

Understanding the physical basis for the side chain conformational preferences of methionine

Alejandro Virrueta^{1,2}, Corey S. O'Hern^{1,2,3,4},
and Lynne Regan^{2,5,6*}

¹Department of Mechanical Engineering & Materials Science,
Yale University;

²Integrated Graduate Program in Physical & Engineering Biology,
Yale University;

³Department of Physics, Yale University;

⁴Department of Applied Physics, Yale University;

⁵Department of Molecular Biophysics & Biochemistry, Yale University;

⁶Department of Chemistry, Yale University.

* Corresponding Author: lynne.regan@yale.edu

Short Title: Computational modeling of methionine residues

Keywords: methionine, hydrophobic amino acids, side chain conformations,
side chain dihedral angles, rotamer prediction,
protein structure prediction, protein-protein interactions

November 11, 2015

1 Abstract

Methionine (Met) is a structurally versatile amino acid most commonly found in protein cores and at protein-protein interfaces. Thus, a complete description of the structure of Met is important for a fundamental understanding of protein structure and design. In previous work, we showed that the hard-sphere dipeptide model is able to recapitulate the side chain dihedral angle distributions observed in high-resolution protein crystal structures for the nine amino acids we have studied to date: Val, Thr, Ser, Leu, Ile, Cys, Tyr, Trp, and Phe. Using the same approach, we are also able to predict the observed χ_1 and χ_2 side chain dihedral angle distributions for Met. However, the form of the side chain dihedral angle distribution $P(\chi_3)$ predicted by the hard-sphere dipeptide model does not match the observed distribution. In this manuscript, we investigate the possible origins of the discrepancy and identify minimal additions to the hard-sphere dipeptide model necessary to quantitatively predict $P(\chi_3)$ of Met. We find that adding weak attractive interactions between hydrogen atoms to the hard-sphere dipeptide model is sufficient to achieve predictions for $P(\chi_3)$ that closely match the observed $P(\chi_3)$ distributions for Met, and its near isosteres norleucine (Nle) and selenomethionine (Mse). We explicitly show that adding weak attractions between hydrogens in the dipeptide model does not negatively affect the agreement between the predicted and observed side chain dihedral angle distribution for Val, as we expect for other amino acids.

2 Introduction

Methionine (Met) is a unique and important amino acid. As the only non-branched, non-aromatic hydrophobic residue, it is structurally versatile and found both in hydrophobic cores and at protein-protein interfaces [1–6]. Thus, the ability to simply and accurately model Met behavior is essential to a fundamental understanding of protein structure and to the design of novel proteins [7–12]. In previous work, using hard-sphere models of dipeptide mimetics, we correctly recapitulated the side chain dihedral angle distributions observed in protein crystal structures for the nine amino acids we have studied to date: Val, Thr, Ser, Leu, Ile, Cys, Tyr, Trp, and Phe [13–15]. The novelty of this work is that we were able to show that local, steric interactions play the dominant role in determining the form of the side chain dihedral angle distributions in proteins. Knowledge-based potentials [16] can also reproduce the observed side chain dihedral angle distributions, but they cannot provide insight into the physical mechanisms that produce them.

Using the hard-sphere dipeptide model, we are also able to predict the observed side chain dihedral distributions for χ_1 ($P(\chi_1)$) and χ_2 ($P(\chi_2)$) for Met as shown in Fig. 1 (left) and (center). However, the hard-sphere dipeptide model is not able to predict the observed side chain dihedral angle distribution $P(\chi_3)$ for Met (Fig. 1 (right)). In this case, the hard-sphere dipeptide model predicts a much flatter $P(\chi_3)$ than the observed distribution, and does not predict the most probable values of χ_3 . Why is the hard-sphere dipeptide model unable to predict the observed $P(\chi_3)$ for Met? In this manuscript, we investigate the possible origins of the discrepancy and identify the minimal additions to the hard-sphere dipeptide model necessary to quantitatively predict $P(\chi_3)$.

We first investigated whether the reported χ_3 side chain dihedral angles for Met residues in protein crystal structures are derived from high-quality electron density data. Met is susceptible to oxidation, and the electron density can often be sparse at the end of the side chain [17–24]. We therefore studied the electron density maps for all Met side chains in a data set of high-resolution protein crystal structures [25–27]. (See Materials and Methods.) We indeed found that for approximately half of the Met residues in the data set, the electron density did not enable accurate positioning of C_γ , S_δ , and C_ϵ in the side chain. Thus, we created another ‘observed’ data set for

χ_3 of Met, which only contained the subset of Met residues for which the electron density map provided accurate positioning of atoms in the side chain that involve χ_3 . The resulting distribution $P(\chi_3)$ was essentially the same as that when all of the Met residues in the high-resolution data set were used, which indicates that the quality of the observed data set is not the origin of the discrepancy.

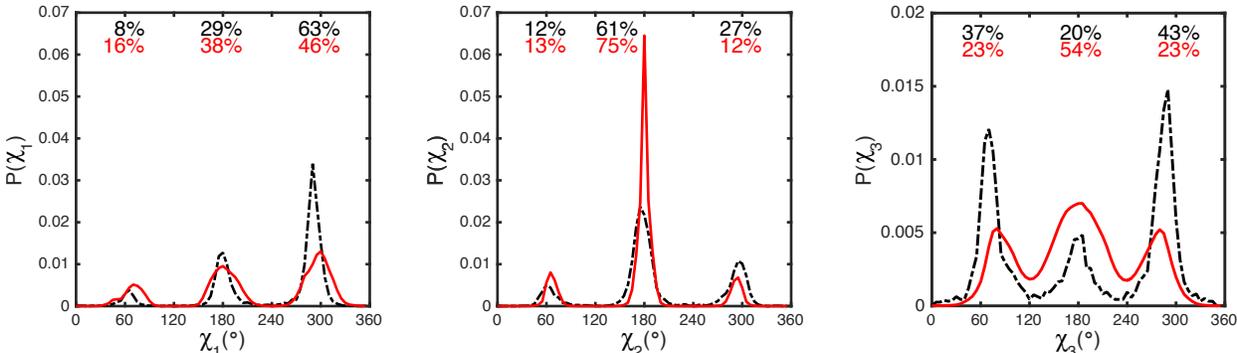


Figure 1: Comparison of the side chain dihedral angle distributions $P(\chi_1)$ (left), $P(\chi_2)$ (center), and $P(\chi_3)$ (right) observed in the Dunbrack 1.7 Å database of high-resolution protein crystal structures (black dashed line) and calculated using the repulsive Lennard-Jones potential (red solid line) for Met dipeptide mimetics. The percentages in each 120° rotamer bin ($0^\circ \leq \chi_{1,2,3} < 120^\circ$, $120^\circ \leq \chi_{1,2,3} < 240^\circ$, and $240^\circ \leq \chi_{1,2,3} < 360^\circ$) are indicated for the observed (top, black) and calculated (bottom, red) distributions. Note that the maximum on the vertical axis in right panel is a factor of 3.5 times smaller than the maxima in the left and center panels to emphasize the differences between the predicted and observed distributions.

Second, because Met possesses the longest aliphatic side chain, it is possible that $P(\chi_3)$ for Met is more strongly influenced by interactions between atoms on different residues than more compact amino acids [22, 28]. Thus, we also calculated $P(\chi_3)$ from the hard-sphere model of Met residues in the context of the protein, including both inter- and intra-residue steric interactions between atoms. However, we do not find significant differences between the $P(\chi_3)$ calculated using the hard-sphere dipeptide model and the model that includes both inter- and intra-residue interactions.

We also studied the near isosteres of Met, norleucine (Nle) and selenomethionine (Mse), in which the sulfur atom at the side chain δ position is either carbon or selenium, respectively. For Nle, we find that $P(\chi_3)$ predicted by the hard-sphere dipeptide model correctly recapitulates $P(\chi_3)$

observed in high-resolution protein crystal structures. In contrast, the hard-sphere dipeptide model is not able to predict the observed $P(\chi_3)$ for Mse. From these studies, we also showed that interactions between the hydrogens on C_ϵ and the hydrogens on C_γ strongly influence the shape of $P(\chi_3)$.

The calculations of $P(\chi_3)$ for Nle and Mse motivated us to investigate the extent to which the addition of short-range attractive interactions between hydrogen atoms influences the predicted $P(\chi_3)$ distribution. We find that when we add weak hydrogen-hydrogen attractions, the predicted distributions $P(\chi_3)$ for Met and Mse quantitatively match the corresponding observed distributions. Moreover, we show that when the hydrogen-hydrogen attractions are included, the predictions for $P(\chi_1)$ and $P(\chi_2)$ for Met, as well as the prediction for $P(\chi_1)$ for Val, are unchanged. Thus, we also expect the predicted side chain dihedral angle distributions for Ser, Leu, Ile, Cys, Tyr, and Phe to remain in general agreement with the corresponding observed distributions. Thus, we have identified a minimal addition to the hard-sphere dipeptide model that can predict the side chain dihedral angle distributions of amino acids, and which also has the potential to be used to calibrate the attractive energy between residues in protein cores [29–32].

