

# Hydration of the peptide backbone largely defines the thermodynamic propensity scale of residues at the C' position of the C-capping box of $\alpha$ -helices

Susan T. Thomas\*, Vakhtang V. Loladze, and George I. Makhatadze†

Department of Biochemistry and Molecular Biology, Penn State University, College of Medicine, Hershey, PA 17033

Communicated by Robert L. Baldwin, Stanford University Medical Center, Stanford, CA, July 23, 2001 (received for review June 21, 2001)

The C' position of the C-capping box is the second residue outside of the helix. Statistical analysis of residue distribution at the C' position in the  $\alpha$ -helices' C-capping box showed that different amino acid residues occur with different probabilities, with the strongest preference being for glycine. To understand the physico-chemical basis for this preference, we studied the effects that 17 amino acid substitutions at the C' position in an  $\alpha$ -helix of ubiquitin have on the stability of this protein. We determined the following rank order of amino acid residues at the C' position with respect to their effect on the stability: Gly>His>Asn>Arg>Lys>Gln>Ala>Phe>Met>Ser>Asp>Glu>Trp>Thr>Pro>Ile>Val. The effect of the amino acid substitutions on the structure also was evaluated by comparing the  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear sequential quantum correlation spectra and showed no significant changes in the structures of the most stable (Gly) and the least stable (Val) variants. The obtained changes in stability highly correlate ( $r = 0.85$ ) with the statistical distribution of the residues at the C' position indicating that the measured thermodynamic propensities are unbiased by secondary interactions. We also found that the measured thermodynamic propensities correlate well with the amide hydrogen exchange data on short model peptides ( $r = 0.85$ ) and the calculated hydration of the peptide backbone ( $r = 0.88$ ). These results combined with the changes in enthalpy and entropy of unfolding of ubiquitin variants suggest that dehydration of the peptide backbone plays a significant role in defining the thermodynamic propensity scale at the C' position of the C-capping box in  $\alpha$ -helices. This propensity scale is useful for protein secondary structure predictions and protein design.

The first basic structural motif of globular proteins proposed was the  $\alpha$ -helix (1). The structure of the  $\alpha$ -helix is characterized by hydrogen bonding patterns between amide hydrogen bond donors and carbonyl oxygen acceptors of residues situated four apart in the sequence. This pattern of hydrogen bonding, however, implies that the four initial amide hydrogen bond donors and the four final carbonyl oxygen hydrogen bond acceptors do not have hydrogen bonding partners. The potential effect of this is fraying of the helix ends. In 1988, Richardson and Richardson (2) as well as Presta and Rose (3) analyzed statistical preferences of the amino acid residues at the ends of  $\alpha$ -helices and revealed the existence of specific capping interactions at both the N and C termini, which compensate for the unsatisfied hydrogen bonds and thus prevent the ends from fraying. According to Aurora *et al.* (4) and Harper and Rose (5), amino acid residues in helices and their flanking residues can be labeled as:

...N"-N'-Ncap-N1-N2-N3-N4-... ..

C4-C3-C2-C1-Ccap-C'-C"-C''... ..

where numbered residues have helical backbone dihedral angles ( $\phi = -64 \pm 7^\circ$ ;  $\psi = -41 \pm 7^\circ$ ). Ncap with dihedrals  $\phi = -94 \pm 15^\circ$  and  $\psi = 167 \pm 5^\circ$  (5) and Ccap with dihedrals  $\phi \approx -81^\circ$  and

$\psi \approx -2^\circ$  (6) are boundary residues that belong both to the helix and the adjacent turn. Residues at the C' position often have positive dihedral angles and often are occupied by a glycine residue (7). What is the physico-chemical basis that defines this preference? One explanation is that Gly introduces flexibility to the peptide backbone, which allows it to adopt a left-handed conformation, i.e., positive  $\phi$  and  $\psi$  dihedral angles, with a minimal loss of energy (2, 3). Serrano *et al.* (8) argue that Gly at the C' position is largely exposed to solvent and substitution of C' Gly by other amino acid residue lead to dehydration of the peptide backbone, affect the strength of the C' to C3 hydrogen bond, and decrease stability. To gain deeper insight we have studied the effects amino acid substitutions at the C' position of the  $\alpha$ -helix in ubiquitin have on the stability of this protein. The C-capping box of the  $\alpha$ -helix of ubiquitin has a sequence—Ile<sub>30</sub>Gln<sub>31</sub>Asp<sub>32</sub>Lys<sub>33</sub>Glu<sub>34</sub>Gly<sub>35</sub>Ile<sub>36</sub>Pro<sub>37</sub>—and shows the characteristics of the  $\alpha_L$  motif (4) with Gly-35 as a C' residue, the backbone-backbone hydrogen bond between residues C' (Gly-35) and C3 (Gln-31) and van der Waals interactions between residues C4 (Ile-30) and C'' (Ile-36). As expected Gly-35 has positive torsional angles,  $\phi = 81.2^\circ$  and  $\psi = 5.3^\circ$  (4, 6, 7, 9, 10).

