

# New Insights into the Interdependence between Amino Acid Stereochemistry and Protein Structure

Alice Qinhuo Zhou,<sup>†\*\*</sup> Diego Caballero,<sup>‡\*\*</sup> Corey S. O'Hern,<sup>‡§\*\*</sup> and Lynne Regan<sup>†¶\*\*\*</sup>

<sup>†</sup>Department of Molecular Biophysics and Biochemistry, <sup>‡</sup>Department of Physics, <sup>§</sup>Department of Mechanical Engineering and Materials Science, <sup>¶</sup>Department of Chemistry, and <sup>\*\*</sup>Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut

**ABSTRACT** To successfully design new proteins and understand the effects of mutations in natural proteins, we must understand the geometric and physicochemical principles underlying protein structure. The side chains of amino acids in peptides and proteins adopt specific dihedral angle combinations; however, we still do not have a fundamental quantitative understanding of why some side-chain dihedral angle combinations are highly populated and others are not. Here we employ a hard-sphere plus stereochemical constraint model of dipeptide mimetics to enumerate the side-chain dihedral angles of leucine (Leu) and isoleucine (Ile), and identify those conformations that are sterically allowed versus those that are not as a function of the backbone dihedral angles  $\phi$  and  $\psi$ . We compare our results with the observed distributions of side-chain dihedral angles in proteins of known structure. With the hard-sphere plus stereochemical constraint model, we obtain agreement between the model predictions and the observed side-chain dihedral angle distributions for Leu and Ile. These results quantify the extent to which local, geometrical constraints determine protein side-chain conformations.

## INTRODUCTION

Researchers in computational protein design seek to create new proteins with desirable properties, such as novel folds, enhanced stability, or tailored binding affinity and specificity (1). Although a number of successes in protein design have been achieved in recent years, the problem is by no means solved (2–12). In a recent study (13), for example, protein domains were designed to bind to a conserved region of the stem of influenza hemagglutinin protein. However, only 3% of the designed structures exhibited any binding when tested experimentally. That work both illustrated the state of the art in computational protein design and highlighted its limitations, as the authors themselves subsequently discussed (14).

There are several issues with current approaches to computational protein design. Current force fields mix knowledge-based and molecular-mechanics-based terms with relative weights that are determined ad hoc and are specific to each design problem (15,16). This approach also results in double counting of some energetic contributions. For example, including a knowledge-based helix propensity term double counts the energetics of van der Waals and hydrogen-bonding interactions. Moreover, many of the molecular-mechanics-based terms (e.g., van der Waals, electrostatics, and solvent-mediated interactions) do not need to be included in all applications. How-

ever, molecular-dynamics force fields have been optimized with all terms present, and with respect to a particular water model, which makes it difficult for researchers to assess the sensitivity of molecular-mechanics force fields to individual energetic terms. Instead of making the force fields more complicated, we seek a computational methodology in which the force fields are simplified to include only the dominant terms that are relevant to a particular application.

Exploration of the limits of a hard-sphere and stereochemical model for protein structure has a long history. More than 40 years ago, Ramakrishnan and Ramachandran (17) identified the allowed backbone conformations of an alanyl dipeptide given hard-sphere and stereochemical constraints. The sterically allowed combinations of the backbone dihedral angles  $\phi$  and  $\psi$  predicted for the alanyl dipeptide match those observed in proteins of known structure.

The influence of steric and packing constraints in proteins has been investigated extensively in both experiments and computational studies (18–31). For example, in experiments, researchers have determined the structural and thermodynamic changes in response to large- to small-cavity-forming mutations and alternative core-packing arrangements. In addition, the Richardson group (32–35) developed a method to assess the quality of protein crystal structures and ameliorate incorrect ones. They found that the highest-resolution structures efficiently fill space with few steric clashes, whereas low-quality structures are less well packed and possess many steric clashes. Dunbrack and colleagues have extensively analyzed the side-chain dihedral angle distributions in high-resolution protein crystal structures (36–39). They emphasized that

Submitted June 10, 2013, and accepted for publication September 16, 2013.

\*Correspondence: [lynne.regan@yale.edu](mailto:lynne.regan@yale.edu)

This is an Open Access article distributed under the terms of the Creative Commons-Attribution Noncommercial License (<http://creativecommons.org/licenses/by-nc/2.0/>), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editor: Jose Faraldo-Gomez.

© 2013 The Authors

0006-3495/13/11/2403/9 \$2.00



<http://dx.doi.org/10.1016/j.bpj.2013.09.018>

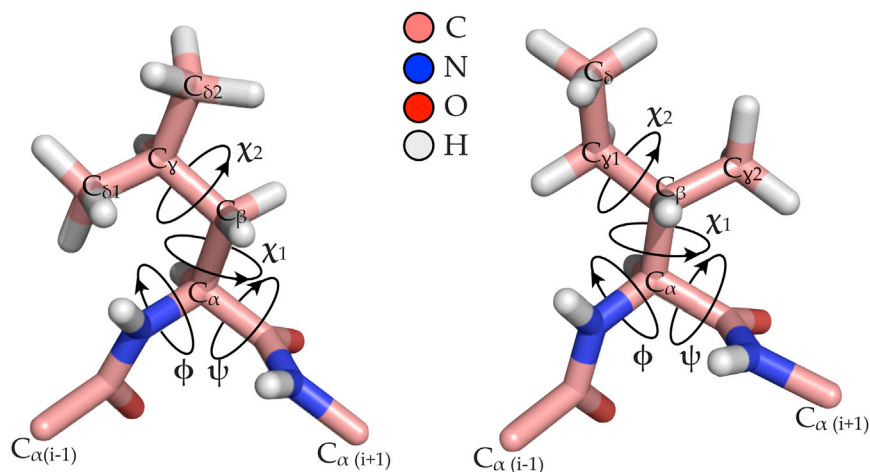


FIGURE 1 Stick representation of Leu (*left*) and Ile (*right*) dipeptide mimetics. The backbone dihedral angles,  $\phi$  and  $\psi$ , and the side-chain dihedral angles  $\chi_1$  and  $\chi_2$  are highlighted, with positive angles indicated by the arrows. The methyl hydrogen atoms were added using the REDUCE program (56). The  $C_\alpha$  atoms of the central, preceding ( $i + 1$ ), and trailing ( $i - 1$ ) amino acids are labeled. Carbon, nitrogen, oxygen, and hydrogen atoms are shaded pink, blue, oxygen, and white, respectively. To see this figure in color, go online.

the side-chain dihedral angle distributions are rotameric, with high probabilities at specific  $\chi_1$  and  $\chi_2$  combinations that depend sensitively on the backbone dihedral angles  $\phi$  and  $\psi$ . They also showed that certain rotamers are rare because of steric repulsions analogous to those that constrain the conformations of hydrocarbon chains.

Backbone (e.g., CMAP and Amber-NMR) and side-chain (e.g., Amber-ILDN) dihedral angle potentials and backbone-dependent rotamer libraries have been developed for implementation into molecular-dynamics simulation packages (40–42). However, even with these corrections, results from CHARMM and Amber still disagree with each other in their predictions for the distributions of the backbone and side-chain dihedral angles for dipeptide mimetics (43,44). Without the CMAP corrections, CHARMM predictions for the backbone dihedral angle distributions can be well

outside the hard-sphere limits of the Ramachandran plot (45,46).