3 Materials and Methods

3.1 Observed data sets

We employ two data sets of protein crystal structures, the ‘Dunbrack 1.0 Å’ and ‘Dunbrack 1.7 Å’ databases [25–27] for our calculations. The 1.0 Å data set is chosen for its high resolution and is used to construct accurate model dipeptide mimetics that are employed in the calculations. The 1.7 Å data set is a much larger data set, which allows us to construct smooth ‘observed’ side chain dihedral angle distributions, against which we can compare our predicted distributions.

The ‘Dunbrack 1.0 Å’ data set is a collection of *ultra-high* resolution protein crystal structures with resolution ≤ 1.0 Å, R-factors ≤ 0.2 , side-chain B-factor per residue ≤ 30 Å², and sequence identity $\leq 50\%$ culled from the Protein Data Bank (PDB) [26, 27]. This data set includes 220 proteins and 778 Met residues. We use the bond length and bond angle combinations in this data set to construct the Met dipeptide mimetics.

The side chain dihedral angle distributions predicted by the hard-sphere model are compared to those obtained from ‘Dunbrack 1.7 Å’ data set of high-resolution protein crystal structures [25]. This data set includes protein crystal structures with resolution ≤ 1.7 Å, R-factors ≤ 0.25 , side-chain B-factors per residue ≤ 30 Å², and sequence identity $\leq 50\%$. The Dunbrack 1.7 Å database contains 792 protein structures with 3817 Met residues.

For Mse and Nle, we created our own databases of structures. We identified 234 protein structures with resolution ≤ 1.7 Å in the PDB that contain 955 Mse residues. However, there are many fewer high-resolution protein structures that contain Nle. There are only 26 Nle residues in 21 protein structures with resolution ≤ 2.7 Å. We supplemented the protein structures with small-molecule crystal structures from the Cambridge Structural Database [33]. We identified 177 and 314 compounds that include Met- and Nle-like side chains, respectively. We will refer to this set of structures as the ‘small-molecule database’. This data set is available on request.

3.2 Stereochemistry of Met, Mse, and Nle dipeptide mimetics

We studied hard-sphere, explicit hydrogen representations of Met (N-acetyl-L-Met-N'-methylamide), Nle (N-acetyl-L-Nle-N'-methylamide), and Mse (N-acetyl-L-Mse-N'-methylamide) dipeptide mimetics, as shown in Fig. 2. A dipeptide mimetic is a single amino acid (labeled i) plus the C_α , C , and O atoms of the preceding amino acid ($i - 1$) and N , H , and C_α atoms of the proceeding amino acid ($i + 1$).

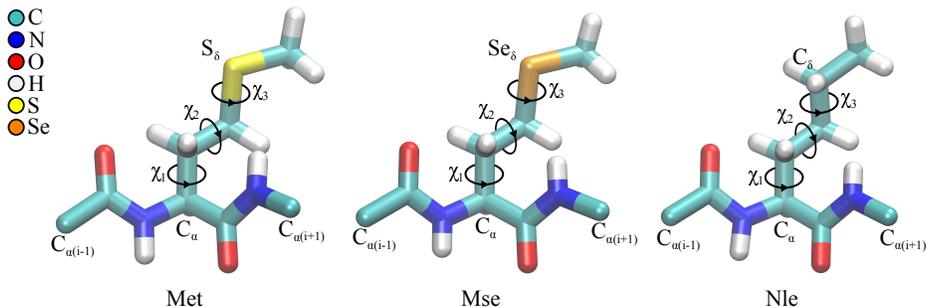


Figure 2: Stick-model representation of Met (left), Mse (center), and Nle (right) dipeptide mimetics. The side-chain dihedral angles χ_1 , χ_2 , and χ_3 and several key atoms are labeled. The residues before ($i - 1$) and after ($i + 1$) the i th central residue are labeled at the C_α atom. These dipeptide mimetics are similar to each other except the atom type and number of hydrogens at the δ atom position on the side chain.

Each atom is represented by a sphere with diameter σ_i , for example, C_{sp^3} : 1.5 Å, C_{sp^2} : 1.3 Å, N : 1.3 Å, and O : 1.4 Å as we used previously in Refs. [13–15, 34, 35]. We also include two types of hydrogen atoms, the amine hydrogen H_N : 1.0 Å and all other hydrogens H : 1.1 Å. We varied the sulfur and selenium atom sizes so that the predicted side chain dihedral angle distributions best fit the observed distributions for Met and Mse. (See Secs. 3.3 and 3.4.) The atom sizes S : 1.75 Å and Se : 1.9 Å (which are similar to tabulated values of van der Waals radii [36–38]) provided the best match between the predicted and observed side chain dihedral angle distributions. Hydrogen atoms were added to the dipeptide mimetics using the program REDUCE [39], which sets the bond lengths for $C-H$ and $N-H$ to 1.1 and 1.0 Å, respectively, and the bond angles to 109.5° and 120° for angles involving C_{sp^2} and C_{sp^3} atoms. Additional dihedral angle degrees of freedom involving hydrogens are chosen to minimize steric clashes [39].

We fixed the bond lengths and angles between all bonded atoms to those observed in the Dun-

brack 1.0 Å or small-molecule database. There is a total of 7 distinct bond lengths and 12 distinct bond angles for Met, Mse, and Nle involving heavy atoms. The observed mean and standard deviations for all of the distinct bond lengths and angles for Met, Mse, and Nle are listed in Tables 1 and 2. To calculate the averages and standard deviations of the bond lengths and bond angles involving backbone atoms, we used 260 of the 778 available Met, 871 of the 955 available Mse, and 23 of the 26 available Nle residues with no missing atoms and without multiple backbone conformations. For Nle, we used 314 side chains from the small-molecule data set to calculate averages and standard deviations for bond lengths and bond angles involving side chain atoms.

Bond type	Average bond length (Å)			Standard deviation of bond length (Å)		
	Met	Mse	Nle	Met	Mse	Nle
C' - C	1.52	1.52	1.52	0.01	0.01	0.02
C' - O	1.24	1.23	1.23	0.01	0.01	0.02
C' - N	1.33	1.33	1.34	0.01	0.01	0.03
C - N	1.46	1.46	1.45	0.01	0.01	0.01
C - C	1.52	1.52	1.50	0.02	0.02	0.07
C - S	1.79	----	----	0.02	----	----
C - Se	----	1.94	----	----	0.04	----

Table 1: Average and standard deviation (in Å) of the 7 distinct bond lengths for Met, Mse, and Nle observed in the Dunbrack 1.0 Å and small-molecule databases.

Angle type	Average bond angle (°)			Standard deviation of bond angle (°)		
	Met	Mse	Nle	Met	Mse	Nle
C - C' - O	120.5	120.5	120.7	1.2	0.9	2.0
O - C' - N	122.6	122.8	122.7	1.2	0.9	1.1
C - C' - N	116.9	116.6	116.6	1.4	1.2	2.1
C' - N - C	121.5	121.3	121.9	1.9	1.6	1.7
N - C - C'	110.5	110.6	110.5	1.4	1.4	3.2
N - C - C	110.8	110.9	109.0	2.0	2.5	4.0
C - C - C'	110.3	110.0	110.2	1.9	1.7	1.5
C - C - C	113.5	113.8	115.2	2.4	1.7	6.5
C - C - S	113.2	---	---	2.7	--	--
C - C - Se	---	112.4	---	--	2.7	--
C - S - C	100.5	---	---	2.3	--	--
C - Se - C	---	98.3	---	--	2.2	--

Table 2: Average and standard deviation (in degrees) of the 12 distinct bond angles for Met, Mse, and Nle observed in the Dunbrack 1.0 Å and small-molecule databases.

3.3 The hard-sphere dipeptide model

To calculate the side chain dihedral angle distributions for a given amino acid, we extract a Met, Mse, or Nle from the Dunbrack 1.0 Å or small-molecule database with a given set of backbone dihedral angles, bond lengths, and bond angles and construct a dipeptide mimetic. We evaluate the total potential energy for the dipeptide mimetic at each side chain dihedral angle combination (χ_1, χ_2, χ_3) , $V(\chi_1, \chi_2, \chi_3) = \sum_{i<j} V_R(r_{ij})$, where

$$V_R(r_{ij}) = \begin{cases} \epsilon_R \left(1 - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)^2, & \text{for } r_{ij} \leq \sigma_{ij} \\ 0, & \text{for } r_{ij} > \sigma_{ij}, \end{cases} \quad (1)$$

is the purely repulsive Lennard-Jones interatomic potential energy, r_{ij} is the separation between the centers of non-bonded atoms i and j , $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, σ_i is the diameter of atom i , and ϵ_R is the strength of the repulsive interactions.

