobic residues, by a gain in solution entropy.[21] Any reliable assessment of the thermodynamics of folding has to rely on an estimation of how much conformational entropy is eliminated by the folding process. It is obvious that the existence of conformational preferences of amino acid residues should substantially reduce the conformational entropy of unfolded peptides and proteins compared with a local random coil scenario, hence increasing the stability of the native state relative to the unfolded state. Local order caused by conformational preferences would also reduce the entropy of intrinsically disordered proteins (IDPs), which should be advantageous for promoting disorder→order transitions caused by IDP binding to biomolecules.[22, 23] However, a quantitative picture connecting residue level order/disorder and conformational entropy has still to emerge and quantitative estimations about the suspected entropy reduction are rare in the literature. To our knowledge, only Baxa et al. undertook an attempt to calculate the loss of conformational entropy based on realistic conformational propensities.[24] These authors selected ubiquitin, a relatively small protein dominated by β-sheet folds, as a suitable model system. They calculated the conformational entropy of the unfolded

protein by biasing MD simulations with information from coil libraries. Thus, the authors arrived at TΔS values between 3.0 and 4.5 kJ/mol·residue, depending on the force field used for the MD simulations. This number was found to be smaller than earlier estimates based on a more random coil-like modeling of the unfolded state (i.e. neglecting conformational preferences).

This paper is aimed at elucidating the reduction of conformational entropy caused by experimentally observed conformational propensities of amino acids in aqueous solution in comparison with an assumed random coil distribution. To this end we utilize conformational distributions of amino acid residues in GxG peptides that have recently been reported by our group.[16, 25-27] The use of GxG host-guest model systems allows one to probe the guest "x" residue in a minimal nearest-neighbor environment, hence allowing one to extract information on intrinsic propensities. Moreover, we use recently reported thermodynamic data for GxG peptides[28] to estimate the temperature dependence of the conformational entropy of the investigated x-residues. We briefly explore the influence of nearest neighbors on the conformational distributions of a limited set of central residues in homo-tripeptides like AAA, VVV, KKK, and DDD.[12, 29] Finally, we estimate the conformational entropy losses caused by a coil→helix transition and compared the result to other computational and experimental values. Our results reveal that the conformational entropies of completely unfolded proteins (i.e. no non-local interactions besides the sporadic formation of turns are present) are substantially (i.e. 1-2 R·T/residue at room temperature) less than what one would expect for a polypeptide with local random coil behavior.

Methods

In a series of studies on unblocked tripeptides we recently obtained normalized conformational distributions $p_x(\varphi,\psi)$ of amino acid residues in unblocked GxG peptides dissolved in water, where x denotes the investigated set of amino acid residues. This was achieved by a consistent analysis of amide I' band profiles (IR, Raman and VCD) and NMR J-coupling constants.[16, 25-27] The conformational distributions were described as superposition of two-dimensional Gaussian functions located at different positions of the Ramachandran plot as follows:[12]

$$p_x(\varphi,\psi) = \sum_{i=1}^{N} \frac{\chi_{x,j}}{2\pi\sqrt{\left|\hat{V}_{x,j}\right|}} e^{-0.5\left(\vec{\rho}_x-\vec{\rho}^{\,0}_{x,j}\right)^T \hat{V}_{x,j}\left(\vec{\rho}_x-\vec{\rho}^{\,0}_{x,j}\right)} \tag{1}$$

with

$$\vec{\rho} = \begin{pmatrix} \varphi_x \\ \psi_x \end{pmatrix} \tag{2}$$

as the position vector for the *j-th* Gaussian sub-distribution in the Ramachandran space and

$$\hat{V}_{x,j} = \begin{pmatrix} \sigma^2_{\varphi x,j} & \sigma^2_{\varphi\psi\, x,j} \\ \sigma^2_{\varphi\psi\, x,j} & \sigma^2_{\psi\, x,j} \end{pmatrix} \tag{3}$$

where $\sigma_{\varphi x,j}$ and $\sigma_{\psi x,j}$ are the halfwidths of these distributions along φ and ψ, respectively. The off-diagonal elements of reflect correlations between the two coordinates. χ_j is the mole fraction of the *jth*-sub-distribution. Each of these sub-distributions is associated with known secondary structure such as right-handed helical, β-strand, pPII and various conformations that appear in canonical turn structures. The respective values for , χ_j, , which define the

conformational ensemble for each residue are all tabulated in recent papers and can be easily obtained.[16, 25-27]

The distributions described by the above formalism can be illustrated by classical Ramachandran plots. Figure 1 exhibits such plots for GAG, GVG, GLG and protonated GDG, which represent four different types of amino acid residues with regard to their intrinsic propensities. Alanine is peculiar in that it exhibits a significantly above the average pPII propensity (i.e. 0.72).[16] Sub-distributions with pPII like conformations cluster in the Ramachandran plot around φ=-70^o and ψ=150^o. Leucine shows a more balanced pPII/β-strand ratio with pPII still being more populated, an observation recently reproduced by DFT calculations for a leucine dipeptide in low dielectric environment. Valine has a slight preference for β-strand because it is heavily entropically favored over pPII. [16, 28] Protonated aspartic acid has an above average propensity for various turn-like conformations stabilized by intrapeptide hydrogen bonding between side chain and amide protons.[26]

In order to calculate the conformational entropy of the amino acid residue x in GxG for distributions such as those shown in Figure 1, we utilize the well know Boltzmann definition of entropy:

$$S_x = R \cdot \int_{-\pi}^{\pi} p_x(\varphi,\psi) \cdot \ln p_x(\varphi,\psi)\, d\varphi\, d\psi \qquad (4)$$

where R is the gas constant.

It should be noted in this context that the thus defined Boltzmann entropy can be different from the so-called Gibbs entropy which is calculated by integrating over all coordinates on which the total probability function depends. In eq. (4), integrations is carried out only over the backbone coordinates. Boltzmann and Gibbs entropy are identical only if all the coordinates involved are independent, so that the probability function can be written as a product function.30

This is certainly not the case for a highly coupled peptide-water system. The Boltzmann entropy defined by eq. (4) contains contributions from other coordinates (solvation, side chain, etc) but does not account for the total entropy of the system.

The result of a numerical integration of eq. (1) depends on the size of the micro-states into which the Ramachandran space is sub-divided.[24] Here, we use rather small sized segments with $\Delta\varphi=\Delta\psi=2.0^{o}$. We found that this lies well in the regime in which S_x depends linearly on $\Delta\varphi$ and $\Delta\psi$. In this paper we use an ideal random coil distribution as references. Figure 2 shows an artificial Ramachandran plot, which mimics a classical local random coil situation. It was produced by a superposition of Gaussian distributions, whose statistical weights, positions and widths are listed in Table 1. Contrary to the experimentally based plots in Figure 2, this random coil distribution assumes substantial sampling of right handed helical and similarly structured turn-like conformations. As in the classical plots of Ramachandran et al.[1] there is no clear distinction between pPII and β-strand, since they display a rather homogeneous sampling of a broad region in the upper left quadrant of the Ramachandran plot. The relative conformational entropy of the considered GxG peptides is thus written as:

$$\Delta S_x = S_{rc} - S_x \qquad [5]$$

where S_{rc} is the entropy of the random coil distribution in Figure 2.

