Published in final edited form as:

J Am Chem Soc. 2011 November 23; 133(46): 18750–18759. doi:10.1021/ja205609c.

Computational Design of Thermostabilizing β -Amino Acid Substitutions

Agustina Rodriguez-Granillo^{‡,1}, Srinivas Annavarapu^{‡,1}, Lei Zhang², Ronald L. Koder^{*,2}, and Vikas Nanda^{*,1}

¹Department of Biochemistry, Robert Wood Johnson Medical School, UMDNJ and the Center for Advanced Biotechnology and Medicine, Piscataway, New Jersey 08854

²Department of Physics, The City College of New York, New York, New York 10031

Abstract

Judicious incorporation of β -amino acids in engineered proteins confer many advantages such as preventing degradation by endogenous proteases, and designing novel structures and functions not accessible to homochiral polypeptides. Glycine to β -alanine substitutions at the carboxy-termini can stabilize α -helices by reducing conformational entropy. Beyond alanine, we propose additional side chain effects on the degree of stabilization conferred by β -amino acid substitutions. A detailed, molecular understanding of backbone and side chain interactions is important for developing rational, broadly applicable strategies in using β -amino acids to increase protein thermostability. Insight from structural bioinformatics combined with computational protein design can successfully guide the selection of stabilizing β -amino acid mutations. Substituting a key glycine in the Trp-Cage mini-protein with β -Gln dramatically stabilizes the fold without altering the protein backbone. Stabilities of individual substitutions can be understood in terms of the balance of intramolecular forces at both the α -helix C-terminus and throughout the protein.

Keywords

computational protein design; β -amino acid; α -helix; hydrogen bonding; molecular dynamics; thermodynamics; NMR; Trp-Cage mini-protein

1. INTRODUCTION

We hypothesize polar β -amino acids can stabilize the carboxy-termini of α -helices through side chain-to-backbone hydrogen bonds. Computational methods are developed to predict the energetic consequences of such substitutions. A major drawback of peptide therapeutics that limits their biomedical application is low thermodynamic stability, which results in weak activity and rapid degradation by endogenous proteases. Globular proteins typically maintain stable native structures through extensive tertiary contacts within hydrophobic cores. Peptides have minimal tertiary contacts, resulting in highly dynamic ensembles that significantly populate partially or fully unfolded states. Strategies to stabilize peptides focus either on enhancing the scant tertiary structure, or favoring secondary structure. Various chemical approaches have been developed to modulate protein stability such as engineering

*Corresponding Author: nanda@cabm.rutgers.edu (VN), rlkoder@gmail.com (RLK).

[‡]These authors contributed equally.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

disulfides¹, metal binding sites² or incorporating fluorine-derivatized amino acids³. Constrained oligoproline derivatives have been used as templates for helix initiation improving both folding kinetics and stability⁴. Various bio-orthogonal chemical approaches have been used to improve peptide stability by linking sidechains⁵ or replacing a backbone hydrogen bond with a carbon-carbon bond⁶.

Our understanding of the role of stereochemistry in protein structure and folding comes from seminal contributions by Ramachandran in the early sixties and seventies^{7,8}. Models of allowed dipeptide conformations based on hard-sphere sterics of atomic interactions have provided a physical basis for understanding the complex structures of natural proteins. High resolution structures of model peptides have provided some glimpse into the increased structural diversity accessible with non-natural amino acids⁹. Work on somatostatin demonstrated that incorporating D-amino acids into small peptides could dramatically enhance their biological activity¹⁰. Introducing one, or at most a few D-amino acids has been used to stabilize β -turns,^{11,12} novel folds¹³ and larger proteins¹⁴. We build on previous work and explore the use of computational design with D-amino acids to stereochemically favor native secondary structure as a strategy for improving peptide thermostability.

Typically, incorporation of D-amino acids into existing proteins and peptides will adversely impact structure and stability. Substituting an L-Ala for a D-Ala in the middle of an α -helix reduces its stability by ~ 1 kcal/mol due to steric clashes between the side chain methyl and the backbone carbonyls of proximal residues in the polypeptide chain.¹⁵ In specific contexts, a Gly to D-Ala substitution at the carboxy-terminus can improve its stability.^{16–18} Glycine plays a unique structural role at the carboxy-terminus of an α -helix, forming a network of hydrogen bonds that cap backbone carbonyls of adjacent residues while directing the chain to the next structural element. These structural 'punctuation marks' have been frequently observed in natural proteins.¹⁹ Glycine lacks a side chain, allowing it to easily assume an α_L conformation ($\phi \sim +65^\circ$, $\psi \sim +42^\circ$) which is energetically unfavorable for the other nineteen L-amino acids.⁸ By replacing Gly with D-Ala, it is possible to stably form the α_L capping structure without the entropic cost of fixing the conformation of glycine. This strategy has been applied to several proteins with mixed results. In a series of Gly to D-Ala substitutions in small proteins, it was found that stability was improved by anywhere from 0.6 to 2 kcal/mol.¹⁶ Equivalent substitutions in ubiquitin showed no increase in stability.¹⁷ In a monomeric α -helix, it was found that D-Ala had no effect, but D-Arg stabilized the helix by ~ 1 kcal/mol.²⁰ These discrepancies indicate that the role of stereochemistry in modulating helix stability is still an open problem. A systematic and rigorous exploration of stereochemical rules will be critical in developing rational and computational methods for engineering enhanced peptide stability.

In this study, we investigate the potential for D-amino acids to stabilize ends of α -helices both through reducing backbone flexibility and promoting novel side chain-to-backbone hydrogen bonds that are not accommodated by equivalent L-stereoisomers. The potential for such interactions were noted in Monte Carlo simulations of poly-LD-alanine peptides, sampling both backbone flexibility and side chain chirality.²¹ In simulations of an eleven residue poly-alanine peptide, (L-Ala)₉-D-Ala-L-Ala and (D-Ala)₉-L-Ala-D-Ala were frequently found. These corresponded to right and left-handed α -helices respectively. Interestingly, the change in chirality at the end of the sequence was due to the incorporation of C-capping motifs (Figure 1A). Note that the methyl group of alanine was proximal to exposed carbonyls of the helix terminus. We hypothesized other amino acids beside alanine might provide additional stability through additional side chain-to-backbone hydrogen bonds. An example of such an interaction was found in the high-resolution structure of an enzyme alanine racemase²², where the imidazole N δ^1 nitrogen of an L-histidine at the C' position made a hydrogen bond to the C1 carbonyl in a preceding α_L turn.

