Caballero, D.; Määttä, Jukka; Zhou, A.; Sammalkorpi, Maria; O'Hern, C.S.; Regan, L.

Intrinsic alpha-helical and beta-sheet preferences: A computational case study of Alanine

Published in:
Protein Science

DOI:
10.1002/pro.2481

Published: 01/01/2014

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Caballero, D., Määttä, J., Zhou, A., Sammalkorpi, M., O'Hern, C. S., & Regan, L. (2014). Intrinsic alpha-helical and beta-sheet preferences: A computational case study of Alanine. *Protein Science*, *23*(7), 970-980. https://doi.org/10.1002/pro.2481

This author preprint is published in final form as: Protein Science 23(7), 970-980 (2014).

http://onlinelibrary.wiley.com/doi/10.1002/pro.2481/epdf
DOI: 10.1002/pro.2481

Intrinsic α-helical and β-sheet conformational preferences: A computational case study of Alanine

Diego Caballero[1,2], Jukka Määttä[3], Alice Qinhua Zhou[2,4], Maria Sammalkorpi[3], Corey S. O'Hern[1,2,5,6], and Lynne Regan[2,4,7]

[1]Department of Physics, Yale University;
[2]Integrated Graduate Program in Physical and Engineering Biology, Yale University;
[3]Department of Chemistry, Aalto University;
[4]Department of Molecular Biophysics & Biochemistry, Yale University
[5]Department of Mechanical Engineering & Materials Science, Yale University;
[6]Department of Applied Physics, Yale University;
[7]Department of Chemistry, Yale University.

April 16, 2014

1 Abstract

A fundamental question in protein science is what is the intrinsic propensity for an amino acid to be in an α-helix, β-sheet, or other backbone dihedral angle (ϕ-ψ) conformation. This question has been been hotly debated for many years because including all protein crystal structures from the protein database increases the probabilities for α-helical structures, while experiments on small peptides observe that β-sheet-like conformations predominate. We perform molecular dynamics (MD) simulations of a hard-sphere model for Ala dipeptide mimetics that includes steric interactions between non-bonded atoms and bond length and angle constraints with the goal of evaluating the role of steric interactions in determining protein backbone conformational preferences. We find four key results. For the hard-sphere MD simulations, we show that 1) β-sheet structures are roughly three and half times more probable than α-helical structures, 2) transitions between α-helix and β-sheet structures only occur when the backbone bond angle τ (N-C_α-C) is greater than 110°, and 3) the probability distribution of τ for Ala conformations in the 'bridge' region of ϕ-ψ space is shifted to larger angles compared to other regions. In contrast, 4) the distributions obtained from Amber and CHARMM MD simulations in the bridge regions are broader and have increased τ compared to those for hard sphere simulations and from high-resolution protein crystal structures. Our results emphasize the importance of hard-sphere interactions and local stereochemical constraints that yield strong correlations between ϕ-ψ conformations and τ.

2 Introduction

The first structure of a protein was solved over 50 years ago [1,2]. Around that time, Ramachandran and colleagues showed that simple 'hard-sphere' models of dipeptides could predict the sterically allowed regions of backbone dihedral angle (ϕ-ψ) space [3]. Most importantly, these allowed regions correspond to the combinations of ϕ and ψ that were observed in the protein crystal structures. There are currently over 80,000 structures deposited in the protein data bank (PDB) [4], and the overwhelming majority of amino acids in those structures have backbone dihedral angle combinations that fall into the regions predicted by Ramachandran and colleagues [5, 6].

Knowing the intrinsic backbone conformational preferences of amino acids is necessary for a fundamental understanding of the dynamics of protein folding. Conversions between α-helix and β-sheet conformations are likely to occur during transitions from unfolded to folded structures. However, despite significant work over the last several decades, there is still no consensus concerning the intrinsic backbone conformational preferences for amino acids. Beginning with Chou and Fasman [7], researchers have sought to determine the relative α-helix and β-sheet propensities for each amino acid by analyzing the frequency that each amino acid occurs in α-helices versus β-sheets in protein crystal structures. However, because α-helices are strongly overrepresented in proteins of known structure, as shown dramatically in Fig. 1(a), these analyses also do not provide the intrinsic probability for a given amino acid to have a particular backbone conformation.

Researchers have tried to circumvent this problem by analyzing the distribution of ϕ-ψ backbone dihedral angles in only 'coil' regions of proteins. Although such a strategy has the potential to identify the intrinsic α-helix and β-sheet preferences, there are a number of issues. How should the coil region be defined? For example, if one eliminates residues on the basis of the backbone ϕ-ψ values, removing those with α-helical ϕ-ψ combinations will obviously decrease the α-helical content. Researchers have recognized these limitations, and have employed other strategies to determine 'true' conformational preferences [8–11] and *vide infra*.

There have also been several experimental studies that have ranked the relative α-helix or β-sheet forming propensities [12–28]. Although such data are informative, these experiments actually measure the relative energy difference between a residue in an α-helix versus the denatured state in a given system or between a residue in a β-sheet versus the denatured state for a different system, so the absolute energy difference between the α-helix and β-sheet conformations cannot be determined.

Our goal is to predict computationally the intrinsic probabilities for an amino acid to adopt particular backbone dihedral angle conformations, which is an area of fervent interest [29–36]. We perform our calculations on a dipeptide mimetic because we are interested in the intrinsic conformational preferences, which are mediated by short-range interactions. The dipeptide mimetic is the simplest model that includes all local interactions but none with distant residues. We chose to study alanine (Ala) because it is one of the simplest residues with no side-chain dihedral angles and its secondary structure propensities have been extensively studied. In experimental studies, Ala has one of the highest α-helix propensities [37]. Further, Ala residues are three times as likely to be found in α-helices compared to β-sheets in protein crystal structures [38]. However, backbone conformations can depend strongly on the environment, for example, whether the residue occurs within a stretch of α-helical order or not. To eliminate such effects, researchers have therefore attempted to measure propensities in extremely short peptides using a number of spectroscopic techniques (Table 1). Most of these experimental studies find that short Ala peptides populate α-helical structures in solution less than 20% of the time. Consistent with these observations, the Wu 'Coil-3' library [10], which is derived from protein crystal structures but only considers residues that occur in neither α-helices nor β-sheets and are not pre-proline or in turns, finds that only 20% of alanines have α-helical ϕ-ψ values. Structures with β-sheet or polyproline II (PPII) ϕ-ψ values are now dominant (Fig. 1(b)).

We present the results for molecular dynamics (MD) simulations on an Ala dipeptide mimetic using a simplified force field that includes only intraresidue stereochemical constraints and hard-sphere interactions. The simplicity of this model allows us to determine to what extent backbone conformational preferences can be explained by the hard-sphere plus stereochemical constraint model alone. In addition, the hard-sphere model of the Ala dipeptide mimetic allows us to run long simulations and directly measure the equilibrium probability distributions of Ala backbone conformations. We find that non-α-helical structures predominate, with equilibrium populations of α-helix conformations totaling less than 25%. For comparison, we also performed MD simulations of Ala dipeptide mimetics using the GROMACS simulation package [48] with recent versions of the Amber [49,50] and CHARMM [51] force fields and their associated optimized explicit water models (see Materials and Methods for details). The resulting ϕ-ψ distributions are different from each other and from our hard-sphere simulations, because of the strong differential contributions of additional terms in these force fields. Our hard-sphere MD simulations also enable us to investigate in detail transitions between α-helix and β-sheet conformations. We find that such transitions only occur when the main-chain bond angle, τ, is large. Interestingly, the Amber and CHARMM force fields do not capture this strong interdependence between transitions between α-helices and β-sheets and the main-chain angle τ. The importance of the value of τ on transitions between α-helices and β-sheets in our hard-sphere MD simulations is consistent with the observation that in proteins of known structure residues that populate the 'bridge region' of ϕ-ψ space, between the α-helix and β-sheet regions, possess larger values of τ [3, 52–55].

3 Results and Discussion

In Fig. 2, we show the probability distribution $P(\phi, \psi)$ for the backbone dihedral angle combinations ϕ-ψ for the thermally equilibrated, hard-sphere model of the Ala dipeptide mimetic. In this model, τ is allowed to sample values from the distribution observed in protein crystal structures. In this plot, we also show the Ramachandran 'outer' and 'normal' limits [3] for $\tau = 110^\circ$ and the regions we designate as α' and β'. (See Table 1 (left) for the definitions of the α' and β' regions.) α' includes both classic α-helix and bridge regions, and β' includes both classic β-sheet and PPII regions. Similar limits have been used by others [56–58].

There are several important features in Fig. 2. $P(\phi, \psi)$ from the hard-sphere simulations largely respects the Ramachandran limits in the α-helix and β-sheet regions. The main discrepancy in this respect is in the Ramachandran plot in the vicinity of α_L ($\phi = 60^\circ$ and $\psi = 60^\circ$). This discrepancy stems from the fact that the Ramachandran, *et al.* outer limits were based on non-additive atomic radii and the size of this allowed region varies strongly with τ. In addition, the occurrence of conformations

in the bridge region outside of the pictured Ramachandran limits for the hard-sphere simulations reflects the sampling of $P(\tau)$ with average $< \tau >= 110^\circ$ (and standard deviation 3.4°), whereas the Ramachandran limits correspond to a single $\tau = 110^\circ$. It is also immediately apparent that the α-helix region is not overwhelmingly populated compared to the β-sheet region, in contrast with Fig. 1(a). Instead, the maximum probabilities in the α-helix and β-sheet regions are comparable and significantly greater than the maximum probability in the bridge region. We find that the probabilities in the α' and β' regions are 26% and 68%, respectively (Table 2). Furthermore, the ϕ-ψ probabilities are relatively uniform within the α' and β' regions. Thus, we can estimate the α' and β' probabilities by the area in ϕ-ψ space that they occupy, which is 31% and 69% of the total ϕ-ψ space, respectively.

We next investigated the correlations between the backbone dihedral angle combinations (ϕ and ψ) and the bond angle τ. In Fig. 3, we show the probability distribution $P(\tau)$ separately for each region, α, β', and bridge (top to bottom), from MD simulations of the hard-sphere model for the Ala dipeptide mimetic. First, we note that in all three regions, $P(\tau)$ from the hard-sphere MD simulations is similar to the distributions observed in high-resolution crystal structures. Second, we find that when the dipeptide mimetic occurs in the α and β' regions, $P(\tau)$ from the hard-sphere MD simulations closely matches the 'ideal' Boltzmann distribution inferred from only the bond-angle potential energy (Eq. 3). By contrast, the $P(\tau)$ from the hard-sphere MD simulations for conformations in the bridge region are shifted significantly to higher bond angles compared to this 'ideal' distribution. Note that the distributions $P(\tau)$ in the bridge region for both the hard-sphere simulations and high-resolution protein crystal structures are narrower than those in the α and β' regions. In addition, in the middle panel of Fig. 3, we show a small shift of $P(\tau)$ (blue) to smaller angles for crystal structures in the Dunbrack database with β' backbone conformations compared to $P(\tau)$ for all structures [54].