3.4 Short-range attractive interactions between hydrogens

We augment the hard-sphere dipeptide model by including short-range attractive interactions $V_A(r_{ij})$ between hydrogen atom pairs. In this case, the total potential energy for the dipeptide mimetic is $V(\chi_1, \chi_2, \chi_3) = \sum_{i<j} V_R(r_{ij}) + \sum'_{i<j} V_A(r_{ij})$, where the second sum only includes pairs of atoms that are both hydrogens and the first sum includes all other atom pairs,

$$V_A(r_{ij}) = \begin{cases} V_R(r_{ij}) - \frac{\epsilon_A}{\epsilon_R} V_C, & \text{for } r_{ij} \leq \sigma_{ij} \\ \frac{\epsilon_A}{\epsilon_R} (V_R(r_{ij}) - V_C), & \text{for } \sigma_{ij} < r_{ij} \leq 2.5\sigma_{ij} \\ 0, & \text{for } r_{ij} > 2.5\sigma_{ij}, \end{cases} \quad (2)$$

ϵ_A is the strength of the attractive interactions and $V_C = V_R(2.5\sigma_{ij})$ is the potential energy shift that enforces that the potential energy tends to zero continuously at $r_{ij} = 2.5\sigma_{ij}$. As shown in Fig. 3, the form of the pair potential in Eq. 2 allows us to increase the strength of the attractive

interactions, while fixing the repulsive core of the potential by increasing the ratio ϵ_A/ϵ_R . In the limit $\epsilon_A/\epsilon_R \rightarrow 0$, $V_A = V_R$ and we recover the the hard-sphere model with only repulsive Lennard-Jones interactions between all atom pairs.

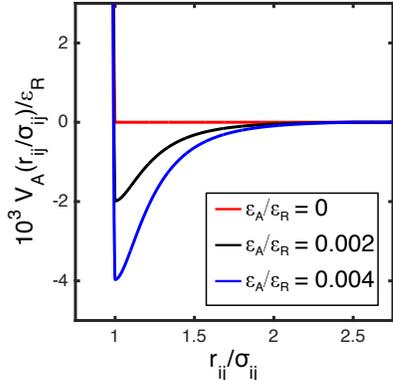


Figure 3: The interatomic potential $V_A(r_{ij})$ for three values of the strength of the attractive interactions $\epsilon_A/\epsilon_R = 0$ (red), 0.002 (black), and 0.004 (blue). When $\epsilon_A/\epsilon_R = 0$, $V_A(r_{ij}) = V_R(r_{ij})$.

3.5 Calculated side chain dihedral angle distributions

For each Met, Mse, or Nle dipeptide mimetic (labeled j), we calculate the Boltzmann weight

$$P_j(\chi_1, \chi_2, \chi_3) \propto e^{-V_j(\chi_1, \chi_2, \chi_3)/k_B T}, \quad (3)$$

where $V_j(\chi_1, \chi_2, \chi_3)$ is the total potential energy of a dipeptide mimetic. We can then calculate the one-dimensional probability distribution $P_j(\chi_3)$ for an individual dipeptide mimetic by integrating over the other two side chain dihedral angles, χ_1 and χ_2 :

$$P_j(\chi_3) = \int P_j(\chi_1, \chi_2, \chi_3) d\chi_1 d\chi_2 = \frac{\int e^{-V_j(\chi_1, \chi_2, \chi_3)/k_B T} d\chi_1 d\chi_2}{\int e^{-V_j(\chi_1, \chi_2, \chi_3)/k_B T} d\chi_1 d\chi_2 d\chi_3}. \quad (4)$$

Similar equations hold for $P_j(\chi_1)$ and $P_j(\chi_2)$. We also calculate the average distributions $P(\chi_1)$, $P(\chi_2)$, and $P(\chi_3)$ for Met, Mse, and Nle by averaging over all dipeptide mimetics of a given type. We set the temperature $k_B T/\epsilon_R < 10^{-3}$ to be sufficiently small so that we are in the hard-sphere limit and $P(\chi_1, \chi_2, \chi_3)$ no longer depends on temperature.

The predictions for $P(\chi_1)$, $P(\chi_2)$, and $P(\chi_3)$ for the hard-sphere dipeptide model will be compared to the corresponding distributions observed in protein crystal structures from the Dun-

brack 1.7 Å database for Met, Mse, and Nle. Note that we obtain approximately the same observed side chain dihedral angle distributions from the Dunbrack 1.7 Å and 1.0 Å databases, but the distributions from the Dunbrack 1.0 Å database have more noise. (See Fig. 1 in Supplementary Information).

3.6 Examination of atom placement relative to the electron density

We also examined the quality of the positioning of three of the atoms (C_γ , S_δ , and C_ϵ) in the Met side chains associated with the χ_3 side chain dihedral angle. We considered the 113 (out of 220) protein structures that contain 300 Met side chains in the Dunbrack 1.0 Å data set for which the electron density is available from the Uppsala Electron Density Server [40]. For each Met side chain, we extracted the observed electron density F_o using PHENIX [41]. We then identified the local maxima of F_o (with observed electron density above three standard deviations) and determined whether they were within the cutoff distance (0.1 Å) of the C_γ , S_δ , and C_ϵ atoms. If each of these three atoms on a given Met side chain could be associated with a local maximum in F_o within the cutoff distance, we considered the side chain conformation to be well-fit to the electron density. Only half of the Met side chains displayed a strong electron density for C_γ , S_δ , and C_ϵ that is well-matched to the model of the atomic positions. Fig. 4 shows two examples of Met side chains, (left) one with an electron density that closely matches the atom placement and (right) one with an electron density that does not match the placement of the atoms.

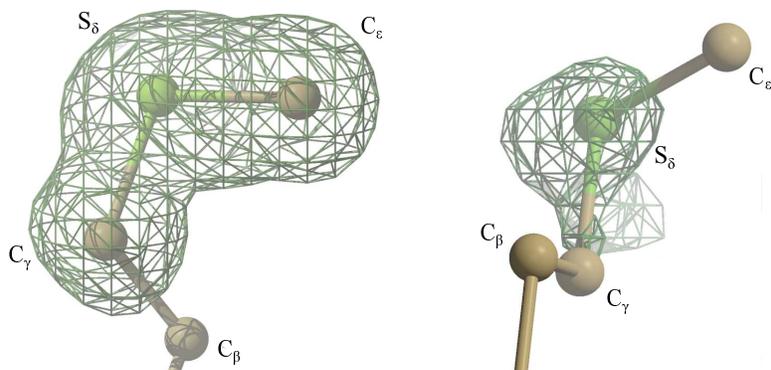


Figure 4: Two examples of Met side chains with observed electron density shown for C_γ , S_δ , and C_ϵ atoms, which are three of the four atoms that define the χ_3 side chain dihedral angle. The left panel shows a Met side chain where the model placement of C_γ , S_δ , and C_ϵ closely matches the observed electron density, and thus the χ_3 side chain dihedral angle is accurately determined. In contrast, for the Met side chain in the right panel, the model placement of S_δ matches the electron density, but there is no observed electron density at the model placement of C_γ and C_ϵ .

4 Results and Discussion

The purely repulsive hard-sphere dipeptide model is able to recapitulate key features (*i.e.* locations of maxima and relative heights of maxima) of the observed side chain dihedral angle distributions $P(\chi_{1,2})$ for all of the uncharged and polar amino acids we have studied [13–15]. In addition, the hard-sphere dipeptide model recovers the overall shape of the observed side chain dihedral angle distributions $P(\chi_1)$ and $P(\chi_2)$ for Met. For example, if the probability is decomposed into three rotamer bins ($0^\circ \leq \chi_{1,2} < 120^\circ$, $120^\circ \leq \chi_{1,2} < 240^\circ$, and $240^\circ \leq \chi_{1,2} < 360^\circ$), the fractional probabilities of the predicted and observed distributions match to within 15% (Fig. 1 (left) and (center)).

However, the prediction for $P(\chi_3)$ for Met from the hard-sphere dipeptide model does not match the observed distribution, for example, the predicted $P(\chi_3)$ is broader than the observed distribution. Also, the hard-sphere dipeptide model predicts 180° as the most probable χ_3 conformation, whereas, the 60° and 300° conformations are the most probable in the observed distribution $P(\chi_3)$ for Met (Fig. 1 (right)).

4.1 Possible reasons for the inability of the hard-sphere dipeptide model to predict $P(\chi_3)$ for Met

There are several reasons for the inability of the hard-sphere model to predict the observed $P(\chi_3)$ for Met. Two of these possibilities are discussed below.

