J|A|C|S
COMMUNICATIONS

Published on Web 09/23/2004

# Exploiting the Right Side of the Ramachandran Plot: Substitution of Glycines by D-Alanine Can Significantly Increase Protein Stability

Burcu Anil,[†] Benben Song,[†] Yuefeng Tang,[†] and Daniel P. Raleigh*[,†,‡]

*Department of Chemistry and Graduate Programs in Biophysics and in Biochemistry and Structural Biology, State University of New York at Stony Brook, Stony Brook, New York 11794*

Received May 17, 2004; E-mail: draleigh@notes.cc.sunysb.edu

A major goal of protein engineering is the enhancement of protein stability. A variety of rational design approaches and clever selection strategies have been employed, but the results are often mixed.[1] Recent advances in solid-phase peptide synthesis and expressed protein ligation have opened the door to the efficient incorporation of noncoded amino acids into globular proteins.[2] These methodological advances provide access to a new range of potential substitutions that can be utilized to rationally enhance protein stability.[2e,3] Here we demonstrate that a significant enhancement in stability can be achieved by replacing glycines that have positive $\varphi$-angles with D-Ala.

Glycines have been the target of mutations designed to increase stability by reducing the entropy of the unfolded state.[4] However, glycine often adopts conformations with positive $\varphi$-angles and substitution by an L-amino acid at such a site can introduce significant strain into the structure.[5] Conversely, non-glycine residues with positive $\varphi$-angles have been substituted with glycine in an attempt to relieve strain and increase stability. In this case the reduction in strain in the folded state can be partially offset by an unfavorable increase in unfolded state entropy as well as more specific folded state effects. Typically, a quite modest increase in stability or even a decrease is observed.[4a,6] An attractive alternative strategy is to replace residues that have positive values of $\varphi$ with D-amino acids. Replacement of an L-amino acid that adopts a positive $\varphi$-angle should enhance protein stability by reducing strain, provided no steric clashes are introduced. However, non-glycine residues with positive $\varphi$ are relatively rare. The most general strategy involves targeting glycines with positive $\varphi$-angles.[7] Replacement of such a residue by a D-amino acid should stabilize the protein by decreasing the entropy of the unfolded state and possibly by more specific native-state effects. There are several reports of the use of D-amino acids to stabilize turns in small $\beta$-hairpins or designed miniature proteins, but no thermodynamic data are available. There are also no reports on the energetic consequences of replacing specific residues in globular proteins with D-amino acids.[8] Here we describe the stabilization of proteins by the site-specific incorporation of D-Ala.

As model systems we chose two globular proteins: an $\alpha$-$\beta$ protein, the N-terminal domain of L9 (NTL9), and an all-$\alpha$ structure, the UBA domain. NTL9 is a 56-residue protein that has been shown to fold via a two-state mechanism.[9] Two variants of NTL9 with single D-Ala substitutions were prepared using solid-phase peptide synthesis. G24 and G34 were chosen for substitution with D-Ala. The mutants are referred to as G24D-Ala and G34D-Ala. G24 with $\varphi = 121°$ is located in the loop between the second $\beta$-strand and the central $\alpha$-helix before the short helix. G34 is located in the loop connecting the central helix and the third $\beta$-strand, and has $\varphi = 100°$ (Figure 1). Replacement with D-Ala does not introduce any steric clash in either of these sites, and the new methyl group is exposed to solvent. The UBA domain is a three-helical domain derived from the human homologue of Rad23A.[11]

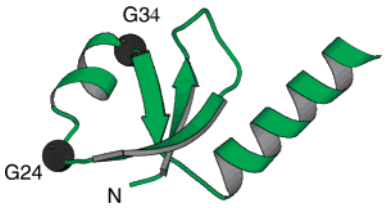


***Figure 1.*** Ribbon diagram of NTL9 (PDB code 1DIV) showing G34 and G24; the N-terminus is labeled. The diagram was created using Molscript.[10]

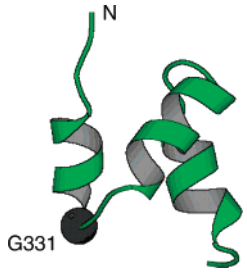


***Figure 2.*** Ribbon diagram of UBA domain (PDB code 1UBA) showing G331; the N-terminus is labeled. The diagram was created using Molscript.[10]

Gly 331 of the UBA domain (residues are numbered according to the full-length protein) is a helix-capping residue located at the end of the first helix and has $\varphi = 121°$ (Figure 2). A variant of UBA with a D-Ala substitution at G331 has been prepared and is denoted G331D-Ala. Modeling indicates that the new methyl group projects into the solvent and makes no interactions. The structures of all three proteins are similar to the respective wild-type as judged by near- and far-UV CD and NMR. The NMR spectra of both NTL9 and UBA exhibit a set of distinctive ring-current-shifted resonances, and these are observed in the mutants. NTL9 also contains a set of $C_\alpha H$ resonances downfield of water, and they are present in the mutants.