We also studied the relationship between τ and transitions between the α and β' regions by calculating $P(\phi, \psi)$ when the average value, $\langle\tau\rangle$, is constrained to be 105°, 110°, or 115° with only 1° standard deviations (see the first and second columns of Fig. 4). For $\langle\tau\rangle = 105^\circ$ (first row) and 110° (second row), transitions between the α and β' regions are never observed over the full simulation, independent of whether the Ala dipeptide mimetic is initialized in the α' (first column) or β' (second column) regions. By contrast, when $\langle\tau\rangle = 115^\circ$ (third row), transitions occur frequently between α and β' and $P(\phi, \psi)$ is independent of the starting values of ϕ and ψ. In the third column, we show an example of the potential energy landscape as a function of ϕ and ψ for different values of τ. We find that the energy barrier in the bridge region begins to decrease for $\langle\tau\rangle = 110^\circ$ and is extremely small for $\langle\tau\rangle = 115^\circ$.

In addition, we performed molecular dynamics simulations of a single Ala dipeptide mimetic in explicit water using the commonly used force fields Amber99sb* [59, 60] and CHARMM27 [61, 62] with and without their respective empirically corrected dihedral angle potentials, Amber99sb-ILDN-NMR [63, 64] and CHARMM27-CMAP [65]. We show the equilibrium probability distributions for the backbone dihedral angle combinations P(ϕ,ψ) from these simulations and from protein crystal structures in Fig. 5. We identify several important features. For Amber99sb* (Fig. 5(a) and (b)), we find that the bridge region is overpopulated compared to proteins of known structure (Fig. 5(e) and (f)), and the α' and β' regions are strongly non-uniform. Also, $P(\phi, \psi)$ for Amber99sb-ILDN-NMR is very similar to the probability distribution for Amber99sb*.

In contrast, CHARMM27 (Fig. 5(c) and (d)) populates the region $-180^\circ < \psi < -60^\circ$, which is sterically disallowed. The CMAP correction prevents sampling of this region. Although the $P(\phi, \psi)$ distributions are different for CHARMM27 and CHARMM27-CMAP (Fig. 5(c) and (d)), the relative populations of structures in the α' and β' regions are similar for both (see Table 2).

Interestingly, we find that the α' and β' propensities are approximately 26% and 72%, respectively, from *both* Amber simulations, which is similar to the results from the hard-sphere model. The CHARMM force field predicts a significantly higher population for α', roughly 50% for both α' and β' both with and without CMAP corrections. Similar differences between the CHARMM and Amber force fields were obtained by Vymetal and Vondrasek [58].

We also studied the correlation between the bond angle τ and backbone dihedral angles ϕ and ψ in the CHARMM and Amber molecular dynamics simulations (Fig. 6). For both force fields, we

observe that the peaks in the bond angle distributions $P(\tau)$ are shifted to larger values, $\tau \approx 113^\circ$, and the distributions are wider than those found in proteins of known structure. While it is possible that $P(\tau)$ for peptides in solution is broader than that from protein crystal structures, there is no obvious reason to expect a shift in the mean of the bond angle distributions when comparing protein crystal data and data from peptides in solution. As suggested from the results in Fig. 4, a shift in the peak of $P(\tau)$ to larger values facilitates transitions between the α' and β' regions. Note that in contrast to the hard-sphere model, the harmonic bond-angle potential energies are centered on $\tau = 110^\circ$ and 107° for Amber and CHARMM, respectively, but other interactions shift the average to larger values of $< \tau > \approx 113^\circ$.

We observe very different behavior for the hard-sphere model. In this case, when $< \tau >$ is 110° or lower, no transitions between α' and β' are observed. Thus, the hard-sphere model predicts that there must be a correlation between a large bond angle τ and the backbone dihedral angles ϕ and ψ when they are in the bridge region. This correlation is also found in protein crystal structures (Fig. 3). In contrast, for the Amber and CHARMM MD simulations of the Ala dipeptide mimetic, the average τ is larger than that observed in protein crystal structures.

What leads to the differences in the sampling of backbone conformations between Amber and CHARMM and the hard-sphere model? The Amber and CHARMM force fields incorporate a large number of inter-dependent terms as well as longer-range interactions, which have been optimized so that these force fields can reproduce many aspects of the behavior of small molecules, proteins and nucleic acids. These terms combine to give an eminently reasonable 'average' treatment of a protein - as evidenced by many successful simulations of protein structure [66, 67]. With the hard-sphere model that we present, we do not attempt to model the complex interactions that occur in large proteins. Instead, we seek to describe the exact stereochemistry of a dipeptide mimetic. The results we present, along with our prior studies of the side-chain dihedral angle distributions of different amino acids [55, 68, 69][AZ, CO, LR submitted to J Mol Biol], make it clear that steric repulsion is the dominant force in specifying the allowed backbone and sidechain conformations of a large set of amino acids. We believe that with Amber and CHARMM, the contribution of steric repulsion is being outweighed by the contributions from other terms in the force field. In other circumstances, where sterics are not necessarily the dominant interaction, the additional terms in the Amber and CHARMM force fields are vital to include.

An additional discovery is the importance of the inter-dependence of ϕ-ψ and τ. Ramachandran had predicted [3] and we showed for protein crystal structures [55] that the distribution of ϕ-ψ angles depends on the value of τ (i.e. the Ramachandran plots for an Ala dipeptide mimetic are different for $\tau = 105^\circ$, 110°, and 115°). The studies we present here expand on that observation, and show that transitions between α-helix and β-sheet conformations require τ to be large. An interesting research direction to pursue in the development of the AMBER and CHARMM force fields is to reweigh the strength of the steric interactions relative to others or implement directly a ϕ-ψ-τ correlation term to ensure that transitions between α-helix and β-sheet backbone conformations occur by increasing the bond angle angle τ.

Despite decades of work, there is still considerable debate concerning the intrinsic propensities for amino acids to adopt α-helix versus β-sheet structures. To address this issue, we performed molecular dynamics simulations of an Ala dipeptide mimetic using a minimal model that includes only stereochemical constraints and hard-sphere interactions between non-bonded atoms. This model predicts probabilities for α-helix and β-sheet structures (26% and 68%, respectively) that are consistent with both random coil libraries and experimental data on short peptides. We also observe a strong correlation between the bond angle τ and transitions between α-helix and β-sheet conformations. For $\langle\tau\rangle < 110^\circ$, such transitions between α-helix and β-sheet do not occur. In contrast, for $\langle\tau\rangle = 115^\circ$, the Ala dipeptide is able to transition from α-helix to β-sheet conformations. However, in MD simulations of the Ala dipeptide mimetic in water using the Amber and CHARMM force fields, we find that the average bond angle $\tau \approx 113^\circ$ (above the average found for high-resolution protein crystal structures) for all ϕ and ψ dihedral angle combinations, which indicates that the other inter-depependent and longer-range interactions outweigh the repulsive steric interactions.

4 Materials and Methods

We studied an all-atom hard-sphere representation of an Ala dipeptide mimetic, N-acetyl-L-Ala-N'-methylamide, as shown in Fig. 7. This Ala dipeptide mimetic is composed of 21 bonds between pairs of atoms and 36 bond angles (including bonds that involve hydrogen atoms). We built our model using stereochemical parameters, *i.e.* the average and standard deviation of the bond lengths (l_{ij}^0 and Δl_{ij}), bond angles (θ_{ijk}^0 and $\Delta\theta_{ijk}$), and ω backbone dihedral angles ($\omega_{ijkl}^0 \approx 0$ and $\Delta\omega_{ijkl}$) obtained for Ala residues in the Dunbrack Database [38]. This culled database is composed of 850 high-resolution, non-homologous protein structures with resolution ≤ 1.7 Å, side chain B-factors per residue < 40 Å^2 (local B-factor filtering), and R-factors ≤ 0.25. This data set includes 16477 Ala residues.

We compare our results in Fig. 1(b) and Table 1 to the Wu 'Coil-3' library [10]. The Coil-3 library includes 6178 protein structures from the PDB with a resolution < 2.0 Å, R-factors < 2.0, and a 50% sequence identity cutoff. The Coil-3 library does not include residues in α-helices or β-sheets. In addition, pre-proline and turn residues are excluded. The total number of Ala residues in the Coil-3 library is 20,761.

The atomic diameters σ_i are: C(sp^3) 1.5Å, C(sp^2) 1.4Å, N 1.4Å, O 1.4Å, and H 1.05Å, which are identical to values employed in previous studies [55, 68, 69], except the oxygen diameter was changed from 1.45Å to 1.4Å to improve sampling in ϕ-ψ space (see Supplementary Material). Hydrogen atoms were added to the structure using the REDUCE software program [70]. Our simulations of the Ala dipeptide mimetic include the following four interaction potentials between spherical atoms i and j: 1) a purely repulsive Lennard-Jones potential,

$$V_{lj} = \epsilon \left(1 - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right)^2 \Theta\left(\sigma_{ij} - r_{ij}\right), \tag{1}$$

where ϵ is the characteristic energy scale of the interaction, r_{ij} is the separation between non-bonded atoms i and j, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, and $\Theta(x)$ is the Heaviside step function that prevents interactions between atoms when they are not in contact; 2) a harmonic potential to constrain the bond lengths,

$$V_{bl} = \frac{K^{l_{ij}}}{2}\left(r_{ij} - l_{ij}^0\right)^2, \tag{2}$$

where $K^{l_{ij}} = T/(\Delta l_{ij})^2$ and T is the temperature in units of the Boltzmann constant; 3) a harmonic potential to constrain the bond angles,

$$V_{ba} = \frac{K^{\theta_{ijk}}}{2}\left(\theta_{ijk} - \theta_{ijk}^0\right)^2, \tag{3}$$

where $K^{\theta_{ijk}} = T/(\Delta\theta_{ijk})^2$; and 4) a harmonic potential to constrain the two ω_{ijkl} dihedral angles (defined by the groups of four atoms C_α^{i-1}-C^{i-1}-N-C_α and C_α-C-N^{i+1}-C_α^{i+1}) to be planar,

$$V_\omega = \frac{K^{\omega_{ijkl}}}{2}\omega_{ijkl}^2, \tag{4}$$

where $K^{\omega_{ijkl}} = T/(\Delta\omega_{ijkl})^2$. Note that the spring constants $K^{l_{ij}}$, $K^{\theta_{ijk}}$, and $K^{\omega_{ijkl}}$ are chosen so that the standard deviations at temperature T of the bond lengths, bond angles, and ω dihedral angles match those for Ala residues from high-resolution protein crystal structures. The total potential energy V is obtained by summing the interactions in Eqs. 1-4 over all non-bonded pairs of atoms, bonds, bond angles, and the two backbone dihedral angles ω for the Ala dipeptide.