Previous work on helix capping with natural amino acids often utilized the Protein Data Bank (PDB) as a resource for identifying common interactions at the ends of helices.²³ Finding similar motifs involving D -amino acids has proved challenging, because only a minute fraction of the structures deposited in the PDB contain D -amino acids.²⁴ Of these cases, many are short cyclic peptides that do not contain helical elements. To address this issue, we implemented a 'mirror-image template' search of the PDB.²⁵ The examination of α_{L} containing structures in the PDB has traditionally been useful in exploring the structural role of side chain stereochemistry.²⁶ It was reasoned that amino acids making side chain hydrogen bonds to the backbone of left-handed turns would be the mirror-image equivalent of D -amino acids interacting with right-handed turns. By normalizing the frequencies of observed amino acids flanking left-handed turns to random expectation, several over-represented and under-represented residues at both the N and C-termini of helices were identified. Glutamine was a particularly promising candidate, both in terms of its propensity (3.7-fold overrepresented) and the presence of C' side chain to helix backbone hydrogen bonds in multiple high resolution protein structures (an example is shown in Figure 1B), suggesting that D -Gln should be stabilizing at the C-terminus of an α_{R} -helix. Conversely, asparagine had a propensity of 0.6, suggesting it was unfavorable at this position. Less than one hundred α_{L} -turns were identified in the previous study, resulting in large statistical uncertainties. It is crucial to test the validity of these motifs in an experimental system as is being done here. Furthermore, the C-termini of α -helices are often at the protein surface, making extensive interactions with solvent.²⁷ Thus, the expected contribution of such hydrogen bonds is debatable, given that formation of the interaction would come at the expense of desolvation.^{17,28}

To test the prediction that certain polar D -amino acid motifs at the C' position can stabilize an α_{R} -helix, we describe the design and characterization of D -amino acid substitutions to the Trp-cage mini-protein TC5b.²⁹ The Trp-cage was chosen as a model system for several reasons. It has been the subject of intense computational and biophysical study, making it a well understood model system for protein folding.^{30,31,32} At twenty amino acids long, the Trp-cage is readily accessible to modification with non-natural amino acids by solid-phase peptide synthesis. Its small size also makes it amenable to molecular simulation, and a significant number of molecular dynamics studies have been performed on it.^{31,33} Unlike many peptides of this size, folding is cooperative and can be described as a two-state process,^{34,35} simplifying thermodynamic interpretations of amino acid substitutions. Previous work on members of the Trp-cage family demonstrated that D -Ala substitutions could dramatically raise protein thermal stability.¹⁸ This study will present the design, thermodynamic stability, solution structure and molecular dynamics characterizations of several Trp-cage D -amino acid variants with the goal of developing computational tools for the design of thermostabilizing substitutions.

2. MATERIALS AND METHODS

2.1 Sample preparation

Peptides were synthesized using solid-phase Fmoc chemistry at the Tufts University Core Facility (<http://www.tucf.org>). N- and C-termini were acetylated and amidated respectively. Peptides were purified to 90% purity by reverse-phase HPLC, and products were verified by mass spectrometry. Sequences in this study are:

TC5b WT	NLYIQWLKD	G GPSSGRPPPS
TC5b G10A	NLYIQWLKD	A GPSSGRPPPS
TC5b G10Q	NLYIQWLKD	Q GPSSGRPPPS
TC5b G10 D A	NLYIQWLKD	D A GPSSGRPPPS

TC5b G10bQ NLYIQWLKD *D*QGPSSGRPPPS
 TC5b G10bN NLYIQWLKD *D*NGPSSGRPPPS

2.2 Thermal and chemical denaturation

2.2.1 Circular Dichroism (CD)—Peptides were dissolved in de-ionized water and concentration was measured at 280 nm using an extinction coefficient of 6990 M⁻¹cm⁻¹. Peptide solutions at 50 μM concentration were prepared by diluting peptide stocks in 10 mM phosphate buffer, pH 7.0. CD experiments were performed on an Aviv Model 400 Spectrometer (Aviv Biologicals, Lakewood, NJ). Ellipticity spectra were collected from 190 to 260 nm in 0.5 nm steps at 25 °C. 1.0 nm bandwidth, one second averaging time and one second equilibration time were used. Molar ellipticity was calculated by multiplying raw values by peptide concentration (50 μM), number of residues (20) and cell path length (1 mm).

For thermal denaturation, peptide solutions were diluted in 10 mM phosphate buffer (pH 7.0) to obtain a concentration of 50 μM. Ellipticity was measured at a wavelength of 222 nm with a bandwidth of 1.0 nm, temperature dead band of 0.15 °C, temperature equilibration time of 1.0 minute and averaging time of 2.0 seconds. Temperatures ranged from 0 °C to 80 °C in 2 °C steps. Temperature melts were conducted for six concentrations of guanidine HCl (GnHCl) for each peptide: 0.0, 1.0, 2.0, 3.0, 4.0 and 5.0 M.

Temperature melt plots obtained for multiple GnHCl concentrations were fit to folded and unfolded baselines. Heat capacity was assumed to remain constant with changes in GnHCl concentration for each Trp-cage variant. Fraction folded $f(T)$ was calculated:

$$f(T) = [\theta(T) - \theta_U(T)] / [\theta_F(T) - \theta_U(T)] \quad \text{eq. 1}$$

where θ , θ_F and θ_U are observed ellipticity, ellipticity of folded form and ellipticity of unfolded form respectively.

Free energy (ΔG), enthalpy (ΔH) and heat capacity (ΔC_p) estimates fit to the experimental data using the Gibbs-Helmholtz equation and the following relationships:

$$\Delta G = \Delta H (1 - T/T_M) - \Delta C_p ((T_M - T) + T \ln(T/T_M)) \quad \text{eq. 2a}$$

$$K = \exp(-\Delta G/(RT)) \quad \text{eq. 2b}$$

$$\alpha = K/(1+K) \quad \text{eq. 2c}$$

$$\theta_t = \alpha(\theta_F - \theta_U) + \theta_U \quad \text{eq. 2d}$$

T_M is the melting temperature, α is the model fraction folded.

2.2.2 Differential Scanning Calorimetry (DSC)—DSC experiments were carried out on an N DSC II (Applied Thermodynamics) calorimeter. The instrument was run using de-ionized water in sample and reference cell from 273°K to 373°K for at least ten scans to ensure stabilization of readings. Trp-cage peptides were dialyzed as described. Solutions of Trp-cage variants were prepared in 10 mM phosphate buffer (pH 7.0) at a concentration of 2 mg/ml. Eight hundred microliters of sample and buffer solutions were loaded into their respective cells while the instrument temperature was between 293°K and 313°K. The pressure was then slowly raised to two atmospheres. Temperature was raised from 273°K to

373°K at a rate of 1.5°K/min. An equilibration time of ten minutes was allowed between scans.

2.3 NMR Spectroscopy and Structure Calculations

Spectra were collected on a Varian Inova 600 MHz spectrometer at 4 °C. Samples were 1.5–2 mM protein in 10 mM potassium phosphate buffer pH, 7.0.

The majority of experiments were conducted using an unlabelled Trp-cage 10_bQ sample with 10% D₂O for frequency lock. ¹H-¹H-NOESY³⁶ and ¹H-¹H-TOCSY³⁷ spectra were collected using WATERGATE solvent suppression and assigned by standard methods. A DIPSI-2³⁸ spinlock was employed for the TOCSY, and mixing times of 60, 150 and 200 ms were used in the NOESY spectra. Spectral widths were 7000 Hz for both direction ¹H.

In order to resolve ambiguous assignments for some of the ¹⁵N-glutamine atoms, a sample was created with the tenth position residue uniformly labeled by ¹⁵N and ¹³C. Due to the prohibitive cost of labeled ¹⁵N-amino acids, a peptide with inverted chirality was synthesized: this peptide was synthesized with all ¹⁵N-amino acids except at position 10, where uniformly ¹⁵N, ¹³C labeled ¹⁵N-glutamine was used. A CD spectrum of the inverso-TC5bG10_bQ was consistent with a folded, mirror-image structure (Figure S1). Sensitivity-enhanced ¹³C and ¹⁵N HSQC³⁹ of this sample were collected in 100% D₂O for ¹³C and 10% D₂O for ¹⁵N. Spectral widths were 6000 Hz for ¹H, 24130 Hz for ¹³C and 2000Hz for ¹⁵N.