Seventeen amino acid substitutions at the Gly-35 position of ubiquitin were incorporated into the yeast ubiquitin sequence and the effects of these substitutions on ubiquitin stability were measured by differential scanning calorimetry (DSC). In addition, the effect the substitutions had on the ubiquitin structure was evaluated by using NMR spectroscopy. We found that there is a large (up to 30°C, depending on the nature of the amino acid residues) decrease in stability, relative to wild-type (WT) ubiquitin. These large decreases in stability do not seem to be caused by the large structural changes upon substitution as shown by comparison of heteronuclear sequential quantum correlation (HSQC) spectra of ubiquitin variants. Analysis of the changes in the thermodynamic parameters that accompany the decrease in stability and the correlational analysis with the data reported in the literature indicate that the decrease in stability is largely defined by the dehydration of the peptide backbone by the side chain of a residue at the C' position.

## Materials and Methods

**Mutagenesis and Expression of Ubiquitin Variants.** Site-directed mutagenesis of the yeast ubiquitin gene was carried out by using the QuickChange Site-Directed Mutagenesis kit (Stratagene).

Abbreviations: DSC, differential scanning calorimetry; WT, wild type; ASA, amount of surface area.

\*Present address: Center for Biomolecular Analysis, Department of Biochemistry, University of Texas, Health Science Center, San Antonio, TX 78229.

†To whom reprint requests should be addressed. E-mail: makhatadze@psu.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

**Table 1. Thermodynamic parameters of unfolding of the position 35 variants of ubiquitin at pH 3.0**

Residue in position 35	$T_m$ , °C	$\Delta H_{cal}(T_m)$ , kJ/mol	$\Delta H(50^\circ\text{C})$ , kJ/mol	$\Delta S(50^\circ\text{C})$ , J/mol·K	$\Delta G(50^\circ\text{C})$ , kJ/mol	$\Delta\Delta G(50^\circ\text{C})$ , kJ/mol	$\Delta\Delta H(50^\circ\text{C})$ , kJ/mol	$-T\cdot\Delta\Delta S(50^\circ\text{C})$ , kJ/mol
GLY	59.4	205	172	516	5.3	0.0	0	0
HIS	58.1	188	160	481	4.6	-0.7	-12	11
ASN	54.1	174	160	488	2.4	-2.9	-12	9
ARG	53.5	157	145	443	1.9	-3.4	-27	24
LYS	53.6	163	150	460	1.4	-3.9	-22	18
GLN	52.1	161	154	473	1.2	-4.1	-18	14
ALA	50.7	145	143	440	0.9	-4.4	-29	25
PHE	50.4	162	161	497	0.5	-4.8	-11	6
MET	50.4	153	152	469	0.5	-4.8	-20	15
SER	48.7	142	147	456	-0.3	-5.6	-25	19
ASP	49.9	154	154	478	-0.4	-5.7	-18	12
GLU	49.4	150	152	472	-0.5	-5.8	-20	14
TRP	46.0	134	148	464	-1.9	-7.2	-24	17
THR	41.3	102	132	420	-3.7	-9.0	-40	31
PRO	39.6	87	123	393	-3.9	-9.2	-49	40
ILE	36.8	90	136	437	-5.2	-10.5	-36	26
VAL	27.5	45	124	402	-5.8	-11.1	-48	37

Estimated experimental errors are  $T_m \pm 0.2^\circ\text{C}$ ,  $\Delta H_{cal}(T_m) \pm 7\%$ .

The presence of mutations was confirmed by sequencing the entire ubiquitin gene with an Applied Biosystems PRISM 310 Genetic Analyzer. Overexpression of the ubiquitin variants was done in the BL21(DE3) strain of *Escherichia coli* as described (11). Each variant was purified to apparent homogeneity as described (12). For reasons discussed previously (13), all described ubiquitin variants contained R63K substitution. Overexpression and purification of the  $^{15}\text{N}$ -labeled WT and G35V ubiquitin variant was done as described (14). To increase the yield, the protein was expressed as a His-tag fusion and purified by using Ni-column and gel filtration chromatography as described (14). Attachment of His tag does not affect the position of resonances on  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum (14).

**DSC.** DSC experiments were performed on a VP-DSC (Microcal, Northampton, MA) instrument at a scan rate of 1 deg/min. The protein concentration for ubiquitin mutants varied between 1.0 and 3.0 mg/ml. Detailed procedures of the sample preparation for the DSC experiment were the same as described (12, 15). The partial molar heat capacity of the protein,  $C_{p,pr}(T)$ , was obtained from the experimentally measured apparent heat capacity difference between the sample (containing protein solution) and reference (containing corresponding buffer solution) cells,  $\Delta C_p^{app}$ , using the following expression:

$$C_{p,pr}(T) = \frac{C_{p,buf}}{\bar{V}_{buf}} \cdot \bar{V}_{pr} - \frac{\Delta C_p^{app}(T)}{m_{pr} \cdot M}, \quad [1]$$

where  $\bar{V}_{pr}$  is the partial molar volume of a protein calculated as described (16),  $C_{p,buf}$  and  $\bar{V}_{buf}$  are the partial molar heat capacity and the partial molar volume of the aqueous buffer, respectively,  $m_{pr}$  is the mass of the protein in the cell, and  $M$  is the molar mass of the protein.

