Intrinsic \(\alpha\)-helical and \(\beta\)-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 Regan2,4,7*

1Department of Physics, Yale University, New Haven, Connecticut 06520
2Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut 06520
3Department of Chemistry, Aalto University, 02150 Espoo, Finland
4Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520
5Department of Mechanical Engineering and Materials Science, Yale University, New Haven, Connecticut 06520
6Department of Applied Physics, Yale University, New Haven, Connecticut 06520
7Department of Chemistry, Yale University, New Haven, Connecticut 06520

Received 17 January 2014; Accepted 17 April 2014
DOI: 10.1002/pro.2481
Published online 00 Month 2014 proteincience.org

ABSTRACT: A fundamental question in protein science is what is the intrinsic propensity for an amino acid to be in an \(\alpha\)-helix, \(\beta\)-sheet, or other backbone dihedral angle (\(\phi\)-\(\psi\)) conformation. This question has been hotly debated for many years because including all protein crystal structures from the protein database, increases the probabilities for \(\alpha\)-helical structures, while experiments on small peptides observe that \(\beta\)-sheet-like conformations predominate. We perform molecular dynamics (MD) simulations of a hard-sphere model for Ala dipeptide mimetics that includes steric interactions between nonbonded 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) \(\beta\)-sheet structures are roughly three and half times more probable than \(\alpha\)-helical structures, (2) transitions between \(\alpha\)-helix and \(\beta\)-sheet structures only occur when the backbone bond angle \(\tau\) (N–C–C–) is greater than 110\(^\circ\), and (3) the probability distribution of \(\tau\) for Ala conformations in the “bridge” region of \(\phi\)-\(\psi\) 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 \(\tau\) 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 \(\phi\)-\(\psi\) conformations and \(\tau\).

Keywords: alanine; backbone conformations; hard-sphere; simulations; Amber; CHARMM; \(\alpha\)-helix; \(\beta\)-sheet; intrinsic propensity; \(T\)-angle

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Science Foundation; Grant number: PHY-1019147. Grant sponsor: National Science Foundation; Grant number: DMR-1307712. Grant sponsor: Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences. Grant sponsor: Facilities and staff of the Yale University Faculty of Arts and Sciences, High Performance Computing Center, National Science Foundation; Grant number: CNS 08-21132. Grant sponsor: the National Science Foundation; Grant number: CNS 08-21132. Grant sponsor: Academy of Finland. Grant sponsor: Emil Aaltonen Foundation. Grant sponsor: Marie Curie Career Integration Grant within the 7th European Community Framework Programme; Grant number: 293861. Grant sponsor: CSC IT Centre for Science.

*Correspondence to: Lynne Regan; Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT. E-mail: lynne.regan@yale.edu
Introduction

The first structure of a protein was solved over 50 years ago.1,2 Around that time, Ramachandran et al. 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 et al.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 Figure 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 used other strategies to determine “true” conformational preferences8–11 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

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 Database38 (16,477 Ala residues extracted from 850 high-resolution, nonhomologous protein structures with resolution ≤ 1.7 Å, side chain B-factors per residue < 40 Å² and R-factors ≤ 0.25, see Materials and Methods). Note the large α-helix peak. (b) Data from the Wu “Coil-3” library10 (20,761 Ala residues extracted from 6178 nonhomologous protein structures with resolution < 2.0 Å and R-factor < 0.2, see Materials and Methods). β-sheet structures now predominate.
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 I). 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 preproline 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 package48 with recent versions of the Amber49,50 and CHARMM51 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