Given the importance of side-chain packing in specifying the stability of protein-protein interfaces (47,48) and protein cores, we argue that for computational approaches to protein design to be successful, one must quantitatively understand the form of the side-chain dihedral angle distributions, i.e., one must explain why particular side-chain dihedral angle combinations are more or less probable. In this work, we present the results of computational studies of Leu and Ile dipeptide mimetics. We explain the observed side-chain dihedral angle probabilities for these uncharged, nonpolar residues using a hard-sphere model with stereochemical constraints (i.e., the bond lengths, bond angles, and  $\omega$  backbone dihedral angles set to experimental values) and no additional energetic terms.

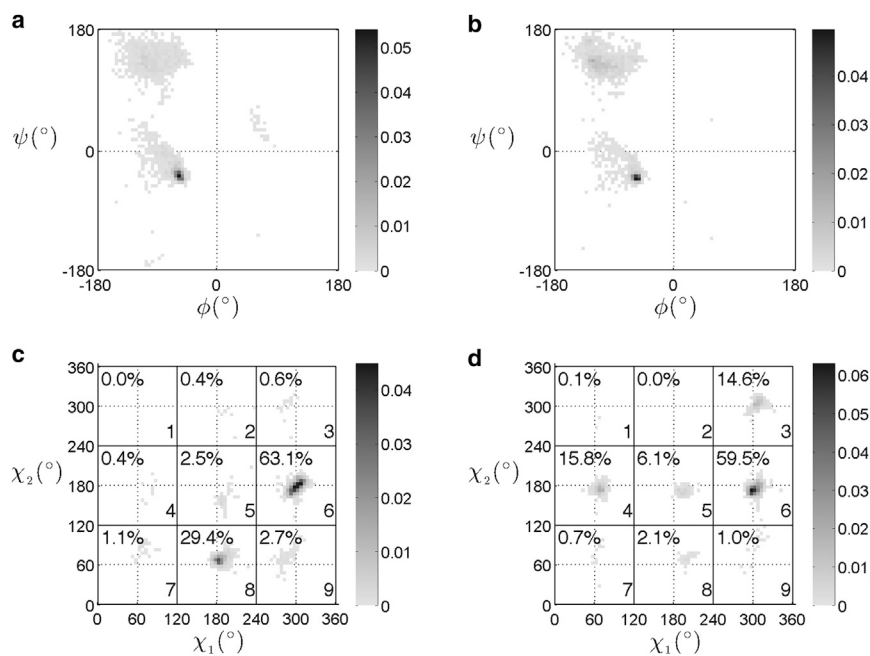


FIGURE 2 Observed probability distributions for the backbone dihedral angles  $P(\phi, \psi)$  (*top*) and side-chain dihedral angles  $P(\chi_1, \chi_2)$  (*bottom*) binned in  $5^\circ \times 5^\circ$  increments for Leu (*left*) and Ile (*right*) from protein crystal structures in the culled Dunbrack database. (*a–d*) The sums of the probability distributions over all  $\phi$  and  $\psi$  in *a* and *b*, or over all  $\chi_1$  and  $\chi_2$ , in *c* and *d* equal one. In *c* and *d*, the probability values within each of the nine  $\chi_1$  and  $\chi_2$  boxes are labeled.

## MATERIALS AND METHODS

Fig. 1 shows stick representations of the Leu and Ile dipeptide mimetics (N-acetyl-leucine-N'-methylamide and N-acetyl-isoleucine-N'-methylamide). Dipeptide conformations for both Leu and Ile are specified by the backbone dihedral angles  $\phi$  and  $\psi$ , side-chain dihedral angles  $\chi_1$  and  $\chi_2$ , 12 bond lengths, 15 bond angles, and two additional backbone dihedral angles  $\omega_1$  and  $\omega_2$  (without rotations of the hydrogen atoms; see the [Supporting Material](#)). We compare the results of our calculations with a subset (structures with resolution  $\leq 1.0$  Å and R factor  $\leq 0.2$ ) of Leu and Ile residues from the PDB provided by Dr. Roland Dunbrack, Jr., extracted from PISCES (49,50). From here on, we will refer to this database as the culled Dunbrack database. Note that this data set is not a subset of the set presented in Shapovalov and Dunbrack (39), even though a similar methodology was used to obtain it. Our selected subset includes 2204 Leu and 1555 Ile residues. The culled Dunbrack database is just one of several high-resolution protein databases that could have been used (33,52).

The culled Dunbrack database, against which we compare our calculations, is carefully curated to include a large number of high-resolution and high-confidence structures. Some researchers have reported that only at resolutions less than  $\sim 0.7$  Å are x-ray crystal structures truly free of refinement bias (53). However, the extremely small number of available ultrahigh-resolution structures (only six) precludes a meaningful statistical analysis. Nevertheless, we performed a side-chain conformational analysis of the 51 Leu and 32 Ile residues in these ultrahigh-resolution structures. We observed no significant differences between these analyses and those based on the culled Dunbrack data set. See the [Supporting Material](#) and Figs. S1 and S2.

Fig. 2 shows the observed probability distributions for the backbone dihedral angles  $P(\phi, \psi)$  and side-chain dihedral angles  $P(\chi_1, \chi_2)$  for Leu (Fig. 2, *a* and *c*) and Ile (Fig. 2, *b* and *d*) from protein crystal structures in the culled Dunbrack database. The probability distributions were binned in  $5^\circ \times 5^\circ$  boxes and normalized separately so that the sum over all  $\phi$  and  $\psi$ , or over all  $\chi_1$  and  $\chi_2$ , equals one. Note that the majority (60%) of Ile residues have side-chain dihedral angles that fall near a single rotamer combination ( $300^\circ, 180^\circ$ ) (box 6). For ease of reference, we decomposed  $\chi_1$  and  $\chi_2$  space into nine boxes, labeled 1–9. The  $\chi_1$  and  $\chi_2$  combinations around ( $300^\circ, 300^\circ$ ) (box 3), ( $60^\circ, 180^\circ$ ) (box 4), ( $180^\circ, 180^\circ$ ) (box 5), and ( $180^\circ, 60^\circ$ ) (box 8) are sometimes observed, whereas the  $\chi_1$  and  $\chi_2$  combinations around ( $60^\circ, 300^\circ$ ) (box 1), ( $180^\circ, 300^\circ$ ) (box 2), ( $60^\circ, 60^\circ$ ) (box 7), and ( $300^\circ, 60^\circ$ ) (box 9) rarely occur (with probabilities  $\leq 1\%$ ). For Leu residues,  $>90\%$  of the side-chain dihedral angles are found with  $\chi_1$  and  $\chi_2$  combinations around ( $300^\circ, 180^\circ$ ) (box 6) and ( $180^\circ, 60^\circ$ ) (box 8). Side-chain dihedral angle combinations around ( $180^\circ, 180^\circ$ ) (box 5) and ( $300^\circ, 60^\circ$ ) (box 9) are sometimes observed, whereas all other  $\chi_1$  and  $\chi_2$  combinations are rarely observed.

