The Power of Hard-Sphere Models: Explaining Side-Chain Dihedral Angle Distributions of Thr and Val

Alice Qinhua Zhou,[†¶] Corey S. O'Hern,[§¶] and Lynne Regan[†‡¶*]
[†]Department of Molecular Biophysics and Biochemistry, [‡]Department of Chemistry, [§]Departments of Mechanical Engineering and Materials Science and Physics, and [¶]Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut

ABSTRACT The energy functions used to predict protein structures typically include both molecular-mechanics and knowledge-based terms. In contrast, our approach is to develop robust physics- and geometry-based methods. Here, we investigate to what extent simple hard-sphere models can be used to predict side-chain conformations. The distributions of the side-chain dihedral angle χ_1 of Val and Thr in proteins of known structure show distinctive features: Val side chains predominantly adopt $\chi_1 = 180^\circ$, whereas Thr side chains typically adopt $\chi_1 = 60^\circ$ and 300° (i.e., $\chi_1 = \pm 60^\circ$ or $g-$ and g^+ configurations). Several hypotheses have been proposed to explain these differences, including interresidue steric clashes and hydrogen-bonding interactions. In contrast, we show that the observed side-chain dihedral angle distributions for both Val and Thr can be explained using only *local* steric interactions in a dipeptide mimetic. Our results emphasize the power of simple physical approaches and their importance for future advances in protein engineering and design.

INTRODUCTION

Ramachandran and Ramachandran (1) calculated the sterically allowed combinations of protein backbone dihedral angles (ϕ and ψ) using hard-sphere models of dipeptides. They found excellent agreement between the allowed ϕ- and ψ-angles they calculated and reported in the Ramachandran plot, and those found in the first crystal structures of peptides and small proteins (2). Since then, researchers have found that the ϕ- and ψ-angles for all amino acids in thousands of proteins fall within the allowed regions of the Ramachandran plot. Indeed, a comparison of the ϕ- and ψ-angles for all amino acids in a protein with the Ramachandran plot is now a routine quality test for new protein structures (3,4).

Surprisingly, few similar calculations have been performed to predict sterically allowed conformations of amino-acid side-chain dihedral angles (5,6). Some studies have employed quantum mechanical calculations to predict side-chain dihedral angle distributions for dipeptides (7,8). However, such calculations are too computationally intensive to be routinely applied to proteins. As of this writing, most structure prediction software employs the statistics of side-chain dihedral angles that occur in proteins of known structure to construct knowledge-based side-chain dihedral angle energy functions (9–11) or rotamer libraries (12–20).

We present the results of our calculations on the side-chain dihedral angle distributions of Val and Thr using a simple hard-sphere model for a dipeptide mimetic analogous to that of Ramachandran and colleagues (Fig. 1). It is essential to include all hydrogen atoms in the model; indeed interatomic clashes that arise from rotations of the hydroxyl and methyl hydrogens within the dipeptide mimetic determine the form of the side-chain dihedral angle distributions. Our predictions for the side-chain dihedral angle distributions for Val and Thr match those observed in proteins of known structure. We chose Val and Thr for the following reasons: 1), both possess a single side-chain dihedral angle, χ_1; 2), Val has an exclusively nonpolar, aliphatic side chain, whereas there is a polar, hydroxyl group on the Thr side chain; and 3), both Val and Thr are β-branched, but their side-chain dihedral angle distributions are distinct (Fig. 2). In particular, the major peak in the side-chain dihedral angle distribution for Val in an α-helix backbone conformation is at $\chi_1 = 180^\circ$, whereas the major peak for Thr in an α-helix conformation is at $\chi_1 = 300^\circ$ (g^+ or -60°). Note that our results provide an alternative explanation for the dihedral angle distribution of Thr in the α-helix conformation. Previous hypotheses invoked interresidue hydrogen-bonding and steric clashes between the Thr side-chain and main-chain atoms on neighboring residues (21–23).

MATERIALS AND METHODS

Database of protein crystal structures

We compare our predictions of the side-chain dihedral angle distributions for Val and Thr to the distributions obtained from the Dunbrack database (20). This database is composed of 850 high-resolution nonhomologous protein structures (sequence identity <50%) solved by x-ray crystallography and deposited in the Protein Data Bank. Only structures with resolution $\leq$1.7 Å, side chain B-factors <30 Å^2, and R-factors $\leq$0.25 are included.

We calculated two sets of probability distributions for Val and Thr in both α-helix and β-sheet backbone conformations from the Dunbrack database (Fig. 2). (The probability distribution of ϕ- and ψ-angles for Thr and Val in the Dunbrack distribution are shown in Fig. S6 in the Supporting Material.) The left panel of Fig. 2 provides the side-chain χ_1 distributions for Thr and Val averaged over all α-helix (or β-sheet) conformations, where the

Submitted December 2, 2011, and accepted for publication January 27, 2012.

*Correspondence: lynne.regan@yale.edu

Editor: Nathan Baker.