Protein concentration was measured spectrophotometrically by using a known extinction coefficient for the WT ubiquitin,  $\epsilon_{1\text{cm},280\text{nm}}^{0.1\%} = 0.149$  (11–13). This extinction coefficient was used for all but G35W ubiquitin variants because none of those amino acid substitutions introduced residues that absorb in the far UV range. The extinction coefficient for the G35W variant was estimated at  $\epsilon_{1\text{cm},280\text{nm}}^{0.1\%} = 0.80$  using the methods of Pace *et al.* (17). Correction for light scattering was taken into account as described (18). Analysis of the heat capacity profiles was done by using nonlinear regression routine NLREG and in-house written

scripts (15). Individual curves for a given ubiquitin variant were fit to a two-state transition model with the heat capacities of the native and the unfolded states, the enthalpy of unfolding,  $\Delta H_{cal}$ , heat capacity of unfolding,  $\Delta C_p$ , and transition temperature,  $T_m$ , as independent variables (12). The standard thermodynamic functions under reference conditions were calculated as:

$$\Delta H_{cal}(T) = \Delta H(T_m) + \Delta C_p \cdot (T - T_m), \quad [2]$$

$$\Delta S(T) = \Delta S(T_m) + \Delta C_p \cdot \ln(T/T_m) = \frac{\Delta H_{cal}(T_m)}{T_m} + \Delta C_p \cdot \ln(T/T_m), \quad [3]$$

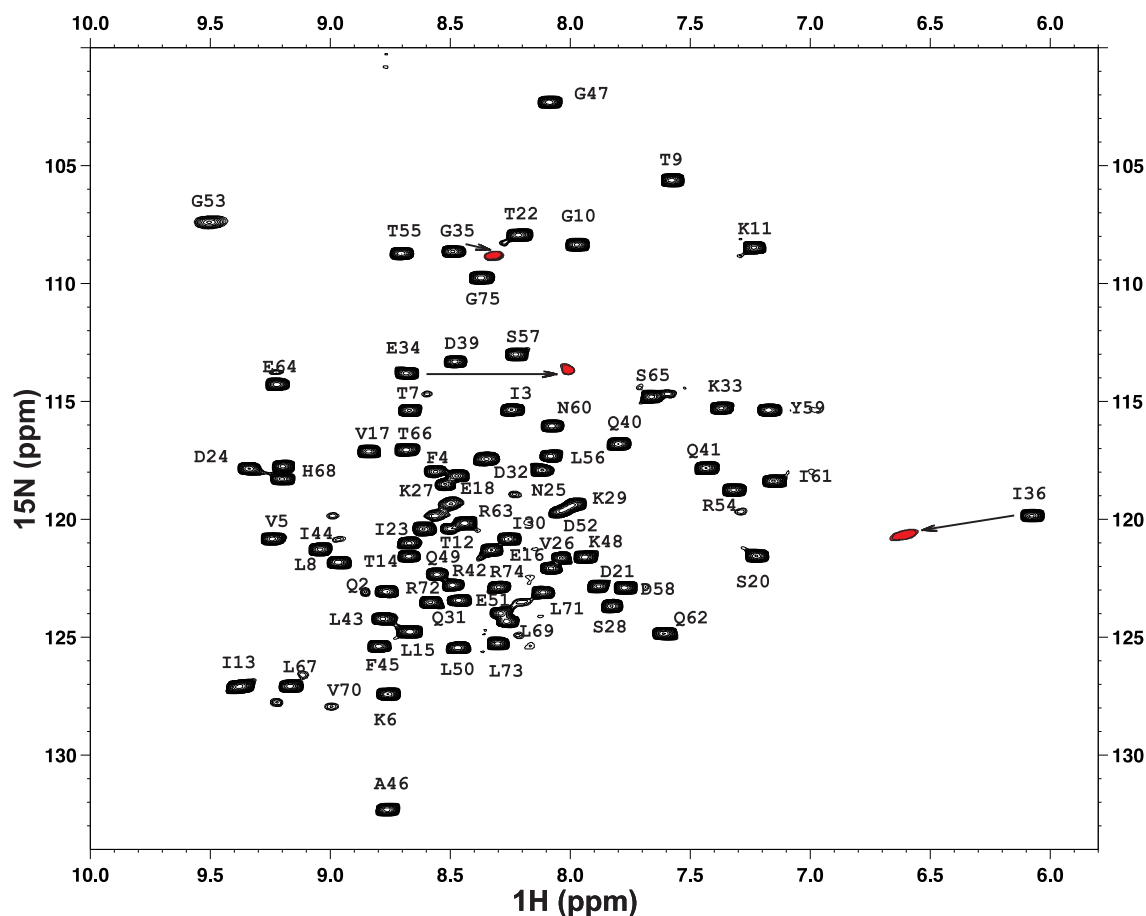
$$\Delta G(T) = (T_m - T) \cdot (\Delta H_{cal}(T_m)/T_m - \Delta C_p) - T \cdot \Delta C_p \cdot \ln(T/T_m), \quad [4]$$

where  $\Delta S(T)$  and  $\Delta G(T)$  are the entropy and Gibbs energy functions of a protein, respectively. As a control for the correctness of the experiments we compared the stability of the G35D variant in the background of true yeast ubiquitin. We found the same effect of this substitution on the stability ( $\Delta T_m = 9^\circ\text{C}$ ;  $\Delta\Delta G = -5.8$  kJ/mol) as in the case of the pseudo-WT protein used for the major part of this study.