To obtain a physical understanding of the observed side-chain dihedral angle distributions of Leu and Ile, we model the atoms in the dipeptide mimetics as hard spheres with specified radii and bond-length, bond-angle, and  $\omega$ -backbone dihedral-angle constraints (54). Using this model, we exhaustively sample all backbone ( $\phi, \psi$ ) and side-chain dihedral angles ( $\chi_1, \chi_2$ ) and determine which angle combinations give rise to steric overlaps and which ones do not. In this context, a steric overlap is defined as a clash between two nonbonded atoms (with both located on the side chain or one on the side chain and the other on the backbone, i.e., we do not consider clashes between backbone atoms) that satisfies  $r_{ij} < (\sigma_i + \sigma_j)/2$ , where  $r_{ij}$  is the center-to-center separation between atoms *i* and *j* with diameters  $\sigma_i$  and  $\sigma_j$ . We then calculate the probability distributions for sterically allowed combinations of the side-chain dihedral angles  $\chi_1$  and  $\chi_2$  for particular values of the backbone dihedral angles  $\phi$  and  $\psi$ . Our calculations involve the following steps: First, we set the atom sizes for hydrogen,  $sp^3$  carbon,  $sp^2$  carbon, nitrogen, and oxygen to be 1.05, 1.5, 1.4, 1.4, and 1.45 Å, respectively. These values were calibrated in our previous studies of the side-chain dihedral angle distributions for Val and Thr (55). We then add the methyl hydrogens and position them using the REDUCE software package (56). To calculate the backbone and side-chain dihedral-angle distribu-

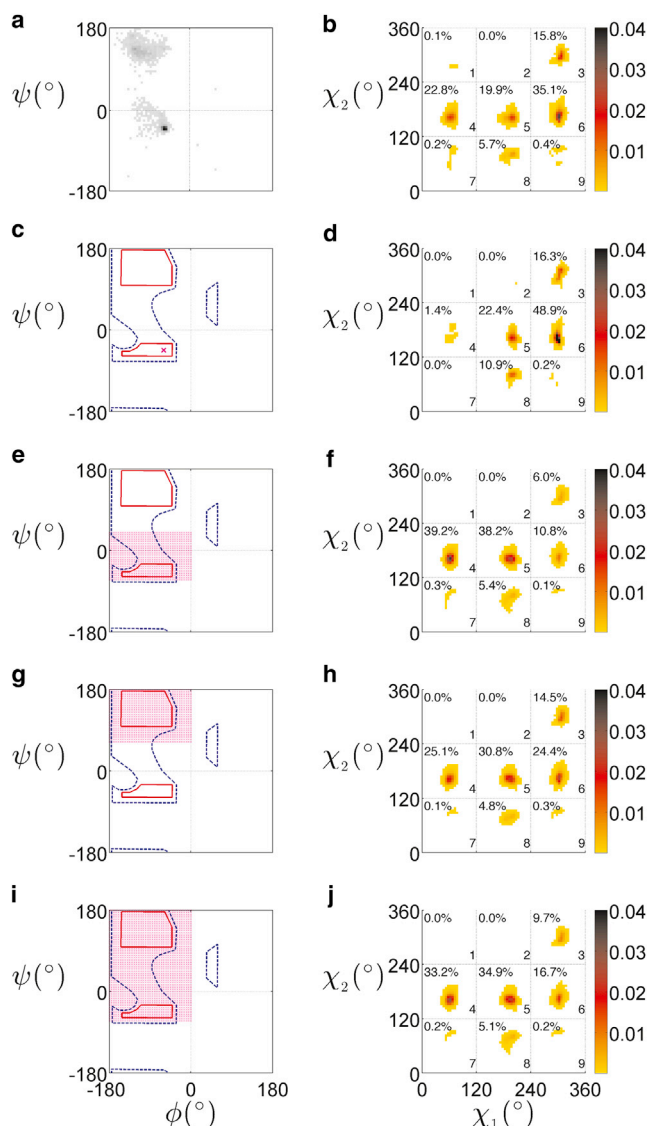


FIGURE 3 Calculated probability distributions of the sterically allowed side-chain dihedral-angle combinations  $\chi_1$  and  $\chi_2$  ( $5^\circ \times 5^\circ$  bins) from the steric plus stereochemical constraint model (in *b, d, f, h*, and *j*) after averaging over all Ile configurations with the  $\phi$  and  $\psi$  backbone dihedral angles given in *a, c, e, g*, and *i*, respectively. Panel *a* shows the distribution of  $\phi$  and  $\psi$  from the culled Dunbrack database. Panel *c* indicates that the dipeptide mimetics derived from the culled Dunbrack database have  $\phi$  and  $\psi$  set to the canonical  $\alpha$ -helix values,  $\phi = -60^\circ$  and  $\psi = -45^\circ$ . Panels *e, g*, and *i* represent uniform sampling of  $\phi$  and  $\psi$  values in the shaded regions that coincide roughly with the  $\alpha$ -helix,  $\beta$ -sheet, and  $\alpha$ -helix plus  $\beta$ -sheet regions of the Ramachandran plot outer limits (dashed line) for  $\tau = 115^\circ$ . Note that sterically allowed conformations can occur outside the Ramachandran outer limits because we are not including clashes between backbone atoms. To see this figure in color, go online.

tions, we discretize the  $\phi$  and  $\psi$  or  $\chi_1$  and  $\chi_2$  plane into  $5^\circ \times 5^\circ$  boxes, and for each box we sum the number of Leu or Ile backbone or side-chain conformations that are sterically allowed. The number of counts in each box normalized by the total number of rotamer combinations sampled gives  $P(\phi, \psi)$  and  $P(\chi_1, \chi_2)$ . Thus, the sum of  $P(\phi, \psi)$  and  $P(\chi_1, \chi_2)$  over all  $\phi$  and  $\psi$ , or over all  $\chi_1$  and  $\chi_2$ , equals one. See the [Supporting Material](#) for additional details of the computational methods.