0006-3495/12/05/2345/8 $2.00

doi: 10.1016/j.bpj.2012.01.061

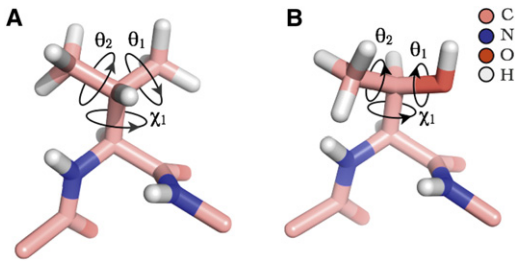

FIGURE 1 Stick representation of (*A*) Val and (*B*) Thr dipeptide mimetics. Both Val and Thr are shown in an α-helix backbone conformation ($\phi = -57°, \psi = -47°$). The side-chain dihedral angle χ_1 (N-C_α-C_β-$C_{\gamma 1}$) for both Val ($\chi_1 = 179.5°$) and Thr ($\chi_1 = 117.7°$), hydrogen dihedral angles $\theta_1 = 180°$ (C_α-C_β-$C_{\gamma 1}$-$H^1_{C_{\gamma 1}}$ on $C_{\gamma 1}$) and $\theta_2 = 0°$ (C_α-C_β-$C_{\gamma 2}$-$H^1_{C_{\gamma 2}}$ on $C_{\gamma 2}$) for Val and $\theta_1 = 180°$ (C_α-C_β-$O_{\gamma 1}$-$H^1_{C_{\gamma 1}}$) and $\theta_2 = 30°$ (C_α-C_β-$C_{\gamma 2}$-$H^1_{C_{\gamma 2}}$) for Thr are indicated.

secondary structure classification was deposited into the Protein Data Bank. This first data set includes 3418 examples of Val in an α-helix, 5020 examples of Val in a β-sheet, 2598 examples of Thr in an α-helix, and 2727 examples of Thr in a β-sheet. The right panel shows the side-chain χ_1 distributions for Thr and Val for conformations with ϕ- and ψ-values over narrow ranges ($\pm 10°$) that are centered on the canonical α-helix and β-sheet ϕ- and ψ-angles ($\phi = -57°, \psi = -47°$ for α-helix and $\phi = -119°, \psi = -113°$ for β-sheet). The second data set includes 2149 examples of Val in an α-helix, 681 examples of Val in a β-sheet, 1453 examples of Thr in an α-helix, and 189 examples of Thr in a β-sheet. The probability distributions for the side-chain dihedral angle χ_1 were calculated using a bin size of 5° and normalized so that the area under the curve equals 1.

Hard-sphere model

The χ_1 dihedral angles of Val and Thr are defined by the clockwise rotation around the C_α-C_β bond (viewed from C_α to C_β) of the backbone atoms N-C_α-C_β-$C_{\gamma 1}$. Note that we use the standard IUPAC-IUB (24) chemistry conventions for the atomic labeling order. Thus, $\chi_1 = 180°$ for Val puts its two C_γ atoms in the same position that the β-branch of Thr would occupy for $\chi_1 = 60°$. Thus, $O_{\gamma 1}$ in Thr is equivalent to $C_{\gamma 2}$ in Val. In Val, the dihedral angle is defined by the clockwise rotation about the C_β-$C_{\gamma 1}$ bond (viewed from C_β to $C_{\gamma 1}$) involving the side-chain atoms C_α-C_β-$C_{\gamma 1}$-$H^1_{O_{\gamma 1}}$, and the θ_2 dihedral angle is defined by the clockwise rotation about the C_β-$C_{\gamma 2}$ bond (viewed from C_β to $C_{\gamma 2}$) involving the side-chain atoms C_α-C_β-$C_{\gamma 2}$-$H^1_{C_{\gamma 2}}$. In Thr, the θ_1 dihedral angle is defined by the clockwise rotation about the C_β-$O_{\gamma 1}$ bond (viewed from C_β to $O_{\gamma 1}$) involving the side-chain atoms C_α-C_β-$O_{\gamma 1}$-$H^1_{O_{\gamma 1}}$, and the θ_2 dihedral angle is defined in the same way as θ_2 for Val. (See Fig. 1 and Fig. S1.)

We built the Val and Thr dipeptide mimetics (*n*-acetyl-L-Val-methylester and *n*-acetyl-L-Thr-methylester) starting from randomly chosen Val and Thr residues from the Dunbrack database, with all hydrogens added using the REDUCE program (25). (See Fig. 1.) The bond lengths and bond angles of the dipeptide mimetics were set to the mean bond length and angles associated with Val and Thr residues in Dunbrack database. (See Tables 1 and 2, and see Table S1, Table S2, Table S3, and Table S4 and Fig. S2 in the Supporting Material). The backbone conformations were set to canonical α-helix ($\phi = -57°, \psi = -47°$) and β-sheet ($\phi = -119°, \psi = -113°$) values.

All bond lengths, bond angles, and backbone dihedral angles are fixed whereas χ_1, θ_1, and θ_2 are varied. The side-chain dihedral angle χ_1 was sampled from 0° to 360° in 5° increments, whereas θ_1 and θ_2 in Val and θ_2 in Thr were sampled from 0° to 120° in 5° increments because of the threefold symmetry of the methyl group. The value θ_1 for the OH group in Thr is sampled from 0° to 360° in 5° increments. Thus, we sampled 41, 472 ($72 \times 24 \times 24$)