Results and Discussion

We used eqs. (1-5) and the distribution parameters in refs.[16, 25-27] to calculate ΔS_x for all thus far investigated GxG peptides and the corresponding Helmholtz energy contributions $T\Delta S_x$ at 298 K. The latter is displayed in Figure 3. The significance of the obtained values is judged by the extent by which they exceed the thermal energy RT indicated by the upper horizontal line. For aspartic acid (D), we calculated $T\Delta S_x$ for the fully protonated and the fully ionized molecule. Apparently, all $T\Delta S_x$ values exceed the thermal energy, though to a different extent. The average deviation from random coil is -3.84kJ/mol. Not surprisingly, the largest deviation from random coil entropy is obtained for alanine with ΔTS_x=-4.72kJ/mol indicating a quite restricted conformational sampling at room temperature. The ΔTS_x varies significantly between residues, with differences reaching up to approximately 2kJ/mol. Alanine as well as aspartatic acid, asparagine, serine, threonine and methionine exhibit $T\Delta S_x$ values above 4kJ/mol, i.e. they lie in the 2*RT regime (marked by the lower dashed horizontal line in Figure 3).

We wondered how the $T\Delta S_x$ depends on temperature. Toal et al. recently showed that an increase in temperature affects nearly exclusively the equilibrium between pPII and β of the central residue of GxG, whereas the population of turns is practically temperature independent.[28] Utilizing the temperature dependence of conformationally sensitive $^3J(H^NH^\alpha)$ coupling constants derived from NMR measurements, these authors were able to extract thermodynamic parameters governing the pPII⇔β transition. From these thermodynamic parameters we were able to calculate the distribution functions $p_x(\phi,\psi)$ as a function of temperature. Subsequently, we we used $p_x(\phi,\psi)$ to calculate ΔS_x for each peptide at various temperatures. In order to consider the increased sampling of conformational space within the individual (pseudo)potential wells for each conformation with rising temperature, we multiplied the Gaussian bandwidths of the

respective sub-distribution with a factor $(T/T_R)^{1/2}$. We used the thus obtained distribution functions in eq. (4) to calculate $T\Delta S_x$ for five different temperatures between 290 and 360 K. As shown in Figure 4, the entropy difference between the conformational distributions of x in GxG and the assumed random coil distribution at room temperature decreases with increasing temperatures reflecting the increased available conformational space for each residue. As shown above, all ΔTS_x values are negative at room temperature, which indicates that the conformational entropies of all amino acid residues are less then that of a random coil, as expected. While the residue specific ΔS_x contributions to the Gibbs free energy are still within the range of thermal energy at room temperature, one should keep in mind that the total reduction of the conformation entropies of IDPs and unfolded proteins would be quite substantial. The temperature at which a given amino acid residue's conformational entropy contribution to the Helmholtz or Gibbs energy reaches the random coil value at room temperature is different for each species. It is highest for isoleucine, for which random coil entropy is not reached within the considered temperature regime. This reflects the rather high β-strand propensity of this amino acid residue at high temperatures.[28] For a majority of the amino acids residues, however, $T\Delta S_x$ becomes even positive at T>340K, indicating a higher degree of disorder. These results suggest that with respect to their conformational entropy unfolded peptides and proteins might indeed become also locally random coil like at high temperatures, even though distributions are still different from those of canonical random coil ensembles owing to the very low occupancy of right handed helical like conformations.

Recently, we determined the enthalpy (energy) and entropy differences between pPII and β-strand of the herein investigated GxG peptides based on NMR and circular dichroism data. The question arises whether there is any relationship between conformational ΔS_x and the earlier

obtained $\Delta S_{pPII-\beta} = S_{pPII} - S_{\beta}$. Interestingly, we found that for a majority of the residues investigated ΔS_x and $\Delta S_{pPII-\beta}$ are somewhat correlated (Figure 5). We obtained a correlation coefficient of 0.62 when omitting isoleucine and valine, which we previously found to have abnormally high ΔS_{pPII-B} values in favor of β-strand conformations.[28] As the experimentally derived entropy difference between pPII and β-strand conformations decreases along the series of amino acid residues, the relative conformational entropy ΔS_x increases, indicating larger deviations from a random coil conformation.

Taken together, the above analysis clearly reveals that the intrinsic conformational entropy is residue specific and substantially lower then suggested by the local random coil model. This indicates that particularly for IDP-type peptides (A_{β}, salmon calcitonin) and relatively small foldable proteins the total conformational entropy might depend on the amino acid composition. Generally, the conformational ensembles of unfolded peptides and proteins can be assumed to exhibit significantly less microstates than expected based on a random coil model. With regard to IDPs, the composition of sequences, which are involved in e.g. protein-protein and protein-DNA interactions, might matter. Conformational entropies of IDPs and unfolded proteins of similar size should differ, because the former are less heterogeneous than the latter with regard to their amino acid composition.[23, 31, 32]

Nearest neighbor interactions. Thus far, our group has conducted and reported only a limited number of studies aimed at exploring nearest neighbor interactions. A more systematic study is currently underway in our laboratory. However, we have determined the conformational distributions of the central residues of some unblocked X_3 homopeptides, namely A_3, V_3, protonated and ionized D_3 and protonated K_3. Our structural analysis has revealed that alanine

neighbors slightly increase the pPII propensity of A, that valine neighbors substantially increase the β-strand propensity of V,[12] that D neighbors reduce the population of asx-turn like structures of D in the protonated state and increase the pPII propensity of D in the ionized state.[33] For K, lysine neighbors creates a rather broad distribution which peaks in between pPII and β-strand in the Ramachandran plot.[29] With respect to conformation entropy, alanine and aspartic acid neighbors of A and D further increase $T\Delta S_x$ to values above 5 kJ/mol. For valine (in VVV) and lysine (KKK), however, the difference to the assumed random coil distribution is substantially reduced. These few examples show that nearest neighbor interactions with like neighbors can do both, increase or decrease the conformational entropies of amino acid residues.

Secondary structure propensities and helix-coil transitions. The results presented thus far clearly indicate that the ideal unfolded state of peptides and proteins carries much less conformational entropy than suggested by the random coil model. The question now arises to what extent our calculated entropy values are corroborated by experimental data. The best systems for such a comparison are peptides that are capable of adopting secondary structures without assistance by tertiary structure scaffold. Thus, the thermodynamic balance does not contain any contributions from non-local interactions besides i→i+4 hydrogen bonding. This condition is met by alanine based polypeptides in water. Chakrabartty et al. determined the Zimm-Bragg parameters for an alanine based host-guest system, where they inserted different amino acid residues into the central part of the peptide.[34] This guest residue is flanked by alanine residues; other non-alanine residues were kept in sufficient distance from the guest residue to avoid any type of side chain – side chain interactions. The Zimm-Bragg parameter s of this guest residue reflects the Gibbs energy associated with its coil⇔helix equilibrium provided that the residue is preceded by an

alanine residue in the helical state.[35] The Gibbs energy is determined by various enthalpic and entropic terms as follows:

$$\Delta G_{helix \leftrightarrow coil,x} = RT \cdot \ln s_x = \Delta H_{v,x} + \Delta H_{H,x} + \Delta H_{HB,x} + \Delta H_{rel,x} - T \cdot \left(\Delta S_x' + \delta S_{cor,x} \right) \quad (6)$$

where ΔH_v denotes the enthalpic difference between coil and helix *in vacuo*, ΔH_H the hydration enthalpy, ΔH_{HB} hydrogen bonding enthalpy, and ΔH_{rel} the difference between the solvent relaxation energy, which reflects the adaption of the solvent after moving a solute from vacuum into water. The entropic term $\Delta S_x'$ can be related to above ΔS_x as follows:

$$\Delta S_x' = \left(S_{\alpha x} - S_{rc} \right) - \Delta S_x \quad (7)$$

where $S_{\alpha x}$ is the conformational entropy of the residue in a right-handed helical conformation, which we calculated by assuming that each residue can sample a conformational space describable by a Gaussian distribution with $\sigma_\varphi = \sigma_\psi = 10 cm^{-1}$. This resembles distributions obtained from crystal structures and MD simulations.[36] Apparently, $\Delta S_x'$ implicitly contains contributions from structural changes including hydrogen bond formation, solvation entropy changes and the entropy associated with solvent relaxation. The term $\delta S_{cor,x}$ accounts for difference between Boltzmann and Gibbs entropy reflecting coupling between e.g. solvent and backbone coordinates.[30]