**NMR Spectroscopy.**  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear sequential quantum correlation spectra were acquired on a Bruker spectrometer (DRX-600) in 95%  $\text{H}_2\text{O}/5\%$   $^2\text{H}_2\text{O}$  as described (14). The solutions were  $\approx 1$  mM for the WT-ubiquitin and 0.2 mM G35V-ubiquitin and contained 5 mM acetate buffer, pH 5.0, with added  $^2\text{H}_2\text{O}$  to yield a 90%  $\text{H}_2\text{O}/10\%$   $^2\text{H}_2\text{O}$  solution.

## Results

**Effects of Amino Acid Substitutions at the C' Position on the Stability of Ubiquitin.** Table 1 summarizes the results of calorimetric studies of 17 variants of ubiquitin with the single-site amino acid substitutions at the C' position of the C-capping box in ubiquitin. Comparison is done under a fixed set of conditions, 50°C and pH 3.0. This pH was selected because at this pH the transition temperature for the WT protein is 60°C (12) and thus more stable variants will have the unfolding transition completed before reaching the upper limit (110°C) of the operating temperature range of the DSC instrument. The reference temperature was set at 50°C because it is an average temperature of



**Fig. 1.**  $^1\text{H}$ - $^{15}\text{N}$  chemical shift correlation spectrum of the WT variant of yeast ubiquitin with the partial assignments of the backbone  $^{15}\text{N}$  amide protons (14). The resonances that are significantly different in G35V variant are shown in red with the arrow indicating which resonances have shifted.

unfolding of all ubiquitin variants studied at pH 3.0, and thus the extrapolation error over the entire set of variants is minimal.

The WT protein that has Gly at position 35 is the most stable protein. All other 16 aa residues at the C' position destabilize the ubiquitin molecule to a different degree. The G35V variant is most dramatically destabilized, with a 32°C lower transition temperature. Other strongly destabilizing variants are G35I ( $\Delta T_m = -23^\circ\text{C}$ ) and G35T ( $\Delta T_m = -18^\circ\text{C}$ ), which are also  $\beta$ -branched amino acid residues.

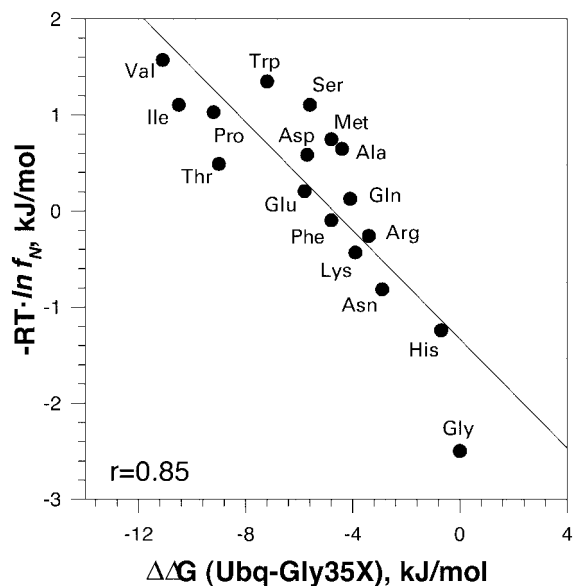
#### Effect of Val Substitution at C' Position on the Structure of Ubiquitin.

Because the G35V variant showed the lowest stability relative to other position 35-substituted proteins we addressed the question of structural similarity between the WT and G35V variant. We used NMR spectroscopy to characterize the structure of the G35V variant. Because this variant is significantly less stable than the WT, the  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear sequential quantum correlation spectrum, despite having intense cross-peaks, was characterized by a broader resonance dispersion relative to the WT protein. Nevertheless, the overall position of resonances in the G35V variant was very similar to that of the WT. The exceptions were found in the cross-peaks corresponding to the residue at the substitution site (C') and the residues before (Ccap-E34) and after (C'-I36). As can be seen from Fig. 1, all three residues in the G35V variant show changes largely along the  $^1\text{H}$ -axes. The changes appear to be consistent with a small rearrangement of the peptide backbone that alters the relative position of the amide protons (19). The dihedral angles for the C' position of

ubiquitin in the WT protein are  $\phi = 81.2^\circ$  and  $\psi = 5.3^\circ$ , well outside the allowable positive dihedral angles for any nonglycine residue. It is conceivable that the dihedral angles of the C' residue in the G35V variant are moving closer to the allowable  $\alpha_L$  region, leading to the changes in the resonances of the adjacent in-sequence residues. Such changes have been previously observed for T4 lysozyme (20), hen egg lysozyme (19), and rop (21). Overall it is apparent that there is no significant difference between the most stable and the least stable position 35 ubiquitin variant and thus we can assume that the structures of the remaining position 35 ubiquitin variants (with G35P to a lesser degree) are also structurally similar to the WT.