(a) *Poorly resolved electron density maps* For several reasons, including the high propensity for oxidation, model placement of the atoms in the Met side chain often does not match the observed electron density. We therefore examined the electron density and placement of atoms in Met side chains from the structures in the Dunbrack 1.0 Å database. We considered the 113 (out of 220) protein structures for which electron density was available from the Uppsala Electron Density Server [40]. This data set includes 300 Met residues. Indeed, only half of these residues displayed strong electron density for C_γ , S_δ , and C_ϵ that is well-matched to the model of the atomic positions. (See Fig. 4 for examples of well-matched and poorly matched side chain atom placements.) However, we found that $P(\chi_3)$ obtained from the subset of accurately positioned Met side chains was nearly identical to $P(\chi_3)$ averaged over all Met side chains (Fig. 5 (left)), regardless of the quality of their electron density.

We also analyzed the side chain dihedral angle distributions of chemical compounds containing the equivalent of a Met side chain (C_β to C_ϵ) from the small-molecule database. (See Materials and Methods.) As shown in Fig. 5 (center), $P(\chi_3)$ calculated from these Met-like small molecules also matched $P(\chi_3)$ from the Dunbrack 1.7 Å database of protein crystal structures. Thus, the quality of the observed dataset is not the origin of the discrepancy between the predicted and observed $P(\chi_3)$ for Met.

(b) *Local protein environment of Met residues* Another possible source of the discrepancy between the predicted and observed $P(\chi_3)$ is that interactions between the Met side chain and atoms in adjacent residues determine the form of $P(\chi_3)$ for Met [22, 28]. We therefore computed the repulsive interactions (Eq. 1) between pairs of atoms in the Met side chain and all other atoms in the protein crystal structure within 7 Å of C_β of Met, and averaged the predicted distribution $P(\chi_3)$ for all Met residues in the Dunbrack 1.7 Å data set. However, this calculated $P(\chi_3)$ for Met

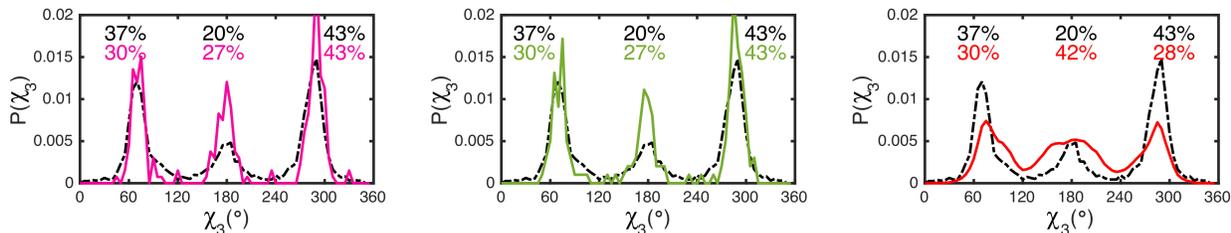


Figure 5: (left) The side chain dihedral angle distribution $P(\chi_3)$ for Met residues from protein crystal structures in the Dunbrack 1.7 Å database that displayed strong electron density for C_γ , S_δ , and C_ϵ that is well-matched to the model of the atomic positions (pink solid line). The percentages in each 120° rotamer bin are indicated for this partial set of structures (bottom, pink). (center) $P(\chi_3)$ for Met-like structures from the Cambridge Structural Database of small-molecule crystal structures (green solid line). The percentages in each 120° rotamer bin are indicated for the small-molecule database (bottom, green). (right) $P(\chi_3)$ for Met dipeptides calculated using the repulsive Lennard-Jones potential including both intra- and inter-residue interactions between atoms in the dipeptide and neighboring residues (red solid line). The percentages in each 120° rotamer bin for the calculated distribution are indicated (bottom, red). In each panel, we also show the observed $P(\chi_3)$ from the Dunbrack 1.7 Å database (black dashed lines) and percentages in each 120° rotamer bin (top, black).

(that includes repulsive intra- and inter-residue atomic interactions) is not significantly different from the predicted $P(\chi_3)$ for the dipeptide mimetic as shown in Fig. 5 (right). Thus, not including interactions between the Met side chain and atoms on neighboring residues does not cause the discrepancy between the predicted and observed $P(\chi_3)$.

4.2 Side chain dihedral angle distributions $P(\chi_3)$ for near isosteres of Met

To gain insight into the form of the side chain dihedral angle distribution $P(\chi_3)$ for Met, we also studied $P(\chi_3)$ for the near isosteres of Met (Nle and Mse) where S_δ is replaced by C_δ and Se_δ , respectively. In Fig. 6 (left), we show the observed $P(\chi_3)$ for Nle from the Dunbrack 1.7 Å database. It possesses a strong peak at $\chi_3 = 180^\circ$ and two minor peaks at 60° and 300° . The hard-sphere dipeptide model for Nle predicts the same form as the observed $P(\chi_3)$ with closely matched fractional probabilities in each of the three 120° rotamer bins.

The observed $P(\chi_3)$ for Mse from the Dunbrack 1.7 Å database (Fig. 6 (right)) is similar to that for Met (Fig. 1 (right)). It is much flatter than $P(\chi_3)$ for Nle, but possesses two small peaks near $\chi_3 = 60^\circ$ and 300° . However, the $P(\chi_3)$ for Mse predicted by the hard-sphere dipeptide model is

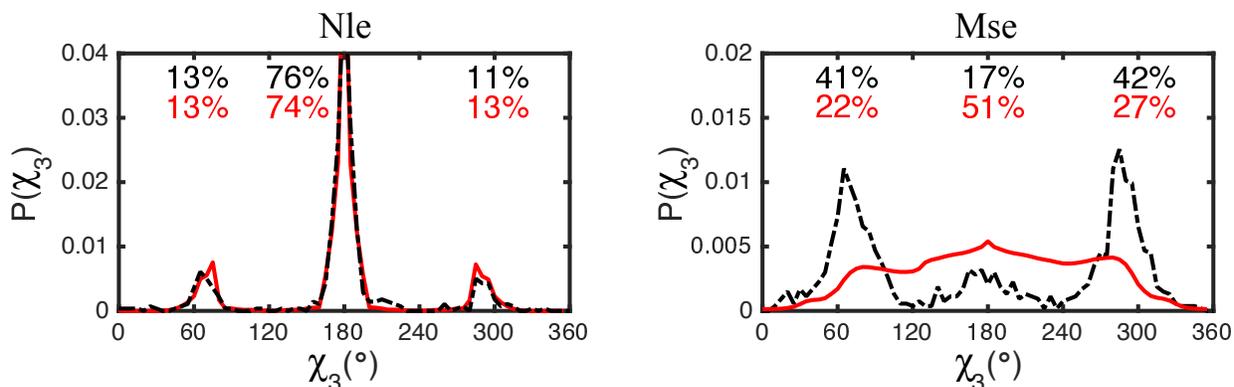


Figure 6: (left) Comparison of the side chain dihedral angle distributions $P(\chi_3)$ for Nle observed in the Dunbrack 1.7 Å database (black dashed line) and calculated using the repulsive Lennard-Jones potential (red solid line). (right) Comparison of the side chain dihedral angle distributions $P(\chi_3)$ for Mse observed in the Dunbrack 1.7 Å database (black dashed line) and calculated using the repulsive Lennard-Jones potential (red solid line). For both panels, the percentages in each 120° rotamer bin are indicated for the observed (top, black) and calculated (bottom, red) distributions.

nearly uniform with a broad maximum near $\chi_3 = 180^\circ$. The hard-sphere dipeptide model predicts $\chi_3 = 180^\circ$ as the most probable rotamer bin for Mse, while conformations in the $\chi_3 = 60^\circ$ and 300° rotamer bins are most probable for the observed distributions.

4.3 Analysis of $P(\chi_3)$ for individual dipeptide mimetics

We analyzed the probability distributions that are calculated for individual Met, Mse, and Nle dipeptide mimetics, instead of averaging over all such dipeptide mimetics in the Dunbrack 1.7 Å database. Surprisingly, the predicted $P_j(\chi_3)$ for approximately 45% of the Met dipeptide mimetics were flat from $\chi_3 = 60^\circ$ to 300° , while the predictions for the remaining 55% yielded peaks near $\chi_3 = 60^\circ$, 180° , and 300° (Fig. 7). However, we found no significant differences in the predicted $P_j(\chi_1)$ and $P_j(\chi_2)$ between the average distribution and distributions obtained from single Met dipeptide mimetics. We hypothesize that the variation in the predicted $P_j(\chi_3)$ distributions is caused by differences in the bond length and bond angle combinations for each individual Met dipeptide mimetic.