It follows from eq. (6) that one could expect some correlation between $\Delta G_{helix \Leftrightarrow coil}$ and ΔS_x if all the other contributions to Gibbs energy vary only modestly with the character of the guest residue's side chain. Figure 6 shows a plot of $-RT*ln(s)$ versus $\Delta S_x'$. A first glance at the data would hardly reveal any significant correlation. However, a closer look suggests the existence of two subsets of data for which a good correlation might exist. The smaller of these subset (called minor set in the following) just contains the branched and bulky residues F,W,V and T; the larger

one (called major set in the following) contains the remaining investigated residues besides alanine which does not fit into any of these two sets. The former have a lower helical propensity. We subjected both sets to linear regressions and obtained r-values of 0.89 and 0.94 for the major and minor set, respectively. Both values indicate rather good correlations. The slopes obtained from the fitting are 563±92 K for the minor and 671±156 K for the major set; the corresponding axis values are 7.96±1.0 kJ/mol and 11.4±1.8 kJ/mol. In the limit of statistical accuracy, the two slopes can be considered as identical. The fit to the minor data sets carry less significance since two of the four data points are nearly identical, so that the regression is practically based on only three data points.

One might be tempted to question the significance of a strategy that seems to arbitrarily partition the considered data set into two subsets and subject each of them to an individual regression analysis. It is perhaps thinkable that any random set of data could be dealt with by such an analysis, which would render the fits in Figure 6 statistically meaningless. To check for this possibility, we created a plot of 15 random numbers (corresponding with the number of data points in Figure 7), which is shown in Figure 7. A regression analysis of all data points yields a r-value of 0.53. This data set can indeed be dissected into subsets that can be subjected to linear fitting. However, as shown in Figure 7, the minimal number of these subsets is three and the maximal number is four. If we require a minimum of four data points per subset, the minimal and maximal number of linear plots for the data in Figure 6 is 2. One of our subsets contains 11 data points, its occurrence would be highly unlikely if the data distribution was random. This underscores the notion that our analysis is statistical significant. The significance of the identity of the amino acid subsets will be described further below.

With the exception of alanine, the correlations obtained for both data sets clearly suggest that the helical propensity actually decreases with decreasing difference between the conformational entropies of the respective residues. This seems to be somewhat surprising, because eq. (6) suggests that the slope of our correlation plots should be negative, in line with the expectation that a reduction in the loss of entropy should yield a higher helical propensity. However, one has to consider the well-established fact that folding/unfolding processes in general and helix⇔coil transitions in particular are governed by enthalpy-entropy compensation, which becomes exact at the folding temperature. In such a case enthalpy and entropy are linearly related and we can assume that a contribution $\Delta H_x'$ is related to $\Delta S_x'$ by:

$$\Delta H_x' = T_c \cdot \Delta S_x' \qquad (8)$$

We modify eq. (6) with eq. (8) as follows:

$$\Delta G_{helix \leftrightarrow coil,x} = \Delta H_{0,x} + (T_c - T)\Delta S_x' + (T_c' - T)\delta S_{cor,x} \qquad (9)$$

Note that $\Delta H_{0,x}$ denotes just the uncompensated enthalpy without any further specification. Eq. (9) considers the possibility that the compensation temperature for $\delta S_{cor,x}$ is different from T_c. $\Delta H_{0,x}$ is the part of the conformational enthalpy which is not subject to enthalpy-entropy compensation. Thus we obtain a correlation with a positive slope, if $T_c > T$. Based on the slopes of the correlation lines in Figure 6 one obtains T_c-values of 837 K for the major and 969 K for the minor data set. If one identifies the axis values obtained from the regression with the sum of $\Delta H_{0,x}$ and $(T_c'-T)\delta S_{cor,x}$, one can solve eq. (9) for T to obtain $\Delta G_{helix \leftrightarrow coil,x}=0$ temperatures between 41K and 268 K for the $\Delta S_x'$ range between -0.014 and 0.01 J/mol·K for the major set. For the minor set, the compensation temperature would be 154 K for $\Delta S_x' = 0.014$ J/mol·K, while it would become negative for 0.01 J/mol·K. These low temperatures are not unreasonable in view of the fact that besides alanine and arginine none of the considered residues exhibits s-values

above 1, which suggests that they are not helix promoting in the absence of any non-local interactions. Hence, one expects the entropy to be dominant at room temperature.

Eq. (9) indicates that a good correlation between ΔG_x' and ΔS_{conf} requires that the enthalpic contributions exhibit only a very limited dependence of the side chain of the analyzed set of residues. This notion apparently applies to the two data sets individually, while two residues that do not belong to the same set differ in term of the above thermodynamic parameters. The member F, V and W of the minor set have in common that the respective side chains are aliphatic and bulky, which is doubtless making particularly their solvation energies different from the those of the major set. The fourth member of the minor set, T, has a polar side chain, which however is as bulky as valine. The only observation that does not fit into the picture is the absence of I in the minor data set. Y lies somewhat between the two regression curves which might reflect the properties of the phenol side chain.

The behavior of alanine deserves some additional comments. Many attempts have been made to explain the high helical propensity of this residue in folded proteins. Results reported in this and in our earlier paper clearly show that alanine is peculiar in the unfolded state as well as it does not fit into any of the categories utilizable for other amino acids (with the exception of G and P, both not considered in this study). First, alanine has an abnormally high pPII propensity in the unfolded state (0.72 in GAG, > 0.8 in AAA, > 0.6 in polyalanines). With regard to GxG, its ΔG(T) curve does not share an iso-equlibrium point with other residues, most likely owing to the fact that its enthalpy-entropy compensation temperature is higher and pure conformational energy difference (i.e. *vacuo*) between pPII and β-strand are still slightly in favor of the former. [28] This notion is supported by the results of DFT calculation on the blocked alanine dipeptide which yielded a predponderance for pPII even in very low dielectric environment.[13]

In this study we show that only alanine has a high helical propensity that involves a comparatively small loss of conformational entropy upon the transition from a coil to a helical distribution. We, in accordance with others, attributed the high pPII propensity to an optimal hydration of the backbone, which is only guaranteed with alanine as side chain (again, we do not consider glycine here).[28] In the helical state alanine guarantees minimal interactions between side chains and is likely to still permit backbone hydration thus minimizing disfavorable enthalpic contributions to the Gibbs energy and allowing the hydrogen bonding contribution to dominate.

Comparison with experimental and theoretical entropy differences. Experimental values for entropic contributions to the helix propensity of amino acid residues are relatively rare. Luo and Baldwin conducted a thermodynamic study from which they determined the Gibbs energy and enthalpy contribution underlying the helix-forming propensity of x=A, L, I and V in Ac-KA_4xA_4KGY-NH_2 in different mixtures of trifluoroethanol and water.[37] The values reported for pure water allow the estimation of the respective entropic contributions to the Gibbs energy at 298 K, namely -3.6 kJ/mol, -3.0 kJ/mol, -3.3 kJ/mol and 4.0 kJ/mol for A, L, I, V. Corresponding $T\Delta S_x'$-values obtained from our calculations are -3.0 kJ/mol, -4.5 kJ/mol, -4.7 kJ/mol and -4.4 kJ/mol. Apparently, the calculated values are in the same order of magnitude as the experimental ones. We do not necessarily expect a perfect agreement of, or a close correlation between, experimental and estimated values, since the alanine context in this peptide can modify the intrinsic propensity of these amino acid and the solvation entropy is not taken into consideration. For a local random coil scenario the respective Gibbs or Helmholtz energy contributions would be much larger (i.e. ~-7.1 kJ/mol) than the above experimental values and thus in substantially lesser agreement with experimental data.