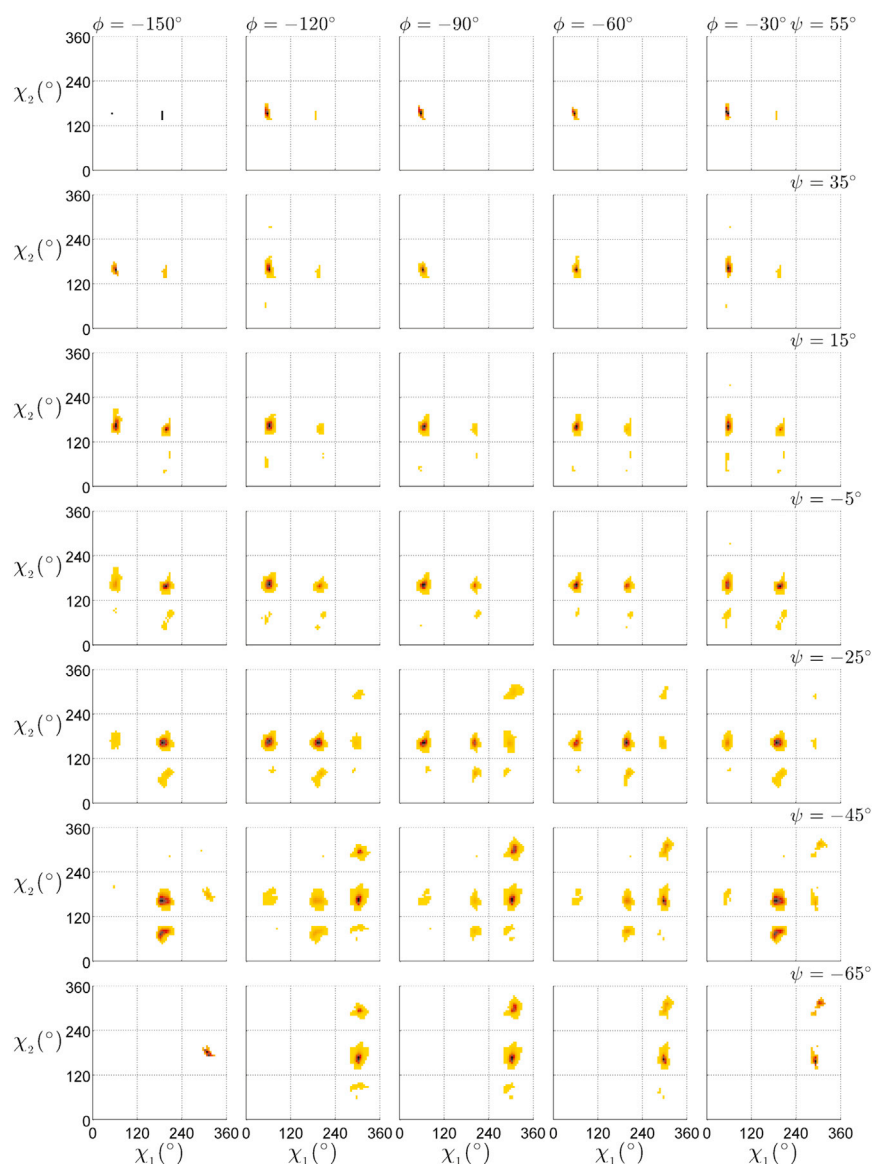


FIGURE 4 Calculated probability distributions  $P(\chi_1, \chi_2)$  of the sterically allowed side-chain dihedral-angle combinations  $\chi_1$  and  $\chi_2$  using the hard-sphere plus stereochemical constraint model for Ile dipeptides extracted from protein crystal structures in the culled Dunbrack database, after setting them to particular values of the backbone dihedral angles  $\phi$  and  $\psi$  indicated in each panel. The sum of the  $P(\chi_1, \chi_2)$  distributions over all  $\chi_1$  and  $\chi_2$  equals one in each panel separately. To see this figure in color, go online.

## RESULTS AND DISCUSSION

Fig. 3 summarizes the results obtained using the hard-sphere plus stereochemical constraint model for Ile dipeptide mimetics. In panels *b*, *d*, *f*, *h*, and *j*, we show the calculated probability distributions  $P(\chi_1, \chi_2)$  of sterically allowed side-chain dihedral-angle combinations  $\chi_1$  and  $\chi_2$  when the backbone dihedral angles  $\phi$  and  $\psi$  are sampled according to the distributions shown in *a*, *c*, *e*, *g*, and *i*, respectively. When  $\phi$  and  $\psi$  are sampled according to the observed Ile dipeptides in the culled Dunbrack database, where the majority of  $\phi$  and  $\psi$  are in the  $\alpha$ -helix region of the Ramachandran plot, the model predicts that the boxes with the most sterically allowed side-chain dihedral-angle combinations  $\chi_1$  and  $\chi_2$  are boxes 6 (35%), 4 (23%), 5 (20%), and 3 (16%), which is similar to the results from the culled Dunbrack database in Fig. 2 *d*, i.e., boxes 6 (60%), 4 (16%),

3 (15%), and 5 (6%). One interesting exception, which we will investigate in future studies, is box 5, for which we predict 20%, whereas the culled Dunbrack database gives 6%. This discrepancy suggests that the Dunbrack database does not uniformly weight the sterically allowed side-chain dihedral-angle combinations. Note that both the calculated and Dunbrack distributions do not populate boxes 1, 2, 7, and 9.

To determine the origin of the high-probability  $\chi_1$  and  $\chi_2$  combinations in box 5 centered around  $(300^\circ, 180^\circ)$ , we investigated how the sampling of the backbone dihedral angles influences the side-chain dihedral-angle distributions. In Fig. 3 *d*, we show the sterically allowed probability distribution  $P(\chi_1, \chi_2)$  for Ile dipeptides derived from the culled Dunbrack database after setting the backbone dihedral angles to canonical  $\alpha$ -helix values  $\phi = -60^\circ$  and  $\psi = -45^\circ$ . Setting the  $\phi$  and  $\psi$  backbone dihedral angles



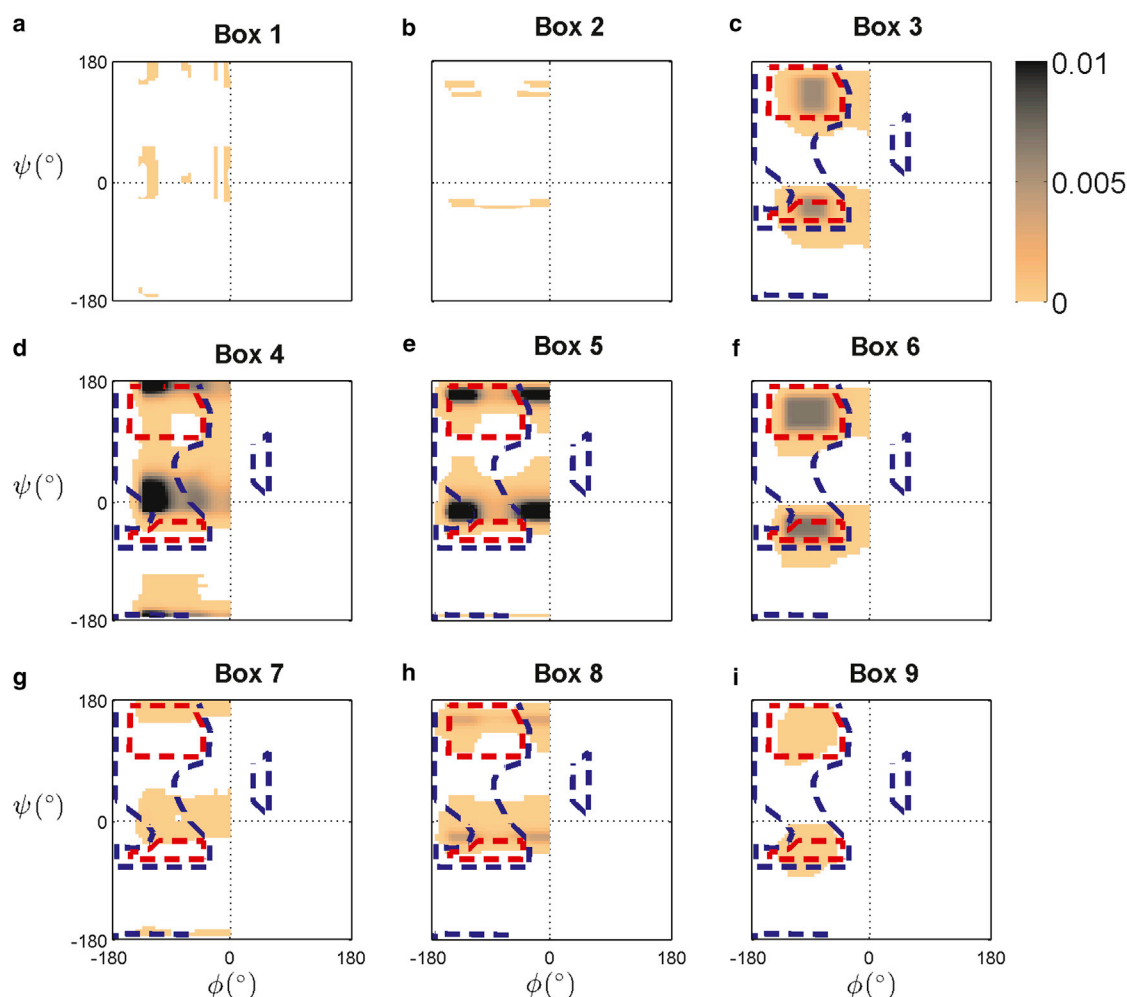


FIGURE 5 Calculated probability distributions  $P(\phi, \psi)$  based on the sterically allowed combinations of Ile side-chain dihedral angles in boxes 1–9 (Fig. 2 *d*) in each panel. The Ramachandran plot inner (red) and outer (blue) limits for  $\tau = 115^\circ$  are indicated. The sums of the distributions  $P(\phi, \psi)$  over all  $\phi$  and  $\psi$  equal one for each panel separately. To see this figure in color, go online.