¹H-¹H NOESY spectra with 150ms mixing times were collected in 100% D₂O using double ¹³C isotope filters, resulting in cross peaks and distance constraints between ¹³C-bound protons only. Similar spectra collected using a ¹⁵N isotope filter for the first proton and a ¹³C isotope filter for the second in 10% D₂O, resulted in cross peaks and distance constraints between ¹⁵N-bound protons and ¹³C-bound protons only. Spectral widths were 6000Hz for ¹H.

Natural abundance sensitivity-enhanced constant time ¹³C-HSQC were performed to confirm the NOESY and TOCSY-based assignments. Spectral widths were 7000 Hz for ¹H and 24130 Hz for ¹³C.

NOE intensities were obtained from NOESY spectra. The constraints were converted to distance and used in simulated annealing protocol using the ARIA⁴⁰ and CNS⁴¹ software packages. Assignments and the constraints list are included in supplementary information.

Fifty structures were generated, and out of which 20 structures were kept. The total energy was used for sort criterion. 425 restraints was used for structure calculation, with zero violations. Out of the used restraints there were 282 unambiguous restrains, 143 ambiguous restraints. The lowest energy structure was deposited in the PDB (<http://www.rcsb.org/pdb> under ascension ID 2LDJ).

2.4 Computational Methods

2.4.1 Design of Trp-cage variants with protCAD—Computational models of Trp-cage ¹⁵N-amino acid variants were created using protCAD (protein Computer Aided Design)⁴² software. This protein design package is versatile and can be used for various design functions such as energy calculations, mutations, changes to backbone and side chain dihedrals, and modeling simulations. Variants of the Trp-cage, where ¹⁵N-amino acids replaced left-handed glycines were modeled using the structure of Trp-cage variant TC5b as a starting point (1L2Y.pdb)²⁹. All rotamers of ¹⁵N-Gln were modeled at position ten of the Trp-Cage using protCAD software and total vdW energy was computed using nonbonding united-

atom parameters from AMBER 95⁴³. Hydrogen bonds were calculated using a side chain Ne to backbone carbonyl oxygen distance cutoff of 3.5 Å (g−, gauche −; t, trans; g+, gauche +).

2.4.2 Molecular Dynamics (MD) simulations—The atomic coordinates for wild-type (WT) Trp-cage were obtained from the PDB, entry 1L2Y²⁹. The following mutations were generated in the WT structure *in silico* with the *leap* module of Amber 10⁴⁴: G10A, G10Q, G10bA, G10bN and G10bQ.

MD simulations were performed for the different protein variants using Amber 10⁴⁴. The initial structures were immersed in a pre-equilibrated truncated octahedral cell of TIP3P explicit water molecules⁴⁵, and counter ions were added to neutralize the systems. Protein atoms were described with the parm99SB^{43,46} force field parameterization. The protonation state of the titratable residues corresponded to the stable form at pH 7.0. Water molecules extended at least 12 Å from the surface of the proteins. Simulations were performed in the NPT ensemble (constant pressure of 1 atm and temperature of 300 K was maintained using the Berendsen coupling scheme⁴⁷) employing periodic boundary conditions. The SHAKE algorithm was employed to keep bonds involving hydrogen atoms at their equilibrium length⁴⁸.

The systems were optimized, heated to 300 K and equilibrated for 200 ps. The structures were then simulated until the root-mean-square deviation (RMSD) as a function of time was stable for at least 50 ns (total simulation time was more than 100 ns for each). The converged last 50 ns were used for data analysis. RMSD, root-mean-square fluctuations (RMSF), average structures and hydrogen bond frequencies were calculated for each of the systems using the *ptraj* module of Amber 10⁴⁴. Cluster analysis was performed with the program *kclust* from MMTSB⁴⁹, using RMSD as the similarity measure, with a cutoff of 2 Å.

2.4.3 Generating Potentials of Mean Force—To describe the thermodynamics of the side chain capping interaction in G10bN and G10bQ, we computed the Helmholtz free energy $A(r)$ along the distance r between the side chain of residue 10 (N8 in G10bN and Ne in G10bQ) and the backbone carbonyl of Leu7, Lys8 and Asp9, yielding a free energy profile or a potential of mean force (PMF)⁵⁰:

$$A(r) = -k_B T \ln \zeta(r) + \text{constant} \quad \text{eq. 3}$$

where k_B is the Boltzmann constant, T is the absolute temperature, and $\zeta(r)$ is the probability of the coordinate taking on a particular value r . The constant was chosen so that the most probable distribution corresponded to a free energy of zero.

2.4.4 Estimating side chain entropy—To estimate the change in entropy of side chain 10 between the folded and unfolded states in G10bN and G10bQ, two additional simulations that represented the unfolded states were performed. These consisted of the tri-peptides GNG and GQG in explicit solvent, using the same parameters as above. Both peptides were simulated for ~160 ns and the last 50 ns were used for data analysis.

The conformational entropy S of the residue 10 side chain in G10bN and G10bQ and the central amino acid in GNG and GQG was estimated using the classical definition:

$$S = -k_B \sum_i p_i \ln p_i \quad \text{eq. 4}$$

where the sum is taken over all conformational states (side chain rotamers) of the system, and p_i is the probability of being at a particular state i . The change in conformational entropy was calculated for different side chain rotamers definitions, including two binning offsets (from 0° to 360° and from 30° to 390°) and three bin sizes (30°, 60° and 90°) for χ_1 and χ_2 in Asn, and χ_1 , χ_2 and χ_3 in Gln (Table S2). The average of these values are reported in Table 3.

The vibrational entropy of the side chain in G10_N and G10_Q and the central amino acid in GNG and GQG was estimated by calculating the quasi-harmonic entropy⁵¹ using the *ptraj* module of Amber 10. In this approach the atomic fluctuation matrix is calculated as the mass-weighted covariance matrix obtained from the snapshots of the MD simulation. The structures obtained from the simulations were spatially superimposed to a common reference structure to exclude all translational and rotational motions exhibited by the molecule.⁵² The entropy was estimated every 250 ps from the last 50 ns of the MD simulations. The calculation was performed for the side chain heavy atoms of residue 10.

3. RESULTS AND DISCUSSION

3.1 Stereochemistry and Structure

The Trp-cage version 5b (TC5b) sequence²⁹ was selected as the starting point for all molecular engineering efforts. Based on the solution NMR structure, Gly 10 is in the α_L conformation at the end of a nine-residue α -helix, making it an ideal candidate for introducing D-amino acid substitutions. As a negative control, two L-amino acid variants were tested: G10A and G10Q. Although it was expected that such substitutions would be inherently destabilizing, we were motivated to test them experimentally for two reasons: to explore the importance of correct α_L favoring stereochemistry at position 10 for protein stability, and previous structural bioinformatics work uncovered several examples of glutamine in an α_L position following a right-handed helix making side chain-backbone hydrogen bonds which could presumably provide stability.

Both of L-variants showed no helical secondary structure as measured by CD at 25°C, instead adopting a fully random-coil state (Figure 2A). No thermal unfolding transition was observed by CD. In contrast, WT TC5b was clearly α -helical, with characteristic minima in the CD spectrum at 208 and 222 nm. A residue capable of easily accommodating an α_L conformation is clearly important for folding, and while L-Gln can contribute an α -helix capping hydrogen bond when introduced in TC5b, it is not sufficient to retain structure.