<table>
<thead>
<tr>
<th>System</th>
<th>Source</th>
<th>α’ (%)</th>
<th>β’ (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Dunbrack database</td>
<td>65</td>
<td>34</td>
<td>38</td>
</tr>
<tr>
<td>A</td>
<td>Wu coil-3 database</td>
<td>24</td>
<td>74</td>
<td>10</td>
</tr>
<tr>
<td>A2</td>
<td>IR</td>
<td>11</td>
<td>89</td>
<td>39</td>
</tr>
<tr>
<td>A2</td>
<td>Raman</td>
<td>18</td>
<td>82</td>
<td>39</td>
</tr>
<tr>
<td>A3</td>
<td>CD</td>
<td>0</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>A3</td>
<td>IR</td>
<td>20</td>
<td>80</td>
<td>41</td>
</tr>
<tr>
<td>A3</td>
<td>NMR: J-coupling</td>
<td>0</td>
<td>100</td>
<td>42</td>
</tr>
<tr>
<td>A3</td>
<td>NMR: J-coupling, Raman</td>
<td>8</td>
<td>92</td>
<td>43</td>
</tr>
<tr>
<td>A3</td>
<td>NMR: J-coupling, CD</td>
<td>0</td>
<td>100</td>
<td>44</td>
</tr>
<tr>
<td>A3</td>
<td>Raman</td>
<td>0</td>
<td>100</td>
<td>45</td>
</tr>
<tr>
<td>GAG</td>
<td>NMR: J-coupling, Raman</td>
<td>10</td>
<td>85</td>
<td>46</td>
</tr>
<tr>
<td>A5</td>
<td>CD</td>
<td>0</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>A5</td>
<td>CD</td>
<td>10</td>
<td>33</td>
<td>47</td>
</tr>
</tbody>
</table>
hard-sphere MD simulations is consistent with the
“normal” limits\(^3\) for we also show the Ramachandran “outer” and
observed in protein crystal structures. In this plot,
is allowed to sample values from the distribution
model of the Ala dipeptide mimetic. In this model,
larger values of \(s\) \(a\) between the
bridge regions (which are separated by a horizontal dashed
green line). The
lines, respectively) for the bond angle
are overlaid on the image. The \(a\) \(\phi = 60^\circ\) and \(\psi = 60^\circ\)). This discrepancy stems from the fact that the Ramachandran \(a\) \(\phi\) \(\psi\) outer limits were based on nonadditive atomic radii and the size of this allowed region varies strongly with \(\tau\). 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
transitions between \(a\)-helices and \(\beta\)-sheets in our
hard-sphere MD simulations is consistent with the
observation that in proteins of known structure resi-
dues that populate the “bridge region” of \(\phi - \psi\) space,
between the \(a\)-helix and \(\beta\)-sheet regions, possess
larger values of \(\tau\).\(^3,\,52–55\)

Results and Discussion
In Figure 2, we show the probability distribution
\(P(\phi, \psi)\) for the backbone dihedral angle combinations
\(\phi - \psi\) for the thermally equilibrated, hard-sphere
model of the Ala dipeptide mimetic. In this model,
\(\tau\) 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 des-
ignate as \(a\) \(a\) and \(\beta\) \(\beta\). (See Table I (top) for the defini-
tions of the \(a\) \(a\) and \(\beta\) \(\beta\) regions.) \(a\) \(a\) includes both
classic \(a\)-helix and bridge regions, and \(\beta\) \(\beta\) includes
both classic \(\beta\)-sheet and PPII regions. Similar limits
have been used by others.\(^56–58\)

There are several important features in Figure 2.
\(P(\phi, \psi)\) from the hard-sphere simulations largely
respects the Ramachandran limits in the \(a\)-helix and
\(\beta\)-sheet regions. The main discrepancy in this
respect is in the Ramachandran plot in the vicinity
of \(a\) \(\phi = 60^\circ\) and \(\psi = 60^\circ\)). This discrepancy stems
from the fact that the Ramachandran \(a\) \(\phi\) \(\psi\) outer limits were based on nonadditive atomic radii and the size of this allowed region varies strongly with \(\tau\). 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

---

**Table II. Probabilities for the Ala dipeptide mimetic to occur in the \(a\) \(a\) and \(\beta\) \(\beta\) regions for the hard-sphere model as well as CHARMM and Amber MD simulations**