to canonical helix values somewhat increases the probability of box 6 from 35% to 49%. This result suggests that one reason for the large number of side-chain dihedral angle combinations near  $(300^\circ, 180^\circ)$  in the culled Dunbrack database is the preponderance of  $\alpha$ -helical structures in the database.

To further investigate the interdependence between the backbone dihedral angles  $\phi$  and  $\psi$  and side-chain dihedral angles  $\chi_1$  and  $\chi_2$ , we also calculated the sterically allowed  $P(\chi_1, \chi_2)$  when uniformly sampling over different regions of  $\phi$  and  $\psi$  space: the  $\alpha$ -helix region (Fig. 3 *e*),  $\beta$ -sheet region (Fig. 3 *g*), and the combined  $\alpha$ -helix and  $\beta$ -sheet regions (Fig. 3 *i*). The calculated results corresponding to each of these sampling methods are shown in Fig. 3, *f*, *h*, and *j*, respectively. Sampling different regions of  $\phi$  and  $\psi$  space in this fashion has dramatic consequences for the sterically allowed side-chain dihedral-angle distributions. For example, we find that box 6 no longer contains the most sterically allowed  $\chi_1$  and  $\chi_2$  combinations when we sample

uniformly over  $\phi$  and  $\psi$  space. Boxes 4 and 5 now contain the largest number of sterically allowed  $\chi_1$  and  $\chi_2$  combinations, with >80% of the total contained in boxes 4, 5, and 6. This result emphasizes that  $\chi_1$  and  $\chi_2$  combinations in box 6 might be overweighted in rotamer libraries that do not account for the high  $\alpha$ -helix content in the Protein Data Bank (PDB).

We present the sterically allowed distributions  $P(\chi_1, \chi_2)$  for the relevant regions of  $\phi$  and  $\psi$  space for Ile in Fig. 4. A close examination of these data makes it clear that the  $\psi$  dependence of  $P(\chi_1, \chi_2)$  is stronger than the  $\phi$  dependence (except perhaps for values near  $\psi = -65^\circ$ ). For values in the range  $35^\circ \leq \psi \leq 55^\circ$  (i.e., the top two rows of Fig. 4), box 4 contains the only sterically allowed  $\chi_1$  and  $\chi_2$  combinations over the full range,  $-180^\circ \leq \phi \leq -30^\circ$ . As  $\psi$  decreases, sterically allowed  $\chi_1$  and  $\chi_2$  combinations populate box 5 as well as box 4. The most diverse collection of sterically allowed  $\chi_1$  and  $\chi_2$  combinations occurs in the range  $-65^\circ \leq \psi \leq -25^\circ$ , with boxes 3,

4, 5, 6, and 8 containing a significant number of sterically allowed combinations. For  $\psi \leq -65^\circ$ , the number of sterically allowed  $\chi_1$  and  $\chi_2$  combinations begins to decrease significantly.

Another illustrative way to display our data is to plot sterically allowed  $\phi$  and  $\psi$  values for each box of  $\chi_1$  and  $\chi_2$  combinations. In Fig. 5, we count the number of sterically allowed  $\chi_1$  and  $\chi_2$  combinations that occur within  $5^\circ \times 5^\circ$  boxes in  $\phi$  and  $\psi$  space for Ile. As expected, we find that there are very few  $\phi$  and  $\psi$  combinations that admit sterically allowed  $\chi_1$  and  $\chi_2$  combinations in boxes 1, 2, 7, and 9. In addition, sterically allowed  $\chi_1$  and  $\chi_2$  combinations that populate boxes 3 and 6 are associated with  $\phi$  and  $\psi$  combinations near canonical  $\alpha$ -helix and  $\beta$ -sheet values. In contrast, sterically allowed  $\chi_1$  and  $\chi_2$  combinations that populate boxes 4 and 5 are associated with the bridge region and elevated  $\psi$  values in the  $\beta$ -sheet region of the Ramachandran plot. This behavior is also found in protein crystal structures from the culled Dunbrack database, as shown in Fig. S3.

We also investigated the influence of correlations among the bond angles, bond lengths, and  $\omega$ -backbone dihedral angles on the distribution of sterically allowed side-chain dihedral angles. In Fig. 6, we analyze the effects of the correlations between the 12 bond lengths, 15 bond angles, and 2  $\omega$ -backbone dihedral angles on the calculated sterically

allowed probability distributions  $P(\chi_1, \chi_2)$  for Ile dipeptides when the backbone dihedral angles are fixed at the  $\alpha$ -helix canonical values  $\phi = -60^\circ$  and  $\psi = -45^\circ$ . In Fig. 6 a, we show the calculated  $P(\chi_1, \chi_2)$  for Ile residues from the culled Dunbrack database with  $\phi$  and  $\psi$  at  $\alpha$ -helix canonical values (same as Fig. 3 d). The correlation coefficients between the bond lengths and bond and dihedral angles for Ile residues from the culled Dunbrack database are shown in Fig. 6 d, with labels given in Table 1. The amplitudes of the fluctuating positive and negative correlations are above random noise (Fig. 6 f). In Fig. 6 b, we show  $P(\chi_1, \chi_2)$  for artificial Ile dipeptide mimetics with bond lengths, bond angles, and  $\omega$ -backbone dihedral angles randomly selected from Gaussian distributions with means, standard deviations (SDs), and multivariate correlations that match those from the culled Dunbrack database. We find that the probability distributions  $P(\chi_1, \chi_2)$  shown in Fig. 6, a and b, are very similar to those obtained from Ile dipeptides constructed without building in multivariate correlations. Thus, correlations in the bond lengths, bond angles, and  $\omega$ -dihedral angles do not strongly influence the distribution of sterically allowed side-chain dihedral angles in dipeptides.

