Predicting the side-chain dihedral angle distributions of non-polar, aromatic, and polar amino acids using hard sphere models

Alice Qinhua Zhou[a,b,c], Corey S. O'Hern[b,d], Lynne Regan[a,b,*]

[a]Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, CT, USA
[b]Integrated Graduate Program in Physical and Engineering Biology (IGPPEB), Yale University, New Haven, CT, USA
[c]Howard Hughes Medical Institute International Research Fellow
[d]Departments of Mechanical Engineering & Materials Science, Applied Physics, and Physics, Yale University, New Haven, CT, USA
*Corresponding author, Lynne.Regan@yale.edu

Abstract

The side-chain dihedral angle distributions of all amino acids have been measured from myriad high-resolution protein crystal structures. However, we do not yet know the dominant interactions that determine these distributions. Here, we explore to what extent the defining features of the side-chain dihedral angle distributions of different amino acids can be captured by a simple physical model. We find that a hard-sphere model for a dipeptide mimetic that includes only steric interactions plus stereochemical constraints is able to recapitulate the key features of the back-bone dependent observed amino acid side-chain distributions of Ser, Cys, Thr, Val, Ile, Leu, Phe, Tyr, and Trp. We also find that for certain amino acids, performing the calculations with the amino acid of interest in the central position of a short α-helical segment improves the match between the predicted and observed distributions. We also identify the atomic interactions that give rise to the differences between the predicted distributions for the hard-sphere model of the dipeptide and that of the α-helical segment. Finally, we point out a case where the hard-sphere plus stereochemical constraint model is insufficient to recapitulate the observed side-chain dihedral angle distribution – namely the distribution $P(\chi_3)$ for Met.

Introduction

Many types of interactions, including steric, van der Waals, and electrostatic, determine the structure of proteins in general and the conformations of amino acid side chains in particular. However, the relative contributions of the different types of interactions are not known. For example, the dominant interactions that control the side-chain dihedral angle distributions for each residue have not been identified, even though there is now a wealth of high-resolution structural data on thousands of proteins. We seek to determine to what extent the key features of the side-chain dihedral angle distributions of different amino acids observed in protein crystal structures can be captured by steric repulsion and stereochemical constraints alone.

A similar question motivated Ramachandran and colleagues to model peptides as hard spheres with stereochemical constraints for the bond lengths and angles [1, 2]. Using such hard-sphere calculations, they identified the backbone dihedral angle combinations (φ and ψ) for an alanyl dipeptide mimetic that are sterically allowed, given physically reasonable values for the atom sizes, bond lengths, and bond angles. They found that large regions of φ and ψ space are disallowed due to steric clashes between atoms in the dipeptide. Moreover, we now know that essentially every amino acid in every protein structure obeys the hard-sphere limits defined by Ramachandran and coworkers [3, 4].

Common computational strategies for calculating the side-chain dihedral angle distributions in proteins involve quantum mechanical calculations [5] and molecular mechanics force fields, such as Amber [6], CHARMM [7], GROMOS [8], and OPLS [9]. However, these simulation methods require prohibitively large computing resources when considering large proteins. In addition, molecular mechanics force fields directly sample the experimentally measured backbone and side-chain dihedral angle distributions using knowledge-based potentials, such as CHARMM-CMAP [10] and χ-CMAP [11], or Amber-ILDN [12]. Thus, the molecular mechanics force fields do not allow one to quantify the distinct contribution of steric repulsion interactions to the side-chain dihedral angle distributions.

In this manuscript, we ask the key question: To what extent, do steric interactions plus stereochemical constraints determine the side-chain dihedral angle distributions for each residue? Specifically, we determine the sterically allowed side-chain dihedral angle distributions for the nonpolar (Leu, Ile, and Val), aromatic (Phe, Tyr, and Trp), and polar residues (Cys, Ser, and Thr) in either α-helix or β-sheet backbone conformations using a hard-sphere model that only includes steric interactions plus stereochemical constraints for the bond lengths and angles. For each residue, we determine which features of the side-dihedral angle distributions can be explained by a hard-sphere model, in the context of an amino acid dipeptide mimetic (Fig. 1). For several amino acids, the major features of the observed distributions are recapitulated by the hard-sphere model for a dipeptide mimetic. For certain other amino acids, some features of the observed distributions are not captured by the hard-sphere model for a dipeptide mimetic. These discrepancies were especially apparent for α-helical backbone conformations. We therefore also performed hard-sphere calculations of side-chain dihedral angle distributions for each amino acid in the context of a nine-residue α-helical segment. In some cases it is clear that certain side-chain dihedral angle combinations are sterically disallowed due to atomic clashes with the surrounding residues in the α-helical segment, not by clashes within the local environment of the dipeptide mimetic. The ability to predict amino acid side-chain dihedral angle distributions using a simple physical model provides an efficient tool for computational protein design strategies that does not rely on knowledge-based potentials for side-chain dihedral angles [13].