The free energy of folding is $-4.17$ kcal/mol for wild-type NTL9, $-6.04$ kcal/mol for G34D-Ala, and $-5.47$ kcal/mol for G24D-Ala, determined by GuHCl denaturation. The stability of G34D-Ala is noticeably higher than that of G24D-Ala. Consequently, we also determined its stability by amide H/D exchange measurements; the value $-6.1$ kcal/mol is in good agreement with the denaturation experiments. Substitution of G34 with D-Ala increases $T_m$ by 7 °C. Substitution of G331 with D-Ala in the UBA domain also resulted in an increase in the stability of that protein, albeit to a lesser extent. The $T_m$ is increased by 9 °C. The free energy of folding of wild-type UBA domain is $-1.34$ kcal/mol. The G331D-Ala variant is $-0.6$ to $-0.7$ kcal/mol more stable (Figure 3). The thermodynamic parameters are summarized in Table 1.

[†] Department of Chemistry.
[‡] Graduate Program in Biophysics and Graduate Program in Biochemistry and Structural Biology.


**10.1021/ja047119i CCC: $27.50** 

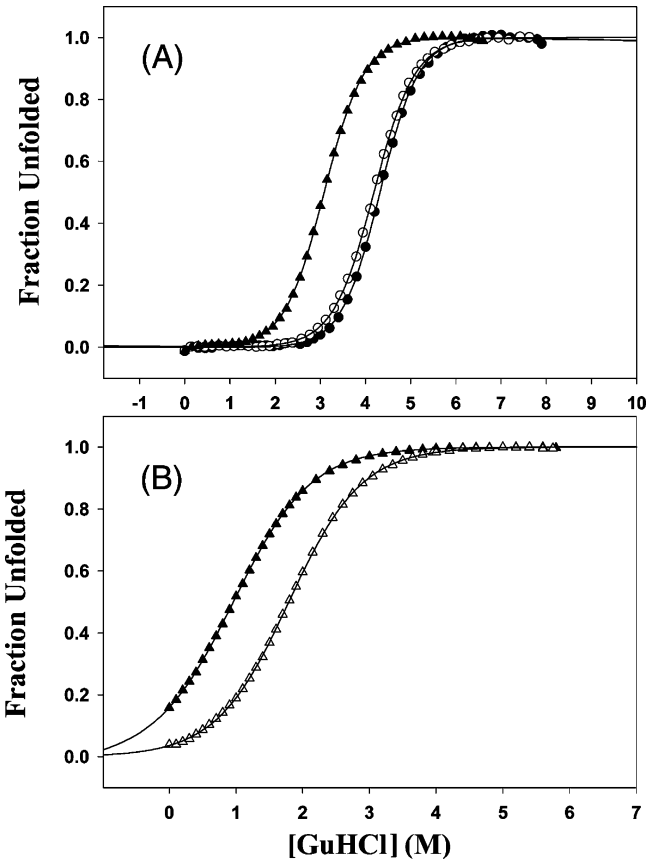


***Figure 3.*** Plots of fraction unfolded versus concentration of denaturant. (A) NTL9 variants (▲) wild-type, (○) G24D-Ala, (●) G34D-Ala. (B) UBA variants (▲) wild-type, (△) G331D-Ala. The solid lines represent the best fits to data.

***Table 1.*** Data for NTL9 and UBA Variants

| protein | $\Delta G_f°(H_2O)$ (kcal mol$^{-1}$) | $m$ (kcal mol$^{-1}$ M$^{-1}$) | $C_m$ (M) | $T_m$ (°C) |
|---|---|---|---|---|
| NTL9 | −4.17 ± 0.06 | 1.35 ± 0.02 | 3.08 ± 0.01 | 79.8 |
| G34D-Ala | −6.04 ± 0.10 | 1.39 ± 0.02 | 4.34 ± 0.01 | 87.1 |
| G24D-Ala | −5.47 ± 0.06 | 1.33 ± 0.02 | 4.11 ± 0.01 | *b* |
| UBA | −1.34 ± 0.06 | 1.12 ± 0.01 | 1.19 ± 0.01 | 65.2 |
| G331D-Ala | −1.94 ± 0.03 | 1.08 ± 0.01 | 1.79 ± 0.01 | 73.7 |

[a] Data for NTL9 and variants were obtained in pH 5.45, 100 mM NaCl buffer. Data for UBA variants were obtained in pH 6.5, 50 mM $K_2HPO_4$, 1 mM DTT buffer. GuHCl denaturations were carried out at 25 °C. Uncertainties represent the standard errors of the fit. $C_m$ is the midpoint and $m$ is the constant of proportionality relating $\Delta G°$ and denaturant concentration. [b] The $T_m$ for G24D-Ala could not be accurately estimated because of lack of a post-transition baseline.

We conclude that D-Ala substitution at a glycine site with a positive $\varphi$-angle is a viable strategy for rationally increasing protein stability, as long as the additional methyl group does not introduce any steric clashes. The extent of stabilization depends on the local details of the site and could also depend on unfolded state structure. It may not be as large if the region containing the glycine of interest is already structured in the unfolded state or if the site is very flexible in the native state. The cost of solvating the new methyl group could also contribute. In the examples here the methyl is exposed, so differences in solvation between folded and unfolded states should be less important. The high frequency of glycines with $\varphi > 0°$ in proteins suggests that substitution with D-Ala at glycine sites meeting the criteria mentioned above could be a general strategy.