<table>
<thead>
<tr>
<th>Source</th>
<th>(a) (%)</th>
<th>(\beta) (%)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard-sphere force field</td>
<td>26</td>
<td>68</td>
<td>current work 55</td>
</tr>
<tr>
<td>Amber99sb + TIP4P-Ew</td>
<td>26</td>
<td>72</td>
<td>59,60</td>
</tr>
<tr>
<td>Amber99sb-ILDN-NMR + TIP4P-Ew</td>
<td>27</td>
<td>72</td>
<td>63,64</td>
</tr>
<tr>
<td>CHARMM27 + TIP3SP</td>
<td>52</td>
<td>47</td>
<td>61,62</td>
</tr>
<tr>
<td>CHARMM27-CMAP + TIP3SP</td>
<td>45</td>
<td>48</td>
<td>65</td>
</tr>
</tbody>
</table>

---

**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 \(\tau = 110^\circ\) and defini-
tions of the \(a\) \(\alpha\) and \(\beta\) \(\beta\) classifications (thick green solid lines)
are overlaid on the image. The \(a\) \(\alpha\) \(\alpha\) region \((-160^\circ < \phi < -20^\circ\)
and \(-120^\circ < \psi < 50^\circ\)) includes both the classic \(a\)-helix and
bridge regions (which are separated by a horizontal dashed
green line). The \(\beta\) \(\beta\) \(\beta\) 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
\(\beta\)-sheet and PPII regions (which are separated by a vertical
dashed green line).

---

**Figure 3.** The probability distribution \(P(\tau)\) of the bond angle \(\tau\)
obtained from the hard-sphere MD simulations (red shading)
of the Ala dipeptide mimetic in each of three separate
regions, \(a\) \(\alpha\) \(\alpha\) \(\alpha\) (top), \(\beta\) \(\beta\) \(\beta\) \(\beta\) \(\beta\) (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.
Ramachandran limits correspond to a single \( \tau = 110^\circ \). It is also immediately apparent that the \( \alpha \)-helix region is not overwhelmingly populated compared to the \( \beta \)-sheet region, in contrast with Figure 1(a). Instead, the maximum probabilities in the \( \alpha \)-helix and \( \beta \)-sheet regions are comparable and significantly greater than the maximum probability in the bridge region. We find that the probabilities in the \( \alpha' \) and \( \beta' \) regions are 26 and 68%, respectively (Table II). Furthermore, the \( \phi \)-\( \psi \) probabilities are relatively uniform within the \( \alpha' \) and \( \beta' \) regions. Thus, we can estimate the \( \alpha' \) and \( \beta' \) probabilities by the area in \( \phi \)-\( \psi \) space that they occupy, which is 31 and 69% of the total \( \phi \)-\( \psi \) space, respectively.

We next investigated the correlations between the backbone dihedral angle combinations (\( \phi \) and \( \psi \)) and the bond angle \( \tau \). In Figure 3, we show the probability distribution \( P(\tau) \) separately for each region, \( \alpha \), \( \alpha' \), 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 \( \alpha \) and \( \beta' \) 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 \( \alpha \) and \( \beta' \) regions. In addition, in the middle panel of Figure 3, we show a small shift of \( P(\tau) \) (blue) to smaller angles for crystal structures in the Dunbrack database with \( \beta' \) backbone conformations compared to \( P(\tau) \) for all structures.

We also studied the relationship between \( \tau \) and transitions between the \( \alpha \) and \( \beta' \) 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°
transitions between the $a$ and $b$ regions are never observed over the full simulation, independent of whether the Ala dipeptide mimetic is initialized in the $a$ (first column) or $b$ (second column) regions. By contrast, when $h_{5,115}/C_{14}$ (third row), transitions occur frequently between $a$ and $b$ and $P(\phi, \psi)$ is independent of the starting values of $\phi$ and $\psi$. In the third column, we show an example of the potential energy landscape as a function of $\phi$ and $\psi$ for different values of $\tau$. We find that the energy barrier in the bridge region begins to decrease for $h_{5,110}/C_{14}$ and is extremely small for $h_{5,115}/C_{14}$.

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

In contrast, CHARMM27 [Fig. 5(c,d)] populates the region $-180 < \psi < -60^\circ$, 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,d)], the relative populations of structures in the $a'$ and $b'$ regions are similar for both (see Table II).