#### Possible Secondary Effect at the C' Position that Might Bias the Propensity.

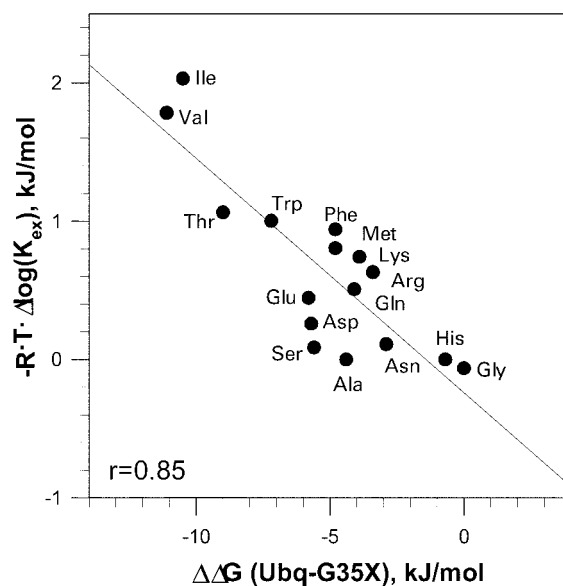
The C' position of the C-capping box in ubiquitin is fully solvent exposed. It was believed for a long time that residues at solvent exposed positions do not have a significant effect on the protein stability (22). This belief was based on the assumption that the only interactions that the surface residue can have are with the solvent, and because in the unfolded state the surface residue interacts with the same solvent, the net change in the interactions between native and unfolded states is zero. However, recently it has been shown that at least one type of interaction, the charge-charge interactions between surface residues, can be an important factor for modulating the protein stability (23). Because some of the C' position variants of ubiquitin incorporate acidic (Asp, Glu) and basic (His, Lys, Arg) residues instead of uncharged glycine, it is important to estimate the possible effects of incorporating charged residues at the C'



**Fig. 2.** Correlation of Gibbs energy change of the C' ubiquitin variants relative to WT  $\Delta\Delta G(\text{Ubq-Gly35X})$  with the normalized probabilities of different residues in the C' position of an  $\alpha$ -helix (7). Normalized probabilities,  $f_N$ , were converted into Gibbs energy as  $-R\cdot T\cdot\ln f_N$ , using 298 K as a reference temperature. Linear correlation coefficient is 0.85.

position on the stability of the ubiquitin molecule. We have shown recently that simple electrostatic calculations can reliably predict the effect of substitutions of solvent exposed charged residues in several proteins including ubiquitin (11, 23). We thus performed similar calculations on position 35 of ubiquitin. It appears that substitution of Gly-35 for either acidic or basic residues at this position has a negligible effect on the charge-charge interactions (the expected  $\Delta\Delta G$  caused by changes in charge-charge interactions between WT and G35E, G35D, G35R, or G35K is less than 0.3 kJ/mol). Thus under the experimental conditions the side chains of Asp and Glu residues are not expected to have perturbed pKa and will be largely protonated. Taken together, it is unlikely that the charge-charge interactions will bias the intrinsic thermodynamic propensities of the residues at the C' position of the C-capping box in ubiquitin. This, of course, does not mean that there will not be other electrostatic perturbations upon substitution with the charged residue at this position. One can expect non-negligible interactions of charged side chains at C' with the helix dipole (24, 25). However, these interactions will exist in all C-capping motifs and thus are considered to be one of the factors defining thermodynamic propensities.

**Correlation with the Statistical Distributions.** Fig. 2 shows the correlation of the thermodynamic propensities of different amino acid residues at the C' position in ubiquitin and statistical analysis of the probabilities of amino acid residues at the C' position. The latter data were compiled by the Rose group and are based on 1,316 helices from 274 known protein structures (7). The correlation coefficient between the two data sets is 0.85, which can be considered to be quite good, keeping in mind that the residue distribution in the known protein structures is not filtered for the additional interactions that involve the residues at the C' position of  $\alpha$ -helices. We thus can assume at least as a first approximation that the propensities observed in the C' position of the  $\alpha$ -helix in ubiquitin are to a large degree defined by some intrinsic properties. Additional support for this comes from the fact that the thermodynamic propensities obtained at the C' position of ubiquitin have good correlation (correlation