MD simulations on L-amino acid substituted variants were consistent with experimental observations: the structure significantly deviated from the WT TC5b throughout the simulations (Figure 3A) and backbone fluctuations significantly increased in both L variants with respect to WT (Figure 3C). Although the average structure of G10A was similar to WT in the production run (Figure 3E), averaging can mask dynamics throughout the entire simulation. Snapshots during this run significantly deviated from a WT conformation (Figure S2). In the case of G10Q, the average structure was also significantly different than WT in the production run (Figure 3E) and snapshots from the entire simulation show that key hydrophobic core interactions between the α -helix and the poly-proline stretch were compromised (Figure S3), resulting in partial unfolding.

Given the lack of structure of G10A and G10Q substitutions, we proceeded to focus on D-amino acid substitutions at this position. G10_DA was predicted to stabilize TC5b by stereochemically favoring the α_L conformation, reducing the entropic cost of folding. Based on previous analysis of the PDB, D-Gln is hypothesized to be the most favorable helix-capping residue²⁵.

TC5b G10_DA, G10_DN and G10_DQ variants were synthesized and examined for structure and stability. CD spectra of all _D-variants were consistent with significant α -helical structure, and had stronger negative ellipticity band at 208 and 222 nm relative to WT TC5b (Figure 2B). Tryptophan fluorescence emission spectra of WT and modified _D-amino acid peptides were nearly identical and consistent with a partially exposed tryptophan at position 6 (Figure S4). MD simulations show that the RMSD of the _D-variants were comparable or even smaller than WT (Figure 3B) and the average structures are closely aligned with TC5b (Figure 3F). The increase in helical content evidenced by CD agreed with the increase in helical content at the N-terminus of the α -helix in all _D-variants, as shown in the average structures from the MD simulations (Figure 3F). This structural change was not observed in the _L-variants.

3.2 Contributions of _D-amino acids to stability

Thermodynamic properties of the Trp-cage were determined by thermal denaturation, following loss of ellipticity at 222 nm as a function of increasing temperature. However, both G10_DA and G10_DQ were partially folded at temperatures greater than 85°C, making it difficult to establish an unfolded-state baseline. To address this, we measured thermal denaturation curves at a series of guanidine hydrochloride (GnHCl) concentrations from 0 M to 5 M, fitting the entire data set globally to the Gibbs-Helmholtz equation to extrapolate estimates of free energy of folding, enthalpy, heat capacity and melting temperature (T_M) (Figures 4, S5).⁵³ Based on this analysis, the stability of TC5b was consistent with previous measurements of this peptide by other groups.⁵⁴ All _D-variants were more stable than WT. Based on the free energy of unfolding at 25°C (Table 1), both G10_DA and G10_DQ were stabilized by ~0.9 kcal/mol with respect to WT, and G10_DN by ~0.6 kcal/mol. Based on T_M , the most stable variant was G10_DQ, with a ΔT_m of ~23°C relative to WT, followed by G10_DA and G10_DN. DSC experiments were carried out on the same variants, although the thermograms could not be fit to obtain thermodynamic parameters due to the broad, shallow transitions (Figure S6). In the absence of distinct folded and unfolded baselines, T_M values were calculated using first derivative plots, and confirm the same trend as the one found by CD (Table 1). The difference between the two biophysical methods is notable for the lower stability WT and G10_DN species. This is attributed to the broad unfolding transitions due to lower ΔH for these two species (Table S1), resulting in a non-linear dependence of unfolded state heat capacity on temperature.^{34,55}

MD simulations also indicated a similar stability trend. As opposed to the _L-variants, all _D-variants were less flexible than WT (Figure 3D). The greatest changes were observed for residues 9 to 15 (which includes the 3_{10} helical turn). Decrease in backbone flexibility (root-mean-square fluctuations - RMSF) may indicate increased stability as the reduced likelihood of random thermal motions away from the most compact, folded conformation. Based on RMSF trends, both G10_DA and G10_DQ are predicted to be more stable than G10_DN, and G10_DN more than WT, as observed experimentally. Conversely, an increase of backbone flexibility, such as the one observed for G10A and G10Q, may increase the chances of random thermal motions away from the folded state, unfolding the overall structure.

MD may be used to interrogate the structure and dynamics of TC5b and variants for the molecular features of capping stability. In WT Trp-Cage, Gly10 participates in a network of backbone-to-backbone hydrogen bond capping interactions with both the α -helix and the right-handed 3_{10} helical turn (Table 2). These interactions include the backbone amide of Gly10 with the backbone carbonyls of Trp6, Leu7 and Lys8 in the α -helix, and the backbone carbonyl of Gly10 with the backbone amide of Ser13 and Ser14 in the 3_{10} helical turn. Because of the nature of the interactions and the proximity of the residues, more than one interaction can be present at the same time. The dominant, or the most frequent, interactions involved Leu7 and Ser13 (Table 2). In all _D-variants, the frequency of these dominant hydrogen bonds increased, and the trend correlated with experimentally observed

stability ($G10_{\text{DQ}} > G10_{\text{DA}} > G10_{\text{DN}} > \text{WT}$). In the case of the Leu7 interaction, the increase ranges from ~6% in $G10_{\text{DN}}$ to ~24% in $G10_{\text{DQ}}$. In the case of the Ser13 interaction, the increase is more dramatic and ranges from ~24% in $G10_{\text{DN}}$ to ~35% in $G10_{\text{DQ}}$. We could speculate that presence of a D-amino acid at position 10 drives the structure to a single state in which these dominant hydrogen bonds are present. The significant increase in the frequency of the hydrogen bond between position 10 and Ser13 may contribute to the increase rigidity of the entire helical turn in all D-variants.

3.3 Structure of TC5b $G10_{\text{DQ}}$

Using protCAD, D-Gln was modeled into the NMR structure of TC5b at position 10 (Figure 5). Over half of the 27 discrete rotamers sampled were sterically allowed, of which four formed a network of hydrogen bonds between D-Gln Ne and helix backbone carbonyls. One rotamer shown in Figure 5 participated in hydrogen bonds with carbonyls of Leu7 and Lys8, capping both C1 and C2 positions.

To solve the solution structure of the D-Gln variant, it was necessary to characterize the sequence enantiomer, as 15N, 13C D-Gln was not readily available. The resulting molecule had the identical, but inverse structure and stability as measured by CD (Figure S1). The NMR structure of this variant showed the presence of similar capping interactions as modeled by protCAD. Low-energy structures in the final ensemble presented a D-Gln10 hydrogen bond with backbone carbonyl from Lys8 (Figure 6A). Structures in which the D-Gln side chain faced away from the backbone were also consistent with experimental restraints but had higher energy scores. The HSQC spectrum of residue 10 in $G10_{\text{DQ}}$ (Figure 6B) presented two well-defined peaks for each of the pairs $C\beta\text{-H}\beta$ and $C\gamma\text{-H}\gamma$, suggesting a unique environment for these atoms and limited mobility of the side chain.

α -chemical shifts for WT, D-Ala and D-Asn variants were the same within experimental error, indicating that all peptides studied had the same fold.

3.4 Simulating contributions of side chain capping to stability

The solution structure of $G10_{\text{DQ}}$ and molecular simulations support the model that D-Gln participates in side chain capping interactions. Based on this, we predict that the increase in stability in $G10_{\text{DQ}}$ relative to WT is due to the presence of these side chain-to-backbone hydrogen bonds. In theory, D-Asn could participate in similar capping interactions, since Asn and Gln side chains have the same functional group, differing only by a methylene in the side chain, and modeling indicates that D-Asn can be incorporated and present side chain-backbone hydrogen bonds. If this is the case, why is $G10_{\text{DQ}}$ more stable than $G10_{\text{DN}}$?