While experimental data with regard to the entropy loss associated with coil→helix transitions are rare, a variety of attempts have been made to estimate the corresponding contributions from conformational entropy. In a very early study Schellman estimated that folding would produce losses in conformational entropy corresponding to TΔS values between 3.5 and 9 kJ/mol.[38] Our values are clearly at the lower end of this region. Yang et al. reported that entropic losses that disfavor the folding into helices and sheets lie in the range between -7 and -9 kJ/mol, indicating that they assumed the unfolded state to be even slightly more random than the ensemble described by our random coil distribution.[39] D'Aquino et al. calculated the conformational entropy of residues of various blocked dipeptides *in vacuo*.[40] They normalized the respective distribution function on the conformational ensemble of the glycine dipeptide. The thus obtained relative entropies are very similar and cluster around -4.2 kJ/mol. This is what one expects for a local random coil distribution. Only for alanine did these authors calculate the conformational entropy loss associated with a coil→helix transition and obtained a value of -5.6 kJ/mol, which lies closer to our random coil than to our GAG based value. The number is substantially larger than the above experimental value of -3.6 kJ/mol. Baldwin used experimental data and an assumed entropy convergence temperature to estimate that the conformational entropy contribution to the overall folding entropy of hen lysozyme is 5.3 kJ/mol.[41] This includes side chain entropies. This value is at the upper limit of the $T\Delta S_x$' values used for the plots in Figure 7.

As indicated earlier, an estimation of folding induced entropic losses has recently been performed by Baxa et al. for ubiquitin.[24] To this end they biased MD simulation performed with an OPLS/AA-L force field with propensities obtained from truncated coil libraries. They showed that this leads to a substantial reduction of conformational entropy in the unfolded state and thus

also to a decrease of the entropic loss associated with secondary structure formation. Figure 8 compares our $T\Delta S_x$ values (at room temperature) with the entropic losses that they calculated for amino acid residues found in helical segments of folded ubiquitin. If one disregards D, S and N, which all exhibit an above the average propensity for turn-like structures, one obtains an excellent correlation (R=0.98) which involves the amino acids A, E, I, K, V and Y. The slope of the regression line is 0.46±0.06, which reflects the fact that the entropic Gibbs/Helmholtz energy contributions reported by Baxa et al. are somewhat larger than our values (the average value is -3.9 kJ/mol), but this is not surprising since the authors considered nearest neighbor interactions in their coil library, which we did not take into account.

What is the general significance of this and related thermodynamic studies on short peptides for the understanding of unfolded peptides/proteins and the energetics of protein folding? With regard to the unfolded state it becomes clear that in principle the total entropy depends on the amino acid sequence, but this might become irrelevant for longer proteins with comparable mixtures of aliphatic, polar and ionizable side chains. The average conformational entropy for these three groups is slightly different; from our data we obtained $<\Delta TS_x>_{aliph}$'=-4.5kJ/mol (V,I,L,M, A is taken out), $<\Delta TS_x>_{arom}$'=-3.9 kJ/mol (Y,W,F), $<\Delta TS_x>_{charged}$'=-4.6 kJ/mol (K,E, and R, D_i is taken out) and $<\Delta TS_x>_{pol}$'=-3.4 kJ/mol (S,T and N). The values for A and D_i (-3.0 and -2.4 kJ/mol) are considerable lower than the average of their respective group. IDPs are known to contain more charged and less aliphatic and aromatic residues than foldable proteins,[23] but the above values suggest that this would not make necessarily a difference with regard to the conformational entropy. However, a preponderance of polar residues combined with a substantial fraction of alanine and aspartic acid residues could produce a peptide or protein with substantially less conformational entropy than one would expect to find for proteins

with a dominance of aliphatic groups and charged groups with longer side chains. Since IDPs have also more polar than aliphatic and aromatic group, the actual entropy gain might be smaller. Generally, the sequence will matter mostly for comparatively short IDPs, IDP segments and unfolded peptides, particularly if the composition is not very heterogeneous. Self-aggregation peptides have an above the average fraction of hydrophobic groups. Bulky neighbors can interact with each other (see VVV in Figure 2), thus affecting the total entropy. Polyalanines will always have rather low conformational entropies which might favor helix formation [34, 42, 43] and self-aggregation alike. [44]

A stronger sequence influence could arise from the conformation and residue dependence of the solvation enthalpy and entropy, which Toal et al. recently reported for the GxG peptides discussed in this study.[28] This should have a significant effect particularly on the stabilization of unfolded states at high temperatures, at which entropically favored β-strand structures become increasingly populated. To illustrate this point we calculated the changes in average enthalpy and entropy associated with a temperature increase from 298 to 350 K by utilizing the following relationships:

$$\langle \xi \rangle = \frac{\xi e^{-\Delta G_{pPII-\beta}/RT}}{1 + e^{-\Delta G_{pPII-\beta}/RT}} \qquad (9)$$

where ξ represents $\Delta H_{pPII-\beta}$ and $\Delta S_{pPII-\beta}$, R is the gas constant and T the absolute temperature. Figure 9 shows the difference ζ between $<\Delta H_{pPII-\beta}>$ and $T<\Delta S_{pPII-\beta}>$ values calculated for 298 and 350 K for all amino acids investigated. The obtained diagram is reminiscent of the large differences between the $\Delta H_{pPII-\beta}$ and $\Delta S_{pPII-\beta}$ values reported by Toal et al.[28] While the changes are in the range between 0 and -5 kJ/mol for most amino acid residues, L, R and particularly I and V stand out in that the data indicate their solvent free energy becomes substantially more

entropic at high temperature. This implies that unfolded states of peptides and proteins with a preponderance of these amino acids (e.g. in aliphatic peptides with a high propensity for self-aggregation) are substantially more entropic and less enthalpic than polypeptides with more polar or ionizable side chains.

We used the data plotted in Figure 4 to calculate the gain in conformational backbone entropy involving the above increase of temperature. Not too surprising we obtained rather similar values. The average $\Delta(TS_x)$ values for all amino acids is 3.19±0.17 kJ/mol. The difference between the $\Delta(TS_x)$ visualized in Figure 4 and the differences between $T<\Delta S_{pPII-\beta}>$ depicted in Figure 9 could well reflect differences between the Boltzmann and Gibbs entropy, which can be expect to predominantly reflect solvation. Moreover, one has to consider that all thermodynamics in Figure 9 are expressed with reference to pPII.

The losses of conformational entropy calculated in this study should be considered as an estimate of how conformational propensities can reduce the entropy of unfolded peptides and proteins. Calculations of more exact values would require the consideration of nearest neighbor interactions. However, the limited data about the influence of nearest neighbors presented in this paper suggest that, while significant, its consideration will not change the entropic contributions to the Helmholtz or Gibbs energy by more than RT. Results form a more systematic investigation of nearest neighbor effects which is currently being conducted in our laboratory rather indicates that they have a drastic influence on the pPII/β-equilibrium and thus on solvation enthalpies and entropies of residues (Toal, Richter, Schwalbe, Schweitzer-Stenner, unpublished). Generally, the statistical entropy derived from conformational propensities constitute an upper limit for unfolded proteins, the real values might be lower. Multiple lines of evidence suggest that even in the unfolded state non-local interactions can temporary stabilize secondary structure like turns,

sheets and helices. An exact estimation of the thermodynamics of protein folding has to take this into account as well.