**Acknowledgment.** We thank Profs. Jane Richardson and Carlos Simmerling for helpful discussions. This work was supported by NIH Grants GM54233 and GM70941.

**Supporting Information Available:** Plots of CD signal at 222 nm vs denaturant concentration showing the direct fit to the experimental data; CD spectra and plots of Cα $^1$H chemical shifts for all proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) (a) Matthews, B. W. *Adv. Protein Chem.* **1995**, *46*, 249−278. (b) Wintrode, P. L.; Arnold, F. H. *Adv. Protein Chem.* **2000**, *55*, 161−225. (c) Malakauskas, S. M.; Mayo, S. L. *Nat. Struct. Biol.* **1998**, *5*, 470−475. (d) Martin, A.; Sieber, V.; Schmid F. X. *J. Mol. Biol.* **2001**, *309*, 717−726. (e) van den Burg, B.; Eijsink, V. G. *Curr. Opin. Biotechnol.* **2002**, *13*, 333−337. (f) Spector, S.; Wang, M.; Carp, S. A.; Robblee, J.; Hendsch, Z. S.; Fairman, R.; Tidor, B.; Raleigh, D. P. *Biochemistry* **2000**, *39*, 872−879. (g) Sanchez-Ruiz, J. M.; Makhatadze, G. I. *Trends. Biotechnol.* **2001**, *19*, 132−135.

(2) (a) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. *Int. J. Pept. Protein Res.* **1992**, *40*, 180−193. (b) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779. (c) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249−289.

(3) (a) Horng, J. C.; Raleigh, D. P. *J. Am. Chem. Soc.* **2003**, *125*, 9286−9287. (b) Arnold, U.; Hinderaker, M. P.; Köditz, J.; Golbik, R.; Ulbrich-Hofmann, R.; Raines, R. T. *J. Am. Chem. Soc.* **2003**, *125*, 7500−7501. (c) Bilgicer, B.; Fichera, A.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 4393−4399. (d) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A., III; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790−2796. (e) Arnold, U.; Hinderaker, M. P.; Nilsson, B. L.; Huck, B. R.; Gellman, S. H.; Raines, R. T. *J. Am. Chem. Soc.* **2002**, *124*, 8522−8523.

(4) (a) Matthews, B. W.; Nicholson, H.; Becktel, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6663−6667. (b) Ganter, C.; Plückthun, A. *Biochemistry* **1990**, *29*, 9395−9402. (c) Stites, W. E.; Pranata, J. *Proteins* **1995**, *22*, 132−140.

(5) (a) Stites, W. E.; Meeker, A. K.; Shortle, D. *J. Mol. Biol.* **1994**, *235*, 27−32. (b) Takano, K.; Yamagata, Y.; Yutani, K. *Proteins* **2001**, *45*, 274−280. (c) Kim, J.; Brych, S. R.; Lee, J.; Logan, T. M.; Blaber, M. *J. Mol. Biol.* **2003**, *328*, 951−961.

(6) (a) Karplus, P. A. *Protein Sci.* **1996**, *5*, 1406−1420. (b) Nicholson, H.; Soderlind, E.; Tronrud, D. E.; Matthews, B. W. *J. Biol. Chem.* **1989**, *210*, 181−193.

(7) (a) Chakrabarti, P.; Pal, D. *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1−102. (b) Hovmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr.* **2002**, *D58*, 768−776. (c) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437−450.

(8) (a) Imperiali, B.; Moats, R. A.; Fisher, S. L.; Prins, T. J. *J. Am. Chem. Soc.* **1992**, *114*, 3182−3188. (b) Haque, T. S.; Gellman, S. H. *J. Am. Chem. Soc.* **1997**, *119*, 2303−2304. (c) Richardson, J. S.; Richardson, D. C.; Tweedy, N. B.; Gernert, K. M.; Quinn, T. P.; Hecht, M. H.; Erickson, B. W.; Yan, Y.; McClain, R. D.; Donlan, M. E. *Biophys. J.* **1992**, *63*, 1185−1209. (d) Struthers, M.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342−345.

(9) Kuhlman, B.; Luisi, D. L.; Evans, P. A.; Raleigh, D. P. *J. Mol. Biol.* **1998**, *284*, 1661−1670.

(10) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.

(11) The construct used consists of residues 319−363 with Q319E and P333E substitutions. The mutations do not alter the structure but increase solubility and avoid problems with pyroglutamic acid formation. (a) Dieckmann, T.; Withers-Ward, E. S.; Jarosinski, M. A.; Liu, C. F.; Chen, I. S.; Feigon, J. *Nat. Struct. Biol.* **1998**, *5*, 1042−1047. (b) Mueller, T. D.; Feigon, J. *J. Mol. Biol.* **2002**, *319*, 1243−1255.

JA047119I