In agreement with the NMR data, MD simulations of TC5b $G10_{\text{DQ}}$ show that D-Gln can hydrogen bond with the backbone carbonyl of Leu7, Lys8 and Asp9 (insets, Figure 7). However, these interactions were not present in the average structure. Cluster analysis of the conformation of D-Gln side chain identified sixteen clusters. Two clusters accounting for ~8.2 % and ~1.9 % of the total population presented strong hydrogen bonding to backbone carbonyls (the heavy atom distance and angle in each cluster centroid structure was 3.1 Å and 144° to Asp9, and 2.8 Å and 172° to Leu7). Clusters that represented ~19.5 % and ~2.3 % of the structures hydrogen bond weakly (3.2 Å and 115° to Asp9, and 3.5 Å and 98° to Lys8).

As opposed to $G10_{\text{DQ}}$, cluster analysis of the conformation of D-Asn10 side chain from structures obtained from the MD simulation of $G10_{\text{DN}}$ identified eleven clusters, of which only one cluster that represented ~1 % of the structures hydrogen bond strongly (3.5 Å and 150° with Lys 8).

The free energy surface along a given coordinate is known as a potential of mean force (PMF). Figure 7 shows the PMF for the distance between the side chain of residue 10 and the backbone carbonyl of Leu7, Lys8 and Asp9 for G10_bN and G10_bQ. These profiles show that in general, the global minimum does not correspond to the capping interaction (distance < 3.5 Å), and is between ~5 and ~7 Å. This agrees with the lack of side chain capping interactions in the average structures.

In the case of the PMF along the distance between ^DGln10 and Asp9, the capping interaction is very close in energy to the global minimum ($\Delta E \sim 0.1$ kcal/mol), with a barrier (E_a) less than 1 kcal/mol, and is present in ~15% of the simulation. In the case of the distance between ^DGln10 and Lys8, the minima are separated by ~1 kcal/mol, with a barrier of ~1.5 kcal/mol, and the hydrogen bond is present in ~4% of the simulation. These low barriers allow rapid interconversion between the different conformations. In the case of the distance between ^DGln10 and Leu7, the minima are separated by ~1.5 kcal/mol with an E_a of forming the capping interaction greater than 2.5 kcal/mol, and the interaction is present in only ~2.2% of the simulation.

In comparison, the barriers of forming the capping interactions in G10_bN are significantly greater (E_a greater than 3.5 kcal/mol). In this case, the PMF of only Lys8 shows a minimum that corresponds to the capping interaction, with a $\Delta E \sim 2.5$ kcal/mol, and this interaction is only present in ~0.8% of the simulation. On the other hand, the hydrogen bonds with Leu7 and Lys9 are present in 0.08 and 0.2% of the simulation, respectively. Therefore, ^D-Asn forms only one of the three possible capping interactions.

Overall, these results suggest that the side chain capping interactions might be important for stability, since G10_bQ is more stable and can make this type of interactions as opposed to G10_bN.

3.5 Capping interactions versus entropic and solvation costs

If the side chain capping interactions are important for stability as suggested above, why is then G10_bA more stable than G10_bN? Even if G10_bN does not certainly engage in capping interactions as frequently as G10_bQ, it does it more frequently than G10_bA, which completely lacks this ability. Therefore, other factors should contribute to the destabilization of G10_bN with respect to G10_bA.

While providing a network of stabilizing hydrogen bonds, helix capping interactions must overcome unfavorable loss of side chain entropy and desolvation of backbone carbonyls in the C3-C1 and Ccap positions. The existence of a barrier is explicitly evident in the PMF plots of side chain-backbone hydrogen bonds. Position 10 in the Trp-cage is located at the end of the helix and is surface exposed. Therefore, this residue will potentially be able to “freely” move and also to interact with the solvent. The backbone carbonyls of residues 7, 8 and 9 are also solvent exposed. For these reasons, the penalty of forming a capping interaction might include the entropic cost of fixing the side chain of residue 10 in a particular conformation and the desolvation of both the side chain and the backbone carbonyl of the residue that is interacting with. The potential barriers of side chain entropy and solvation are treated separately.

Conformational restraints make capping interactions unfavorable in G10_bN. Somewhat surprisingly, although the backbone fluctuation of residue 10 is smaller in G10_bQ (Figure 3D), the side chain fluctuation (measured as the mass-weighted RMSD with respect to the average structure) is greater than in G10_bN (Figure 8). Since ^D-Asn is in a relatively fixed conformation, it is more difficult for this residue to engage in capping interactions. On the

other hand, D-Gln10 side chain is more flexible and thus, it is more probable for this residue to adopt conformations that would allow capping interactions.

To estimate the entropic cost (ΔS) of fixing the side chain in the folded state, we estimated the change in conformational and vibrational entropy between the unfolded and folded states. The folded state is considered as the ensemble of structures from the MD simulation of the corresponding variant, G10_{bN} or G10_{bQ} . Inspired by previous work⁵⁶, the unfolded state was considered as the ensemble of structures from the MD simulation of a Gly-X-Gly peptide, where X = Asn or Gln. Although D-Asn is more rigid in G10_{bN} than D-Gln in G10_{bQ} (Figure 8), the ΔS upon folding is similar in both cases (Table 3). As suggested previously⁵⁷, there is essentially no change in the vibrational entropy upon folding, and the entropic cost of folding is mainly due to a loss of available conformations or rotamers. In any case, the entropic cost is small, probably because both side chains are fully solvent exposed. Similarly, we can assume that the side chain ΔS upon folding equals to zero for D-Ala in G10_{bA} , since there are no side chain rotamers for Ala.

In Table 4, we can see that presence of a D-amino acid in position 10 does not greatly reduce the backbone solvation of Leu7, Lys8 or Asp9, since G10_{bA} has only slightly lower solvation than WT. However, presence of side chain-to-backbone capping interactions reduces backbone solvation of these three residues. For example, in all variants including WT, there is on average more than one water molecule near ($<3\text{\AA}$) the carbonyl of Leu7 when there is no side chain capping interaction. When this interaction is formed in G10_{bQ} , less than 1 water molecule solvates the carbonyl of Leu7. Similarly, backbone desolvation penalties are also observed for Lys8 and Asp9 in G10_{bQ} , and for Lys8 in G10_{bN} . In terms of the side chain of residue 10, there is only a desolvation penalty for D-Gln10 interaction with Leu7 in G10_{bQ} (data not shown). To note, in G10_{bQ} , the side chain capping interaction with Leu7 has the highest energy barrier in the free energy profiles (Figure 7). Therefore, a higher desolvation penalty of G10_{bQ} when it interacts with Leu7 might explain the higher energy cost of this interaction.

4. CONCLUSIONS

The largest per-residue gains in Trp-Cage stability have resulted from D-amino acids substitutions at α_{L} -glycines as compared to other molecular strategies: computational optimization of core interactions⁵⁸, stabilizing proline by stereospecific fluorination, adding helix favoring amino acids³² and amino-terminus capping interactions⁵⁴. Andersen and colleagues substituted D-Ala at position 10 in TC10b variant, raising its T_m from 56 to 72°C.¹⁸ A similar magnitude of stabilization was found here for the TC5b, suggesting D-amino acid substitutions can be combined with other molecular strategies in an additive way leading to hyperstable peptides.