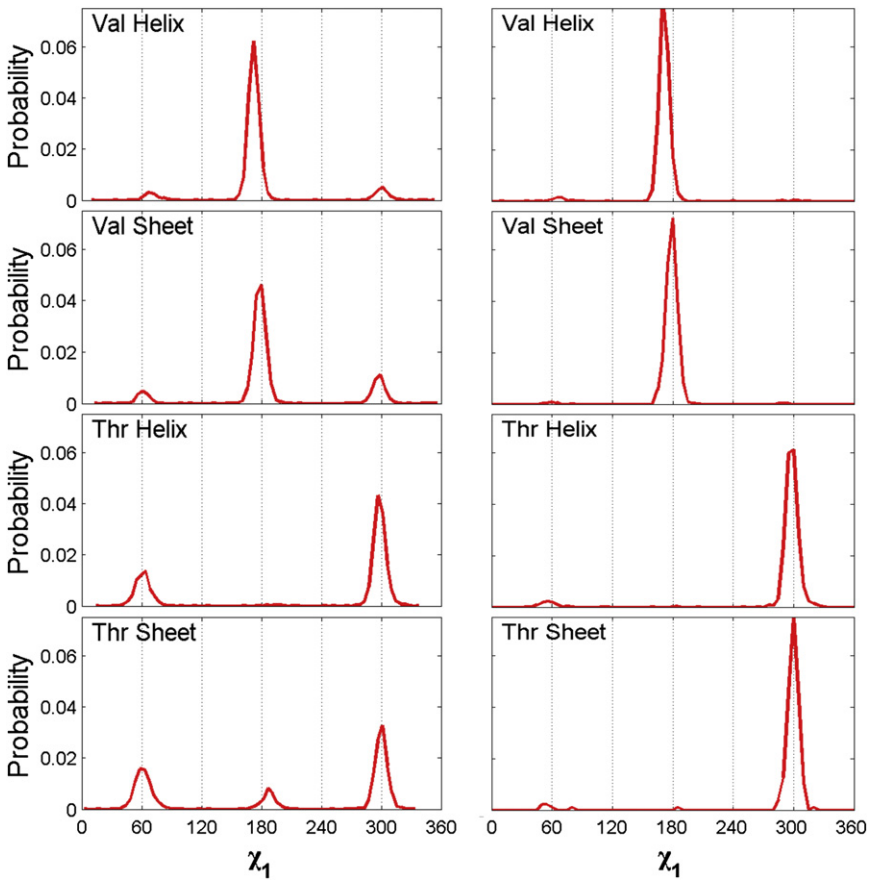

FIGURE 2 Probability distributions of side-chain dihedral angles for Val and Thr. (*Left panel*) Side-chain χ_1 distributions for Thr and Val averaged over all α-helix (or β-sheet) conformations with ϕ- and ψ-angles within the Dunbrack secondary structure classification scheme. (*Right panel*) Side-chain χ_1 distributions for conformations of Thr and Val in the Dunbrack database with ϕ- and ψ-values over narrow ranges ($\pm 10°$) that are centered on the canonical α-helix and β-sheet ϕ- and ψ-angles ($\phi = -57°, \psi = -47°$ for α-helix and $\phi = -119°, \psi = -113°$ for β-sheet).

TABLE 1 Mean bond lengths for the seven bond types in Val and Thr from the Dunbrack database

	Mean bond length (Å)	
Bond type	Val	Thr
C′–O	1.23	1.23
C′–N	1.33	1.33
C_α–C′	1.53	1.53
C_α–N	1.46	1.46
C_α–C_β	1.55	1.54
C_β–$C_{\gamma 1}$	1.52	—
C_β–$O_{\gamma 1}$	—	1.43
C_β–$C_{\gamma 2}$	1.52	1.52

C′ indicates the backbone carbonyl carbon. (See the *top panel* of Fig. S2 for the bond length distributions.)

and 124, 416 ($72 \times 24 \times 72$) combinations of χ_1, θ_1, and θ_2 in Val and Thr, respectively. The positions of the nonrotatable hydrogens, including those in the NH group and on C_α and C_β, are determined by chemistry.

The set of conformations that arise from the dihedral angle rotations, χ_1, θ_1, and θ_2, are obtained using the following computational technique. The new position $\vec{r}' = (x', y', z')$ of an atom originally located at $\vec{r} = (x, y, z)$ following a rotation by an angle w (where w can be χ_1, θ_1, or θ_2) about a unit vector $\hat{n} = (n_x, n_y, n_z)$ (i.e., from C_α to C_β for χ_1 in both Val and Thr, from C_β to $C_{\gamma 1}$ for θ_1 in Val, from C_β to $O_{\gamma 1}$ for θ_1 in Thr, and from C_β to $C_{\gamma 2}$ for θ_2 in both Val and Thr) is given by

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{Q} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{1}$$

where

$$\mathbf{Q} = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2\,(q_1\,q_2 - q_0\,q_3) & 2\,(q_0\,q_2 + q_1\,q_3) \\ 2\,(q_1\,q_2 + q_0\,q_3) & (q_0^2 - q_1^2 + q_2^2 - q_3^2) & 2\,(-q_0\,q_1 + q_2\,q_3) \\ 2\,(-q_0\,q_2 + q_1\,q_3) & 2\,(q_0\,q_1 + q_2\,q_3) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix}, \tag{2}$$

and where $q_0 = \cos w/2$, $q_1 = n_x \sin w/2$, $q_2 = n_y \sin w/2$, and $q_3 = n_z \sin w/2$. The unit vectors $\hat{n}$ are $(\vec{r}_{C_\alpha} - \vec{r}_{C_\beta})/|\vec{r}_{C_\alpha} - \vec{r}_{C_\beta}|$ for χ_1 in both Val and Thr, $(\vec{r}_{C_\beta} - \vec{r}_{C_{\gamma 1}})/|\vec{r}_{C_\beta} - \vec{r}_{C_{\gamma 1}}|$ for θ_1 in Val, $(\vec{r}_{C_\beta} - \vec{r}_{O_{\gamma 1}})/|\vec{r}_{C_\beta} - \vec{r}_{O_{\gamma 1}}|$ for θ_1 in Thr, and $(\vec{r}_{C_\beta} - \vec{r}_{C_{\gamma 2}})/|\vec{r}_{C_\beta} - \vec{r}_{C_{\gamma 2}}|$ for θ_2 in both Val and Thr. Because the dihedral angle rotations about $\hat{n}_{\chi_1}$, $\hat{n}_{\theta_1}$, and $\hat{n}_{\theta_2}$ commute, the order in which the rotations are performed does not affect the final location $\vec{r}'$.

We determine for each combination of χ_1, θ_1, and θ_2 the distance between all pairs of nonbonded atoms. If the distance between the centers of all nonbonded atoms satisfies

$$r_{ij} \geq d_{ij}, \tag{3}$$

where d_{ij} is the sum of their radii, we consider the conformation with this particular χ_1, θ_1, and θ_2 as "allowed". The proportion of θ_1 and θ_2 combinations that yield allowed conformations was tabulated for each χ_1 sampled. From this fraction, we calculate the probability distribution of allowed side-chain dihedral angles for Val and Thr.