We performed implicit-solvent Langevin dynamics [71] simulations of the Ala dipeptide mimetic by numerically integrating

$$m_i \frac{d^2\vec{r}_i}{dt^2} = \xi\frac{d\vec{r}_i}{dt} + \vec{\Gamma}_i - \frac{\partial V}{\partial \vec{r}_i} \tag{5}$$

for the atomic positions $\vec{r}_i$, where m_i is the mass of atom i, the Gaussian-distributed, δ-function correlated random forces $\vec{\Gamma}_i$ on atom i obey $\langle \vec{\Gamma}_i(t) \cdot \vec{\Gamma}_j(t') \rangle = 2\xi T \delta_{ij} \delta(t-t')$, and $\delta(x)$ (δ_{ij}) is the Dirac (Kronecker) δ-function. We implemented a modified Velocity Verlet algorithm to integrate Eq. 5 with a time step $\Delta t = 10^{-4}\ t_0$, where $t_0 = \sigma_H \sqrt{m_H/\epsilon}$, and damping parameter $\xi = 5\epsilon t_0/\sigma_H^2$.

The initial atomic velocities were drawn from a Maxwell-Boltzmann distribution at temperature T^*, where $T^* = T/\epsilon \approx 10^{-2}$. The ratio T/ϵ determines the average amount of overlap (*i.e.* pair separations that satisfy $r_{ij} < \sigma_{ij}$) between non-bonded atoms that occurs in the simulations. In the $(T/\epsilon) \to 0$ limit, the system explores only sterically allowed conformations. We show in Figs. S3 and S4 that the average number of overlaps between pairs of nonbonded atoms becomes nonzero above the characteristic temperature T^*, which is the temperature of the simulations, and thus our simulations are carried out in the limit of hard-sphere interactions. To determine the equilibration time for the hard-sphere simulations, we measured the average time, t_r, required to make transitions from α' to β' or from β' to α' (Fig. S1). We then equilibrated the Ala dipeptide for more than $100t_r$ before measuring conformational statistics. We calculate the probability distribution of backbone dihedral angles by binning combinations of ϕ and ψ over $5^\circ \times 5^\circ$ intervals accumulated over statistically different time points.

We also performed simulations of the Ala dipeptide mimetic in explicit water using the protein force fields Amber99sb-ILDN-NMR and CHARMM27-CMAP within the GROMACS 4.5.5 simulation package [48, 72]. Amber99sb-ILDN-NMR refers to the Amber99sb* force field [59, 60] combined with the ILDN side-chain optimization [63] and NMR corrections [64]. The NMR corrections optimize the dihedral angle potentials independently to match the ϕ and ψ values observed in NMR experiments of proteins. The CHARMM27-CMAP combines the CHARMM27 force field [61, 62] with the CMAP knowledge-based correction [65] so that the backbone dihedral angle correlations match those found in a curated database of high-resolution protein crystal structures.

The Amber and CHARMM force-field MD simulations were carried out in the isobaric-isothermal (NPT) ensemble using a stochastic velocity rescaling thermostat [73] and Parrinello-Rahman barostat [74]. The temperature and pressure were maintained at $T = 303$ K and $P = 1$ atm, respectively, using a coupling constant of 2 ps for the thermostat and barostat. Periodic boundary conditions were applied to a $3 \times 3 \times 3$ nm^3 box that contained approximately 880 water molecules. The long-range electrostatic interactions were calculated using the particle-mesh Ewald method [75] with a real-space cut-off of 1 nm. The van der Waals interactions were smoothly decreased to zero between 0.7 and 0.9 nm. The bond lengths were constrained using the linear constraint solver (LINCS) algorithm [76]. The equations of motion were integrated for a total time of 500 ns using the leap-frog algorithm with a time step of 2 fs. The ψ decorrelation times are ≈ 120 ps and 150 ps for Amber and CHARMM, respectively (Fig. S2), which indicates that our simulations are sufficiently long for the dipeptide mimetic to sample the relevant dihedral angle space. For the simulations with the Amber and CHARMM force fields, we employed the Ewald-corrected 4-point water model (TIP4P-Ew) [77] and TIP3P water models [78], respectively.

5 Acknowledgments

We thank Prof. Y. Wu (Hong Kong University of Science and Technology) for providing his 2013 Alanine coil library as well as Dr. R. L. Dunbrack, Jr. (Fox Chase Cancer Center) for the used version of the Dunbrack database. We thank R. L. Dunbrack as well as J. Richardson and D. Richardson for useful insights and discussion.

We also acknowledge support from the National Science Foundation grant No. PHY-1019147 and the Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences. A. Q. Zhou is a Howard Hughes Medical Institute International Research Fellow. This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center and by the National Science Foundation under grant No. CNS 08-21132 that partially funded acquisition of the facilities. This research was also supported by Academy of Finland,

Emil Aaltonen Foundation, and a Marie Curie Career Integration Grant within the 7th European Community Framework Programme under grant agreement 293861. Computational resources by CSC IT Centre for Science, Finland, are gratefully acknowledged.

References

[1] Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181:662–666.

[2] Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT (1960) Structure of Haemoglobin: A three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by x-ray analysis. Nature 185:416–422.

[3] Ramakrishnan C, Ramachandran GN (1965) Stereochemical criteria for polypeptide and protein chain conformations. Biophys J 5:909–933.

[4] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The protein data bank: A computer-based archival file for macromolecular structures. J Mol Biol 112:535–542.

[5] Headd JJ, Immormino RM, Keedy DA, Emsley P, Richardson DC, Richardson JS (2009) Autofix for backward-fit sidechains: Using MolProbity and real-space refinement to put misfits in their place. J Struct Funct Genomics 10:83–93.

[6] Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystalogr 26:283–291.

[7] Chou PY, Fasman G (1974) Prediction of protein conformation. Biochemistry 13:222–245.

[8] Jha AK, Colubri A, Zaman MH, Koide S, Sosnick TR, Freed KF (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. Biochemistry 44:9691–9702.

[9] Beck DAC, Alonso DOV, Inoyama D, Daggett V (2008) The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. Proc Natl Acad Sci 105:12259–12264.

[10] Jiang F, Han W, Wu YD (2013) The intrinsic conformational features of amino acids from a protein coil library and their applications in force field development. Phys Chem Chem Phys 15:3413–3428.

[11] Fitzkee NC, Fleming PJ, Rose GD (2005) The protein coil library: A structural database of nonhelix, nonstrand fragments derived from the PDB. Proteins 58:852–854.

[12] Chakrabartty A, Schellman JA, Baldwin RL (1991) Large differences in the helix propensities of alanine and glycine. Nature 351:586–588.

[13] Horovitz A, Matthews JM, Fersht AR (1992) α-helix stability in proteins. II. Factors that influence stability at an internal position. J Mol Biol 227:560–568.

[14] Zhou NE, Kay CM, Sykes BD, Hodges RS (1993) A single-stranded amphipathic α-helix in aqueous solution: Design, structural characterization, and its application for determining α-helical propensities of amino acids. Biochemistry 32:6190–6197.

[15] Kim CA, Berg JM (1993) Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. Nature 362:267–270.

[16] Blaber M, Zhang XJ, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW (1994) Determination of α-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. J Mol Biol 235:600–624.

[17] Creamer TP, Rose GD (1994) α-helix-forming propensities in peptides and proteins. Proteins 19:85–97.

[18] Minor DL, Kim PS (1994) Measurement of the β-sheet-forming propensities of amino acids. Nature 367:660–663.

[19] Bai Y, Englander SW (1994) Hydrogen bond strength and β-sheet propensities: The role of a side chain blocking effect. Proteins 18:262–266.

[20] Smith CK, Withka JM, Regan L (1994) A thermodynamic scale for the β-sheet forming tendencies of the amino acids. Biochemistry 33:5510–5517.

[21] Smith CK, Regan L (1995) Guidelines for protein design: The energetics of β sheet side chain interactions. Science 270:980–982.

[22] Otzen DE, Fersht AR (1995) Side-chain determinants of β-sheet stability. Biochemistry 34:5718–5724.

[23] Luque I, Mayorga OL, Freire E (1996) Structure-based thermodynamic scale of α-helix propensities in amino acids. Biochemistry 35:13681–13688.

[24] Myers JK, Pace CN, Scholtz JM (1997) A direct comparison of helix propensity in proteins and peptides. Proc Natl Acad Sci USA 94:2833–2837.

[25] Bahar I, Kaplan M, Jernigan RL (1997) Short-range conformational energies, secondary structure propensities, and recognition of correct sequence-structure matches. Proteins 29:292–308.

[26] Pal D, Chakrabarti P (2000) β-sheet propensity and its correlation with parameters based on conformation. Acta Crystallogr D 56:589–594.

[27] Avbelj F, Baldwin RL (2002) Role of backbone solvation in determining thermodynamic β propensities of the amino acids. Proc Natl Acad Sci USA 1309-1313:99.

[28] Miller JS, Kennedy RJ, Kemp DS (2002) Solubilized, spaced polyalanines: A context-free system for determining amino acid α-helix propensities. J Am Chem Soc 124:945–962.

[29] Drozdov AN, Grossfield A, Pappu RV (2004) Role of solvent in determining conformational preferences of alanine dipeptide in water. J Am Chem Soc 126:2574–2581.

[30] Kwac K, Lee KK, Han JB, Oh KI, Cho M (2008) Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: Comparisons with IR and vibrational circular dichroism spectra. J Chem Phys 128:105106–105106.

[31] Ishizuka R, Huber GA, McCammon JA (2010) Solvation effect on the conformations of alanine dipeptide: Integral equation approach. J Phys Chem Lett 1:2279–2283.