Summary

Based on recently reported conformational distributions of amino acid residues of GxG peptides in water we calculated the relative conformational entropy of selected x-residues thereby using an assumed (local) random coil like distribution as reference. We found that the conformational preferences reported in our earlier studies lead to a substantial decrease of conformational entropy. The corresponding contributions to the Gibbs/Helmholtz energy at room temperature lie between 1 and 2 RT. The entropy reduction effect is most pronounced for alanine and protonated aspartic acid. Upon increasing temperature the backbone entropy of GxGs approaches the entropy of the assumed random coil distribution. We used the obtained conformational backbone entropies to calculate the respective loss associated with a coil → helix transitions and obtained rather good correlations with the respective Gibbs energy of helix formation (i.e. -RT·ln s) for two different subsets of amino acid residues. We found that our calculated entropy losses are in qualitative agreement with experimental values reported for amino acid residues in an alanine based host-guest system. Moreover, we showed that the backbone entropy losses derived in this study correlates well with corresponding data obtained with coil library guided MD simulations for amino acid residues found in helical segments of ubiquitin. Finally, we showed that our results indicate a quite substantial increase of solvation and conformational entropy with increasing temperature. Taken together, we show that conformational propensities of amino acid residues in GxG can be used for deriving a realistic estimation of the backbone entropy of unfolded peptides and proteins.

Acknowledgement

Siobhan E. Toal is supported by a research assistant fellowship provided by the College of Arts and Sciences. The authors like to thank George Rose for a very helpful discussion about a proper understanding of conformational entropies. The authors do not declare any conflict of interest.

Table 1: List of parameters used to construct a (local) random coil like distribution. The first four columns list the locations and widths of the Gaussian sub-distributions in the Ramachandran space. The fight column list the mole fractions of the respective sub-distribution.

Conformation	$\varphi[^o]$	$\psi[^o]$	$\sigma_\varphi[^o]$	$\sigma_\psi[^o]$	χ
pPII	-70	150	20	20	0.18
β_T	-115	160	20	20	0.15
β_1	-115	134	20	20	0.15
β_2	-140	125	20	20	0.15
type I β	-60	0	20	20	0.1
α_r	-48	-57	20	20	0.195
α_l	60	50	10	20	0.075

Figure legends

Figure 1: Conformational distributions of GxG peptides derived from amide I' profiles and NMR J-coupling constants. Upper panel: GAG (left), GVG (center); lower panel: GLG (left) and protonated GDG (right). Taken from ref. 16 and modified.

Figure 2: Ramachandran plot of a random coil like distribution of conformational ensembles. The respective parameters are listed in table 1. Taken from Toal et al. and modified.45

Figure 3: Display of the difference between the Gibbs energy contributions of the conformational entropies of the indicated amino acid residues and a hypothetical random coil distribution. Single letters denote amino acid residues in GxG peptides, triples of letters indicate the properties of the central residue in tri-homopeptides. The index i indicates the ionized state of the residue. The upper solid and lower dashed horizontal lines indicate the region between RT and 2RT for T=298.15 K.

Figure 4: Relative Helmholtz energies $T\Delta S_x = T(S_x - S_{coil})$ at 298° K(dark grey), 301°K, 311°K, 340°K, and 370°K (light grey) derived from conformational distributions of the central x-residues in GxG peptides

Figure 5: Relationship between $\Delta S_x = S_x - S_{coil}$ and the experimentally derived $\Delta S_{pPII-\beta}$.28 The solid line reflects a linear least squares fit with a correlation coefficient of 0.62.

Figure 6: Comparison of the per-residue Gibbs energy of helix formation (-RT ln(s), s is the Zimm-Bragg parameter)34 and the Gibbs energy contribution of the conformational entropy loss of 15 amino acid residues upon a transition from a coil-like to a right handed helical conformation. The temperature value is 298 K. The conformational entropy loss was calculated as the sum of backbone side chain entropy changes as described in the text. The solid lines represent linear regression to two clusters of data.

Figure 7: Graphic representation of a statistical distribution of numbers. The abscissa reflects the numbering of the attempts to create an arbitrary number in the indicated region with MATLAB. The ordinate is the value of the obtained number. Solid lines visualize different option for selective linear relationships of data sub-sets. The red line results from a linear regression to the entire set of numbers.

Figure 8: ΔS_x-entropy losses obtained from backbone conformational entropies of GxG peptides plotted versus corresponding conformational entropy losses of amino acid residues in helical segments of ubiquitin obtained from MD simulations restricted by propensities derived from truncated coil libraries. The solid line represents a linear fit to a restricted number of data points. The correlation coefficient is 0.98.

Figure 9: Plot of differences ζ between $<\Delta H_{pPII-\beta}>$ (black bars) and T $<\Delta S_{pPII-\beta}>$ (grey bars) of GxG peptides calculated for temperatures of 350 and 298 K by utilizing the $\Delta H_{pPII-\beta}$ and $\Delta S_{pPII-\beta}$ values reported by Toal et al.28