We varied the five atom sizes in increments of 0.05 Å over the ranges indicated in Table 3. Thus, 6×10^4 different combinations of atomic radii were tested. For each combination of atom sizes and dipeptide model (Val in α-helix, Val in β-sheet, Thr in α-helix, and Thr in β-sheet), we scored the predicted χ_1 probability distributions to determine how well they match the observed distributions in the right panel of Fig. 2.

We considered a spectrum of scoring functions to determine the optimal values for the five atomic radii. For example, we considered the scoring function

$$S_1 = \sum_{\chi_1 = 0^\circ}^{360^\circ} \left(P_{\text{obs}}(\chi_1) - P_{\text{calc}}(\chi_1)\right)^2, \tag{4}$$

where $P_{\text{obs}}(\chi_1)$ is the observed probability distribution of χ_1 from the Dunbrack database in the right panel of Fig. 2, and $P_{\text{calc}}(\chi_1)$ is our prediction.

TABLE 2 Mean bond angle for the 10 bond angles in Val and Thr from the Dunbrack database

	Mean bond angle (°)	
Bond angle	Val	Thr
C′($i-1$)–N–C_α	121.7	121.0
N–C_α–C′	110.2	111.2
N–C_α–C_β	111.2	109.4
C′–C_α–C_β	109.3	111.0
C_α–C′–O	120.7	120.7
C_α–C′–N($i+1$)	116.2	116.5
O–C′–N($i+1$)	122.9	122.7
$C_{\gamma 1}$–C_β–$C_{\gamma 2}$	110.2	—
$O_{\gamma 1}$–C_β–$C_{\gamma 2}$	—	109.2
$C_{\gamma 1}$–C_β–C_α	110.2	—
$O_{\gamma 1}$–C_β–C_α	—	109.5
$C_{\gamma 2}$–C_β–C_α	110.0	110.6

C′ indicates the backbone carbonyl carbon; C′($i-1$) is the carbonyl carbon in the residue preceding Val or Thr; N($i+1$) is the nitrogen atom in the residue following Val or Thr. (See the *bottom panel* of Fig. S2 for the bond angle distributions.)

Smaller scores indicate closer correspondence between P_{obs} and P_{calc}. However, S_1 over-weights matches of the highest peaks.

We also considered the Matthew's correlation coefficient S_2 (26). We first convert the probabilities at each χ_1 to binary numbers $p(\chi_1) = 1$ (for $P(\chi_1) > \varepsilon$) and $p(\chi_1) = 0$ (for $P(\chi_1) < \varepsilon$), where $\varepsilon = 6 \times 10^{-5}$ is the smallest nonzero value for the probability in the Dunbrack database. We define

$$S_2 = \frac{(TP)(TN) - (FP)(FN)}{\sqrt{(TP + FN)(TN + FP)(TP + FP)(TN + FN)}}, \tag{5}$$

TABLE 3 Atom sizes screened in the hard-sphere model of Val and Thr dipeptide mimetics

Atom	Range of radii (Å)	Number of sizes tested
H	0.90–1.10	5
N	1.25–1.60	8
O	1.10–1.55	10
C	1.30–2.00	15
C′	1.30–1.75	10

For each combination of atom sizes (using increments of 0.05 Å), we calculated the interatomic distances of nonbonded atoms to determine steric clashes.

where $TP = \sum_{\chi_1} p_{obs} p_{calc}$ is the total number of true-positive events, $TN = \sum_{\chi_1} \bar{p}_{obs} \bar{p}_{calc}$ is the total number of true-negative events, $FP = \sum_{\chi_1} \bar{p}_{obs} p_{calc}$ is the total number of false-positive events, and $FN = \sum_{\chi_1} p_{obs} \bar{p}_{calc}$ is the total number of false-negative events. The value $\bar{p}$ is the logical NOT operator that changes the zero bit to 1 and the one bit to zero. $S_2 = 1$ represents a perfect prediction and $S_2 = 0$ represents a random prediction. We find that S_2 over-weights matches between the smallest peaks, which are subject to large statistical fluctuations due to limited data.

We found that using a third scoring function S_3 ameliorates the problems associated with the first two scoring functions. We define

$$S_3 = \sum_{\chi_1=0^\circ}^{360^\circ} \frac{P_{obs}(\chi_1) + P_{calc}(\chi_1)}{2} e^{-\frac{(P_{obs}(\chi_1) - P_{calc}(\chi_1))^2}{8\sigma^2}}. \quad (6)$$

The maximum score for a perfect match between the observed and calculated distributions is 1. The minimum score ($S_3^{min} \approx e^{-1/(8\,\sigma^2)} \approx 0$) occurs when there is no overlap between P_{obs} and P_{calc}. The prefactor $P_{obs} + P_{calc}$ of the scoring function weights the matches between the high probability peaks more than matches between lower probability subpeaks. The Gaussian factor with $\sigma = 0.04$ was included to appropriately weight matches of peaks of differing heights.

By applying the scoring function S_3, we obtain the best overall match between the observed and calculated side-chain probability distributions for Val and Thr (in α-helix and β-sheet backbone conformations) with the following radii: 1.05 Å for hydrogen, 1.5 Å for sp^3 carbon, 1.4 Å for sp^2 carbon, 1.4 Å for nitrogen, and 1.45 Å for oxygen (Fig. 3). We optimized over all four distributions at once because we sought a generally applicable result rather than multiple, optimal atom sizes that depend on amino-acid type and backbone conformation. In Fig. 5, we show a comparison of P_{calc} and P_{obs} for the optimal set of atom sizes. In contrast, we show P_{calc} for typical nonoptimized sets of atom sizes in Fig. S3.