[32] García-Prieto FF, Galván IF, Aguilar MA, Martín ME (2011) Study on the conformational equilibrium of the alanine dipeptide in water solution by using the averaged solvent electrostatic potential from molecular dynamics methodology. J Chem Phys 135:194502–194502.

[33] Cruz V, Ramos J, Martínez-Salazar J (2011) Water-mediated conformations of the alanine dipeptide as revealed by distributed umbrella sampling simulations, quantum mechanics based calculations, and experimental data. J Phys Chem B 115:4880–4886.

[34] Nerenberg PS, Head-Gordon T (2011) Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides. J Chem Theory Comput 7:1220–1230.

[35] Georgoulia PS, Glykos NM (2011) Using J-coupling constants for force field validation: Application to hepta-alanine. J Phys Chem B 115:15221–15227.

[36] Toal S, Meral D, Verbaro D, Urbanc B, Schweitzer-Stenner R (2013) pH-independence of trialanine and the effects of termini blocking in short peptides: A combined vibrational, NMR, UVCD, and molecular dynamics study. J Phys Chem B 117:3689–3706.

[37] Marqusee S, Robbins VH, Baldwin RL (1989) Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci 86:5286–5290.

[38] Dunbrack RL (2002) Rotamer libraries in the 21st century. Curr Opin Struct Biol 12:431–440.

[39] Grdadolnik J, Grdadolnik SG, Avbelj F (2008) Determination of conformational preferences of dipeptides using vibrational spectroscopy. J Phys Chem B 112:2712–2718.

[40] Eker F, Griebenow K, Schweitzer-Stenner R (2003) Stable conformations of tripeptides in aqueous solution studied by UV circular dichroism spectroscopy. J Am Chem Soc 125:8178–8185.

[41] Woutersen S, Pfister R, Hamm P, Mu YG, Kosov DS, Stock G (2002) Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. J Chem Phys 117:6833–6840.

[42] Graf J, Nguyen PH, Stock G, Schwalbe H (2007) Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study. J Am Chem Soc 129:1179–1189.

[43] Schweitzer-Stenner R (2009) Distribution of conformations sampled by the central amino acid residue in tripeptides inferred from amide I band profiles and NMR scalar coupling constants. J Phys Chem B 113:2922–2932.

[44] Oh KI, Lee KK, Park EK, Yoo DG, Hwang GS, Cho M (2010) Circular dichroism eigenspectra of polyproline II and β-strand conformers of trialanine in water: Singular value decomposition analysis. Chirality 22:E186–E201.

[45] Sharma B, Asher SA (2010) UV resonance raman investigation of the conformations and lowest energy allowed electronic excited states of tri- and tetraalanine: Charge transfer transitions. J Phys Chem B 114:6661–6668.

[46] Hagarman A, Measey TJ, Mathieu D, Schwalbe H, Schweitzer-Stenner R (2010) Intrinsic propensities of amino acid residues in GxG peptides inferred from amide I band profiles and nmr scalar coupling constants. J Am Chem Soc 132:540–551.

[47] Firestine AM, Chellgren VM, Rucker SJ, Lester TE, Creamer TP (2008) Conformational properties of a peptide model for unfolded α-helices. Biochemistry 47:3216–3224.

[48] Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: Fast, flexible, and free. J Comput Chem 26:1701–1718.

[49] Case DA, Darden TA, III TEC, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Goetz AW, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh MJ, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2012) AMBER 12 University of California, San Francisco.

[50] Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package. WIREs Comput Mol Sci 3:198–210.

[51] Brooks BR, III CLB, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: The biomolecular simulation program. J Comp Chem 30:1545–1615.

[52] Schaefer L, Newton SQ, Cao M, Peeters A, Alsenoy CV, Wolinski K, Momany FA (1993) Evaluation of the dipeptide approximation in peptide modeling by ab initio geometry optimizations of oligopeptides. J Am Chem Soc 115:272–280.

[53] Jiang X, Cao M, Teppen B, Newton SQ, Schaefer L (1995) Predictions of protein backbone structural parameters from first principles: Systematic comparisons of calculated N-Cα-C' angles with high-resolution protein crystallographic results. J Phys Chem 99:10521–10525.

[54] Malathy-Sony SM, Saraboji K, Sukumar N, Ponnuswamy MN (2006) Role of amino acid properties to determine backbone τ (N-Cα-C') stretching angle in peptides and proteins. Biophys Chem 120:24–31.

[55] Zhou AQ, O'Hern CS, Regan L (2011) Revisiting the Ramachandran plot from a new angle. Prot Sci 20:1166–1171.

[56] Best RB, Buchete N, Hummer G (2008) Are current molecular dynamics force fields too helical? Biophys J 95:L07–L09.

[57] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725.

[58] Vymetal J, Vondrásek J (2013) Critical assessment of current force fields. short peptide test case. J Chem Theory Comput 9:441–451.

[59] Wang J, Cieplak P, Kollman P (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21:1049–1074.

[60] Best R, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. J Phys Chem B 113:9004–9015.

[61] Mackerell AD, Feig M, III CLB (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:14001415.

[62] Bjelkmar P, Larsson P, Cuendet M, Hess B, Lindahl E (2010) Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. J Chem Theory Comput 6:459–466.

[63] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958.

[64] Li DW, Brüschweiler R (2010) NMR-based protein potentials. Angew Chem Int Ed 49:6778–6780.

[65] MacKerell AD, Feig M, CL Brooks III (2004) Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc 126:698–699.

[66] Bowman GR, Voelz VA, Pande VS (2011) Taming the complexity of protein folding. Curr Op Struct Biol 21:4–11.

[67] Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330:341–346.

[68] Zhou AQ, O'Hern CS, Regan L (2012) The power of hard-sphere models: Explaining side-chain dihedral angle distributions of Thr and Val. Biophys J 102:2345–2352.

[69] Zhou AQ, Caballero D, O'Hern CS, Regan L (2013) New insights into the interdependence between amino acid stereochemistry and protein structure. Biophys J 105:2403–2411.

[70] Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and Glutamine: Using hydrogen atom contacts in the choice of sidechain amide orientation. J Mol Biol 285:1735–1747.

[71] Honeycut JD, Thirumalai D (1992) The nature of folded states of globular proteins. Biopolymers 32:695–709.

[72] Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447.

[73] Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101.

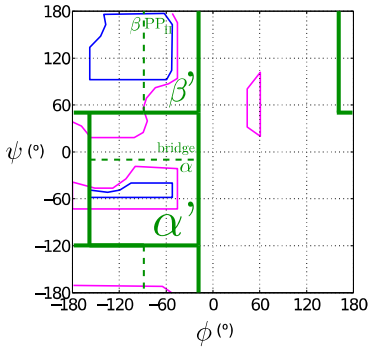
[74] Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52:7182–7190.

[75] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh ewald method. J Chem Phys 103:8577–8593.

[76] Hess B, Bekker H, Berendsen HJ, Herman JC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472.

[77] Horn H, Swope W, Pitera J, Madura J, Dick T, Hura G, Head-Gordon T (2004) Development of an improved four-site water model for biomolecular simulations: Tip4p-ew. J Chem Phys 120:9665.

[78] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, JGao, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, J MW, Wiórkiewicz-Kuczera, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616.



System	Source	α' (%)	β' (%)	Ref.
A	Dunbrack Database	65	34	[38]
A	Wu Coil-3 Database	24	74	[10]
A_2	IR	11	89	[39]
A_2	Raman	18	82	[39]
A_3	CD	0	100	[40]
A_3	IR	20	80	[41]
A_3	NMR: J-coupling	0	100	[42]
A_3	NMR: J-coupling, Raman	8	92	[43]
A_3	NMR: J-coupling, CD	0	100	[44]
A_3	Raman	0	100	[45]
GAG	NMR: J-coupling, Raman	10	85	[46]
A_5	CD	0	46	[47]
A_5	CD	10	33	[47]

Table 1: (left) 'Normal' and 'outer' Ramachandran hard-sphere limits [3] (blue and pink solid lines, respectively) for the bond angle $\tau = 110^\circ$ are overlaid on definitions of the α' and β' classifications (green solid lines). The α' region ($-160^\circ < \varphi < -20^\circ$ and $-120^\circ < \psi < 50^\circ$) includes both the classic α-helix and bridge regions (which are separated by a dashed green line). The β' region ($-180^\circ < \varphi < -20^\circ$ and $50^\circ < \psi < 180^\circ$, $-180^\circ < \varphi < -200^\circ$ and $-180^\circ < \psi < -120^\circ$, $160^\circ < \varphi < 180^\circ$ and $50^\circ < \psi < 180^\circ$) includes both the classic β-sheet and PPII regions (which are separated by a dashed green line). (right) Propensities for Ala residues

and short peptides to occur in α' and β' secondary structure classifications from the Dunbrack database [38], 'Wu Coil-3' library [10], and several experimental measurements. 'A' is Alanine, 'G' is Glycine, 'IR' is infrared spectroscopy, 'Raman' is Raman spectroscopy, 'CD' is circular dichroism, 'NMR' is nuclear magnetic resonance. A_i indicates a peptide with i amino acids.

Source	α' (%)	β' (%)	Ref.
Hard-sphere Force Field	26	68	[current work][55]
Amber99sb + tip4p-ew	26	72	[59][60]
Amber99sb-ildn-nmr + tip4p-ew	27	72	[63][64]
CHARMM27 + tip3sp	52	47	[61][62]
CHARMM27-CMAP + tip3sp	45	48	[65]

Table 2: Probabilities for the Ala dipeptide mimetic to occur in the α' and β' regions for the hard-sphere model as well as CHARMM and Amber MD simulations.

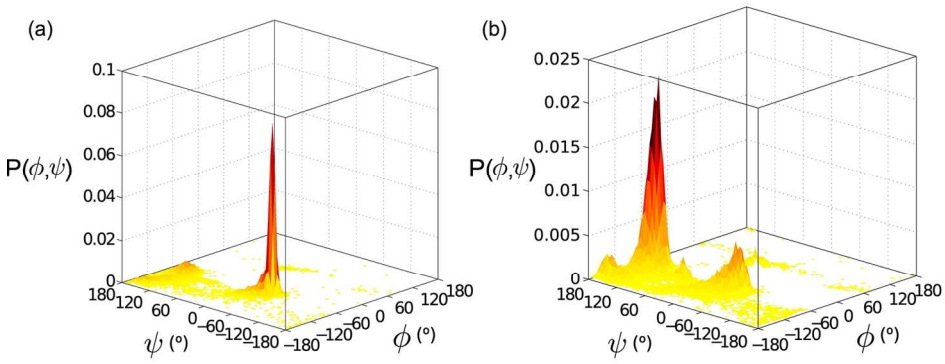


Figure 1: Probability distribution P(φ,ψ) of Ala backbone dihedral angles φ and ψ in proteins of known structure, shown for clarity in 3D. P(φ,ψ) is normalized so that its integral over all φ and ψ is unity. (a) Data from the Dunbrack Database [38] (16477 Ala residues extracted from 850 high-resolution, non-homologous protein structures with resolution ≤ 1.7Å, side chain B-factors per residue < 40Å2 and R-factors ≤ 0.25, see Materials and Methods). Note the large α-helix peak. (b) Data from the Wu 'Coil-3' library [10] (20761 Ala residues extracted from 6178 non-homologous protein structures with resolution < 2.0Å and R factor < 0.2, see Materials and Methods). β-sheet structures now predominate.