Interestingly, we find that the $a'$ and $b'$ 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 $a'$, roughly 50%, for both $a'$ and $b'$ both with and without CMAP corrections. Similar differences between the CHARMM and Amber force fields were obtained by Vymetal and Vondrasek. We also studied the correlation between the bond angle $\tau$ and backbone dihedral angles $\phi$ and $\psi$ in the CHARMM and Amber MD 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. Although 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 Figure 4, a shift in the peak of $P(\tau)$ to larger values facilitates transitions between the $\alpha'$ and $\beta'$ regions. Note that in contrast to the hard-sphere model, the harmonic bond-angle potential energies are centered on $\tau = 110^\circ$ and $107^\circ$ 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 $\alpha'$ and $\beta'$ are observed. Thus, the hard-sphere model predicts that there must be a correlation between a large bond angle $\tau$ and the backbone dihedral angles $\phi$ and $\psi$ 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 $\tau$ 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 interdependent 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. 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 mimic. The results we present, along with our prior studies of the side-chain dihedral angle distributions of different amino acids, (AZ, CO, LR, submitted), make it clear that steric repulsion is the dominant force in specifying the allowed backbone and side-chain 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 interdependence of $\phi$, $\psi$, and $\tau$. Ramachandran had predicted and we showed for protein crystal structures that the distribution of $\phi$-$\psi$ angles depends on the value of $\tau$ (i.e., the Ramachandran plots for an Ala dipeptide mimic are different for $\tau = 105, 110$, and $115^\circ$). The studies we present here expand on that observation, and show that transitions between $\alpha$-helix and $\beta$-sheet conformations require $\tau$ 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 $\phi$-$\psi$-$\tau$ correlation term to ensure that transitions between $\alpha$-helix and $\beta$-sheet backbone conformations occur by increasing the bond angle $\tau$.

Despite decades of work, there is still considerable debate concerning the intrinsic propensities for amino acids to adopt $\alpha$-helix versus $\beta$-sheet structures. To address this issue, we performed MD simulations of an Ala dipeptide mimetic using a minimal model that includes only stereochemical constraints and hard-sphere interactions between nonbonded atoms. This model predicts probabilities for $\alpha$-helix and $\beta$-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 $\tau$ and transitions between $\alpha$-helix and $\beta$-sheet conformations. For $<\tau> < 110^\circ$, such transitions between $\alpha$-helix and $\beta$-sheet do not occur. In contrast, for $<\tau> = 115^\circ$, the Ala dipeptide is able to transition from $\alpha$-helix to $\beta$-sheet conformations. However, in MD simulations...
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 \( \phi \) and \( \psi \) dihedral angle combinations, which indicates that the other interdependent and longer-range interactions outweigh the repulsive steric interactions.

### 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 Figure 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 stereocchemical parameters, that is, the average and standard deviation of the bond lengths (\( \bar{r}_i \) and \( \Delta r_i \)), bond angles (\( \theta_{ijk} \) and \( \Delta \theta_{ijk} \)), and \( \omega \) backbone dihedral angles (\( \phi_{ijkl} \approx 0 \) and \( \Delta \phi_{ijkl} \)) obtained for Ala residues in the Dunbrack Database.\(^{38}\) This cull database is composed of 850 high-resolution, nonhomologous protein structures with resolution \( \leq 1.7 \) Å, side chain B-factors per residue <40 Å\(^2\) (local B-factor filtering), and R-factors <0.25. This dataset includes 16,477 Ala residues.

We compare our results in Figure 1(b) and Table I 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 \( \alpha \)-helices or \( \beta \)-sheets. In addition, proline and turn residues are excluded. The total number of Ala residues in the Coil-3 library is 20,761.