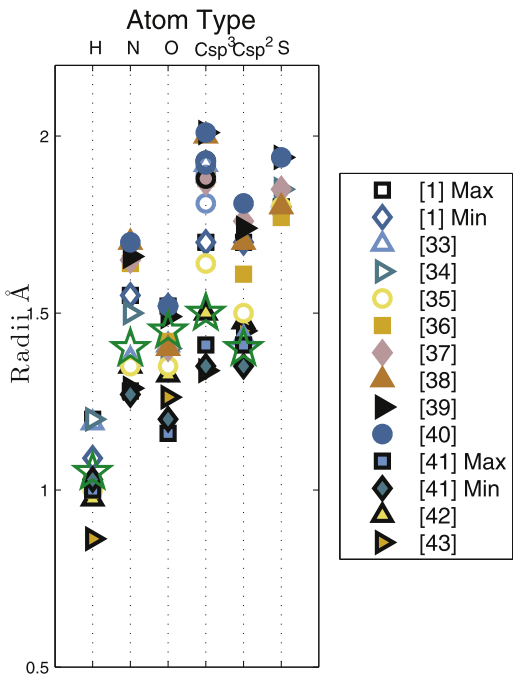

FIGURE 3 Reported values of atomic radii from the literature. Atomic radii from 12 different sources are indicated for hydrogen, nitrogen, oxygen, carbon (sp^3 and sp^2), and sulfur. (*Stars*) The atom sizes that yield the best match between our calculated and the observed distributions of side-chain dihedral angles for Val and Thr. In our model, all carbons are sp^3 except for the main chain carbonyl carbon (C′), which is sp^2. Note references (1,33–43) in the right panel.

The sensitivity of the score S_3 to variations in the atomic sizes is shown in Fig. S4 and Fig. S5. In all cases, the values selected for the atomic radii correspond to minima in S_3. The score S_3 is most sensitive to the size of the Csp^2 and Csp^3 atom sizes. We also confirmed that our hard-sphere model with the optimized set of atomic radii captures the essential features of the Ramachandran map of backbone dihedral angles ϕ and ψ of the alanyl dipeptide mimetic (*n*-acetyl-L-Ala-methylester).

RESULTS

We investigated the extent to which simple hard-sphere models for dipeptide mimetics can predict the allowed conformations of amino-acid side chains. Specifically, we calculate the side-chain dihedral angle distribution $P(\chi_1)$ for Val and Thr in canonical α-helix and β-sheet backbone conformations. We compare the results of our calculations with the distributions of the side-chain dihedral angle χ_1 for Val and Thr observed in a database of high-resolution nonhomologous protein crystal structures (20). Such comparisons are the benchmark for our calculations. The side-chain dihedral angle distributions of χ_1 for Val and Thr from the Dunbrack database are shown in Fig. 2. The left panel shows the χ_1 distributions obtained from all residues designated as either α-helix or β-sheet and the right panel shows the distribution for the canonical α-helix and β-sheet ϕ- and ψ-values ($\phi = -57^\circ$, $\psi = -47^\circ$ for α-helix and $\phi = -119^\circ$, $\psi = 113^\circ$ for β-sheet).

The fact that the two χ_1 distributions in Fig. 2 have small but significant differences indicates that the side-chain dihedral angle distributions vary to some extent with the backbone conformation. A comprehensive study of the extent of such variations will be the subject of future work. In this article, we focus on calculations of the sterically allowed side-chain dihedral angle conformations for the canonical α-helix and β-sheet ϕ- and ψ-values.

The distributions for Val in both the α-helix and β-sheet conformations possess a major peak at $\chi_1 = 180^\circ$ and minor peaks at 60° and 300° that are ϕ- and ψ-dependent. In contrast, the χ_1 distribution for Thr displays a major peak at 300° with minor peaks at 60° and 300° that are ϕ- and ψ-dependent.

It has been suggested that the absence of the peak in the side-chain dihedral angle distribution at $\chi_1 = 180^\circ$ for Thr in an α-helix conformation is because Thr at position i can hydrogen-bond with residues at $(i - 4)$ when $\chi_1 = 60^\circ$ or 300° (21). It has also been suggested that the $\chi_1 = 180^\circ$ conformation is disallowed because it leads to steric clashes between the methyl C_γ of Thr at position i and the main-chain oxygen of residue $(i - 3)$ (22,23). Whereas these explanations of the side-chain dihedral angle distributions emphasize *interresidue* constraints, our studies focus on local steric constraints.

We employ a hard-sphere representation of Val and Thr dipeptide mimetics with fixed bond lengths and angles (see Tables 1 and 2). The dipeptide mimetics possess a central amino acid whose side chain is either Val or Thr,

which is flanked by "half of an amino-acid backbone" on either side (Fig. 1 and Fig. S1). To specify the geometry of the dipeptide mimetic, we set the values of the 7 bond lengths and 10 bond angles for Val and Thr to be the mean values of those bond lengths and angles observed in the Dunbrack database. For both Val and Thr dipeptide mimetics, the backbone conformations were set to either canonical α-helix or β-sheet ϕ- and ψ-angles. Our model includes explicit hydrogen atoms (10 on Val and 8 on Thr) that are initially placed using the REDUCE software package (25). The side-chain methyl and hydroxyl groups are subsequently rotated by the dihedral angles θ_1 and θ_2 to sample different conformations (Fig. 1).