Given the success of the Trp-Cage designs, it is important to consider whether this approach can be generally applied. Placing D-Ala to the C-terminus of a monomeric α -helix did not enhance stability,²⁰ suggesting that much of the stability gain comes from reducing flexibility of turns between secondary structure elements. Stereochemical stabilization of loops/turns between α -helices⁵⁹ and β -strands^{12,60} is an established approach to improving fold stability. Gly 10 is a hinge between the α -helical and poly-proline segments of the Trp-Cage fold. In all C' L-Gln sites identified in the mirror-image template search of the PDB, the capping residue is located at link between the α_{L} turn and another secondary structural element.²⁵ Much of the enhanced stability of D-amino acid substitutions may come from favoring an intermediate in folding that brings elements together.

In addition to location within the fold, a few studies have shown that the choice of D -amino acid influences magnitude of stabilization.^{17,20,61} A series of C-capping auxiliaries terminated by an amino or guanidino group varied in their contributions to the stability of a monomeric α -helix depending on the alkyl-chain length.²⁰ It was suggested both helix-dipole interactions and specific hydrogen bonds were responsible for enhancing stability. This was confirmed in the case of a D -Arg substitution, where a 0.4 kcal/mol increase in stability was observed when electrostatic interactions were screened by 2M NaCl. Of the ten L -Gln capped α_{L} -helices in the mirror-image template PDB search, six participated in sidechain-to-backbone hydrogen bonds, and the remaining sidechains were involved in tertiary packing interactions. D -Gln may therefore be a generally applicable C-capping moiety. It may find particular utility in stabilizing monomeric helices where capping hydrogen bonds provide additional stability - an assertion that will require experimental validation.

One of the broader goals of this study was to develop protocols for predicting the stability of D -amino acid substitutions by MD and mirror-template bioinformatic methods. Overall, the experimental thermal stabilities agree with the stability trend predicted by the simulations: $\text{G10}_{\text{bQ}} > \text{G10}_{\text{bA}} > \text{G10}_{\text{bN}} > \text{WT} > \text{G10}_{\text{A/Q}}$. We observed that the magnitude of backbone fluctuations of amino acids surrounding the substitution correlates with stability: the most stable variants are also the most rigid. A reduction in backbone motion would result from a deeper energy minimum corresponding to the folded, native state.

On the other hand, some portion of the stability conferred by D -amino acid substitutions presumably results from promoting the formation of native-like transition state intermediates during folding, and destabilizing the unfolded state by reducing configurational entropy. To effectively model these phenomena, it will be necessary to thoroughly sample partially folded and unfolded states in the conformational ensemble. In G-X-G tripeptide simulations, the α_{L} conformation was significantly populated when $\text{X}=\text{L-Asn}$, but unoccupied for L-Gln (Figure S7), consistent with helix-propensity studies⁶². The folding pathway plays an important role in addition to the native state ensemble effects studied here.

Carrying out these simulations on larger peptides and small proteins is becoming tractable with replica-exchange methods⁶³ utilizing high-performance computing resources. Such intensive simulations are currently impractical as a direct step in sequence-space search algorithms utilized during computational protein design. Instead, they provide a better molecular understanding of the subtle contributions of solvation, sidechain entropy, electrostatics and hydrogen bonding. This can be used to refine design rules for incorporating stabilizing non-natural amino acid substitutions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the office of Information Services and Technology/High Performance and Research Computing at UMDNJ for access to computational resources, and Hsin Wang, of the Department of Chemistry, the City College of New York, for assistance with NMR measurements.

Funding Sources

VN, SA and ARG were supported by a Seed Grant from the UMDNJ Foundation, NSF MCB-0940187 and NIH R01 GM089949-01. LZ and RLK gratefully acknowledge support by the following grants: NSF MCB-0920448, infrastructure support from NIH P41 GM-66354 to the New York Structural Biology Center, and the NIH National Center for Research Resources to CCNY (NIH 5G12 RR03060).

ABBREVIATIONS

CD	circular dichroism
NMR	nuclear magnetic resonance
DSC	differential scanning calorimetry
MD	molecular dynamics
PMF	potential of mean force
WT	wild-type

REFERENCES

1. Zhou NE, Kay CM, Hodges RS. *Biochemistry*. 1993; 32:3178. [PubMed: 8457578]
2. Giblin MF, Wang N, Hoffman TJ, Jurisson SS, Quinn TP. *Proc Natl Acad Sci*. 1998; 95:12814. [PubMed: 9788997]
3. Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA. *Angew Chem Int Ed Engl*. 2001; 40:1494. [PubMed: 11317312] Tang Y, Ghirlanda G, Vaidehi N, Kua J, Mainz DT, Goddard IW, DeGrado WF, Tirrell DA. *Biochemistry-US*. 2001; 40:2790. Naduthambi D, Zondlo NJ. *J Am Chem Soc*. 2006; 128:12430. [PubMed: 16984189]
4. Kemp DS, Boyd JG, Muendel CC. *Nature*. 1991; 352:451. [PubMed: 1861726]
5. Bracken C, Gulyas J, Taylor JW, Baum J. *J Am Chem Soc*. 1994; 116 6431=6432. Walensky LD, Kung AL, Escher I, Malia TJ, Barbuto S, Wright RD, Wagner G, Verdine GL, Korsmeyer SJ. *Science*. 2004; 305:1466. [PubMed: 15353804]
6. Chapman RN, Dimartino G, Arora PS. *J Am Chem Soc*. 2004; 126:12252. [PubMed: 15453743]
7. Chandrasekaran, R.; Ramachandran, GN. 2nd American Peptide Symposium. First ed.. Lande, S., editor. Vol. Vol. 2. Cleveland, OH: Gordon and Breach; 1970. p. 195
8. Ramachandran GN, Ramakrishnan C, Sasisekharan V. *J Mol Biol*. 1963; 7:95. [PubMed: 13990617]
9. Karle IL, Gopi HN, Balaram P. *P Natl Acad Sci USA*. 2003; 100:13946. Karle IL. *Biopolymers*. 2001; 60:351. [PubMed: 12115146] Banerjee A, Raghothama SR, Karle IL, Balaram P. *Biopolymers*. 1996; 39:279. [PubMed: 8756509] Karle IL, Flippenanderson JL, Uma K, Balaram P. *Biopolymers*. 1993; 33:401. [PubMed: 8461451] Karle IL, Flippenanderson JL, Uma K, Balaram P. *Biochemistry*. 1989; 28:6696. [PubMed: 2790024]
10. Veber DF, Holly FW, Nutt RF, Bergstrand SJ, Brady SF, Hirschmann R, Glitzer MS, Saperstein R. *Nature*. 1979; 280:512. [PubMed: 460433] Veber DF, Holly FW, Paleveda WJ, Nutt RF, Bergstrand SJ, Torchiana M, Glitzer MS, Saperstein R, Hirschmann R. *Proc Natl Acad Sci U S A*. 1978; 75:2636. [PubMed: 208068]
11. Stanger HE, Gellman SH. *J Am Chem Soc*. 1998; 120:4236. Imperiali B, Fisher SL, Moats RA, Prins TJ. *J Am Chem Soc*. 1992; 114:3182. Yan YB, Erickson BW, Tropsha A. *J Am Chem Soc*. 1995; 117:7592.
12. Haque TS, Little JC, Gellman SH. *J Am Chem Soc*. 1996; 118:6975.
13. Rana S, Kundu B, Durani S. *Chem Commun*. 2005:207. Rana S, Kundu B, Durani S. *Chem Commun*. 2004:2462. Struthers MD, Cheng RP, Imperiali B. *Science*. 1996; 271:342. [PubMed: 8553067]
14. Valiyaveetil FI, Sekedat M, Mackinnon R, Muir TW. *Proc Natl Acad Sci*. 2004; 101:17045. [PubMed: 15563591]
15. Hermans J, Anderson AG, Yun RH. *Biochemistry*. 1992; 31:5646. [PubMed: 1610812] Fairman R, Anthony-Cahill SJ, DeGrado WF. *J Amer Chem Soc*. 1992; 114:5458.
16. Anil B, Song B, Tang Y, Raleigh DP. *J Amer Chem Soc*. 2004; 126:13194. [PubMed: 15479052]
17. Bang D, Gribenko AV, Tereshko V, Kossiakoff AA, Kent SB, Makhatazde GI. *Nature chemical biology*. 2006; 2:139.
18. Williams DV, Barua B, Andersen NH. *Organic & biomolecular chemistry*. 2008; 6:4287. [PubMed: 19005584]