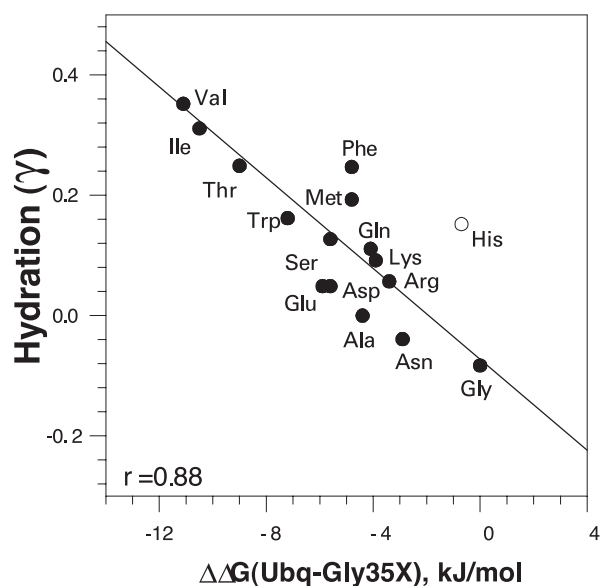
**Fig. 3.** Correlation of Gibbs energy change of the C' ubiquitin variants relative to WT  $\Delta\Delta G(\text{Ubq-Gly35X})$  with the rates of hydrogen exchange of amide protons in short model peptides. The average of base and acid catalysis of the H/D exchange rates relative to Ala, reported by Bai *et al.* (29) for the left (from the side chain) amide proton,  $\log K_{\text{ex}}(X) - \log K_{\text{ex}}(\text{Ala})$ , was converted to the relative Gibbs energy as  $\Delta\Delta G = -R\cdot T\cdot\log[K_{\text{ex}}(X)/K_{\text{ex}}(\text{Ala})]$  using 298 K as a reference temperature. Linear correlation coefficient is 0.85.

coefficients 0.94 and 0.78) with the effects of amino acid substitutions at the C' position of two  $\alpha$ -helices in barnase (8). It must be noted that barnase double-mutant cycle experiments revealed certain bias due to the secondary interactions of residues at C' with the residues close in sequence (8, 26). Our data on G35D are in excellent agreement with the effects of similar substitution in T4-lysozyme for which Gray and Matthews (27) reported a decrease in stability of  $-5.9$  kJ/mol at pH 2.0, comparable with  $-5.7$  kJ/mol in ubiquitin (Table 1). The glycine residue in T4 lysozyme is also in the C' position and has positive dihedrals  $\phi = 76.3^\circ$  and  $\psi = 19.1^\circ$ .

## Discussion

We have shown that the amino acid substitutions at the C' position of  $\alpha$ -helix in ubiquitin have rather dramatic effects on the stability of this protein. What are the factors that define this thermodynamic propensity scale? What makes Gly at this position thermodynamically the most favorable residue? Glycine does not have a side chain and its higher flexibility in the unfolded state means a larger loss of entropy upon folding and thus entropically any substitution of Gly must lead to an increase in stability (28). We observe, however, that all residues at position 35 render a protein with lower stability than the Gly-containing WT. We propose that the addition of the side chain at the solvent exposed Gly would reduce the hydration of the protein backbone and the effect of burying the polar peptide backbone would lead to destabilization. Three different facts support this explanation.

**Correlation Between the Hydrogen Exchange Rates of Amide Proton and Thermodynamic Propensity at C' Position.** Fig. 3 shows the plot of  $\Delta\Delta G$  of the residues at the C' position in ubiquitin and the  $\Delta\Delta G$  obtained from the amide proton exchange rates of different amino acid residues in a short model peptide (29). The original exchange rates were converted into Gibbs energy by using the standard relationship  $\Delta\Delta G = -R\cdot T\cdot\Delta\log(K_{\text{ex}})$ . One needs to keep in mind that the hydrogen exchange in a short peptide



**Fig. 4.** Correlation of Gibbs energy change of the C' ubiquitin variants relative to WT  $\Delta\Delta G(\text{Ubq-Gly35X})$  with the parameter  $\gamma$  (31), which represents relative hydration of the peptide backbone. Linear correlation coefficient is 0.88.

occurs very fast and is thus difficult to measure, leading to a relatively large error. Nevertheless, the two data sets clearly correlate ( $r = 0.85$ ). The difference in the amide exchange rates for different amino acid residues was interpreted in terms of steric blocking effects of the side chain on the accessibility of the backbone amide for the H/D exchange (30). In other words, the difference in amide exchange rates is defined by the accessibility to the solvent. Because this correlates with the thermodynamic propensities of the residues at the C' position of the  $\alpha$ -helix in ubiquitin, we can argue that the same effect, i.e., solvent accessibility of the peptide backbone, contributes to the thermodynamic propensity scale at the C' position.

**Correlation Between the Backbone Hydration Parameter and the Thermodynamic Propensity at C' Position.** Energetics of the backbone hydration have been evaluated by Avbelj (31) using a computational approach. The plot of  $\Delta\Delta G$  of the residues at the C' position in ubiquitin and the parameter  $\gamma$  that describes the hydration is shown in Fig. 4. The overall correlation coefficient is 0.85. However, the largest challenge in the calculations that introduces considerable uncertainty is usually the presence of charged amino acid residues. For example, if the histidine residue is omitted from the analysis, the correlation coefficient increases to 0.88. These correlations again can be interpreted as the difference in the hydration of the peptide backbone is caused by the difference in the steric blocking by different side chains, thus contributing to the thermodynamic propensity of the residues at the C' position of the  $\alpha$ -helix.