We then calculate whether a particular combination of χ_1, θ_1, and θ_2 is sterically allowed (no clashes) or disallowed (one or more clashes) for each combination of sizes for all five atom types (H, N, O, C(sp^3), and C(sp^2)). Fig. 3 shows the ranges of radii for each atom type that have been reported in the literature. These values include measurements from gas phase experiments, small molecule and protein crystal structures, and quantum mechanics calculations, and thus represent reasonable ranges for our study. Therefore, in our calculations, we varied atom sizes in increments of 0.05Å over the following ranges: H = 0.9–1.1 Å, N = 1.25–1.6 Å, O = 1.1–1.55 Å, Csp^3 = 1.3–2.0 Å, and Csp^2 = 1.3–1.75 Å (Table 3). For all combinations of atom sizes (6×10^4 in total), we sampled χ_1, θ_1, and θ_2 in 5° increments. We considered several scoring functions and selected S_3 in Eq. 6 to evaluate our results. We describe the different scoring functions and the reason for choosing S_3 in the Materials and Methods section. We thus identified the combination of atom sizes (see *star symbols* in Fig. 3) that gives the best match between our predictions and the observed χ_1 distributions for Val and Thr in the α-helix and β-sheet conformations. In Fig. 4, we describe our method of calculating the χ_1 distribution using Val in an α-helix conformation as an example. For a particular combination of atom sizes, we identify the number of atomic clashes for each combination of χ_1, θ_1, and θ_2. For each 5° χ_1 increment, we calculate the fraction of θ_1, and θ_2 combinations that have no steric clashes (*top panel* of Fig. 4), and plot this fraction (as a normalized probability distribution) versus χ_1 (*bottom panel* of Fig. 4).

In Fig. 5, we compare the predicted $\chi 1$ probability distributions for Val and Thr in the α-helix and β-sheet conformations to the corresponding χ_1 distributions for residues with the canonical α-helix and β-sheet ϕ- and ψ-values taken from the Dunbrack database. Our calculations predict the strongest peaks of the observed side-chain distributions, i.e., the strong peaks at 180° for Val in the α-helix and β-sheet conformations and at 300° for Thr in the α-helix and β-sheet conformations.

Thus, we have demonstrated that using a hard-sphere model of a dipeptide mimetic with a single set of atom sizes is sufficient to describe the major attributes of the side-chain dihedral angle distributions for Val and Thr in both α-helix

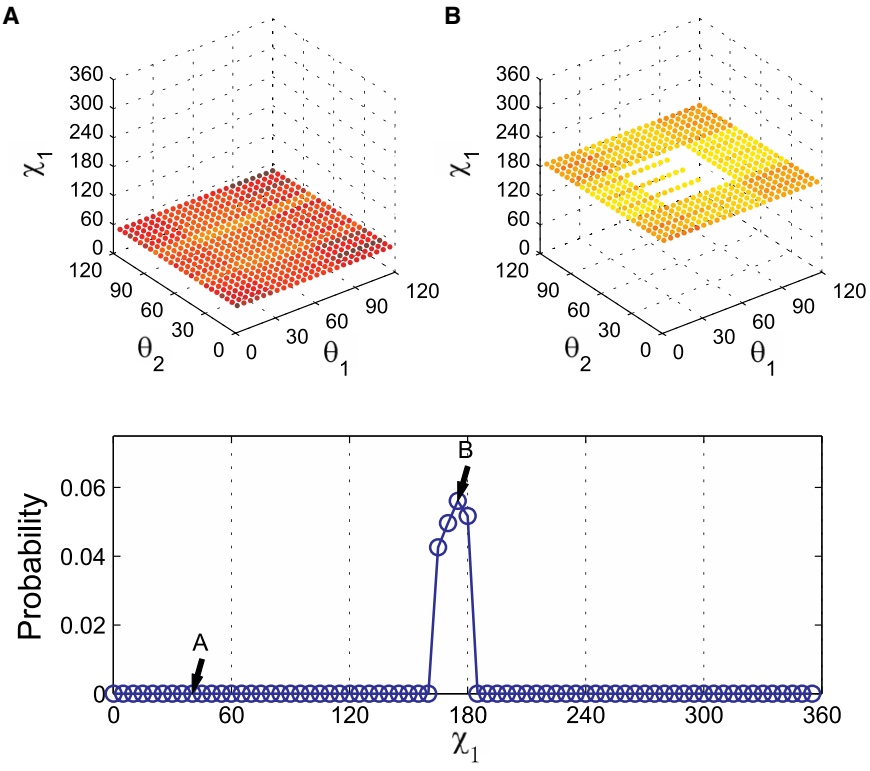

FIGURE 4 Illustration of the prediction of the χ_1 probability distribution using Val in the α-helix backbone conformation. We sampled χ_1 from 0° to 360° and θ_1 and θ_2 from 0° to 120° every 5°. Each data point is colored from white (no clashes) to dark red (highest number of clashes). (*A*, *top panel*) The layer at $\chi_1 = 40°$ with no white data points corresponds to probability $P_{calc} = 0$ as indicated in the bottom panel. (*B*, *top panel*) The layer at $\chi_1 = 175°$ with approximately one-eighth of all possible θ_1 and θ_2 allowed combinations corresponds to point *B* in the lower panel. The same calculations are performed for all sampled values of χ_1 (and shown as the *blue points* in the *lower panel*). The probability P_{calc} is normalized so that the area below the curve is 1. (The set of atom sizes used to calculate P_{calc} is indicated by the *star symbols* in Fig. 3.)