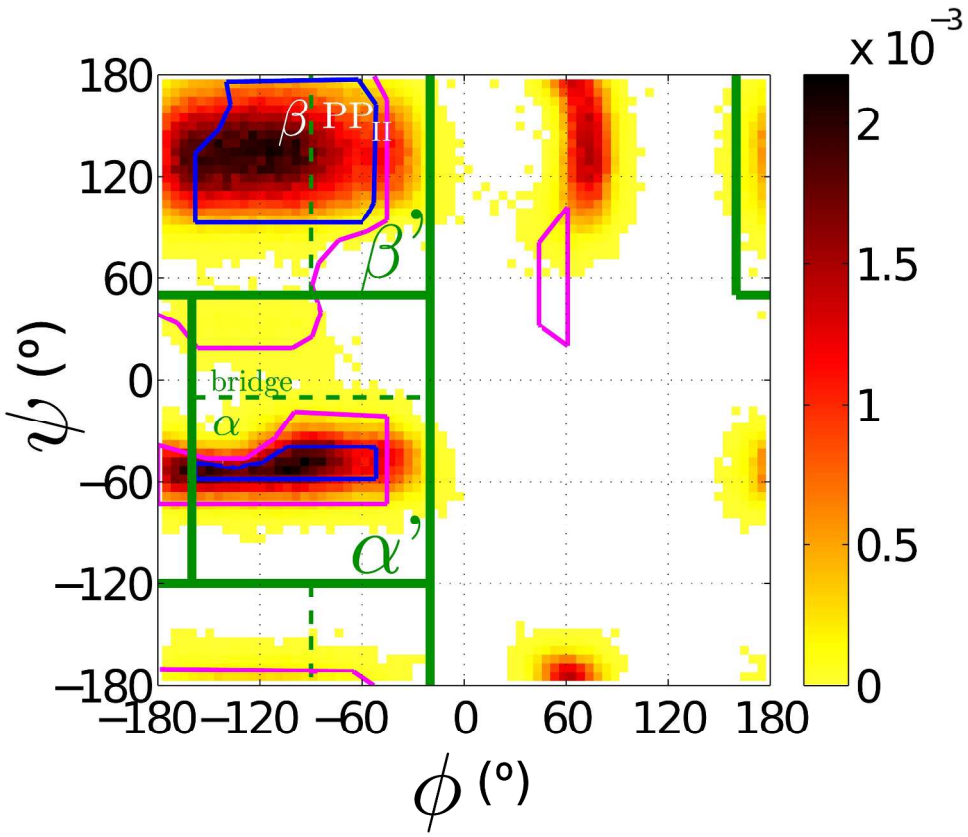


Figure 2: The probability distribution P(φ, ψ) for the hard-sphere model of the Ala dipeptide mimetic. The normal and outer Ramachandran hard-sphere limits [3] (blue and pink solid lines, respectively) for the bond angle τ = 110◦ [3] and definitions of the α′ and β′ classifications (thick green solid lines) are overlaid on the image. The α′ region (−160◦ < φ < −20◦ and −120◦ < ψ < 50◦) includes both the classic α-helix and bridge regions (which are separated by a horizontal dashed green line). The β′ region(−180◦ < φ < −20◦ and 50◦ < ψ < 180◦, −180◦ < φ < −200◦ and −180◦ < ψ < −120◦, 160◦ < φ < 180◦ and 50◦ < ψ < 180◦) includes both the classic β-sheet and PPII regions (which are separated by a vertical dashed green line).

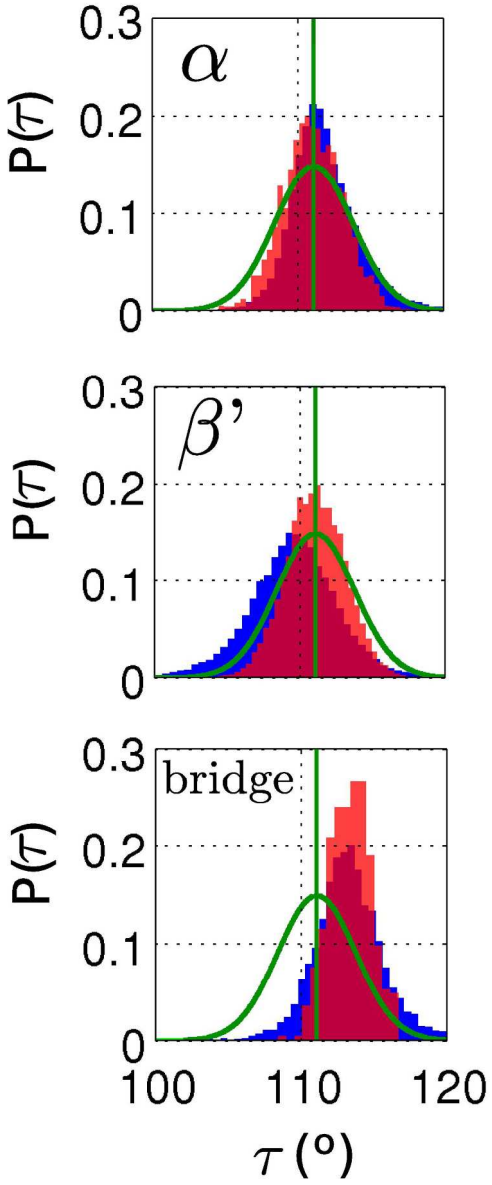


Figure 3: The probability distribution P(τ) of the bond angle τ obtained from the hard-sphere MD simulations (red shading) of the Ala dipeptide mimetic in each of three separate regions, α (top), β' (middle), and bridge (bottom), compared to an 'ideal' P (τ) (green solid line) inferred from a Boltzmann distribution only including the bond-angle potential energy (Eq. 3). P(τ) in each of the three regions obtained from the database of high-resolution protein crystal structures (blue shading) is also shown. The vertical line indicates the average of the 'ideal' distribution.

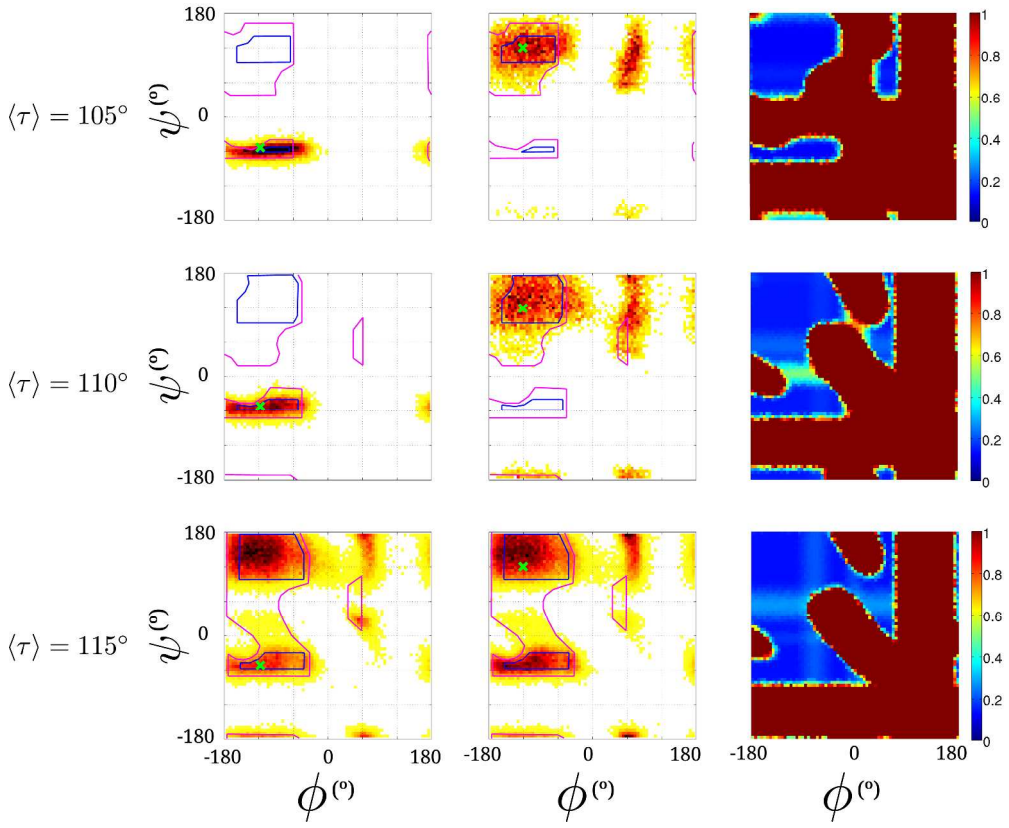


Figure 4: The distribution of the backbone dihedral angles P(φ,ψ) from hard-sphere MD simulations of an Ala dipeptide mimetic (left two colums). Each row corresponds to structures with average bond angles ⟨τ⟩ = 105◦ ± 1◦, 110◦ ± 1◦, and 115◦ ± 1◦, respectively. The normal and outer Ramachandran hard-sphere limits [3] (blue and pink solid lines, respectively) [3] are overlaid on P(φ,ψ) in the first two columns. The MD simulations were initialized in α-helix (first column) and β-sheet (second column) conformations indicated by the green '×' and run at temperature T = 10−2ε. The third column gives the average potential energy (in units of ε, see Eq. 1 in Materials and Methods) for the hard-sphere model of the Ala dipeptide mimetic at each φ and ψ for each average bond angle ⟨τ⟩.