**The Changes in the Thermodynamic Parameters of Unfolding of the C' Ubiquitin Variant Are Consistent with the Thermodynamics of Dehydration of the Peptide Backbone.** Thermodynamic parameters of hydration of the peptide backbone have been previously evaluated by using model compound data (32). It was shown that the hydration of the peptide backbone is accompanied by negative enthalpy and negative entropy changes (33). The changes in entropy and enthalpy, however, are such that the Gibbs energy of hydration is negative (33). These characteristic properties are seen in the  $\Delta\Delta H(50^\circ\text{C})$  and  $-T\Delta\Delta S(50^\circ\text{C})$  values of the ubiquitin variants (Table 1). Indeed, all C' ubiquitin variants have lower

enthalpy of unfolding than the WT protein in which Gly in C' provides the largest backbone exposure. The changes in enthalpy are largest for the least stable C' variants, all of which have  $\beta$ -branched side chains, and thus are expected to have larger steric blocking effects on the peptide backbone from water (30). The changes in the entropic factor on substitutions at the C' position follow a similar pattern, i.e., in all cases the changes in the  $-T\Delta\Delta S(50^\circ\text{C})$  are significant. Again, the increase in the entropic term is the largest for  $\beta$ -branched amino acid residues Thr, Ile, and Val. The loss in enthalpy is nevertheless larger than the gain in entropy leading to an overall decrease in  $\Delta\Delta G(50^\circ\text{C})$ . This pattern is clearly consistent with the thermodynamic data on the hydration of the peptide backbone and in particular with the enthalpy change upon hydration of the peptide backbone. Analysis of the experimental data on the transfer of different amides into water gives an estimate for  $\Delta H_{\text{hyd}}(50^\circ\text{C}) = -61$  kJ/mol (33), which is consistent with the electrostatic calculations on *N*-methylacetamide that give a value of  $-51$  kJ/mol (34). Both of these estimates are higher than the result of the calculations done on a more realistic model, an Ala-based model peptide in extended conformation,  $-33$  kJ/mol (34). For comparison, the observed averaged change in  $\Delta\Delta H(50^\circ\text{C})$  upon substitution of Gly to any other residue is  $-26 \pm 12$  kJ/mol (Table 1). It is important to note the rather small spread in the  $\Delta\Delta H(50^\circ\text{C})$  values for nonglycine residues. This observation is consistent with the estimates for the peptide backbone burial in different amino acid residues relative to Gly. Calculations of Creamer *et al.* (35) show that any residue other than Gly have a very similar amount of surface area (ASA) exposed (average  $\text{ASA} = 25 \pm 2 \text{ \AA}^2$  with the spread from 20 to 29  $\text{ \AA}^2$ ). In contrast the Gly residues has 65  $\text{ \AA}^2$  of backbone ASA exposed (35). If indeed hydration enthalpy can be scaled to the surface area, the predicted decrease in the enthalpy caused by the difference in hydration upon burial of peptide backbone in Gly and non-Gly residues will be

$$\Delta\Delta H_{\text{hyd}}(50^\circ\text{C}) = \frac{-61 \text{ kJ/mol}}{65 \text{ \AA}^2} \times (65 \text{ \AA}^2 - 25 \text{ \AA}^2) = -37 \text{ kJ/mol},$$

which corresponds reasonably well to the experimentally observed changes in  $\Delta\Delta H(50^\circ\text{C})$ . It is remarkable that although changes in the enthalpies of hydration as calculated from the surface area changes for the peptide backbone averaged over all variants closely predict the average changes in the enthalpies of unfolding of the ubiquitin variants, the correlation between  $\Delta\Delta H(50^\circ\text{C})$  and  $\Delta\text{ASA}$  for the peptide backbone for individual variants, even excluding glycine, is poor ( $r = 0.55$ , data not shown). This finding is possibly caused by two factors: (i) surface area of peptide backbone is very small and thus is prone to the relatively large errors because of the small variations in the conformation, and (ii) the enthalpy of hydration does not scale well with the surface area when the changes in the ASA are very small (see ref. 34 for the theoretical evidence of this effect). The decrease in the entropy of unfolding of ubiquitin variants is also consistent with the idea of significant contribution from the dehydration of polar peptide backbone, although by default the other factors such as side-chain configurational entropy also will be in play.

Thus it appears that the decrease in the stability of the C' ubiquitin variants is largely enthalpic, which is consistent with the changes in thermodynamic properties upon dehydration of the polar peptide backbone. Obviously the dehydration of the peptide backbone is not the only parameter that contributes to the observed differences in  $\Delta\Delta G$  values. Other factors such as burial of the  $C_\beta$  atom (e.g., ref. 36), difference in the intrinsic configuration entropy (e.g., ref. 37), and side-chain charge or side-chain dipole interactions with the helix dipole (e.g., ref. 38) will contribute as well. Nevertheless, it seems that the major

factor in the thermodynamic propensity at the C' position is defined by the dehydration of the peptide backbone. This is not the first time that the specific interactions of the peptide backbone with the solvent have been implicated in the modulation of protein stability. Interactions between water and peptide backbone were shown to be important determinants of helix propensities (39). The osmophobic effect is also believed to affect protein stability (40), primarily because of the unfavorable solvation of the peptide backbone by the osmolytes.