Figure 1

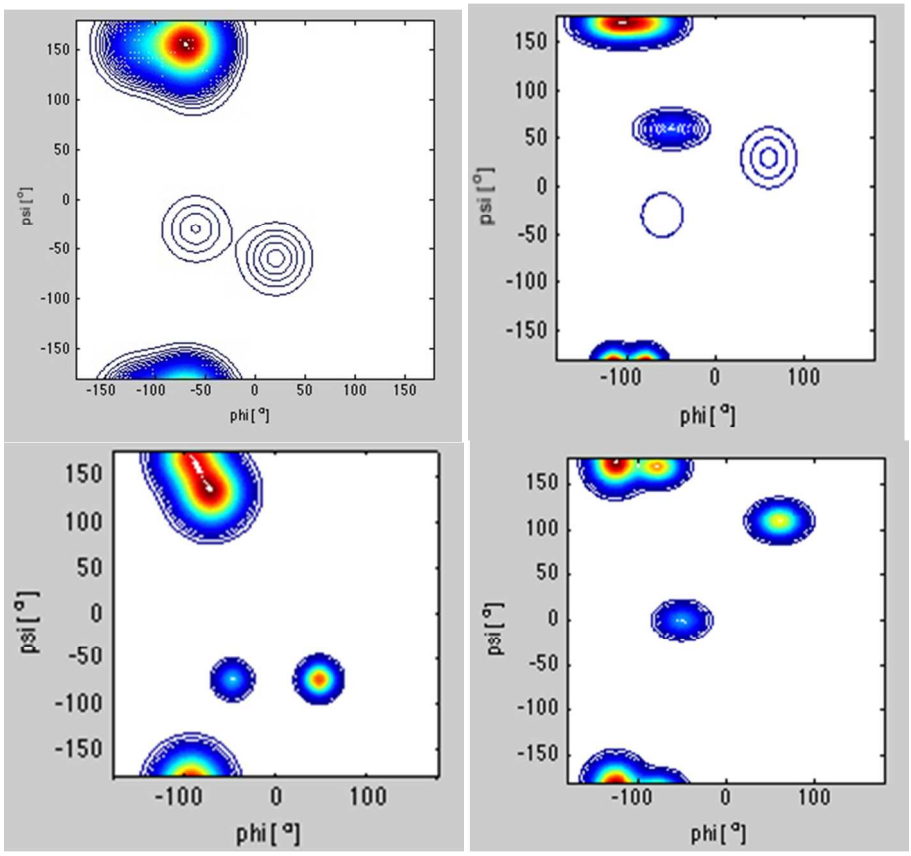


Figure 2

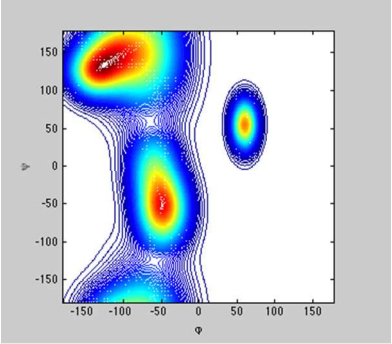


Figure 3

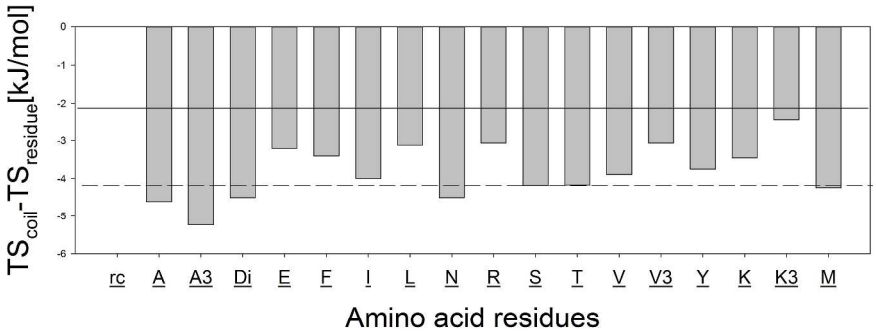


Figure 4

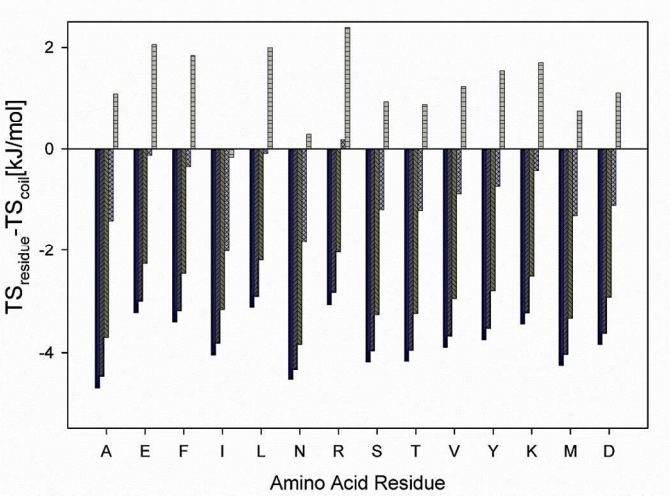


Figure 5

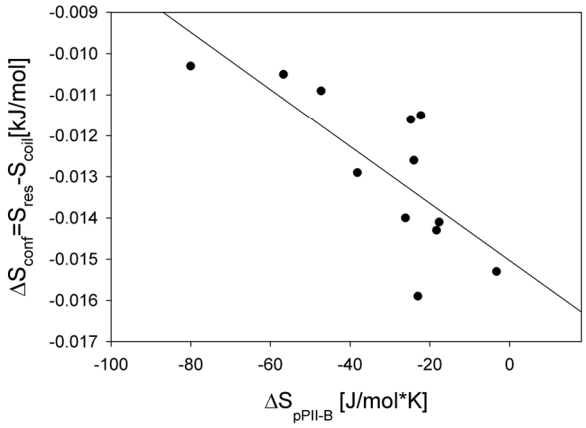


Figure 6

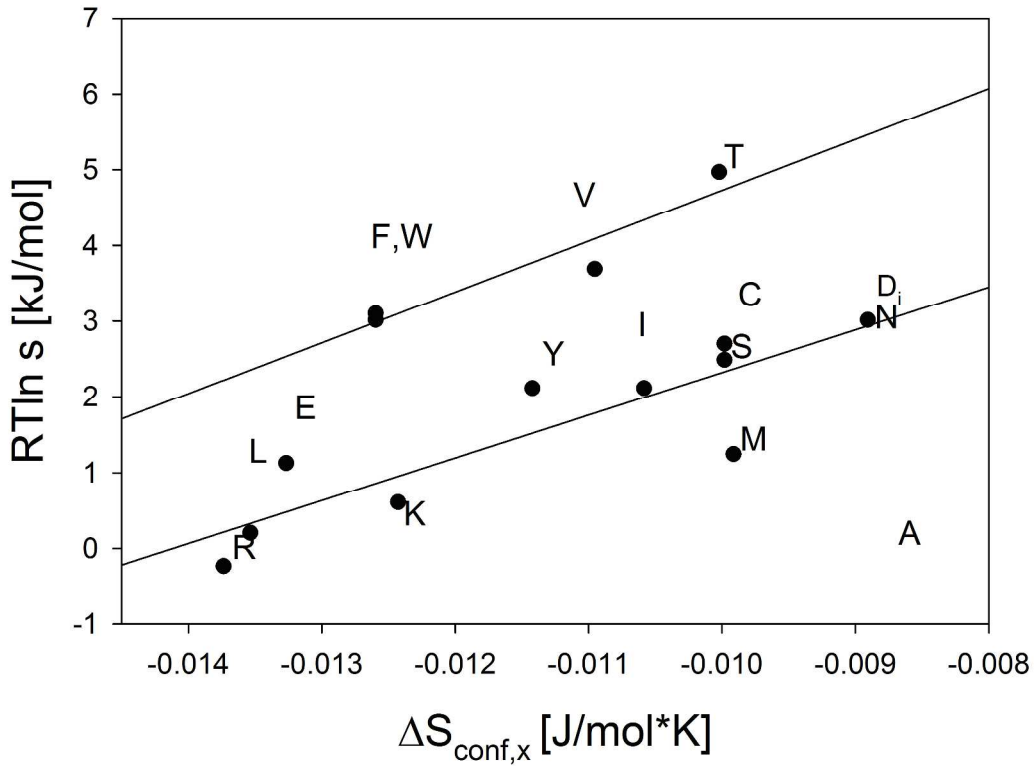


Figure 7

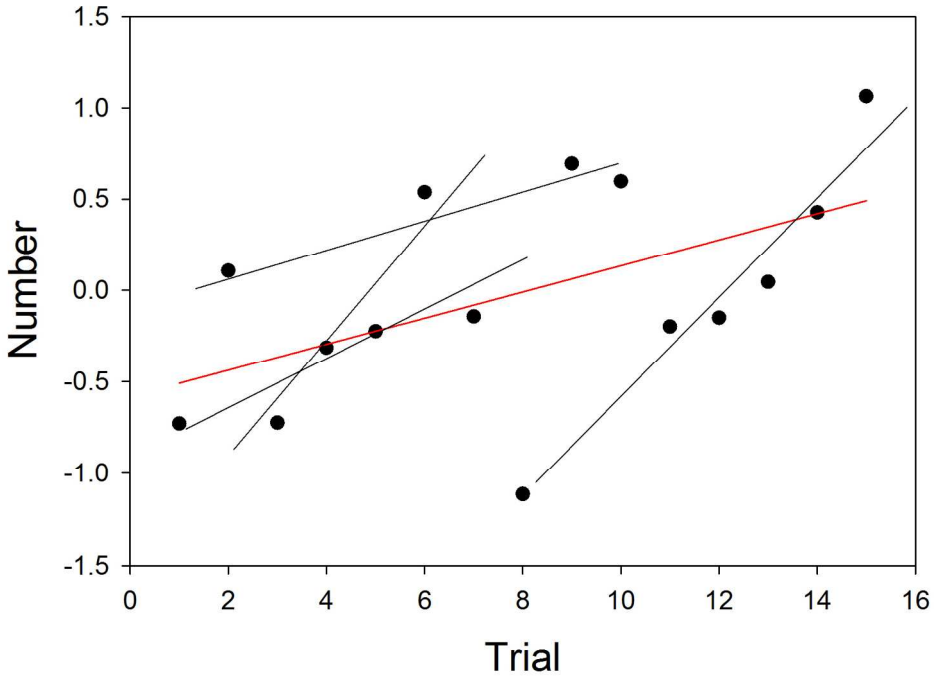


Figure 8

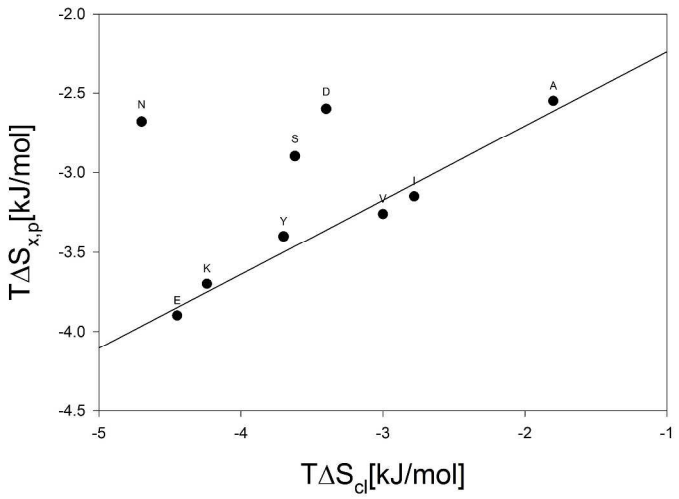
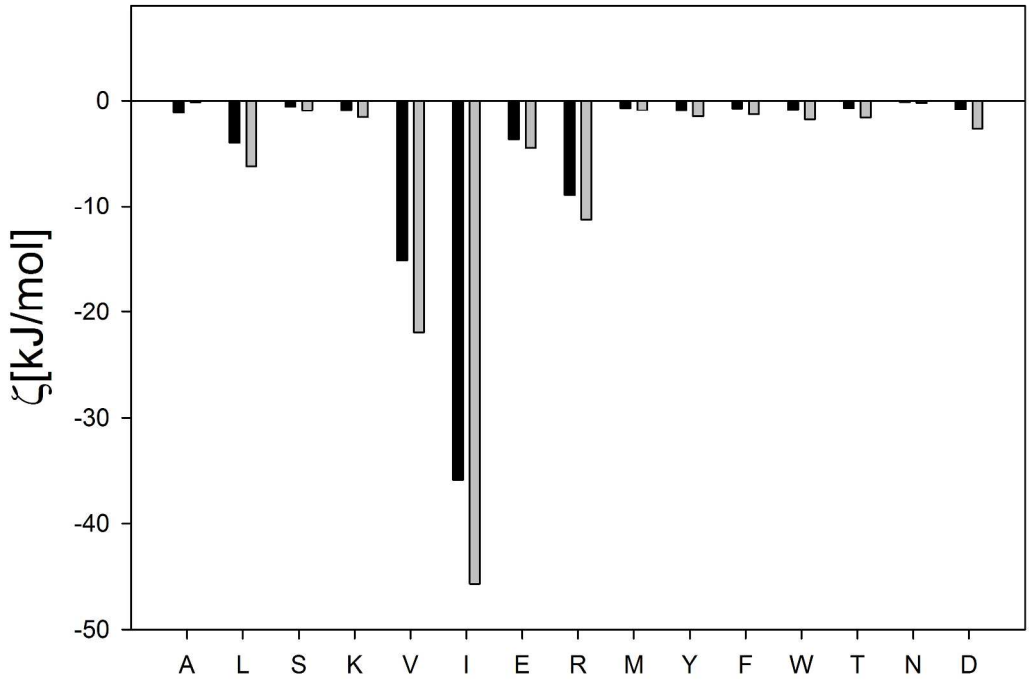


Figure 9



Bibliography

1. G. N. Ramachandran, C. Ramachandran and V. Sasisekharan, J. Mol. Biol., 1963, 7, 95-99.

.

5. Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin and N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 9190-9195.
6. I. H. McColl, E. W. Blanch, L. Hecht, N. R. Kallenbach and L. D. Barron, *J Am Chem Soc*, 2004, **126**, 5076-5077.
7. K. Chen, Z. Liu and N. R. Kallenbach, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 15352-15357.
8. R. Schweitzer-Stenner and T. Measey, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 6649–6654.
9. I. C. Dragomir, T. J. Measey, A. M. Hagarman and R. Schweitzer-Stenner, *J Phys Chem B*, 2006, **110**, 13235-13241.
10. S. Toal, A. Omidi and R. Schweitzer-Stenner, *J. Am. Chem. Soc.*, 2011, **133**, 12728.
11. R. P. Bywater and V. Veryazov, *Naturwissenschaften*, 2013, **100**, 853-859.
12. P. M. Cowan and S. Mc Gavin, *Nature*, 1955, **176**, 501-503.
13. F. Avbelj and R. L. Baldwin, *Proc Natl Acad Sci U S A*, 2004, **101**, 10967-10972.