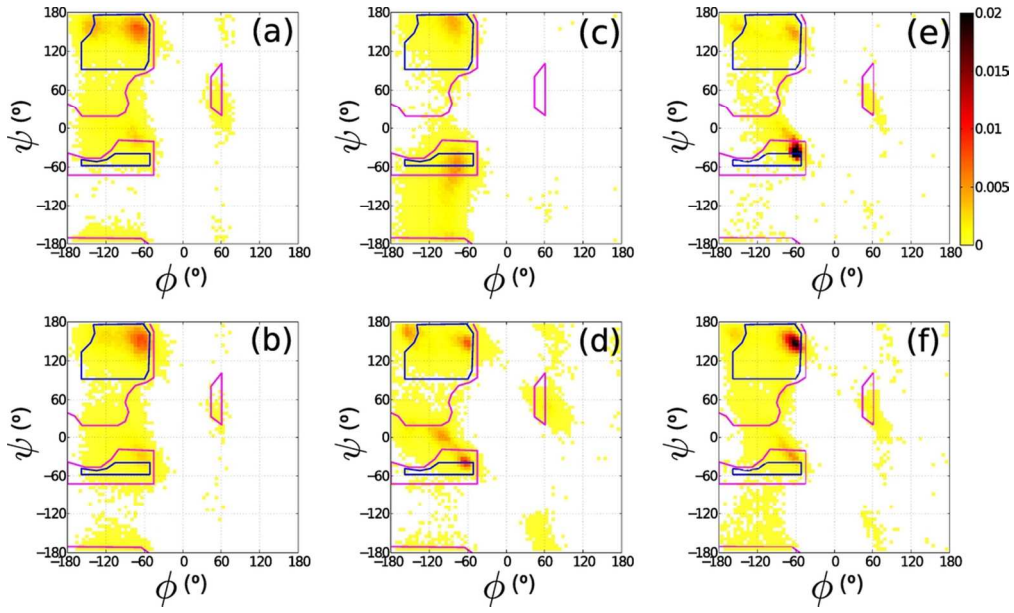


Figure 5: Probability distributions P(φ,ψ) for the backbone dihedral angles φ and ψ obtained from MD simulations of an Ala dipeptide mimetic using recent versions of the CHARMM and Amber force fields, their associated optimzied water models, and with and without the 'ildn-nmr' and 'CMAP' dihedral angle potential corrections: (a) Amber99sb + TIP4P-ew, (b) Amber99sb-ildn-nmr + TIP4P-ew, (c) CHARMM27 + TIP3SP, and (d) CHARMM27-CMAP+TIP3SP. Subpanels (e) and (f) correspond to the Ala φ-ψ distributions from the Dunbrack Database [38] and the Wu 'Coil-3' library [10], respectively. The Ramachandran hard-sphere [3] normal and outer limits (pink and blue lines, respectively) for $\tau = 110^{\circ}$ are overlaid on each panel. The Amber and CHARMM MD simulations were thermally equilibrated at 303 K and sampled for 500 ns.

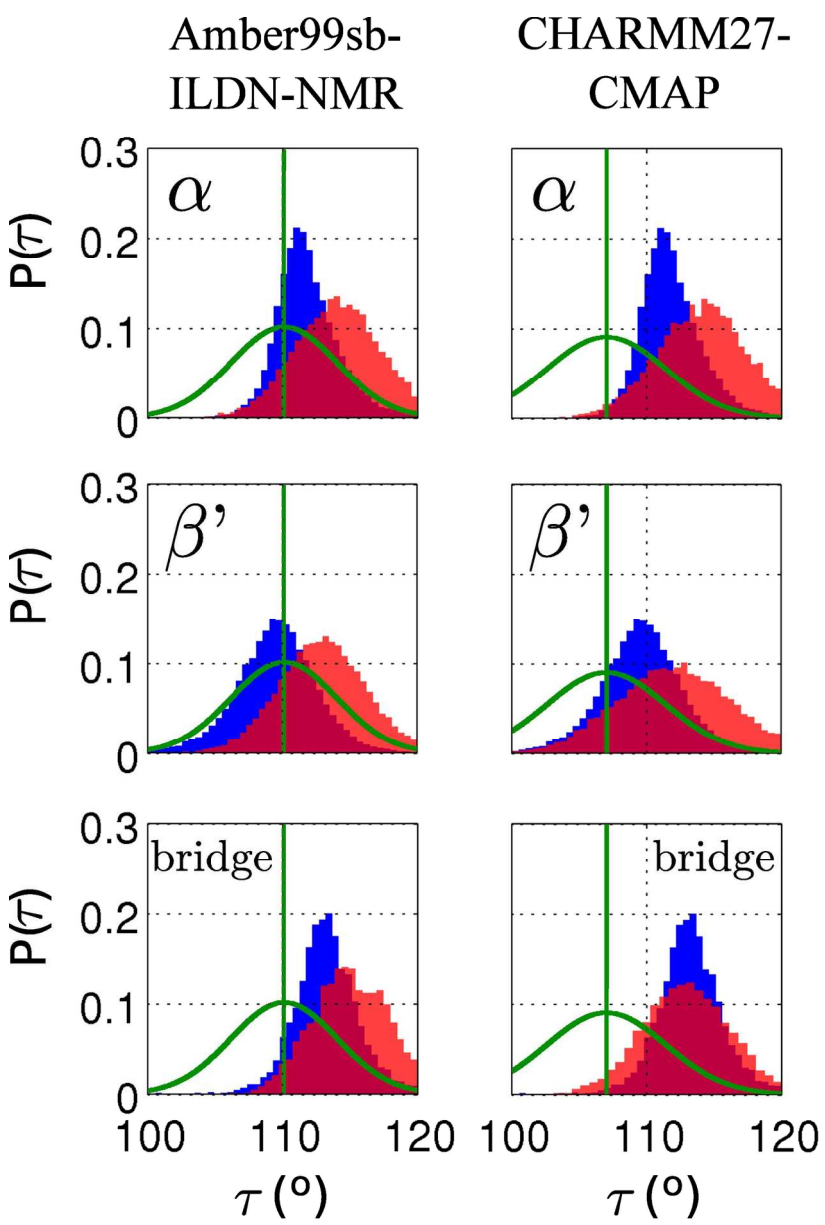
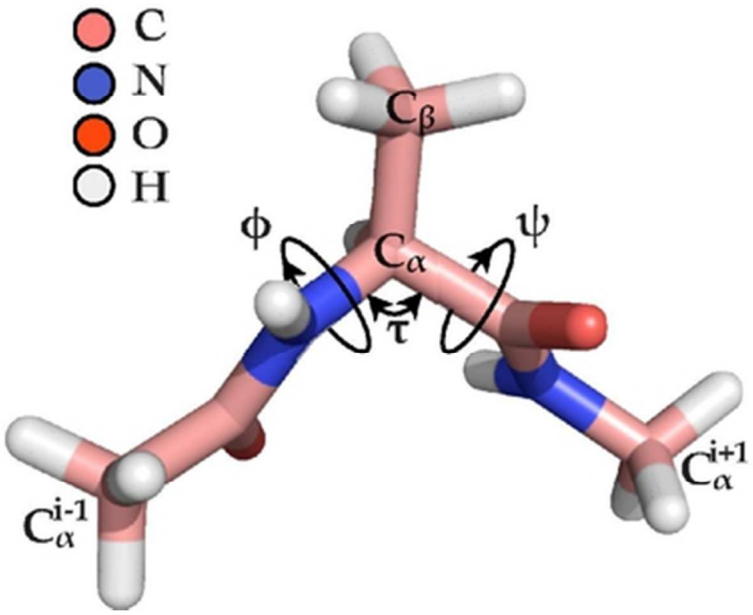


Figure 6: The probability distribution P(τ) of the bond angles τ obtained from the Amber99sb-ildn-nmr + TIP4P-ew (left) and CHARMM27-CMAP+TIP3SP (right) MD simulations (red shading) of the Ala dipeptide mimetic in each of three separate regions, α (top), β' (middle), and bridge (bottom), compared to an 'ideal' P(τ) (green solid line) inferred from a Boltzmann distribution only including the bond-angle potential energy. P(τ) in each of the three regions obtained from the database of high-resolution protein crystal structures (blue shading) is also shown. The vertical line indicates the average of the 'ideal' distribution.



Stick representation of the Ala dipeptide mimetic, N-acetyl-L-Ala-methylamide. The backbone dihedral angles φ and ψ and bond angle τ are indicated. The backbone atoms Cα, Cβ, and Cαi±1 are also labelled.

Table 1: (left) 'Normal' and 'outer' Ramachandran hard-sphere limits [3] (blue and pink solid lines, respectively) for the bond angle $\tau = 110^\circ$ are overlaid on definitions of the α' and β' classifications (green solid lines). The α' region ($-160^\circ < \phi < -20^\circ$ and $-120^\circ < \psi < 50^\circ$) includes both the classic α-helix and bridge regions (which are separated by a dashed green line). The β' region ($-180^\circ < \phi < -20^\circ$ and $50^\circ < \psi < 180^\circ$, $-180^\circ < \phi < -200^\circ$ and $-180^\circ < \psi < -120^\circ$, $160^\circ < \phi < 180^\circ$ and $50^\circ < \psi < 180^\circ$) includes both the classic β-sheet and PPII regions (which are separated by a dashed green line). (right) Propensities for Ala residues and short peptides to occur in α' and β' secondary structure classifications from the Dunbrack database [38], 'Wu Coil-3' library [10], and several experimental measurements. 'A' is Alanine, 'G' is Glycine, 'IR' is infrared spectroscopy, 'Raman' is Raman spectroscopy, 'CD' is circular dichroism, 'NMR' is nuclear magnetic resonance. A_i indicates a peptide with i amino acids.

Table 2: Probabilities for the Ala dipeptide mimetic to occur in the α' and β' regions for the hard-sphere model as well as CHARMM and Amber MD simulations.

Figure 1: Probability distribution $P(\phi,\psi)$ of Ala backbone dihedral angles ϕ and ψ in proteins of known structure, shown for clarity in 3D. $P(\phi,\psi)$ is normalized so that its integral over all ϕ and ψ is unity. (a) Data from the Dunbrack Database [38] (16477 Ala residues extracted from 850 high-resolution, non-homologous protein structures with resolution ≤ 1.7Å, side chain B-factors per residue < 40Å^2 and R-factors ≤ 0.25, see Materials and Methods). Note the large α-helix peak. (b) Data from the Wu 'Coil-3' library [10] (20761 Ala residues extracted from 6178 non-homologous protein structures with resolution < 2.0Å and R factor < 0.2, see Materials and Methods). β-sheet structures now predominate.