We find qualitatively similar results for Leu dipeptides, with a few noteworthy differences. In Fig. 2 c, we show that the most  $\chi_1$  and  $\chi_2$  combinations from Leu residues in the culled Dunbrack database occur in boxes 6 and 8

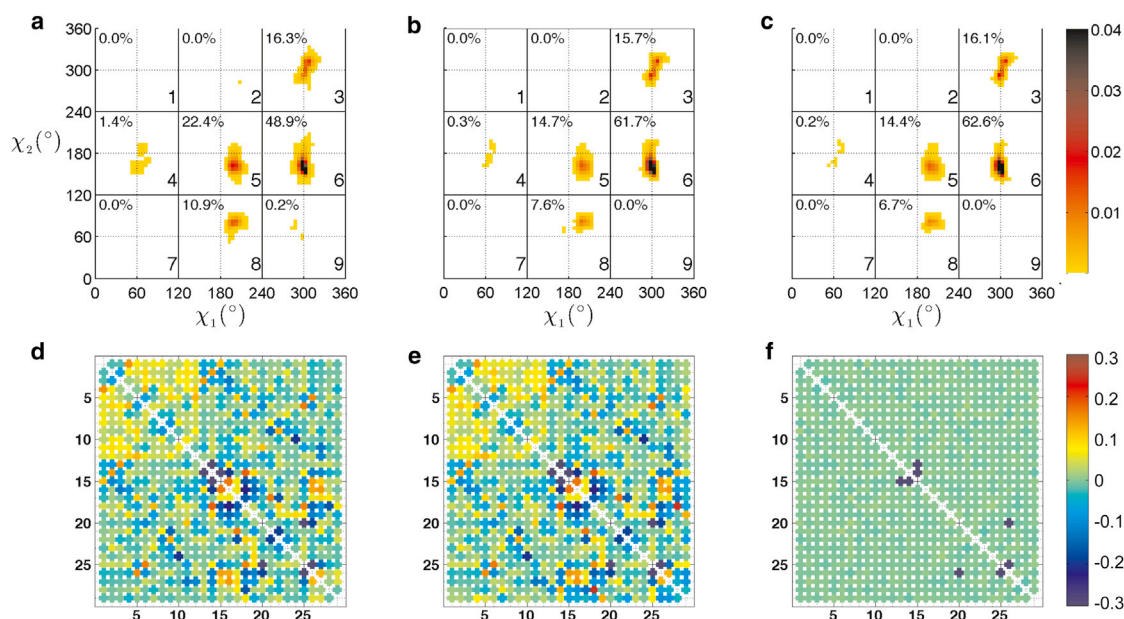


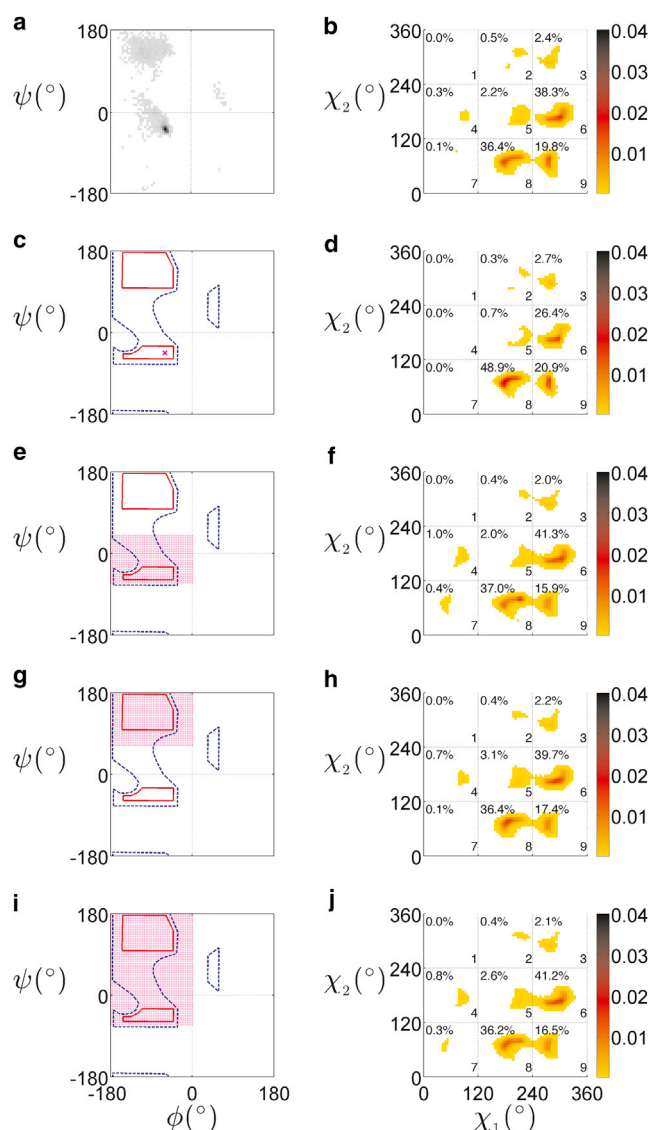
FIGURE 6 (a–c) Calculated probability distribution  $P(\chi_1, \chi_2)$  of the sterically allowed combinations of  $\chi_1$  and  $\chi_2$  for (a) Ile dipeptides extracted from protein crystal structures in the culled Dunbrack database with backbone dihedral angles for all residues rotated to  $\phi = -60^\circ$  and  $\psi = -45^\circ$  (same as Fig. 3 c); (b) 8970 randomly generated Ile dipeptide mimetics with the backbone dihedral angles rotated to  $\phi = -60^\circ$  and  $\psi = -45^\circ$ , bond lengths, bond angles, and dihedral angles  $\omega$  (from residues  $i$  and  $i + 1$ ) chosen randomly from Gaussian distributions with the same mean, SD, and multivariate correlations as found in the culled Dunbrack database; and (c) 8970 randomly generated Ile dipeptide mimetics with backbone dihedral angles set to  $\phi = -60^\circ$  and  $\psi = -45^\circ$  and bond lengths, bond angles, and  $\omega$ -dihedral angles chosen randomly from Gaussian distributions with only means and SDs that match the culled Dunbrack database. Panels d–f show the correlation coefficients between the 12 bond lengths, 15 bond angles, and two backbone  $\omega$ -dihedral angles from the Ile dipeptide mimetics employed to calculate the probability distributions  $P(\chi_1, \chi_2)$  in a–c, respectively. The axes labels in d–f index the bond lengths, bond angles, and dihedral angles as shown in Table 1. To see this figure in color, go online.

**TABLE 1** Indexes that label the 12 bond lengths, 15 bond angles, and two backbone  $\omega$ -dihedral angles that characterize the Ile dipeptide mimetic and appear in Fig. 6, d–f

Index	Name
1	$C_{\alpha}^{i-1} - C^{i-1}$
2	$C^{i-1} - O^{i-1}$
3	$C^{i-1} - N^i$
4	$N - C_{\alpha}$
5	$C_{\alpha} - C$
6	$C - O$
7	$C_{\alpha} - C_{\beta}$
8	$C_{\alpha} - C_{\gamma 2}$
9	$C_{\beta} - C_{\gamma 1}$
10	$C_{\gamma 1} - C_{\delta}$
11	$C^i - N^{i+1}$
12	$N^{i+1} - C_{\alpha}^{i+1}$
13	$C_{\alpha}^{i-1} - C^{i-1} - O^{i-1}$
14	$C_{\alpha}^{i-1} - C^{i-1} - N^i$
15	$O^{i-1} - C^{i-1} - N^i$
16	$C^{i-1} - N^i - C_{\alpha}^i$
17	$N - C_{\alpha} - C_{\beta}$
18	$N - C_{\alpha} - C$
19	$C - C_{\alpha} - C_{\beta}$
20	$C_{\alpha} - C - O$
21	$C_{\alpha} - C_{\beta} - C_{\gamma 1}$
22	$C_{\alpha} - C_{\beta} - C_{\gamma 2}$
23	$C_{\gamma 1} - C_{\beta} - C_{\gamma 2}$
24	$C_{\beta} - C_{\gamma 1} - C_{\delta}$
25	$C_{\alpha}^i - C^i - N^{i+1}$
26	$O^i - C^i - N^{i+1}$
27	$C^i - N^{i+1} - C_{\alpha}^{i+1}$
28	$C_{\alpha}^{i-1} - C^{i-1} - N^i - C_{\alpha}^i$
29	$C_{\alpha}^i - C^i - N^{i+1} - C_{\alpha}^{i+1}$