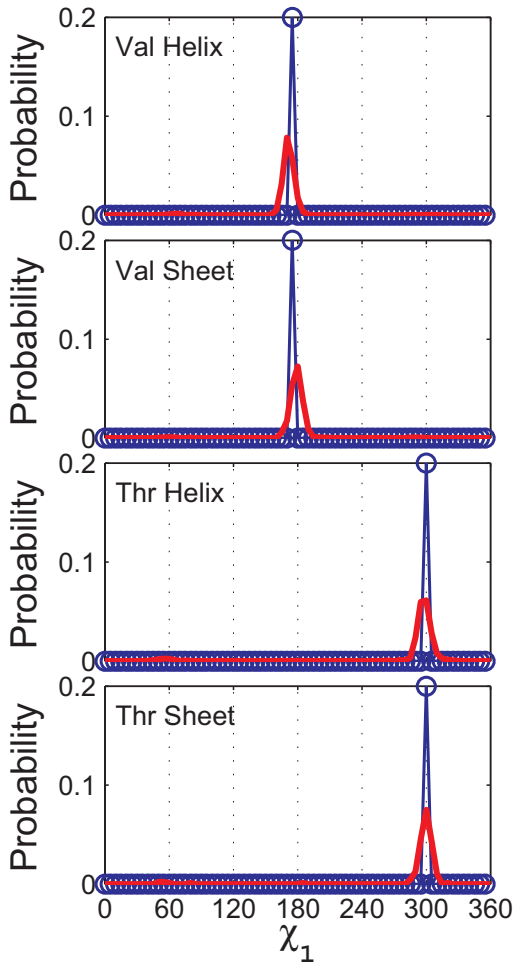

FIGURE 5 Comparison of the probability distributions of the observed and calculated side-chain dihedral angles for Val and Thr. Probability distributions of the side-chain dihedral angle χ_1 from the Dunbrack distribution (20) (*red lines*) are compared to predictions from our hard-sphere model (*blue circles*) for Val (α-helix and β-sheet) and Thr (α-helix and β-sheet). The calculation of the probability distribution for χ_1 for the hard-sphere model is described in the caption to Fig. 4. Both P_{calc} and P_{obs} are normalized so that the area below the curve is 1. (The set of atom sizes used to calculate P_{calc} is indicated by the *star symbols* in Fig. 3.)

and β-sheet backbone conformations. Note that no additional features are required to describe the side-chain dihedral angle distribution for Thr.

DISCUSSION

We have shown that we can predict the key features of the side-chain dihedral angle distributions of Val and Thr in both α-helix and β-sheet backbone conformations using only steric interactions. No steric interactions other than those involving atoms in the dipeptide mimetic are considered. Also, the side-chain OH group of Thr and backbone CO and NH groups are modeled as hard spheres with no polarity and hydrogen-bonding interactions.

Val possesses a simple aliphatic side chain and thus one might not be surprised that a steric model can predict its side-chain dihedral angle distributions (Fig. 5). In contrast, Thr possesses a polar side chain, and thus one might expect that a repulsive hard-sphere model with only intradipeptide interactions would not be able to predict its side-chain dihedral angle distributions. However, the data show that we can predict the side-chain dihedral angle distributions for both Thr and Val. In particular, we predict the absence of major peaks at 180° and the presence of major peaks at 300° and subpeaks at 60° for Thr in both α-helix and β-sheet backbone conformations.

The prevailing two explanations in the literature for the observed side-chain dihedral angle distribution for Thr in the α-helix conformation are the following: 1), The 60° and 300° side-chain dihedral angles are favored because in these conformations the Thr side chain (residue i) can hydrogen bond with the backbone carbonyl oxygen at location $(i-3)$ or $(i-4)$ (21,22). Indeed, McGregor et al. (21) state, "Hydrogen bonding is almost certain to explain why these residues [Ser and Thr] show a shift toward g^+ [300°] in the α-helix." 2), For Thr with $\chi_1 = 180°$, there is an interresidue clash between the methyl C_γ on the Thr side chain at residue i and the main-chain oxygen of residue $(i-3)$ (21,22). However, we argue that these reasons cannot be the fundamental explanations for the observed χ_1 distributions for Thr: 1), The 60° and 300° side-chain dihedral angles are highly populated in both the α-helix and β-sheet conformations. Thus, the hydrogen-bonding based explanation cannot be the fundamental reason for the observed χ_1 distributions because $\chi_1 = 60°$ and 300° do not position the side chain for hydrogen bonding in the β-sheet backbone conformation; and 2), our calculations show that we do not need to include nonlocal interresidue steric clashes (i.e., we do not include residues $(i-3)$ or $(i-4)$ in the model) to predict the observed χ_1 distributions. We believe that our conclusion that steric constraints determine the χ_1 distributions for Thr (without the necessity to include any explicit energetic contribution from side-chain H-bonding) is of fundamental importance. However, once the side chain for Thr in the α-helix conformation is in one of its sterically preferred conformations, it is well positioned to hydrogen bond.

Thus, our key finding is that only local steric constraints among atoms within a dipeptide are sufficient to predict the observed side-chain dihedral angle distributions. Having demonstrated the power of this approach (27), we will perform similar studies of all amino-acid side chains to determine which side-chain dihedral angle distributions can be predicted using simple hard-sphere models of dipeptide mimetics and which require models with additional features. We envision that our hard-sphere model will not be able to capture the behavior of the side-chain dihedral angle distributions for all polar and charged residues. Determining how far the steric model can take us, and when

additional features must be included to reproduce observed behavior, is of key importance and will be the subject of future work.

Our modeling methods do not rely on knowledge-based and statistical approaches that overemphasize average properties of the side-chain dihedral angle distributions. There is ample precedent for this approach in studies of organic molecules (28–31). Instead, we will develop a fundamental understanding of the side-chain dihedral angle distributions, which will allow us to better rank and compare individual protein designs. In addition, our approach will allow us to predict the conformations of peptide foldamers (32), for which there is limited structural data.

SUPPORTING MATERIAL

Six figures and four tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00287-1.

We thank R. L. Dunbrack, Jr. for providing the 2005 version of the Dunbrack database of protein structures. We acknowledge J. Richardson and P. Loria for insights and discussions and thank A. Miranker, S. Mochrie, and Y. Xiong for their critical reading of the manuscript.

We also acknowledge support from the National Science Foundation grant Nos. DMR-1006537 and PHY-1019147 and the Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences.