We find that the minima near $\chi_3 = 120^\circ$ and 240° in the predicted peaked distributions $P_j(\chi_3)$ are predominantly caused by clashes between the C_ϵ and hydrogens on C_γ , as well as clashes

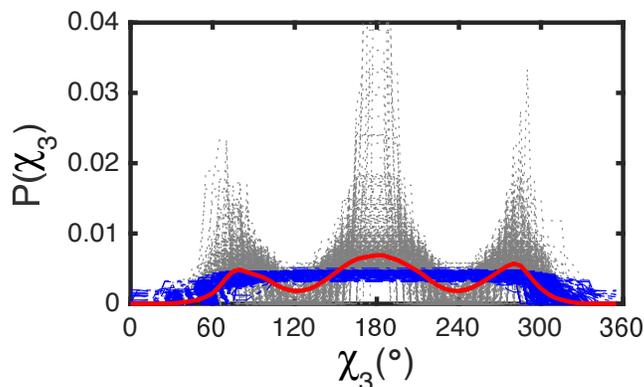


Figure 7: The predicted side chain dihedral angle distribution $P_j(\chi_3)$ from the hard-sphere dipeptide model for individual Met dipeptide mimetics. The distributions are decomposed into two groups: peaked (gray dotted lines) and flat (blue dashed lines). Of the 778 Met dipeptide mimetics examined, 431 (55%) gave ‘peaked’ distributions and 347 (45%) gave ‘flat’ distributions. The solid red line indicates $P(\chi_3)$ averaged over all of the individual Met dipeptide mimetics.

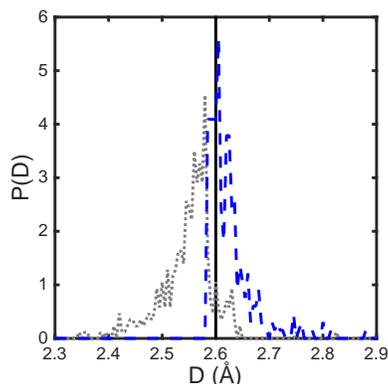


Figure 8: The probability distribution $P(D)$ for the separation D between C_ϵ and the hydrogens on C_γ (when $\chi_3 = 120^\circ$) for the Met dipeptide mimetics that yield ‘peaked’ (gray, dotted) and ‘flat’ (blue, dashed) predicted $P_j(\chi_3)$ distributions. The vertical solid line at the sum of the carbon and hydrogen atomic radii, $D = 2.6 \text{ \AA}$ indicates the threshold below which clashes between C_ϵ and the hydrogens on C_γ will occur.

between hydrogens on C_ϵ and hydrogens on C_γ . These clashes do not occur for dipeptide mimetics with predicted ‘flat’ $P_j(\chi_3)$ distributions.

We measured the distance, D , between C_ϵ and the hydrogens on C_γ at $\chi_3 = 120^\circ$, and examined the distribution $P(D)$ for Met dipeptide mimetics that yield predicted peaked and flat $P_j(\chi_3)$ (Fig. 8). For the dipeptide mimetics with predicted flat $P_j(\chi_3)$ distributions, most of the separations satisfy $D > (\sigma_C + \sigma_H)/2 = 2.6 \text{ \AA}$, and thus C_ϵ and the hydrogens on C_γ do not clash at $\chi_3 = 120^\circ$. In contrast, for the dipeptide mimetics with predicted peaked $P_j(\chi_3)$ distributions, most of the separations satisfy $D < 2.6 \text{ \AA}$, which indicates that clashes occur between C_ϵ and the hydrogens on C_γ at $\chi_3 = 120^\circ$.

We performed a similar analysis of the separation D for individual Nle and Mse residues in

Fig. 9. Due to the differences in the side chain bond lengths and angles (Fig. 2 in the Supplementary Information), $P(D)$ is qualitatively different for Nle, Met and Mse. We find a mean value of $D = 2.5 \text{ \AA}$ for Nle dipeptide mimetics, and approximately 99% of them have a value of $D < 2.6 \text{ \AA}$ at $\chi_3 = 120^\circ$. Thus, for essentially all examples of Nle, there is a clash between C_ϵ and one of the hydrogens on C_γ when $\chi_3 = 120^\circ$ (and $\chi_3 = 240^\circ$), which is responsible for the deep minima in $P(\chi_3)$ near $\chi_3 = 120^\circ$ (and 240°). (See Fig. 6 (left).) Conversely, we find a mean value of $D = 2.76 \text{ \AA}$ for Mse dipeptide mimetics, and approximately 95% of them possess $D > 2.6 \text{ \AA}$ for $\chi_3 = 120^\circ$. $P(\chi_3)$ for Mse is therefore even flatter than $P(\chi_3)$ for Met (Fig. 6 (right)). Nle and Mse thus represent two extremes, and the behavior of Met is intermediate between Nle and Mse. Approximately 55% of Met residues have a value of D below the threshold value of 2.6 \AA , and exhibit peaked distributions, and approximately 45% of Met residues have a value of D above 2.6 \AA , and exhibit flat distributions. The same argument applies to individual Nle and Mse residues. As shown in Fig. 3 in the Supplementary Information, 99% and 95% of Nle and Mse dipeptide mimetics display ‘peaked’ and ‘flat’ $P_j(\chi_3)$ distributions, respectively.

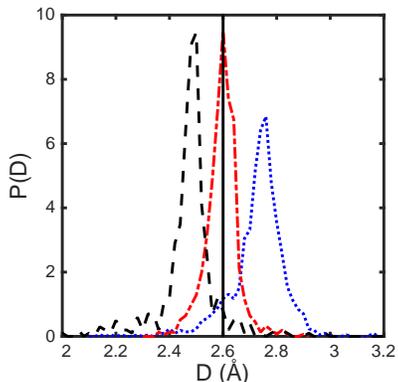


Figure 9: The probability distribution $P(D)$ for the separation D between the C_ϵ and the hydrogens on C_γ at $\chi_3 = 120^\circ$ for all Nle (black dashed line), Met (red dash-dotted line), and Mse (blue dotted line) dipeptide mimetics. The vertical line at $D = (\sigma_C + \sigma_H)/2 = 2.6 \text{ \AA}$ indicates the threshold below which the C_ϵ and the hydrogens on C_γ overlap.

4.4 Generalization of the hard-sphere dipeptide model to include attractive interactions between hydrogen atoms

The above analyses characterized the presence (or absence) and position of minima in the predicted $P(\chi_3)$ for Met, Nle, and Mse based on clashes between side chain atoms. However, even with the understanding of the minima in the predicted $P(\chi_3)$, the relative values of the peaks in the predicted

$P(\chi_3)$ for Met and Mse are still incorrect. Other researchers have proposed that the positioning of χ_3 in Met is influenced by attractive atomic interactions, particularly between hydrogens associated with the β and ϵ carbons [21,22]. We therefore explored the effect on the side chain dihedral angle distributions of increasing the attraction strength (ϵ_A/ϵ_R in Eq. 2) between hydrogen atoms in the dipeptide mimetic model.

We calculated $P(\chi_3)$ for Met, Mse, and Nle dipeptide mimetics over the range from $\epsilon_A/\epsilon_R = 0$ to 0.006 in steps of 0.001. (See Fig. 3.) In Fig. 10, we show the predictions for $P(\chi_1)$ (top row), $P(\chi_2)$ (middle row), and $P(\chi_3)$ (bottom row) for the Met dipeptide model for three values of ϵ_A/ϵ_R : (left) 0, (center) 0.002, and (right) 0.004. As ϵ_A/ϵ_R increases, the predicted peak in $P(\chi_3)$ near $\chi_3 = 180^\circ$ decreases, while the peaks near $\chi_3 = 60^\circ$ and 300° increase. We find that the predicted $P(\chi_3)$ for the dipeptide model with hydrogen-hydrogen attraction strength $\epsilon_A/\epsilon_R \approx 0.002$ closely matches the observed $P(\chi_3)$ for Met. The predicted $P(\chi_1)$ and $P(\chi_2)$ do not change significantly from their respective hard-sphere predictions over the same range of ϵ_A/ϵ_R and both agree with the observed distributions.

In Fig. 11 (right), we demonstrate that the optimal value of the hydrogen-hydrogen attraction strength $\epsilon_A/\epsilon_R = 0.002$ for Met also yields a prediction for $P(\chi_3)$ for Mse that closely matches the observed distribution. In Fig. 6, we found that the predicted $P(\chi_3)$ from the hard-sphere dipeptide model agreed quantitatively with the observed $P(\chi_3)$ for Nle. The hydrogen-hydrogen attractive interactions (with strength $\epsilon_A/\epsilon_R = 0.002$) do not significantly change the predicted $P(\chi_3)$ for Nle (Fig. 11 (left)). The predicted and observed $P(\chi_3)$ for Nle both possess a strong peak near $\chi_3 = 180^\circ$ and minor peaks near $\chi_3 = 60^\circ$ and 300° . We also show in Fig. 4 in the Supplementary Information that nearly all of the predicted $P_j(\chi_3)$ for individual Met, Nle, and Mse dipeptide mimetics are peaked when we include hydrogen-hydrogen attractive interactions with strength $\epsilon_A/\epsilon_R = 0.002$. In contrast, without hydrogen-hydrogen attractive interactions, nearly half of the predicted $P_j(\chi_3)$ distributions for Met dipeptide mimetics (Fig. 7) and nearly all of the predicted $P_j(\chi_3)$ distributions for Mse dipeptide mimetics are flat (Fig. 3 in the Supplementary Information).