(Fig. 7, *f* and *h*), totaling >92% of the side-chain conformations. In this figure we plot the sterically allowed distributions  $P(\chi_1, \chi_2)$  for our model when we employ different sampling methods for  $\phi$  and  $\psi$ . When we sample  $\phi$  and  $\psi$  according to the culled Dunbrack database or when we set  $\phi$  and  $\psi$  to canonical  $\alpha$ -helix values, we find that 75% of the sterically allowed  $\chi_1$  and  $\chi_2$  combinations are found in boxes 6 and 8 (Fig. 7, *f* and *h*). An interesting difference between  $P(\chi_1, \chi_2)$  obtained from the culled Dunbrack database and that predicted from our model is that side-chain conformations in box 9 (Fig. 7, *i*) are more abundant in the model. This abundance occurs despite syn-pentane interactions (Dunbrack) that lead to strong overlaps between backbone and side-chain  $C_{\delta}$  atoms for  $\chi_1 \geq 300^\circ$ . In future studies, we will investigate whether structures in coil libraries more frequently populate the sterically allowed conformations in box 9 (Fig. 7, *i*). In contrast to the results for Ile, the specific method used to sample  $\phi$  and  $\psi$  does not strongly influence the calculated  $P(\chi_1, \chi_2)$  for Leu, i.e., uniform sampling of  $\phi$  and  $\psi$  in *f*, *h*, and *j* gives results qualitatively similar to those obtained by sampling  $\phi$  and  $\psi$  according to the culled Dunbrack distribution.

The sterically allowed distributions  $P(\chi_1, \chi_2)$  for the relevant regions of  $\phi$  and  $\psi$  space are plotted for Leu in Fig. S4. Again, we find that the  $\psi$  dependence of  $P(\chi_1, \chi_2)$  is some-



**FIGURE 7** (a–j) Calculated probability distributions of the sterically allowed side-chain dihedral-angle combinations  $\chi_1$  and  $\chi_2$  ( $5^\circ \times 5^\circ$  bins) from the steric plus stereochemical constraint model (in *b*, *d*, *f*, *h*, and *j*) after averaging over all Leu configurations with  $\phi$  and  $\psi$  backbone dihedral angles given in *a*, *c*, *e*, *g*, and *i*, respectively. See Fig. 3 for additional information. To see this figure in color, go online.

what stronger than the  $\phi$  dependence. For values in the range  $35^\circ \leq \psi \leq 55^\circ$  (i.e., the top two rows of Fig. S4), the model predicts few sterically allowed  $\chi_1$  and  $\chi_2$  combinations, with most occurring in box 6. As  $\psi$  decreases, sterically allowed  $\chi_1$  and  $\chi_2$  combinations populate more boxes, with most occurring in 6, 8, and 9. We also find sterically allowed  $\chi_1$  and  $\chi_2$  combinations that bridge boxes 5 and 6, as well as boxes 8 and 9, which suggests that these conformations enable transitions between rotamers (57).

In Fig. S5, we count the number of sterically allowed  $\chi_1$  and  $\chi_2$  combinations that occur within  $5^\circ \times 5^\circ$  boxes in  $\phi$  and  $\psi$  space for Leu. For the rare  $\chi_1$  and  $\chi_2$  combinations (e.g., boxes 2, 3, 4, 5, and 7), the  $\phi$  and  $\psi$  combinations

that admit sterically allowed  $\chi_1$  and  $\chi_2$  combinations are fairly uniform. In contrast, the highly probable sterically allowed  $\chi_1$  and  $\chi_2$  combinations that populate boxes 6, 8, and 9 for the most part are associated with  $\phi$  and  $\psi$  combinations in the canonical  $\alpha$ -helix and  $\beta$ -sheet regions of the Ramachandran plot, although some conformations in box 6 exist in the bridge region. This predicted behavior is also found in the protein structures from the culled Dunbrack database (Fig. S6).

We also performed similar side-chain conformational analyses on the Leu and Ile residues in ultrahigh-resolution structures, and these gave results similar to those obtained with the calculations described above. See the [Supporting Material](#) and Figs. S1 and S2.

## CONCLUSIONS

In summary, we have enumerated the sterically allowed side-chain dihedral-angle combinations for Leu and Ile dipeptide mimetics using a hard-sphere plus stereochemical constraint model. We find that the regions of the sterically allowed probability distributions  $P(\chi_1, \chi_2)$  correspond to side-chain dihedral-angle combinations that are observed in proteins of known structure. Thus, we emphasize that, in many cases, modeling steric and stereochemical constraints alone can quantitatively describe side-chain conformational statistics. The discrepancies between the side-chain dihedral-angle distributions calculated from our model and those extracted from the PDB are likely due to the particular nonuniform weighting of the sterically allowed side-chain conformations in the PDB and will be investigated in future studies.

Our complete enumeration approach may be contrasted with methods that rely exclusively on the PDB, which are overweighted by the  $\phi$  and  $\psi$  combinations that occur frequently in structures deposited in the PDB. In contrast, with our model, we can interrogate side-chain conformations that are rarely sampled in the PDB as well as the highly probable ones. We are now in a position to calculate the side-chain dihedral-angle distributions for all other dipeptide mimetics and predict side-chain conformations in the context of proteins.

## SUPPORTING MATERIAL

Six figures, and supplemental information are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(13\)01066-7](http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)01066-7).

We thank R.L. Dunbrack, Jr., for providing a new high-resolution set of structures from the PDB, as well as for thought-provoking discussions. We also thank Jane and David Richardson for their valuable insights.

This work was supported by the National Science Foundation (grants DMR-1006537 and PHY-1019147) and the Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences. Alice Qinhua Zhou is a Howard Hughes Medical Institute International Research Fellow.