**Concluding Remarks.** In this article we have established the propensity scale for the C' position of the C-capping box of  $\alpha$ -helices. Analysis of thermodynamic parameters and their

correlation with some structural properties suggests that dehydration of the peptide backbone plays a significant role in defining these propensities, thus underlying the fundamental importance of water for helix formation. In addition, these results can be directly used in protein engineering and design: this propensity scale provides a tool for modulation of protein stability by amino acid substitution at the C' position of  $\alpha$ -helices.

We thank Chris Falzone for the NMR experiments and George Rose for thoughtful comments and Marimar Lopez for discussions of the manuscript. This work was supported by National Institutes of Health Grant GM54537 (to G.I.M.).

1. Pauling, L., Corey, R. B. & Branson, H. R. (1951) *Proc. Natl Acad. Sci. USA* **37**, 205–210.
2. Richardson, J. S. & Richardson, D. C. (1988) *Science* **240**, 1648–1652.
3. Presta, L. G. & Rose, G. D. (1988) *Science* **240**, 1632–1641.
4. Aurora, R., Srinivasan, R. & Rose, G. D. (1994) *Science* **264**, 1126–1130.
5. Harper, E. T. & Rose, G. D. (1993) *Biochemistry* **32**, 7605–7609.
6. Preissner, R. & Bork, P. (1991) *Biochem. Biophys. Res. Commun.* **180**, 660–665.
7. Aurora, R. & Rose, G. D. (1998) *Protein Sci.* **7**, 21–38.
8. Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R. (1992) *J. Mol. Biol.* **227**, 544–559.
9. Milner-White, E. J. (1988) *J. Mol. Biol.* **199**, 503–511.
10. Bork, P. & Preissner, R. (1991) *Biochem. Biophys. Res. Commun.* **180**, 666–672.
11. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999) *Biochemistry* **38**, 16419–16423.
12. Thomas, S. T. & Makhatadze, G. I. (2000) *Biochemistry* **39**, 10275–10283.
13. Makhatadze, G. I., Lopez, M. M., Richardson, J. M., 3rd & Thomas, S. T. (1998) *Protein Sci.* **7**, 689–697.
14. Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. (2001) *Protein Sci.* **10**, 1343–1352.
15. Makhatadze, G. I. (1998) in *Current Protocols in Protein Chemistry*, eds. Coligan, J. E., Dunn, B. M., Ploegh, H. L., Speicher, D. W. & Wingfield, P. T. (Wiley, New York), Vol. 2, pp. 7.9.1–7.9.14.
16. Makhatadze, G. I., Medvedkin, V. N. & Privalov, P. L. (1990) *Biopolymers* **30**, 1001–1010.
17. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995) *Protein Sci.* **4**, 2411–2423.
18. Winder, A. F. & Gent, W. L. (1971) *Biopolymers* **10**, 1243–1251.
19. Masumoto, K., Ueda, T., Motoshima, H. & Imoto, T. (2000) *Protein Eng.* **13**, 691–695.
20. Nicholson, H., Soderlind, E., Tronrud, D. E. & Matthews, B. W. (1989) *J. Mol. Biol.* **210**, 181–193.
21. Predki, P. F., Agrawal, V., Brunger, A. T. & Regan, L. (1996) *Nat. Struct. Biol.* **3**, 54–58.
22. Jaenicke, R. (1999) *Prog. Biophys. Mol. Biol.* **71**, 155–241.
23. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001) *Trends Biotechnol.* **19**, 132–1354.
24. Matthews, B. W. (1993) *Annu. Rev. Biochem.* **62**, 139–160.
25. Rohl, C. A. & Baldwin, R. L. (1998) *Methods Enzymol.* **295**, 1–26.
26. Sancho, J., Serrano, L. & Fersht, A. R. (1992) *Biochemistry* **31**, 2253–2258.
27. Gray, T. M. & Matthews, B. W. (1987) *J. Biol. Chem.* **262**, 16858–16864.
28. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6663–6667.
29. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993) *Proteins* **17**, 75–86.
30. Bai, Y. & Englander, S. W. (1994) *Proteins* **18**, 262–266.
31. Avbelj, F. (2000) *J. Mol. Biol.* **300**, 1335–1359.
32. Makhatadze, G. I. & Privalov, P. L. (1993) *J. Mol. Biol.* **232**, 639–659.
33. Makhatadze, G. I. & Privalov, P. L. (1995) *Adv. Protein Chem.* **47**, 307–425.
34. Avbelj, F., Luo, P. & Baldwin, R. L. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10786–10791. (First Published September 12, 2000; 10.1073/pnas.200343197)
35. Creamer, T. P., Srinivasan, R. & Rose, G. D. (1997) *Biochemistry* **36**, 2832–2835.
36. Blaber, M., Zhang, X. J. & Matthews, B. W. (1993) *Science* **260**, 1637–1640.
37. Creamer, T. P. & Rose, G. D. (1994) *Proteins* **19**, 85–97.
38. Marqusee, S. & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8898–8902.
39. Luo, P. & Baldwin, R. L. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4930–4935.
40. Liu, Y. & Bolen, D. W. (1995) *Biochemistry* **34**, 12884–12891.