Results

We seek to develop a predictive understanding of the side-chain dihedral angle distributions of several classes of amino acids. We employ a simple approach: modeling amino acids as hard-spheres with stereochemical constraints on the bond lengths and angles. Briefly, using Val for illustration, our strategy is as follows. We first identify all occurrences of the amino acid of interest in the 1.0 Å Dunbrack database of protein crystal structures. (See *Materials and Methods* for a description of the Dunbrack database of high-resolution protein crystal structures.) For example, there are 424 Val with α-helical φ-ψ combinations in the database as shown in Table S1 in *Supplementary Material*. Each of these Val residues possesses slightly different bond lengths and angles. We fix the original backbone dihedral angles φ and ψ, and for each of these Val residues, we calculate whether or not a particular side-chain dihedral angle, χ_1, is sterically allowed or disallowed. We then sum these results for all 424 Val with α-helical backbone conformations to obtain a probability distribution P (χ_1) for each χ_1 for Val (α-helix). We compare the predicted side-chain dihedral angle distributions to the observed distributions obtained from the1.7Å (not 1.0Å) Dunbrack database to have sufficient data (850 versus 221 protein structures) to calculate the probability distribution for χ_1. We follow a similar procedure for Ser, Cys, Thr, Ile, Leu, Phe, Tyr, and Trp.

To improve the agreement between our predictions of the side-chain dihedral angle distributions and the observed distributions for certain amino acids, we performed hard-sphere calculations with the target residue positioned in the center of a nine-residue α-helical segment and the side-chains of the other eight residues in the segment removed (i.e. only the backbone remains). See *Materials and Methods* for a more complete description of the computational methods we employ.

Below, we organize the results into "amino acid types" for clarity. Specifically, we group Ser and Cys; Val and Thr; Ile and Leu; and then Phe, Tyr, and Trp. For each residue, we calculate the side-chain dihedral angle distributions for both α-helical and β-sheet backbone conformations for a dipeptide mimetic as described for Val above. We discuss which features of the observed side-chain dihedral distributions our hard-model can reproduce, and which features it cannot. In several cases, we also compare the results for the side-chain dihedral angle distribution obtained from the dipeptide mimetic with calculations for the same amino acid in the center of a nine-residue α-helical segment.

Cys and Ser

In Fig. 2, we compare the results of our hard-sphere calculations of the side-chain dihedral angle distributions P(χ_1) for Ser and Cys in a dipeptide mimetic to the observed distributions for both α-helical and β-sheet backbone conformations. By considering rotations about a single carbon bond in ethane, one would expect three highly probable conformations at χ_1=60°, 180°, and 300°. The C_α-C_β bonds in Cys and Ser are analogous to the carbon-carbon bond in ethane, but Cys and Ser have asymmetric substituents attached to C_α and C_β . The observed distribution for Ser (α-helix) shows three peaks at χ_1=60°, 180°, and 300°. However, the peak at χ_1=60° is greatly diminished for Cys (α-helix) and nearly absent for both Ser and Cys (β-sheet). Our predictions from the hard-sphere model for a dipeptide mimetic match the observed side-chain distributions for both Ser and Cys in both α-helix and β-sheet backbone conformations. The hard-sphere model predicts that χ_1=60° is more strongly disfavored for Ser (β-sheet) compared to Ser (α-helix) and χ_1=60° is strongly disfavored for Cys (both α-helix and β-sheet).