14. K. Rybka, S. E. Toal, D. J. Verbaro, D. Mathieu, H. Schwalbe and R. Schweitzer-Stenner, *Proteins*, 2013, **81**, 968-983.
15. , !!! INVALID CITATION !!!
16. A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky, *FEBS J*, 2005, **272**, 5129-5148.
17. M. C. Baxa, E. J. Haddadain, A. K. Jha, K. F. Fred and T. R. Sosnick, *J. Am. Chem. Soc*, 2012, **134**, 15929-15936.
18. S. E. Toal, D. J. Verbaro and R. Schweitzer-Stenner, *J. Phys. Chem B*, 2014, **118**, 1309-1318.
19. R. Schweitzer-Stenner, *J. Phys. Chem. B*, 2009, **113**, 2922-2932.
20. A. Hagarman, T. J. Measey, D. Mathieu, H. Schwalbe and R. Schweitzer-Stenner, *J Am Chem Soc*, 2010, **132**, 540-551.
21. A. Hagarman, D. Mathieu, S. Toal, T. J. Measey, H. Schwalbe and R. Schweitzer-Stenner, *Chemistry*, 2011, **17**, 6789-6797.
22. R. Schweitzer-Stenner, A. Hagarman, S. Toal, D. Mathieu and H. Schwalbe, *Proteins*, 2013, **81**, 955-967.
23. J. J. Hopfield, *Proc. Natl. Acad. Sci. USa*, 1974, **9**, 3640-3644.
24. V. N. Uversky, C. J. Oldfield and K. A. Dunker, *J. Mol. Recognit.*, 2005, **18**, 343-384.
25. L. Duitch, S. Toal, T. J. Measey and R. Schweitzer-Stenner, *J Phys Chem B*, 2012, **116**, 5160-5171.
26. D. Verbaro, D. Mathieu, S. E. Toal, H. Schwalbe and R. Schweitzer-Stenner, *J. Phys. Chem. B.*, 2012, **116**.
27. A. Chakrabartty, T. Kortemme and R. L. Baldwin, *Prot. Sci.*, 1994, **3**, 843-852.
28. T. P. Creamer and G. D. Rose, *Proteins*, 1994, **19**, 85-97.
29. S. Gnanakaran and A. E. Garcia, *J. Phys. Chem. B*, 2003, **107**, 12555-12557.