## REFERENCES

1. Kortemme, T., L. A. Joachimiak, ..., D. Baker. 2004. Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* 11:371–379.
2. Shandler, S. J., I. V. Korendovych, ..., W. F. DeGrado. 2011. Computational design of a  $\beta$ -peptide that targets transmembrane helices. *J. Am. Chem. Soc.* 133:12378–12381.
3. Korendovych, I. V., D. W. Kulp, ..., W. F. DeGrado. 2011. Design of a switchable eliminase. *Proc. Natl. Acad. Sci. USA.* 108:6823–6827.
4. Grove, T. Z., C. O. Osuji, ..., L. Regan. 2010. Stimuli-responsive smart gels realized via modular protein design. *J. Am. Chem. Soc.* 132: 14024–14026.
5. Cortajarena, A. L., J. Wang, and L. Regan. 2010. Crystal structure of a designed tetratricopeptide repeat module in complex with its peptide ligand. *FEBS J.* 277:1058–1066.
6. Cortajarena, A. L., T. Y. Liu, ..., L. Regan. 2010. Designed proteins to modulate cellular networks. *ACS Chem. Biol.* 5:545–552.
7. Chen, T. S., H. Palacios, and A. E. Keating. 2013. Structure-based redesign of the binding specificity of anti-apoptotic Bcl-x(L). *J. Mol. Biol.* 425:171–185.
8. Humphris-Narayanan, E., E. Akiva, ..., T. Kortemme. 2012. Prediction of mutational tolerance in HIV-1 protease and reverse transcriptase using flexible backbone protein design. *PLOS Comput. Biol.* 8: e1002639.
9. King, N. P., W. Sheffler, ..., D. Baker. 2012. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science.* 336:1171–1174.
10. Roberts, K. E., P. R. Cushing, ..., B. R. Donald. 2012. Computational design of a PDZ domain peptide inhibitor that rescues CFTR activity. *PLOS Comput. Biol.* 8:e1002477.
11. Privett, H. K., G. Kiss, ..., S. L. Mayo. 2012. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. USA.* 109: 3790–3795.
12. Murphy, G. S., J. L. Mills, ..., B. Kuhlman. 2012. Increasing sequence diversity with flexible backbone protein design: the complete redesign of a protein hydrophobic core. *Structure.* 20:1086–1096.
13. Fleishman, S. J., T. A. Whitehead, ..., D. Baker. 2011. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science.* 332:816–821.
14. Fleishman, S. J., and D. Baker. 2012. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell.* 149:262–273.
15. Rohl, C. A., C. E. Strauss, ..., D. Baker. 2004. Protein structure prediction using Rosetta. *Methods Enzymol.* 383:66–93.
16. Guntas, G., C. Purbeck, and B. Kuhlman. 2010. Engineering a protein-protein interface using a computationally designed library. *Proc. Natl. Acad. Sci. USA.* 107:19296–19301.
17. Ramakrishnan, C., and G. N. Ramachandran. 1965. Stereochemical criteria for polypeptide and protein chain conformations. *Biophys. J.* 5:909–933.
18. Tsai, J., R. Taylor, ..., M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290:253–266.
19. Lee, C., and M. Levitt. 1991. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature.* 352:448–451.
20. Chen, J., Z. Lu, ..., W. E. Stites. 2004. Proteins with simplified hydrophobic cores compared to other packing mutants. *Biophys. Chem.* 110:239–248.
21. Benítez-Cardoza, C. G., K. Stott, ..., S. E. Jackson. 2004. Exploring sequence/folding space: folding studies on multiple hydrophobic core mutants of ubiquitin. *Biochemistry.* 43:5195–5203.
22. Willis, M. A., B. Bishop, ..., A. T. Brunger. 2000. Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure.* 8:1319–1328.



23. Johnson, E. C., G. A. Lazar, ..., T. M. Handel. 1999. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure*. 7:967–976.
24. Baldwin, E., J. Xu, ..., B. W. Matthews. 1996. Thermodynamic and structural compensation in “size-switch” core repacking variants of bacteriophage T4 lysozyme. *J. Mol. Biol.* 259:542–559.
25. Harbury, P. B., B. Tidor, and P. S. Kim. 1995. Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl. Acad. Sci. USA*. 92:8408–8412.
26. Buckle, A. M., K. Henrick, and A. R. Fersht. 1993. Crystal structural analysis of mutations in the hydrophobic cores of barnase. *J. Mol. Biol.* 234:847–860.
27. Sandberg, W. S., and T. C. Terwilliger. 1991. Energetics of repacking a protein interior. *Proc. Natl. Acad. Sci. USA*. 88:1706–1710.
28. Munson, M., S. Balasubramanian, ..., L. Regan. 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.* 5:1584–1593.
29. Munson, M., R. O'Brien, ..., L. Regan. 1994. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci.* 3:2015–2022.
30. Lim, W. A., and R. T. Sauer. 1989. Alternative packing arrangements in the hydrophobic core of  $\lambda$  repressor. *Nature*. 339:31–36.
31. Ponder, J. W., and F. M. Richards. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* 193:775–791.
32. Word, J. M., S. C. Lovell, ..., D. C. Richardson. 1999. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* 285:1711–1733.
33. Lovell, S. C., J. M. Word, ..., D. C. Richardson. 2000. The penultimate rotamer library. *Proteins*. 40:389–408.
34. Keedy, D. A., C. J. Williams, ..., J. S. Richardson. 2009. The other 90% of the protein: assessment beyond the C $\alpha$ s for CASP8 template-based and high-accuracy models. *Proteins*. 77(Suppl 9):29–49.
35. Headd, J. J., R. M. Immormino, ..., J. S. Richardson. 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.
36. Dunbrack, Jr., R. L., and M. Karplus. 1994. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.* 1:334–340.
37. Bower, M. J., F. E. Cohen, and R. L. Dunbrack, Jr. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* 267:1268–1282.
38. Dunbrack, Jr., R. L., and F. E. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6:1661–1681.
39. Shapovalov, M. V., and R. L. Dunbrack, Jr. 2011. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. 19:844–858.
40. MacKerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 2004. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126:698–699.
41. Li, D.-W., and R. Brüschweiler. 2010. NMR-based protein potentials. *Angew. Chem. Int. Ed. Engl.* 49:6778–6780.
42. Lindorff-Larsen, K., S. Piana, ..., D. E. Shaw. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 78:1950–1958.
43. Beauchamp, K. A., Y.-S. Lin, ..., V. S. Pande. 2012. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* 8:1409–1414.
44. Vymětal, J., and J. Vondrášek. 2013. Critical assessment of current force fields. Short peptide test case. *J. Chem. Theory Comput.* 9: 441–451.
45. Hu, H., M. Elstner, and J. Hermans. 2003. Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine “dipeptides” (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. *Proteins*. 50:451–463.
46. Mackerell, Jr., A. D., M. Feig, and C. L. Brooks, 3rd. 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.* 25:1400–1415.
47. Reichmann, D., O. Rahat, ..., G. Schreiber. 2007. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* 17:67–76.
48. Schreiber, G., and A. E. Keating. 2011. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* 21:50–61.
49. Wang, G., and R. L. Dunbrack, Jr. 2003. PISCES: a protein sequence culling server. *Bioinformatics*. 19:1589–1591.
50. Wang, G., and R. L. Dunbrack, Jr. 2005. PISCES: recent improvements to a PDB sequence culling server. *Nucleic Acids Res.* 33(Web Server issue):W94–W98.
51. Reference deleted in proof.
52. Lovell, S. C., I. W. Davis, ..., D. C. Richardson. 2003. Structure validation by C $\alpha$  geometry:  $\phi$ ,  $\psi$  and C $\beta$  deviation. *Proteins*. 50:437–450.
53. Tronrud, D. E., and P. A. Karplus. 2011. A conformation-dependent stereochemical library improves crystallographic refinement even at atomic resolution. *Acta Crystallogr. D Biol. Crystallogr.* 67:699–706.
54. Zhou, A. Q., C. S. O'Hern, and L. Regan. 2011. Revisiting the Ramachandran plot from a new angle. *Protein Sci.* 20:1166–1171.
55. Zhou, A. Q., C. S. O'Hern, and L. Regan. 2012. The power of hard-sphere models: explaining side-chain dihedral angle distributions of Thr and Val. *Biophys. J.* 102:2345–2352.
56. Word, J. M., S. C. Lovell, ..., D. C. Richardson. 1999. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285:1735–1747.
57. Petrella, R. J., and M. Karplus. 2001. The energetics of off-rotamer protein side-chain conformations. *J. Mol. Biol.* 312:1161–1175.