19. Schellman, C. Protein Folding. Jaenicke, R., editor. New York: Elsevier/North-Holland; 1980. p. 53Presta, LG.; Rose, GD. Science. Vol. 240. New York, N.Y: 1988. p. 1632Richardson, JS.; Richardson, DC. Science. Vol. 240. New York, N.Y: 1988. p. 1648Aurora, R.; Srinivasan, R.; Rose, GD. Science. Vol. 264. New York, N.Y: 1994. p. 1126
20. Schneider JP, DeGrado WF. J Am Chem Soc. 1998; 120:2764.
21. Nanda V, Degrado WF. J Amer Chem Soc. 2004; 126:14459. [PubMed: 15521766]
22. Watanabe A, Yoshimura T, Mikami B, Hayashi H, Kagamiyama H, Esaki N. J Biol Chem. 2002; 277:19166. [PubMed: 11886871]
23. Serrano L, Fersht AR. Nature. 1989; 342:296. [PubMed: 2812029] Aurora R, Rose GD. Protein Sci. 1998; 7:21. [PubMed: 9514257] Gunasekaran K, Nagarajaram HA, Ramakrishnan C, Balaram P. J Mol Biol. 1998; 275:917. [PubMed: 9480777] Lu M, Shu W, Ji H, Spek E, Wang L, Kallenbach NR. J Mol Biol. 1999; 288:743. [PubMed: 10329176] Altschuler EL. Medical Hypotheses. 2001; 56:478. [PubMed: 11339850]
24. Mitchell JB, Smith J. Proteins. 2003; 50:563. [PubMed: 12577262]
25. Annavarapu S, Nanda V. BMC Struct Biol. 2009; 9:61. [PubMed: 19772623]
26. Kleywegt GJ. J Mol Biol. 1999; 285:1887. [PubMed: 9917419] Watson JD, Milner-White EJ. J Mol Biol. 2002; 315:171. [PubMed: 11779237]
27. Engel DE, DeGrado WF. J Mol Biol. 2004; 337:1195. [PubMed: 15046987]
28. Thomas ST, Loladze VV, Makhatadze GI. Proc Natl Acad Sci. 2001; 98:10670. [PubMed: 11535835]
29. Neidigh JW, Fesinmeyer RM, Andersen NH. Nat Struct Biol. 2002; 9:425. [PubMed: 11979279]
30. Qiu L, Pabit SA, Roitberg AE, Hagen SJ. J Amer Chem Soc. 2002; 124:12952. [PubMed: 12405814] Nikiforovich GV, Andersen NH, Fesinmeyer RM, Frieden C. Proteins. 2003; 52:292. [PubMed: 12833552] Chowdhury S, Lee MC, Xiong G, Duan Y. J Mol Biol. 2003; 327:711. [PubMed: 12634063] Ota M, Ikeguchi M, Kidera A. Proc Natl Acad Sci. 2004; 101:17658. [PubMed: 15591340] Steinbach PJ. Proteins. 2004; 57:665. [PubMed: 15390266] Kaledin M, Brown A, Kaledin AL, Bowman JM. J Chem Phys. 2004; 121:5646. [PubMed: 15366988]
31. Zhou R. Proc Natl Acad Sci. 2003; 100:13280. [PubMed: 14581616] Pitera JW, Swope W. Proc Natl Acad Sci. 2003; 100:7587. [PubMed: 12808142]
32. Lin JC, Barua B, Andersen NH. J Amer Chem Soc. 2004; 126:13679. [PubMed: 15493925]
33. Simmerling C, Strockbine B, Roitberg AE. J Amer Chem Soc. 2002; 124:11258. [PubMed: 12236726] Snow CD, Zagrovic B, Pande VS. J Amer Chem Soc. 2002; 124:14548. [PubMed: 12465960] Kannan S, Zacharias M. Proteins. 2009; 76:448. [PubMed: 19173315] Day R, Paschek D, Garcia AE. Proteins. 2010; 78:1889. [PubMed: 20408169]
34. Streicher WW, Makhatadze GI. Biochemistry-U.S. 2007; 46:2876.
35. Wafer LN, Streicher WW, Makhatadze GI. Proteins. 2010; 78:1376. [PubMed: 20112418]
36. Jeener J, Meier BH, Bachmann P, Ernst RR. J Chem Phys. 1979; 71:4546.
37. Braunschweiler L, Ernst RR. J Magn Reson. 1983; 53:521.
38. Rucker SP, Shaka AJ. Mol Phys. 1989; 68:509.
39. Bodenhausen G, Ruben DJ. Chem Phys Lett. 1980; 69:185.
40. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. Bioinformatics. 2007; 23:381. [PubMed: 17121777]
41. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Acta Crystallogr. 1998; 54:905.
42. Summa, CM. University of Pennsylvania School of Medicine; 2002.
43. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J Am Chem Soc. 1995; 117:5179.
44. Case, DA., et al. San Francisco: University of California; 2008.
45. Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. Journal of Chemical Physics. 1983; 79:926.