We also calculate $P(\chi_1)$ for Val using the dipeptide model with hydrogen-hydrogen attractions

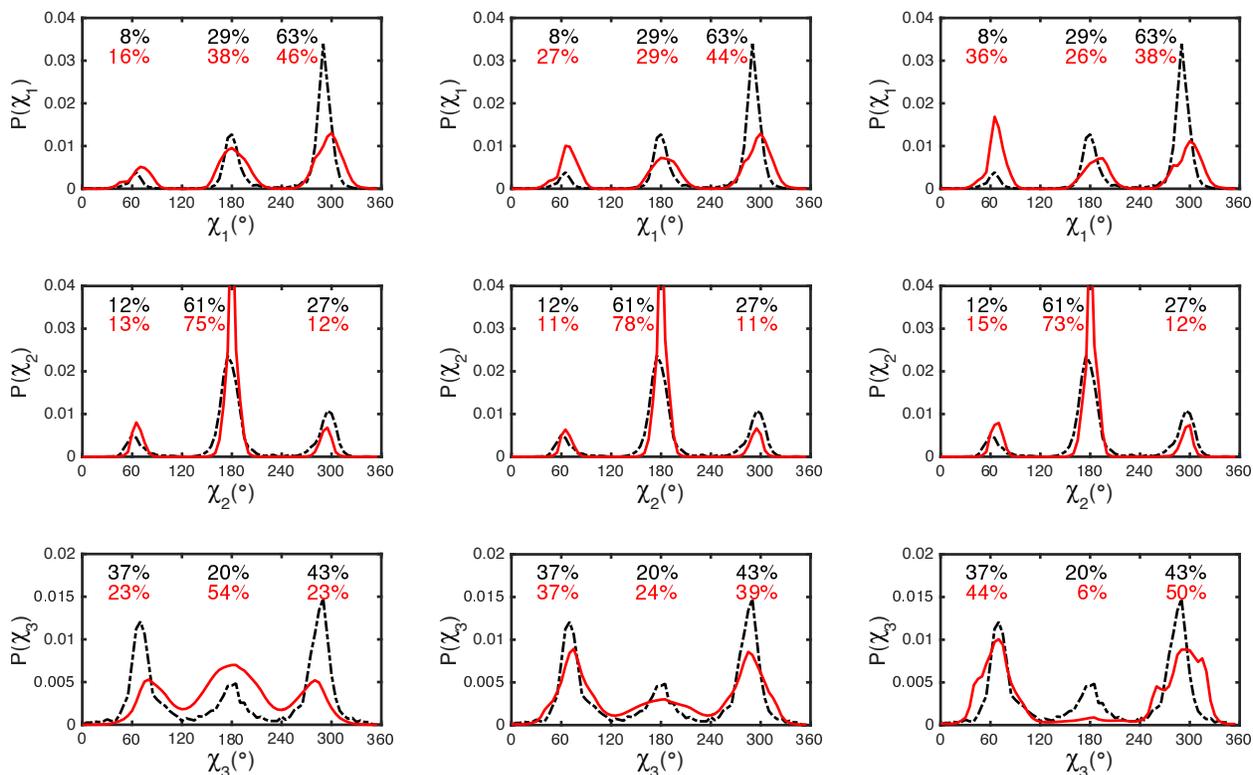


Figure 10: The predicted side chain dihedral angle distributions (top row) $P(\chi_1)$, (middle row) $P(\chi_2)$, and (bottom row) $P(\chi_3)$ for Met from the hard-sphere dipeptide model (red solid lines) with attraction strength (left column) $\epsilon_A/\epsilon_R = 0$, (center column) 0.002, and (right column) 0.004. We also show the observed $P(\chi_1)$, $P(\chi_2)$, and $P(\chi_3)$ for Met from the Dunbrack 1.7 Å database (black dotted lines). The (black) red percentages give the (observed) predicted fractional probabilities in each 120° rotamer bin.

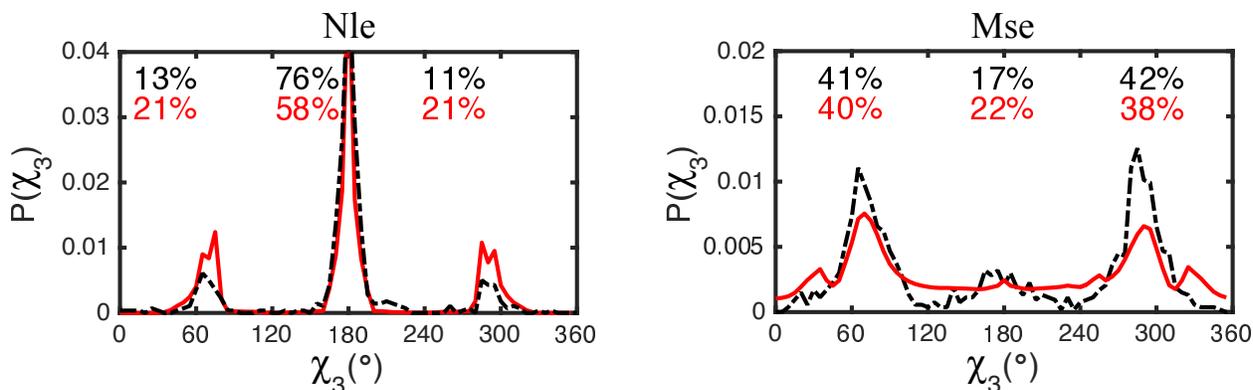


Figure 11: The side chain dihedral angle distributions $P(\chi_3)$ predicted by the dipeptide model with hydrogen-hydrogen attractions with $\epsilon_A/\epsilon_R = 0.002$ (red solid lines) for (left) Nle and (right) Mse dipeptide mimetics. We compare the predicted $P(\chi_3)$ to the observed distributions from the Dunbrack 1.7 Å database (black dashed lines). The (black) red percentages give the (observed) predicted fractional probabilities in each 120° rotamer bin.

with strength $\epsilon_A/\epsilon_R = 0.002$ (Fig. 12). We find that the predicted $P(\chi_1)$ from the dipeptide model that includes short-range attractions between hydrogens is not significantly different from the predicted $P(\chi_1)$ for the hard-sphere dipeptide model without hydrogen-hydrogen attractions. For both predictions, $P(\chi_1)$ possesses a major peak near $\chi_3 = 180^\circ$ and two minor peaks near $\chi_3 = 60^\circ$ and 300° , and the fractional probabilities in each 120° rotamer bin differ by $\approx 10\%$ or less from the observed values. In light of these results, we also expect the predicted side chain dihedral angle distributions $P(\chi_{1,2})$ for Ser, Leu, Ile, Cys, Tyr, and Phe to remain in general agreement with the corresponding observed distributions.

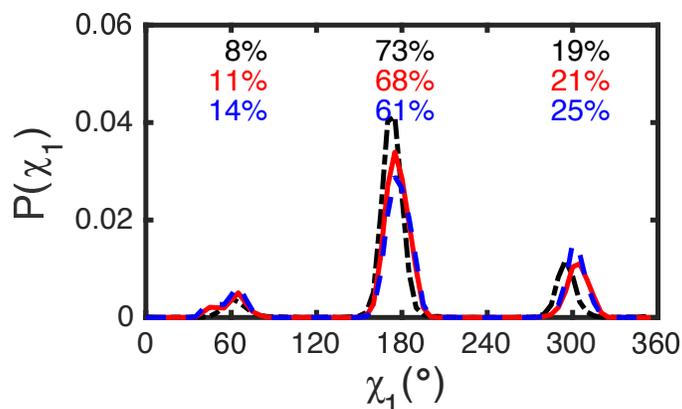


Figure 12: The side chain dihedral angle distribution $P(\chi_1)$ for Val predicted by the dipeptide model with hydrogen-hydrogen attractions with strength $\epsilon_A/\epsilon_R = 0$ (red solid line) and 0.002 (blue dashed line). We also plot the observed $P(\chi_1)$ for Val from the Dunbrack 1.7 Å database (black dashed line). The black (red, blue) percentages give the observed (predicted with $\epsilon_A/\epsilon_R = 0, 0.002$) fractional probabilities in each 120° rotamer bin.