We identified the steric clashes that determine the form of the allowed side-chain dihedral angle distributions for Cys and Ser (Fig. S1 in *Supplementary Material*). We tracked how the separations between the 47 pairs of atoms change as we vary χ_1 for both Cys and Ser dipeptides

in both α-helix and β-sheet backbone conformations. For each amino acid in each backbone conformation, we studied dipeptide models that sample bond length and bond angle combinations from the 1.0Å database. Clashes between the side-chain O_γ and backbone N^{i+1} and H^{i+1} atoms (where H^{i+1} is the hydrogen on N^{i+1}) disallow 50% of the Ser dipeptides at χ_1=60°. Similar calculations show that clashes between the O_γ-O, S_γ-N^{i+1}, and S_γ-H^{i+1} atom pairs disallow nearly all Ser (β-sheet) and Cys (α-helix) residues at χ_1=60°. In contrast, nearly all Cys (α-helix) residues at χ_1=180° are sterically allowed.

Val and Thr

As with Cys and Ser, one might expect three highly populated side-chain conformations near χ_1=60°, 180°, and 300° for Val and Thr. Indeed, three peaks are observed in the probability distributions $P(\chi_1)$ for Val and Thr (α-helix), but not for Val and Thr (β-sheet). The calculated $P(\chi_1)$ for the β-sheet conformations of both the Val and Thr dipeptide mimetics are similar to the observed $P(\chi_1)$ (Fig. 3 (b) and (e)), *i.e.* we find one strong peak at χ_1=180° for Val and at χ_1=300° for Thr [14]. Note that in the IUPAC nomenclature [15] the definition of χ_1 is different for Thr and Val with respect to the methyl branch. Therefore, χ_1=180° in Val is equivalent to χ_1=300° in Thr.

Because the hard-sphere model for the dipeptide mimetic overpredicts the peaks at χ_1=60° and 300° for Val (α-helix) as well as the peaks at χ_1=60° and 180° for Thr (α-helix) as shown in Fig. 3 (a) and (d), we repeated the calculation in a short α-helical segment to determine if inter-residue interactions were influencing the observed distributions. The calculated $P(\chi_1)$ from the hard-sphere model for Val and Thr in α-helical segments more closely matches the observed distributions. In particular, the calculated peaks at χ_1=300° for Val and χ_1=60° and 180° for Thr are reduced, as shown in Fig. 3. The probability near χ_1=300° for Val is reduced due to clashes between the side-chain methyl -$C_{\gamma 2}H_3$ group (at location i) with the backbone carbonyl group at location i-4. The probability near χ_1=180° for Thr is reduced due to clashes between the side-chain -$C_{\gamma 2}H_3$ group (at location i) with the backbone carbonyl groups at locations i-3 and i-4. The probability near χ_1=60° is reduced by similar clashes with the backbone carbonyl group at location i-4 only.

Leu and Ile

Our previous hard-sphere calculations [16] of the side-chain dihedral angle distributions $P(\chi_1, \chi_2)$ for Leu and Ile in a dipeptide mimetic capture the main features of the observed distributions for both α-helix and β-sheet backbone conformations (Fig. 4). However, there are a few subtle differences. For example, in the observed distribution for Leu (α-helix) conformations near χ_1=300°, χ_2=300° (box 9) are rarely populated, representing only 2% of observed conformations. However, our hard-sphere calculations for Leu (α-helix) in the dipeptide mimetic give 21% for the probability in box 9. In addition, for Ile (α-helix), we predict 25% in box 5 (Fig. 4), whereas the observed probability is 2%.

We therefore also calculated the $P(\chi_1, \chi_2)$ for the hard-sphere model of Leu and Ile in an α-helical segment. For Leu, the predicted probability for conformations in box 9 decreases from 21% to 5% (left panels of Fig 4 (b) and (c)), which more closely matches the the observed probability. The decrease in probability in box 9 is caused by clashes between the methyl group (-$C_{\delta 1}H_3$) of Leu at location i and the backbone carbonyl group (C=O) of the residue at location i-4 (Fig. 5). For Ile, the predicted probabilities for conformations in boxes 5 and 6 change from 18% to 4% and 53% to 78%, respectively (right panels of Fig. 4 (b) and (c)), both of which more closely match the observed probabilities (which are 1% for Box 5 and 85% for Box 6). As shown in Fig.