REFERENCES

1. Ramakrishnan, C., and G. N. Ramachandran. 1965. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* 5:909–933.
2. Kendrew, J. C., G. Bodo, …, D. C. Phillips. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* 181:662–666.
3. Laskowski, R. A., M. W. MacArthur, …, J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26:283–291.
4. Davis, I. W., A. Leaver-Fay, …, D. C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35(Web Server issue):W375–W383.
5. Yun, R. H., and J. Hermans. 1991. Conformational equilibria of valine studied by dynamics simulation. *Protein Eng.* 4:761–766.
6. Makowski, M., E. Sobolewski, …, H. A. Scheraga. 2008. Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. IV. Pairs of different hydrophobic side chains. *J. Phys. Chem. B.* 112:11385–11395.
7. Jacobson, M. P., G. A. Kaminski, …, C. S. Rapp. 2002. Force field validation using protein side chain prediction. *J. Phys. Chem. B.* 106: 11673–11680.
8. Kaminski, G. A., R. A. Friesner, …, W. L. Jorgensen. 2001. Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B.* 105:6474–6487.
9. Brooks, B. R., R. E. Bruccoleri, …, M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
10. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
11. Van Der Spoel, D., E. Lindahl, …, H. J. Berendsen. 2005. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26:1701–1718.
12. Lovell, S. C., J. M. Word, …, D. C. Richardson. 2000. The penultimate rotamer library. *Proteins.* 40:389–408.
13. 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.
14. Ding, F., S. Yin, and N. V. Dokholyan. 2010. Rapid flexible docking using a stochastic rotamer library of ligands. *J. Chem. Inf. Model.* 50:1623–1632.
15. Gordon, D. B., G. K. Hom, …, N. A. Pierce. 2003. Exact rotamer optimization for protein design. *J. Comput. Chem.* 24:232–243.
16. Shah, P. S., G. K. Hom, and S. L. Mayo. 2004. Preprocessing of rotamers for protein design calculations. *J. Comput. Chem.* 25:1797–1800.
17. Havranek, J. J., and D. Baker. 2009. Motif-directed flexible backbone design of functional interactions. *Protein Sci.* 18:1293–1305.
18. Jiang, L., B. Kuhlman, …, D. Baker. 2005. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins.* 58:893–904.
19. Shandler, S. J., M. V. Shapovalov, …, W. F. DeGrado. 2010. Development of a rotamer library for use in β-peptide foldamer computational design. *J. Am. Chem. Soc.* 132:7312–7320.
20. Dunbrack, Jr., R. L., and F. E. Cohen. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* 6: 1661–1681.
21. McGregor, M. J., S. A. Islam, and M. J. E. Sternberg. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* 198:295–310.
22. Gray, T. M., and B. W. Matthews. 1984. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* 175:75–81.
23. Chamberlain, A. K., and J. U. Bowie. 2004. Analysis of side-chain rotamers in transmembrane proteins. *Biophys. J.* 87:3460–3469.
24. Richardson, J. S. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34:167–339.
25. 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.
26. Carugo, O. 2007. Detailed estimation of bioinformatics prediction reliability through the fragmented prediction performance plots. *BMC Bioinformatics.* 8:380–386.
27. Zhou, A. Q., C. S. O'Hern, and L. Regan. 2011. Revisiting the Ramachandran plot from a new angle. *Protein Sci.* 20:1166–1171.
28. Hoffmann, R. W., M. Stahl, …, G. Frenking. 1998. Conformation design of hydrocarbon backbones: a modular approach. *Chemistry.* 4:559–566.
29. Wiberg, K. B., and M. A. Murcko. 1988. Rotational barriers: 2. Energies of alkane rotamers. An examination of *gauche* interactions. *J. Am. Chem. Soc.* 110:8029–8038.
30. Förster, H., and F. Vögtle. 1977. Steric interactions in organic chemistry: spatial requirements of substituents. *Angew. Chem. Int. Ed. Engl.* 16:429–441.
31. Beckelhaupt, F. M., and E. J. Baerends. 2003. The case for steric repulsion causing the staggered conformation of ethane. *Angew. Chem. Int. Ed.* 42:4183–4188.
32. Horne, W. S., and S. H. Gellman. 2008. Foldamers with heterogeneous backbones. *Acc. Chem. Res.* 41:1399–1408.
33. Bondi, A. 1964. Van der waals volumes and radii. *J. Phys. Chem.* 68:441–452 and http://periodictable.com/Properties/A/VanDerWaalsRadius.v.html Accessed December 4, 2011.
34. Element data and radii, Cambridge Crystallographic Data Centre, http://www.ccdc.cam.ac.uk/products/csd/radii Accessed December 4, 2011.

35. Seeliger, D., and B. L. de Groot. 2007. Atomic contacts in protein structures. A detailed analysis of atomic radii, packing, and overlaps. *Proteins.* 68:595–601.

36. Pauling, L. 1948. The Nature of the Chemical Bond, 2nd ed. Cornell University Press, Ithaca, NY 187–193.

37. Porter, L. L., and G. D. Rose. 2011. Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc. Natl. Acad. Sci. USA.* 108:109–113.

38. Tsai, J., R. Taylor, …, M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290:253–266.

39. Chothia, C. 1975. Structural invariants in protein folding. *Nature.* 254: 304–308.

40. Richards, F. M. 1974. The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* 82:1–14.

41. Li, A. J., and R. Nussinov. 1998. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins.* 32:111–127.

42. Momany, F. A., L. M. Carruthers, and H. A. Scheraga. 1974. Intermolecular potentials from crystal data. III Determination of empirical potentials and application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines and amides. *J. Phys. Chem.* 78:1595–1630.

43. Allinger N. L. and Y. H. Yuh. 1980. Quantum Chemistry Program Exchange, Indiana University, Program 395.