5 Conclusion

In previous work, we demonstrated the power of the hard-sphere model applied to dipeptide mimetics to predict the side chain dihedral angle distributions observed in proteins. This manuscript focuses on a case in which the hard-sphere dipeptide model is no longer sufficient to predict the side chain dihedral angle distributions observed in proteins—namely $P(\chi_3)$ for Met. We explore possible reasons for the discrepancy between the prediction for $P(\chi_3)$ from the hard-sphere dipeptide model and the observed distributions and seek a minimal change to the hard-sphere model that can recapitulate $P(\chi_3)$ for Met. We show that adding weak attractive interactions between hydrogen atoms can successfully reproduce $P(\chi_3)$ for Met. In addition, we examine the behavior of the near isosteres of Met, Nle and Mse, and learn why repulsive interactions alone are sufficient to predict the side chain dihedral angle distribution of Nle (but not of Met and Mse). With this result, we gain an improved physical understanding of the observed side-chain dihedral angle distributions in proteins. This work also represents an important first step in deconvolving attractive van der Waals interactions and solvent-mediated hydrophobic interactions in protein cores and at protein-protein interfaces.

6 Acknowledgments

We acknowledge support from the Ford Foundation Pre-Doctoral Fellowship program (A.V.), National Science Foundation Graduate Research Fellowships program (A.V.), NSF grant Nos. DMR-1307712 (L.R.) and PHY-1522467 (C.O. and L.R.), and the Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences (A.V., C.O., and L.R.). We thank Prof. James Fraser for motivating us to investigate the quality of the protein crystal structures and the electron density of methionine side chains. We also thank Dr. Jimin Wang for his advice and guidance on the analysis of the electron density of Met side chains. We thank Profs. Jane and David Richardson for their continued interest in this work and for many enlightening discussions. We thank Prof. Roland L. Dunbrack, Jr. for stimulating discussions. This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center, and by the NSF under grant No. CNS 08-21132 that partially funded acquisition of the facilities.

References

- [1] N. D. Younan, R. C. Nadal, P. Davies, D. R. Brown, and J. H. Viles. Methionine oxidation perturbs the structural core of the prion protein and suggests a generic misfolding pathway. *J. Biol. Chem.*, 287(34):28263–75, 2012.
- [2] O. Carugo. Stereochemistry of the interaction between methionine sulfur and the protein core. *Biol. Chem.*, 380(4):495–8, 1999.
- [3] N. C. Gassner, W. A. Baase, J. D. Lindstrom, B. K. Shoichet, and B. W. Matthews. Isolation and characterization of multiple-methionine mutants of T4 lysozyme with simplified cores. *Tech. Prot. Chem.*, 8:851863, 1997.
- [4] T. A. Larsen, A. J. Olson, and D. S. Goodsell. Morphology of protein-protein interfaces. *Structure*, 6(4):421 – 427, 1998.
- [5] S. Ovchinnikov, H. Kamisetty, and D. Baker. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife*, 3:e02030, 2014.
- [6] L. L. Conte, C. Chothia, and J. Janin. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.*, 285:2177–2198, 1999.

- [7] D. A. Keedy, C. J. Williams, J. J. Headd, W. B. Arendall 3rd, V. B. Chen, G. J. Kapral, R. A. Gillespie, J. N. Block, A. Zemla, D. C. Richardson, and J. S. Richardson. The other 90% of the protein: Assessment beyond the Calphas for CASP8 template-based and high-accuracy models. *Proteins*, 77 Suppl 9:29–49, 2009.
- [8] P. Gainza, K. E. Roberts, I. Georgiev, R. H. Lilien, D. A. Keedy, C. Y. Chen, F. Reza, A. C. Anderson, D. C. Richardson, J. S. Richardson, and B. R. Donald. Osprey: Protein design with ensembles, flexibility, and provable algorithms. *Method. Enzymol.*, 523:87–107, 2013.
- [9] A. Leaver-Fay, M. J. O’Meara, M. Tyka, R. Jacak, Y. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, J. J. Gray, T. Kortemme, J. S. Richardson, J. J. Havranek, J. Snoeyink, D. Baker, and B. Kuhlman. Scientific benchmarks for guiding macromolecular energy function improvement. *Method. Enzymol.*, 523:109–43, 2013.
- [10] S. Bromberg and K. A. Dill. Side-chain entropy and packing in proteins. *Prot. Sci.*, 3(7):997–1009, 1994.
- [11] P. D. Thomas and K. A. Dill. Statistical potentials extracted from protein structures: How accurate are they? *J. Mol. Biol.*, 257(2):457–69, 1996.
- [12] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–55, 1990.
- [13] A. Q. Zhou, C. S. O’Hern, and L. Regan. The power of hard-sphere models: Explaining side-chain dihedral angle distributions of Thr and Val. *Biophys. J.*, 102:2345–2352, 2012.
- [14] A. Q. Zhou, D. Caballero, C. S. O’Hern, and L. Regan. New insights into the interdependence between amino acid stereochemistry and protein structure. *Biophys. J.*, 105:2403–2411, 2013.
- [15] A. Q. Zhou, C. S. O’Hern, and L. Regan. Predicting the side-chain dihedral angle distributions of non-polar, aromatic, and polar amino acids using hard-sphere models. *Proteins*, 82:2574–2584, 2014.
- [16] M. R. Betancourt. Comparison between molecular dynamic based and knowledge based potentials for protein side chains. *J. Comput. Biol.*, 17(7):943–52, 2010.
- [17] H. Weissbach, F. Etienne, T. Hoshi, S. H. Heinemann, W. T. Lowther, B. Matthews, G. St. John, C. Nathan, and N. Brot. Peptide methionine sulfoxide reductase: Structure, mechanism of action, and biological function. *Arch. Biochem. Biophys.*, 397(2):172178, 2002.
- [18] A. Masato, F. Kiichi, and S. Uchiyama. Suppression of methionine oxidation of a pharmaceutical antibody stored in a polymer-based syringe. *J. Pharm. Sci.*, 2015.
- [19] P. T. Lang, J. M. Holton, J. S. Fraser, and T. Alber. Protein structural ensembles are revealed by redefining x-ray electron density noise. *Proc. Natl. Acad. Sci. USA*, 111(1):237–42, 2014.
- [20] P. T. Lang, H. L. Ng, J. S. Fraser, J. E. Corn, N. Echols, M. Sales, J. M. Holton, and T. Alber. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Prot. Sci.*, 19(7):1420–31, 2010.

- [21] J. M. Word, S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, J. S. Richardson, and D. C. Richardson. Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.*, 285:1711–1733, 1999.
- [22] G. L. Butterfoss, J. S. Richardson, and J. Hermans. Protein imperfections: Separating intrinsic from extrinsic variation of torsion angles. *Acta Cryst.*, D61:88–98, 2005.
- [23] N. C. Gassner, W. A. Baase, and B. W. Matthews. A test of the ‘jigsaw puzzle’ model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl. Acad. Sci. USA*, 93(22):12155–8, 1996.
- [24] S. Fortier. *Direct Methods for Solving Macromolecular Structures*. Nato Science Series C.: Springer Netherlands, 2013.
- [25] R. L. Dunbrack and F. E. Cohen. Bayesian statistical analysis of protein side-chain rotamer preferences. *Prot. Sci.*, 6:1661–1681, 1997.
- [26] G. Wang and R. L. Dunbrack Jr. PISCES: A protein sequence culling server. *Bioinformatics*, 19:1589–1591, 2003.
- [27] G. Wang and R. L. Dunbrack Jr. PISCES: Recent improvements to a PDB sequence culling server. *Nucleic Acids Res.*, 33:W94–W98, 2005.
- [28] G. L. Butterfoss and J. Hermans. Boltzmann-type distribution of side-chain conformation in proteins. *Prot. Sci.*, 12(12):2719–31, 2003.
- [29] C. Lee and M. Levitt. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. *Nature*, 352:448–451, 1991.
- [30] J. Chang and S. I. Sandler. Interatomic lennard-jones potentials of linear and branched alkanes calibrated by gibbs ensemble simulations for vapor-liquid equilibria. *J. Chem. Phys.*, 121:7474, 2004.
- [31] H. Iijima, J. B. Dunbar Jr., and G. R. Marshall. Calibration of effective van der Waals atomic contact radii. *Proteins*, 2:330–339, 1987.
- [32] F. Caillieza and P. Pernot. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *J. Chem. Phys.*, 134, 2011.
- [33] F. H. Allen. The cambridge structural database: A quarter of a million crystal structures and rising. *Acta Crystallogr.*, B58:380–388, 2002.
- [34] A. Q. Zhou, C. S. O’Hern, and L. Regan. Revisiting the Ramachandran plot from a new angle. *Prot. Sci.*, 20:1166–1171, 2011.
- [35] D. Caballero, J. Määttä, A. Q. Zhou, M. Sammalkorpi, L. Regan, and C. S. O’Hern. Intrinsic α -helical and β -sheet conformational preferences: A computational case study of alanine. *Prot. Sci.*, 23:970–980, 2014.
- [36] S. S. Batsanov. Van der Waals radii of elements. *Inorg. Mater.*, 37:871, 2001.