The atomic diameters \( \sigma_i \) are: C(sp\(^3\)) 1.5 Å, C(sp\(^2\)) 1.4 Å, N 1.4 Å, O 1.4 Å, and H 1.05 Å, which are identical to values used in previous studies,\(^{55,68,69}\) except the oxygen diameter was changed from 1.45 to 1.4 Å to improve sampling in \( \phi-\psi \) space (see Supporting Information). 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,

\[
V_{ij} = c \left( 1 - \frac{\sigma_{ij}}{r_{ij}} \right)^6 \Theta (\sigma_{ij} - r_{ij}),
\]

where \( c \) is the characteristic energy scale of the interaction, \( r_{ij} \) is the separation between nonbonded 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,

\[
V_{hi} = \frac{K_{hi}}{2} (r_{ij} - r_{ij}^0)^2,
\]

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

\[
V_{ba} = \frac{K_{ba}}{2} (\theta_{ijkl} - \theta_{ijkl}^0)^2,
\]

where \( K_{ba} = T/(\Delta \theta_{ijkl})^2 \); and (4) a harmonic potential to constrain the two \( \omega_{ijkl} \) dihedral angles (defined by the groups of four atoms \( C_{\alpha}^{-1}-C_i^{-1}-N-C_{\alpha} \) and \( C_{\alpha}^{-1}-C_i^{-1}-C_{\alpha}^{-1} \)) to be planar,

\[
V_{\text{Planar}} = \frac{K_{\text{Planar}}}{2} \omega_{ijkl}^2.
\]

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

\[
m_i \frac{d^2 \vec{r}_i}{dt^2} = -\zeta \frac{d\vec{r}_i}{dt} + \vec{\Gamma}_i - \frac{\partial V}{\partial \vec{r}_i},
\]

for the atomic positions \( \vec{r}_i \), where \( m_i \) is the mass of atom \( i \), the Gaussian-distributed, \( \delta \)-function correlated random forces \( \vec{\Gamma}_i \) on atom \( i \) obey \( \langle \vec{\Gamma}_i(t) \cdot \vec{\Gamma}_j(t') \rangle = 2\zeta T \delta_{ij} \delta(t - t') \), and \( \delta(x) \) (\( \delta_{ij} \)) is the Dirac (Kronecker) \( \delta \)-function. We implemented a modified Velocity Verlet algorithm to integrate Eq. (5) with a time step \( \Delta t = 10^{-4} \) \( \Delta t_0 \), where \( \Delta t_0 = \sigma_H \sqrt{m_H/c} \), and damping parameter \( \zeta = 5c t_0 / \sigma_H^2 \).
The initial atomic velocities were drawn from a Maxwell-Boltzmann distribution at temperature $T'$, where $T' = T/c \approx 10^{-2}$. The ratio $T/c$ determines the average amount of overlap (i.e., pair separations that satisfy $r_{ij} < \sigma_{ij}$) between nonbonded atoms that occurs in the simulations. In the $(T/c) \rightarrow 0$ limit, the system explores only sterically allowed conformations. We show in Supporting Information Figures 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 performed in the limit of hard-sphere interactions. To determine the equilibration time for the hard-sphere simulations, we measured the average time, $t_n$, required to make transitions from $\alpha'$ to $\beta'$ or from $\beta'$ to $\alpha'$ (Supporting Information Fig. S1). We then equilibrated the Ala dipeptide for more than $100t_n$ before measuring conformational statistics. We calculate the probability distribution of backbone dihedral angles by binning combinations of $\phi$ and $\psi$ 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 force field Amber99sb-ILDN-NMR and CHARMM27-CMAP within the GROMACS 4.5.5 simulation package. Amber99sb-ILDN-NMR refers to the Amber99sb* force field combined with the ILDN side-chain optimization and NMR corrections. The NMR corrections optimize the dihedral angle potentials independently to match the $\phi$ and $\psi$ values observed in NMR experiments of proteins. The CHARMM27-CMAP combines the CHARMM27 force field with the CMAP knowledge-based correction 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 performed in the isobaric–isothermal ensemble using a stochastic velocity rescaling thermostat and Parrinello and Rahman barostat. 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 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 algorithm. 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 $\psi$ decorrelation times are $\approx 120$ ps and 150 ps for Amber and CHARMM, respectively (Supporting Information 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 used the Ewald-corrected four-point water model (TIP4P-Ew) and TIP3P water models, respectively.

Acknowledgments

The authors 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. The authors thank R.L. Dunbrack as well as J. Richardson and D. Richardson for useful insights and discussion. A.Q. Zhou is a Howard Hughes Medical Institute International Research Fellow.

References