5, the methyl -$C_{\gamma 2}H_3$ group of the β-branched Ile clashes with the backbone C=O group of both the i-3 and i-4 residues when χ_1<200°, which eliminates conformations in box 5.

The observed and calculated probability distributions for the side chain dihedral angles P(χ_1, χ_2) for Leu and Ile (β-sheet) are shown in Fig. 4. The observed and calculated probability distributions for Leu (β-sheet) both have high probabilities in boxes 6 and 8. However, the hard-sphere model for the Leu dipeptide mimetic (β-sheet) slightly overpredicts conformations in box 9 (11% versus 2%). This slight overprediction is likely because the side chain of Leu at position i would clash with the side chain of the residue at i-2 when it is in a β-sheet structure rather than in a dipeptide mimetic.

For Ile (β-sheet), both the observed and predicted distributions P(χ_1, χ_2) have a strong peak in box 6 and a minor peak in box 3. The predicted and observed distributions match well without considering neighboring residues in segments of β-sheet secondary structure, because Ile is β-branched and clashes between backbone and side-chain atoms within a dipeptide eliminate the (χ_1, χ_2) combinations equivalent to those in box 9 for Leu.

Phe, Tyr, Trp

In Fig. 6, we compare the calculated dihedral angle distributions P(χ_1, χ_2) for Phe, Tyr, and Trp dipeptide mimetics in α-helical and β-sheet backbone conformations with the observed distributions. For Phe, Tyr, and Trp in both α-helical and β-sheet backbone conformations the predicted and observed distributions are qualitatively similar. To quantitatively compare the distributions, we decompose (χ_1, χ_2) space into 3 regions: 0°<χ_1<120°, 120°<χ_2<240°, and 240°<χ_1<360° with 0°<χ_2<180°. For the three aromatic residues, the predicted and observed side-chain dihedral angle distributions agree quantitatively for β-sheet backbone conformations. In contrast, for the three aromatic residues, the hard-sphere dipeptide model overpredicts the probability in the region 240°<χ_1<360° in α-helical backbone conformations by ≈ 20%.

In Fig. 6, we compare the predicted and observed distributions P(χ_1, χ_2) for Phe, Tyr, and Trp in the context of the nine-residue α-helical segment. We find that the distributions calculated in this setting more closely match the observed distributions for aromatic residues. In particular, the predicted probabilities in the region 240°<χ_1<360° decrease from ≈ 40% to 20% for Phe and Tyr and from ≈ 45% to 30% for Trp, which are all within ±5° of the values in the observed distributions. We show in Fig. 7 that clashes occur between atoms on the aromatic side chain and the carbonyl group of the residue at i−4 when the aromatic residues have χ_1=300° and occur in an α-helical segment. In contrast, these atomic clashes do not occur in the context of a dipeptide mimetic. Based on their examination of the side-chain conformations of Phe in the 61 protein crystal structures available at that time, Sternberg and colleagues [17] suggested that such clashes limit the χ_1=300° conformation of Phe, which the current work quantitatively supports.

Discussion and Conclusions

We have investigated to what extent the hard-sphere plus stereochemical constraint model can recapitulate the side-chain dihedral angle distributions observed in high-resolution protein crystal structures, for Ser, Cys, Val, Thr, Leu, Ile, Phe, Tyr, and Trp. We find that for all of these amino acids, a hard-sphere model of a dipeptide mimetic can explain the main features of the observed side-chain dihedral angle distributions for both α-helix and β-sheet backbone conformations. We also show that in some cases, the match between the predictions from the hard-sphere model and the observed distributions can be improved if we consider the amino acid side chains in the context of an α-helical segment rather than in a dipeptide mimetic.