46. Wang J, Cieplak P, Kollman PA. *J Comp Chem*. 2000; 21:1049. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. *Proteins*. 2006; 65:712. [PubMed: 16981200]
47. Berendsen HJ, Postma JP, van Gunsteren WF, Di Nola A, Haak JR. *J Chem Phys*. 1984; 81:3684.
48. Ryckaert JP, Ciccotti G, Berendsen HJC. *Journal of Computational Physics*. 1977; 23:327.
49. Feig M, Karanicolas J, Brooks CL 3rd. *J Mol Graph Modeling*. 2004; 22:377.
50. Leach, AR. *Molecular Modelling. Principles and Applications*. Addison, Wesley Longman, editor. England: Pearson Education Limited; 2001.
51. Schlitter J. *Chem Phys Lett*. 1993; 215:617. Andricioaei I, Karplus M. *J Chem Phys*. 2001; 115:6289. Brooks BR, Janei D, Karplus M. *Journal of computational chemistry*. 2004; 16:1522.
52. Dixit SB, Andrews DQ, Beveridge DL. *Biophys J*. 2005; 88:3147. [PubMed: 15731390]
53. Walsh ST, Sukharev VI, Betz SF, Vekshin NL, DeGrado WF. *J Mol Biol*. 2001; 305:361. [PubMed: 11124911]
54. Barua B, Lin JC, Williams VD, Kummeler P, Neidigh JW, Andersen NH. *Protein Eng Des Sel*. 2008; 21:171. [PubMed: 18203802]
55. Makhatadze GI, Privalov PL. *J Mol Biol*. 1993; 232:639. [PubMed: 8393940]
56. Creamer TP. *Proteins*. 2000; 40:443. [PubMed: 10861935]
57. Karplus M, Ichiye T, Pettitt BM. *Biophys J*. 1987; 52:1083. [PubMed: 3427197]
58. Bunagan MR, Yang X, Saven JG, Gai F. *J Phys Chem B*. 2006; 110:3759. [PubMed: 16494434]
59. Lahr SJ, Engel DE, Stayrook SE, Maglio O, North B, Geremia S, Lombardi A, DeGrado WF. *J Mol Biol*. 2005; 346:1441. [PubMed: 15713492] Effimov AV. *Curr Opin Struc Biol*. 1993; 3:379.
60. Sibanda BL, Thornton JM. *Nature*. 1985; 316:170. [PubMed: 4010788]
61. Narayana N, Phillips NB, Hua QX, Jia W, Weiss MA. *J Mol Biol*. 2006; 362:414. [PubMed: 16930618]
62. O'Neil KT, DeGrado WF. *Science*. 1990; 250:646. [PubMed: 2237415] Hovmoller S, Zhou T, Ohlson T. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:768. [PubMed: 11976487]
63. Okamoto Y, Sugita Y. *Chem Phys Lett*. 1999; 314:141.

Figure 1. C-capping of an α -helix. **(A)** C-capping motif from a poly-Ala (D)₉(L)₂ peptide obtained from Monte Carlo simulations of poly-LD-Ala peptides in a previous study (ref 21). **(B)** Example of a natural L -Gln C' capping motif in an α_L turn (residues 297 to 301 of 2J6L).

Figure 2.

Far-UV CD spectra of the Trp-cage variants (A), WT, G10A and G10Q; (B) WT, G10bA, G10bN and G10bQ at 25 °C. MRE: mean residue ellipticity.

Figure 3.

(**A, B**): RMSD (in Å, with respect to the first structure) of the backbone heavy atoms as a function of the simulation time (in nanoseconds) for WT and the _L-variants (**A**) and WT and the _D-variants (**B**). The last 50 ns from each simulation were selected for data analysis (production run). (**C, D**) Average fluctuations (RMSF in Å) of backbone heavy atoms as a function of residue for the production run for WT and the _L-variants (**C**) and WT and the _D-variants (**D**). The secondary structure elements are indicated (α_R : right-handed α -helix, 3–10: right-handed 3–10 helical turn, pro: poly-proline stretch). (**E, F**) Superimposition of the average structures (front and side views) for WT and the _L-variants (**E**) and WT and the _D-variants (**F**). WT (black), G10A (orange), G10Q (cyan), G10bA (red), G10bN (green) and G10bQ (blue).

Figure 4.

(A) Thermally-induced unfolding of G10_bQ Trp-cage at different GuHCl concentrations (from dark blue to red: 0.0, 1.0, 2.0, 3.0, 4.0, 5.0 M) monitored by far-UV CD at 222 nm. (B) The melts for G10_bQ (other variants in Fig. S5) were globally fit to obtain the fraction unfolded as a function of temperature (WT, black; G10_bA, red; G10_bN, green; G10_bQ, blue). MRE: mean residue ellipticity.

Figure 5.
Lowest energy rotamer of ${}^{\text{D}}$ Gln modeled onto the TC5b structure²⁹.

Figure 6. (A) One low-energy NMR structure of G10bQ showing a Gln side chain to backbone capping interaction with Lys8. (B) HSQC spectrum of DQ10 in G10bQ.

Figure 7.

Helmholtz free energy (A , kcal/mol - eq.3) as a function of the distance between residue 10 side chain (N δ in G10_{DN}, green; and Ne in G10_{DQ}, blue) and the backbone carbonyl of Leu7, Lys8 and Asp9 obtained from the MD simulations. Representative snapshots from the simulation of G10_{DQ} in which the different capping hydrogen bond interactions are formed are shown.

Figure 8.

Mass-weighted rmsd (in Å, with respect to the average structure) of the side chain heavy atoms of residue 10 as a function of the production run (in ns). G10_bN (green) and G10_bQ (blue) variants.

Table 1

Thermodynamic parameters of Trp-cage variants, obtained from DSC and CD thermal unfolding experiments.

	DSC	CD	
	T _M (°C)	T _M (°C)	ΔG _U 25°C (kcal/mol)
WT	36	46 ± 3	0.8 ± 0.2
G10bA	68	66.6 ± 0.2	1.7 ± 0.1
G10bN	63	56 ± 5	1.4 ± 0.4
G10bQ	69	69.3 ± 0.5	1.7 ± 0.2

Table 2

Time-fraction (%) of backbone-to-backbone hydrogen bond (HB) capping interactions donated by backbone amide (N) of residue 10 to positions 6–8 and backbone carbonyl (O) of residue 10 accepted from amides of residues 13,14. The HB distance and angle cutoffs were 3.5 Å and 120°, respectively.

donor acceptor	WT	G10bA	G10bN	G10bQ	G10A	G10Q
X10.N-	W6.O	36	24	30	24	0
	L7.O	63	73	67	78	40
	K8.O	9	7	7	4	42
X10.O-	S13.N	66	86	82	89	74
	S14.N	12	6	18	3	25
						0

Table 3

Entropic cost (ΔS) of fixing the side chain of D -Asn and D -Gln upon folding of the Trp-cage G10 $_D$ N and G10 $_D$ Q variants. The conformational entropy is given by the side chain rotamers, and the vibrational entropy is given by the quasi-harmonic entropy. Values are expressed as $T\Delta S$, where T is 300 K, in kcal/mol. Reported values are the mean over four calculations (Table S2).

	conformational	vibrational
D -Asn	0.29	0.01
D -Gln	0.32	-0.04

Backbone carbonyl (BB.O) solvation of Leu7, Lys8 and Asp9 of the different Trp-cage variants from the MD simulations.

Table 4

	number of waters time-fraction of HB(%)				
	L7	K8	D9	L7	D9
WT	1.3 ± 0.5	2.2 ± 0.7	2.0 ± 0.8	116	153
G10bA	1.2 ± 0.4	2.0 ± 0.7	1.9 ± 0.7	107	149
G10bN	1.1 ± 0.4	1.9 ± 0.6 (1.4 ± 0.6)	2.0 ± 0.8	103	156
G10bQ	1.2 ± 0.5 (0.5 ± 0.6)	1.9 ± 0.7 (1.3 ± 0.6)	1.9 ± 0.7 (1.2 ± 0.6)	103	134

The results from the first 3 columns are expressed as the mean number of waters ($\pm$ standard deviation) within 3 Å of the corresponding atoms. When a side chain-to-backbone hydrogen bond capping interaction exists (G10bN with K8, and G10bQ with L7, K8 and D9), the values reported correspond to the cases in which there is no capping interaction, and between brackets are the values when the interaction is formed. The capping interaction was defined to be present and absent when the distance between residue 10 side chain (N6 in G10bN and Ne in G10bQ) and the BB.O of the hydrogen bond (HB) acceptor was smaller than 3.5 Å and greater than 4 Å, respectively. Time-fraction (%) of HBs defined between solvent and BB.O; cutoffs were 3.5 Å and 120°. Values greater than 100 reflect multiple solvent donors.