- [37] A. Bondi. Van der Waals volumes and radii. *J. Phys. Chem.*, 68:441, 1964.
- [38] S. Alvarez. A cartography of the van der Waals territories. *Dalton T.*, 42:8617, 2013.
- [39] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson. Asparagine and Glutamine: Using hydrogen atom contacts in the choice of sidechain amide orientation. *J. Mol. Biol.*, 285:1735–1747, 1999.
- [40] G. J. Kleywegt, M. R. Harris, J. Y. Zou, T. C. Taylor, A. Whlby, and T. A. Jones. The Uppsala electron-density server. *Acta Crystallogr.*, D60:2240–2249, 2004.
- [41] P. D. Adams, P. V. Afonine, G. Bunkczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. Phenix: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr.*, D66:213–221, 2010.

1 Supplementary Information

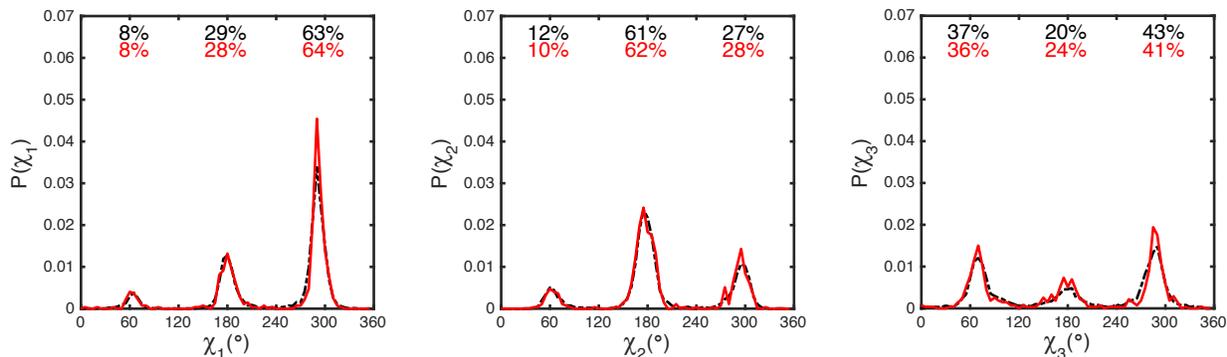


Figure 1: Comparison of the Met side chain dihedral angle distributions $P(\chi_1)$ (left), $P(\chi_2)$ (center), and $P(\chi_3)$ (right) observed in the Dunbrack 1.0 Å (red solid line) and 1.7 Å (black dashed line) databases of protein crystal structures. The percentages in each 120° rotamer bin are indicated for the observed Dunbrack 1.7 Å (top, black) and Dunbrack 1.0 Å (bottom, red) distributions.

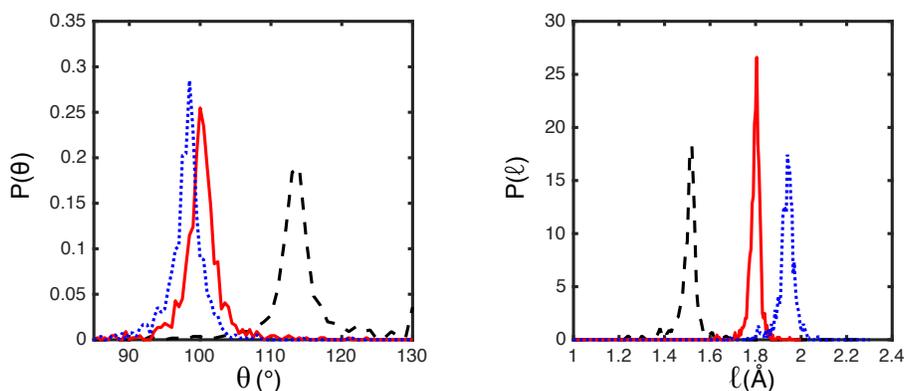


Figure 2: (left) Distribution $P(\theta)$ of the side chain bond angle θ (in degrees) formed by the atoms in the γ , δ , and ϵ locations in Met (C-S-C; red solid line), Mse (C-Se-C; blue dotted line), and Nle (C-C-C; black dashed line) side chains. (right) Distribution $P(\ell)$ of the bond lengths ℓ (in Å) for bonds between the δ atom and γ or ϵ atoms in Met (C-S; red solid line), Mse (C-Se; blue dotted line), and Nle (C-C; black dashed line) side chains.

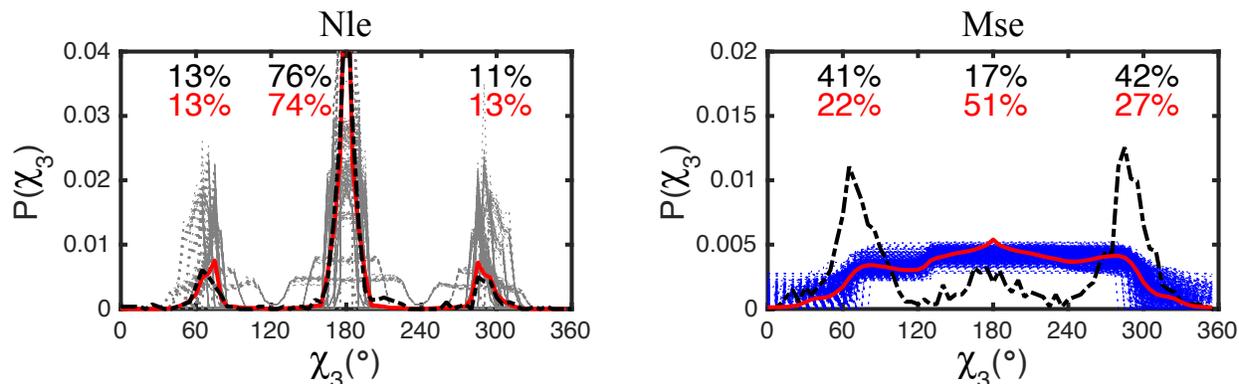


Figure 3: The predicted side chain dihedral angle distribution $P_j(\chi_3)$ from the hard-sphere dipeptide model for individual (left) Nle (gray dotted lines) and (right) Mse residues (blue dotted lines). The solid red line indicates $P(\chi_3)$ averaged over all of the individual dipeptide mimetics. The black dashed line is the observed $P(\chi_3)$ from the Dunbrack 1.7 Å database. For Nle, all of the predicted distributions from individual dipeptide mimetics are peaked, whereas 94% are flat for Mse. The fractional probabilities in each 120° rotamer bin are given for the predicted distribution $P(\chi_3)$ averaged of all dipeptides (red for Nle; blue for Mse) and observed in the Dunbrack 1.7 Å database (black).

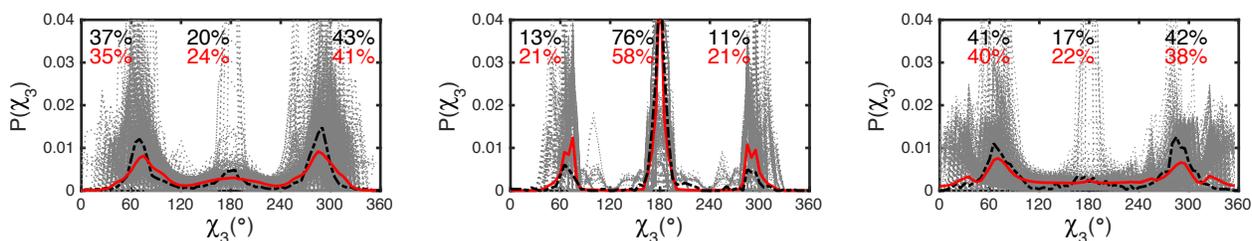


Figure 4: Side chain dihedral angle distributions $P_j(\chi_3)$ for individual Met (left), Nle (center), and Mse (right) dipeptide mimetics (gray dotted lines) computed using the dipeptide model with hydrogen-hydrogen attractive interactions with strength $\epsilon_A/\epsilon_R = 0.002$. The predicted $P(\chi_3)$ averaged over all dipeptide mimetics (solid red line) and $P(\chi_3)$ observed in the Dunbrack 1.7 Å database (black dashed line) are also shown. The percentages in each 120° rotamer bin are indicated for the observed Dunbrack 1.7 Å (top, black) and predicted average (bottom, red) distributions.