Figure 2: The probability distribution $P(\phi, \psi)$ for the hard-sphere model of the Ala dipeptide mimetic. The normal and outer Ramachandran hard-sphere limits [3] (blue and pink solid lines, respectively) for the bond angle $\tau = 110^\circ$ [3] and definitions of the α' and β' classifications (thick green solid lines) are overlaid on the image. The α' region ($-160^\circ < \phi < -20^\circ$ and $-120^\circ < \psi < 50^\circ$) includes both the classic α-helix and bridge regions (which are separated by a horizontal dashed green line). The β' region($-180^\circ < \phi < -20^\circ$ and $50^\circ < \psi < 180^\circ$, $-180^\circ < \phi < -200^\circ$ and $-180^\circ < \psi < -120^\circ$, $160^\circ < \phi < 180^\circ$ and $50^\circ < \psi < 180^\circ$) includes both the classic β-sheet and PPII regions (which are separated by a vertical dashed green line).

Figure 3: The probability distribution $P(\tau)$ of the bond angle τ obtained from the hard-sphere MD simulations (red shading) of the Ala dipeptide mimetic in each of three separate regions, α (top), β' (middle), and bridge (bottom), compared to an 'ideal' $P(\tau)$ (green solid line) inferred from a Boltzmann distribution only including the bond-angle potential energy (Eq. 3). $P(\tau)$ in each of the three regions obtained from the database of high-resolution protein crystal structures (blue shading) is also

shown. The vertical line indicates the average of the 'ideal' distribution.

Figure 4: The distribution of the backbone dihedral angles $P(\phi,\psi)$ from hard-sphere MD simulations of an Ala dipeptide mimetic (left two colums). Each row corresponds to structures with average bond angles $\langle\tau\rangle = 105^\circ \pm 1^\circ$, $110^\circ \pm 1^\circ$, and $115^\circ \pm 1^\circ$, respectively. The normal and outer Ramachandran hard-sphere limits [3] (blue and pink solid lines, respectively) [3] are overlaid on $P(\phi,\psi)$ in the first two columns. The MD simulations were initialized in α-helix (first column) and β-sheet (second column) conformations indicated by the green '×' and run at temperature $T = 10^{-2}\varepsilon$. The third column gives the average potential energy (in units of ε, see Eq. 1 in Materials and Methods) for the hard-sphere model of the Ala dipeptide mimetic at each ϕ and ψ for each average bond angle $\langle\tau\rangle$.

Figure 5: Probability distributions $P(\phi,\psi)$ for the backbone dihedral angles ϕ and ψ obtained from MD simulations of an Ala dipeptide mimetic using recent versions of the CHARMM and Amber force fields, their associated optimzied water models, and with and without the 'ildn-nmr' and 'CMAP' dihedral angle potential corrections: (a) Amber99sb + TIP4P-ew, (b) Amber99sb-ildn-nmr + TIP4P-ew, (c) CHARMM27 + TIP3SP, and (d) CHARMM27-CMAP+TIP3SP. Subpanels (e) and (f) correspond to the Ala ϕ-ψ distributions from the Dunbrack Database [38] and the Wu 'Coil-3' library [10], respectively. The Ramachandran hard-sphere [3] normal and outer limits (pink and blue lines, respectively) for $\tau = 110^\circ$ are overlaid on each panel. The Amber and CHARMM MD simulations were thermally equilibrated at 303 K and sampled for 500 ns.

Figure 6: The probability distribution $P(\tau)$ of the bond angles τ obtained from the Amber99sb-ildn-nmr + TIP4P-ew (left) and CHARMM27-CMAP+TIP3SP (right) MD simulations (red shading) of the Ala dipeptide mimetic in each of three separate regions, α (top), β' (middle), and bridge (bottom), compared to an 'ideal' $P(\tau)$ (green solid line) inferred from a Boltzmann distribution only including the bond-angle potential energy. $P(\tau)$ in each of the three regions obtained from the database of high-resolution protein crystal structures (blue shading) is also shown. The vertical line indicates the average of the 'ideal' distribution.

Figure 7: Stick representation of the Ala dipeptide mimetic, N-acetyl-L-Ala-methylamide. The backbone dihedral angles ϕ and ψ and bond angle τ are indicated. The backbone atoms C_α, C_β, and $C_\alpha^{i\pm1}$ are also labelled.

Intrinsic α-helical and β-sheet preferences: A computational case study of Alanine - Supplementary Material

Diego Caballero[1,2], Jukka Määttä[3], Alice Q. Zhou[2,4], Maria Sammalkorpi[3], Corey S. O'Hern[1,2,5,6], and Lynne Regan[2,4,7]

[1]Department of Physics, Yale University;
[2]Integrated Graduate Program in Physical and Engineering Biology, Yale University;
[3]Department of Chemistry, Aalto University;
[4]Department of Molecular Biophysics & Biochemistry, Yale University
[5]Department of Mechanical Engineering & Materials Science, Yale University;
[6]Department of Applied Physics, Yale University;
[7]Department of Chemistry, Yale University.

April 16, 2014

1 Equilibration of the MD simulations of the Ala dipeptide mimetic

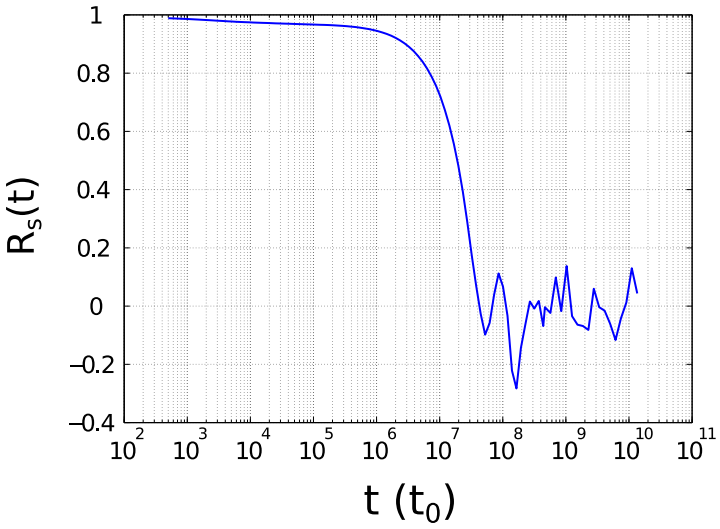


Figure S1: Autocorrelation function $R_s(t)$ that measures transitions from the α to the β' region and vice versa in the Ramachandran map as a function of time t (in units of an atomic collision time t_0) for the hard-sphere model of the Ala dipeptide mimetic with a bond angle τ distribution extracted from Dunbrack Database [1].

To ensure that the molecular dynamics (MD) simulations of the Ala dipeptide mimetic were thermally equilibrated, we employed two measurements of the structural relaxation times. For the first, we defined a binary variable s that can assume the two values, 0 or 1, depending on ψ:

$$s = \begin{cases} 1 & : \psi > 50^\circ \\ 0{,}1 & : -10^\circ < \psi < 50^\circ \\ 0 & : \psi < -10^\circ \end{cases} \quad (1)$$

The value of s for ψ in the bridge region $-10^\circ < \psi < 50^\circ$ changes from 0 to 1 if the system makes a transition from the α to the β' region, otherwise it remains 0. Similarly, s will remain 1 unless the system makes a transition from region β' to α. Thus, the autocorrelation function of s

$$R_s(t) = \frac{\langle (s(t) - \langle s \rangle)^2 \rangle}{\sigma_s^2}, \tag{2}$$

where the angle brackets denote an average over time t and $\sigma_s = \langle s^2 \rangle - \langle s \rangle^2$, measures the average transition time from α to β' and vice versa. In Fig. S1, we show $R_s(t)$ for the hard-sphere model for the Ala dipeptide mimetic with $\langle \tau \rangle = 115^\circ$. Correlations in s decay to zero within a decorrelation time $t_r \approx 10^8 t_0$. We then equilibrated the hard-sphere systems for $100t_r$ before measuring the α' and β' propensities.

For the second measure, we determined the time required for the backbone dihedral angles ϕ and ψ to become decorrelated by calculating the autocorrelation function [2]

$$C(t) = \langle \cos[\psi(t')] \cos[\psi(t'+t)] + \sin[\psi(t')] \sin[\psi(t'+t)] \rangle, \tag{3}$$

where the angle brackets indicate an average over time origins t', and a similar function can be defined for ϕ. To exract a decorrelation time t_r, Eq. 3 can be approximated using a single exponential decay

$$C(t) \approx S_D^2 + (1 - S_D^2) \exp\left(-\frac{t}{t_r}\right), \tag{4}$$

where $S_D^2 = C(\infty)$. In Fig. S2, we show $C(t)$ (along with fits to Eq. 4) for both ϕ and ψ using the Amber and CHARMM molecular dynamics simulations of the Ala dipeptide mimetic. We find that $t_r \approx 117$ ps and 153 ps from ψ decorrelations for Amber and CHARMM, respectively. Note that the ϕ correlation function cannot be fit by a single exponential relaxation for CHARMM. We integrated the Amber and CHARMM MD simulations for more than 500 ns to allow the systems to become independent of their initial conditions.

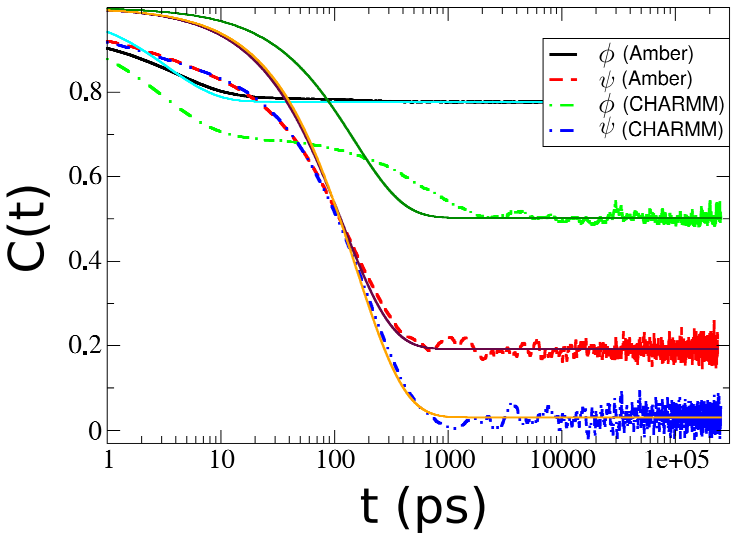


Figure S2: Backbone dihedral angle correlation functions $C(t)$ (with associated fits to Eq. 4) for the Amber and CHARMM molecular dynamics simulations of the Ala dipeptide mimetic.