30. P. Luo and R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1999, **1999**, 4930-4935.
31. J. A. Schellman, *C.R. Trav. Lab. Carlsberg, Ser. Chim.*, 1955, **29**, 230-259.
32. A. S. Yang and B. Honig, *J. Mol. Biol.*, 1995, **252**, 351-365.
33. R. L. Baldwin, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 8096-8072.
34. V. N. Uversky, *Eur. J. Biochem.* , 2002, **269**, 2-12.
35. V. Uversky and A. K. Dunker, in *Protein and Peptide Folding, Misfolding, and Non-Folding*, ed. R. Schweitzer-Stenner, Wiley & Sons, Hoboken, 2012, pp. 3-54.
36. S. Marqusee, V. H. Robbins and R. L. Baldwin, *Proc Natl Acad Sci U S A*, 1989, **86**, 5286-5290.
37. T. J. Measey and R. Schweitzer-Stenner, in *Protein and Peptide Folding, Misfolding and Non-Folding*, ed. R. Schweitzer-Stenner, Wiley & Sons, Hoboken, 2012, pp. 309-350.
38. M. Gruebele and D. Thirumalai, *J. Chem. Phys*, 2013, **139**, 12701-12701-12701-12709.

Bibliography

Bibliography

Bibliography

Bibliography

Bibliography

Bibliography

1. G. N. Ramachandran, C. Ramachandran and V. Sasisekharan, J. Mol. Biol., 1963, 7, 95-99.
2. P. J. Flory, Statistical Mechanics of Chain Molecules, Wiley & Sons, New York, 1969.
3. C. Tanford, Adv. Prot. Chem., 1968, 23, 121-282.
4. A. K. Jha, A. Colubri, M. H. Zaman, S. Koide, T. R. Sosnick and K. F. Freed, Biochemistry, 2005, 44, 9691-9702.
5. A. K. Jha, A. Colubri, K. F. Freed and T. R. Sosnick, Proc Natl Acad Sci U S A, 2005, 102, 13099-13104.
6. J. DeBartolo, A. Jha, K. Freed, F. and T. R. Sosnick, in Proteins and Peptides. Folding, Misfolding and Unfolding., ed. R. Schweitzer-Stenner, Wiley & Sons, Chichester, 2012, pp. 79-98.
7. Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin and N. R. Kallenbach, Proc. Natl. Acad. Sci. USA, 2002, 99, 9190-9195.
8. I. H. McColl, E. W. Blanch, L. Hecht, N. R. Kallenbach and L. D. Barron, J Am Chem Soc, 2004, 126, 5076-5077.
9. K. Chen, Z. Liu, C. Zhou, W. C. Bracken and N. R. Kallenbach, Angew Chem Int Ed Engl, 2007, 46, 9036-9039.
10. J. Graf, P. H. Nguyen, G. Stock and H. Schwalbe, J Am Chem Soc, 2007, 129, 1179-1189.
11. R. Schweitzer-Stenner and T. Measey, Proc. Natl. Acad. Sci. USA, 2007, 104, 6649–6654.
12. R. Schweitzer-Stenner, J. Phys. Chem. B, 2009, 113, 2922-2932.
13. R. P. Bywater and V. Veryazov, Naturwissenschaften, 2013, 100, 853-859.
14. S. Toal, D. Meral, D. Verbaro, B. Urbanc and R. Schweitzer-Stenner, J Phys Chem B, 2013, 117, 3689-3706.
15. P. M. Cowan and S. Mc Gavin, Nature, 1955, 176, 501-503.
16. R. Schweitzer-Stenner, A. Hagarman, S. Toal, D. Mathieu and H. Schwalbe, Proteins, 2013, 81, 955-967.
17. C. J. Penkett, C. Redfield, I. Dodd, J. Hubbard, D. L. McBay, D. E. Mossakowska, R. A. Smith, C. M. Dobson and L. J. Smith, J Mol Biol, 1997, 274, 152-159.
18. F. Avbelj and R. L. Baldwin, Proc Natl Acad Sci U S A, 2004, 101, 10967-10972.

19. Z. Shi, K. Shen, Z. Liu and N. R. Kallenbach, Chem. Rev., 2006, 106, 1877-1897.

20. R. Schweitzer-Stenner, Mol. BioSyst., 2012, in press.

21. K. A. Dill, Biochemistry, 1990, 29, 7133-7155.

22. C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky and A. K. Dunker, BMC Genomics, 2008, 9, 1-20.

23. V. Uversky and A. K. Dunker, in Protein and Peptide Folding, Misfolding, and Non-Folding, ed. R. Schweitzer-Stenner, Wiley & Sons, Hoboken, 2012, pp. 3-54.

24. M. C. Baxa, E. J. Haddadain, A. K. Jha, K. F. Fred and T. R. Sosnick, J. Am. Chem. Soc, 2012, 134, 15929-15936.

25. A. Hagarman, T. J. Measey, D. Mathieu, H. Schwalbe and R. Schweitzer-Stenner, J Am Chem Soc, 2010, 132, 540-551.

26. A. Hagarman, D. Mathieu, S. Toal, T. J. Measey, H. Schwalbe and R. Schweitzer-Stenner, Chemistry, 2011, 17, 6789-6797.

27. K. Rybka, S. Toal, D. Verbaro, D. Mathieu and R. Schweitzer-Stenner, Proteins: Structure, Function and Genetics, 2013, DOI: 10.1002/prot.24226.

28. S. E. Toal, D. J. Verbaro and R. Schweitzer-Stenner, J Phys Chem B, 2014, 118, 1309-1318.

29. D. Verbaro, D. Mathieu, S. E. Toal, H. Schwalbe and R. Schweitzer-Stenner, J. Phys. Chem. B., 2012, 116.

30. E. T. Jaynes, A.J. Phys., 1965, 33, 391-398.

31. V. N. Uversky, Eur. J. Biochem., 2002, 269, 2-12.

32. V. N. Uversky, C. J. Oldfield and K. A. Dunker, J. Mol. Recognit., 2005, 18, 343-384.

33. L. Duitch, S. Toal, T. J. Measey and R. Schweitzer-Stenner, J. Phys. Chem. B, 2012, in press.

34. A. Chakrabartty, T. Kortemme and R. L. Baldwin, Prot. Sci., 1994, 3, 843-852.

35. B. H. Zimm and J. K. Bragg, J. Chem. Phys., 1959, 31, 526.

36. T. P. Creamer and G. D. Rose, Proteins, 1994, 19, 85-97.

37. P. Luo and R. L. Baldwin, Proc. natl. Acad. Sci. USA, 1999, 96, 4930-4935.

38. J. A. Schellman, Compt. Rend. Trav. Lab. Carlsburg, Ser. Chim., 1955, 29, 230-259.

39. C. Yang and B. Honig, J. Mol. Biol., 1995, 252, 351-365.

40. J. A. D'Aquino, J. Gomez, V. J. Hilser, K. H. Lee, L. M. Amzel and E. Freire, Proteins, 1996, 25, 143-156.

41. R. L. Baldwin, Proc. Natl. Acad. Sci. USA, 1986, 83, 8069-8072.

42. S. Marqusee and R. L. Baldwin, Proc Natl Acad Sci U S A, 1987, 84, 8898-8902.

43. S. Marqusee, V. H. Robbins and R. L. Baldwin, Proc Natl Acad Sci U S A, 1989, 86, 5286-5290.

44. T. J. Measey and R. Schweitzer-Stenner, in Protein and Peptide Folding, Misfolding and Non-Folding, ed. R. Schweitzer-Stenner, Wiley & Sons, Hoboken, 2012, ch. 11, pp. 309-350.

45. S. E. Toal, D. J. Verbaro, D. Meral, B. Urbanč and R. Schweitzer-Stenner, submitted for publication, 2012.