2 Hard-Sphere Temperature

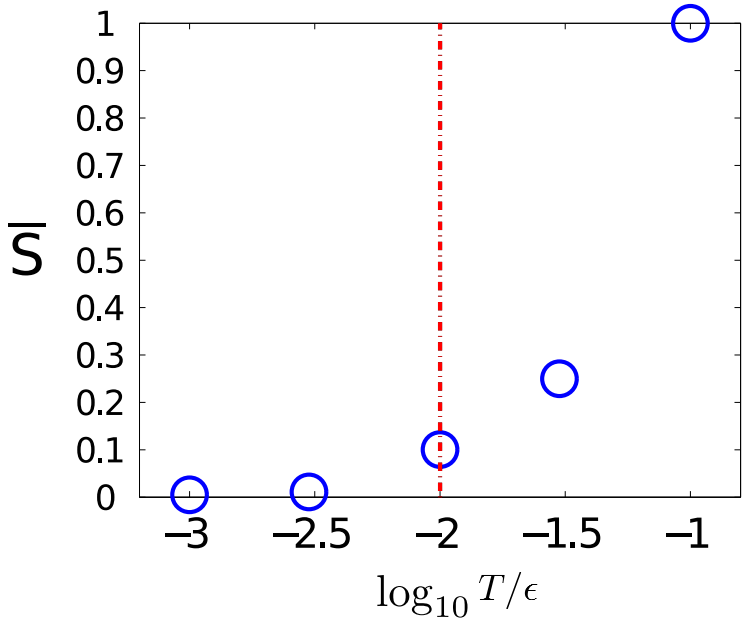


Figure S3: The time-averaged value of the parameter $\overline{S}$, which quantifies the degree to which the MD simulations with non-bonded purely repulsive Lennard-Jones interactions mimic hard-sphere interactions, plotted versus temperature T/ϵ. $\overline{S}$ is normalized so that the maximum value is 1 at the highest temperature studied. The vertical line at $T/\epsilon = 10^{-2}$ indicates the temperature at which most of our MD simulations were performed.

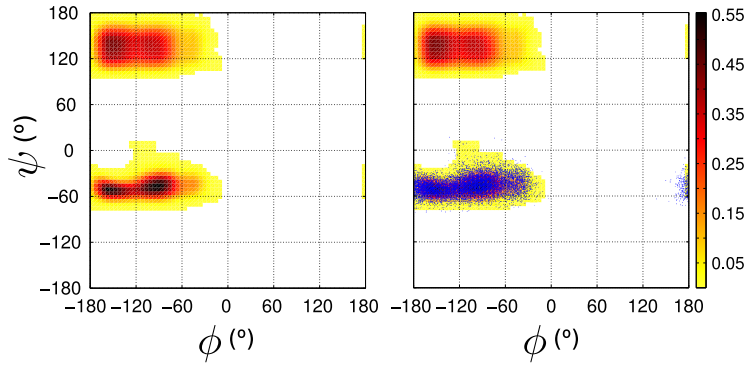


Figure S4: (left) Probability distribution $P(\phi, \psi)$ of sterically allowed backbone dihedral angle combinations for Ala dipeptide mimetics at $\tau = 105°$. (right) Backbone dihedral angle combinations ϕ and ψ (blue dots) sampled during a long MD simulation run with purely repulsive Lennard-Jones interactions between non-bonded atoms with with $\langle\tau\rangle = 105°$ at $T/\epsilon = 10^{-2}$ initialized in an α-helix backbone conformation and overlaid on $P(\phi, \psi)$. Increasing probability is indicated by a color scale from white to yellow to black.

For the MD simulations that employ repulsive Lennard-Jones interactions plus stereochemical constraints, we must choose an appropiate temperature for conducting the simulations. We were interested in performing the simulations at temperatures $T < T_{\rm HS}$, in the low temperature regime, where the non-bonded interactions behave as hard-sphere interactions with minimal interatomic overlaps. To quantify T_{HS}, we defined the following metric that includes weights for both the number and severity of interatomic clashes:

$$\overline{S} = \langle 1 - S_{ij} \rangle \frac{N_d}{N_t}, \tag{5}$$

where N_d/N_t is the fraction of frames in the simulation where at least one nonbonded interatomic overlap is recorded, the severity of the overlap between atoms i and j is

$$S_{ij} = \frac{r_{ij}}{\sigma_{ij}}\Theta(\sigma_{ij} - r_{ij}), \tag{6}$$

where r_{ij} is the separation between atoms i and j, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, σ_i is the diameter of atom i, $\Theta(x)$ is the Heaviside step function, and the angle brackets denote an average over all non-bonded interactions between atoms i and j.

In Fig. S3, we show the time-averaged $\overline{S}$ as a function of temperature T/ϵ. $\overline{S}$ begins to increase strongly for $T/\epsilon \gtrsim 10^{-2.5}$. We thus performed the MD simulations with nonbonded repulsive Lennard-Jones interactions at $T/\epsilon = 10^{-2}$ so that the simulations were approximately in the hard-sphere regime.

Another method to quantify the extent to which the MD simulations with purely repulsive Lennard-Jones interactions mimic hard-sphere behavior is shown in Fig. S4. Using the same method described in Ref. [3], we first calculated the probability distribution $P(\phi, \psi)$ for all of the sterically allowed backbone dihedral angle combinations for Ala dipeptide mimetics (with $\tau \sim 105°$). We then compared the sterically allowed $P(\phi, \psi)$ to the ϕ and ψ combinations sampled by the MD simulations at $\langle\tau\rangle = 105°$, which shows that the simulations explore only the sterically allowed conformations.

3 Ramachandran plot as a function of the size of Oxygen

In Refs. [4, 3, 5], we calibrated the atomic sizes so that the distributions of sterically allowed dihedral angle combinations for Leu, Ile, Thr, and Ser matched those observed in protein crystal structures. This calibration yielded an Oxygen size of 1.45Å. However, by reducing the size of oxygen by 0.05Å, we find that the sterically allowed backbone conformation space does not change significantly, yet we are able to dramatically increase the transition rate between α-helix and β-sheet backbone conformations, and vice-versa, in the Ala dipeptide mimetic. In Fig. S5, we show the probability $P(\phi, \psi)$ for sterically allowed dihedral angle combinations for the Ala dipeptide mimetic as a function of the radius of the oxygen atom, 1.45Å, 1.40Å, 1.35Å, 1.30Å, and 1.25Å. We find that the distributions are very similar for 1.45Å and 1.40Å, and thus we chose 1.40Å to increase the configurational smapling of ϕ-ψ space.

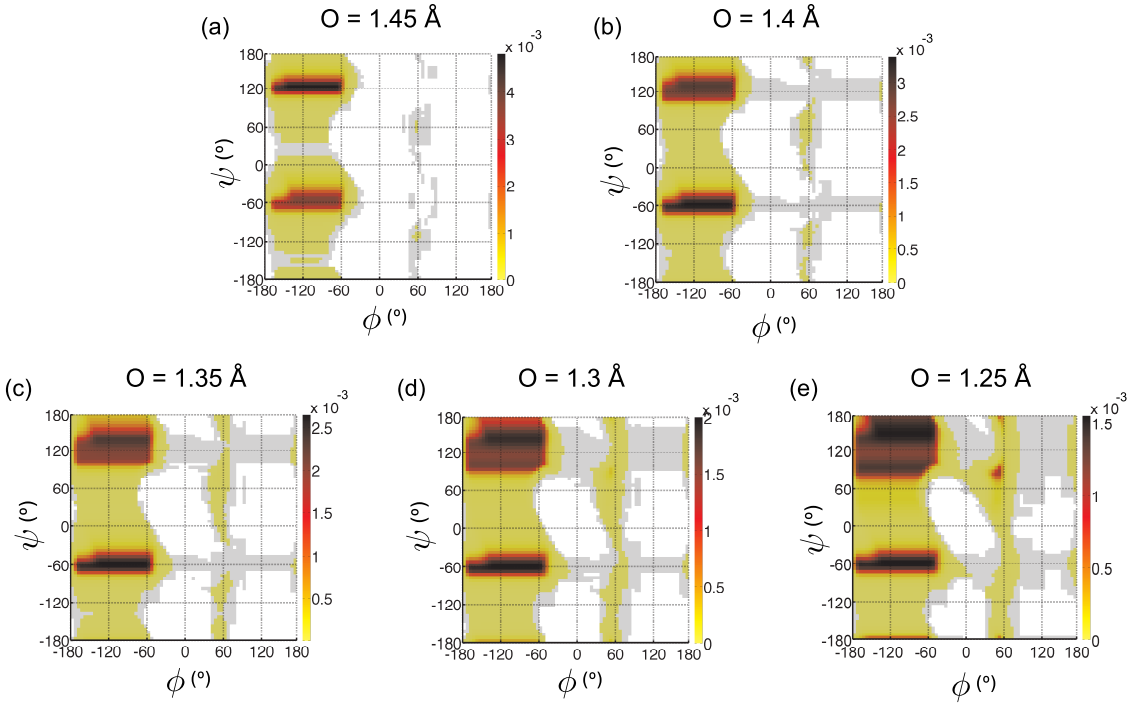


Figure S5: The probability distribution of sterically allowed backbone dihedral angle combinations ϕ and ψ for the Ala dipeptide mimetic as a function of the radius of the oxygen atom, 1.45Å, 1.40Å, 1.35Å, 1.30Å, and 1.25Å in panels (a)-(e). The probability on each panel is the sum of sterically allowed configurations as extracted from all the Alanines in the Dunbrack database [1]. Increasing probability is indicated by the color scale from white to gray to orange to black.

References

[1] Dunbrack RL (2002) Rotamer libraries in the 21st century. Curr Opin Struct Biol 12:431–440.

[2] van der Spoel D, Berendsen HJ (1977) Molecular dynamics simulations of Leu-enkephalin in water and DMSO. Biophys J 72:2032–2041.

[3] Zhou AQ, O'Hern CS, Regan L (2012) The power of hard-sphere models: Explaining side-chain dihedral angle distributions of Thr and Val. Biophys J 102:2345–2352.

[4] Zhou AQ, O'Hern CS, Regan L (2011) Revisiting the Ramachandran plot from a new angle. Prot Sci 20:1166–1171.

[5] Zhou AQ, Caballero D, O'Hern CS, Regan L (2013) New insights into the interdependence between amino acid stereochemistry and protein structure. Biophys J 105:2403–2411.