When considering which amino acids to model using only hard-sphere interactions plus stereochemical constraints, we first chose the non-polar amino acids, because it seemed most likely that their behavior would be well-captured by the hard-sphere model. This proved to be the case for the majority of non-polar residues. The exception is Met (Fig. 8 and Table S1 in *Supplementary Material*). In Fig. 9, we compare the observed side-chain dihedral angle distributions, $P(\chi_1)$, $P(\chi_2)$, and $P(\chi_3)$, with those calculated using the hard-sphere model for Met dipeptides in α-helix backbone conformations (Table S1 in in *Supplementary Material*). Note that the hard-sphere model successfully recapitulates the observed $P(\chi_1)$ and $P(\chi_2)$ distributions for Met, but not $P(\chi_3)$. The observed $P(\chi_3)$ shows strong peaks at χ_3=60° and 300°, whereas the predicted $P(\chi_3)$ is nearly flat over the range $60°<\chi_3<300°$ [4]. The difference between the observed and predicted distributions persists when we vary the sulfur atomic size over the range from1.45 to 2.0Å,and also when we study Met in the context of an α-helical segment.

We also investigated the amino acid Norleucine (Nle), a structural analog of Met where the thioether group (-S-) is replaced by the methylene group ($-CH_2-$) (Fig. 8). The observed distribution $P(\chi_3)$ for Nle is different from that of Met in that χ_3=180° is the most probable value of χ_3. In Fig. 9, we show that the $P(\chi_1)$, $P(\chi_2)$, and $P(\chi_3)$ distributions from the hard-sphere dipeptide model for Nle agree qualitatively with the observed distributions. (Note that the observed Nle distributions possess large fluctuations due to the small number of Nle in the protein data bank (PDB).) In future studies, we will investigate why the hard-sphere plus stereochemical constraint model can recapitulate $P(\chi_3)$ for Nle, but not for Met.

Unexpectedly, we discovered that the hard-sphere plus stereochemical constraint model is also able to recapitulate the behavior of the polar side chains Ser and Thr in both the α-helix and β-sheet backbone conformations without including hydrogen-bonding interactions in the model. For example, our results show that χ_1=300° for Thr is sterically allowed, whereas χ_1=60° and 180° are disallowed by repulsive steric interactions. The χ_1=300° conformation positions the Thr side-chain to hydrogen bond with the backbone [31,32]. But even in the absence of H-bonding, the hard sphere model predicts that χ_1=300° is the most populated conformation. We have not yet investigated charged side chains (Lys, Arg, Glu, and Asp) or residues with amide side chains (Asn and Gln).

In summary, we found that side-chain dihedral angle distributions of many amino acids can be accurately modeled using hard-sphere interactions plus stereochemical constraints alone. The success of the hard sphere plus stereochemical constraint model in predicting amino acid side-chain conformation is a strong foundation from which to investigate more complex systems and phenomena, such as the packing and thermodynamic stability of protein cores and protein-protein interactions.

Materials and methods 4.1 Databases of protein crystal structures

Our calculations of the sterically allowed side-chain dihedral angle distributions are compared to the observed side-chain dihedral angle distributions obtained from a database of high-resolution protein crystal structures provided by Dr. Roland Dunbrack, Jr. [18, 19]. This database is composed of 850 non-homologous protein structures with resolution 1.7 Å or less, side chain B-factors per residue 30 $Å^2$ or less, and R-factors 0.2 or less.

We used a higher-resolution (1.0 Å or less) set of 221 structures to fix the bond lengths, bond angles, and ω backbone dihedral angles of the dipeptide mimetics for our calculations. The higher-resolution set limits the refinement bias on the bond lengths, bond angles, and ω dihedral angles associated with the crystal structure determination, but there are too few structures in this

set to provide meaningful comparisons with the calculated side-chain dihedral angle distributions. The numbers of each residue type in both databases are displayed in Table S1 in the *Supplementary Material*. In addition, we identified 53 protein crystal structures containing Nle in the PDB. 24 of the 53 crystal structures are non-homologous. After removing structures with resolution higher than1.7Å, 11 Nle amino acids remained.

Hard-sphere plus stereochemical constraint model of dipeptide mimetics and α-helical segments

Fig. 1 shows a stick representation of the Ala dipeptide mimetic as well as nine other residues (Val, Thr, Phe, Ser, Cys, Tyr, Leu, Ile, and Trp). The φ backbone dihedral angle is defined by the clockwise rotation around the N- C_α bond (viewed from N to C_α) involving the backbone atoms C-N- C_α-C. The ψ backbone dihedral angle is defined by the clockwise rotation about the C_α-C bond (viewed from C_α to C) involving the backbone atoms N-C_α-C-N. The definitions of the ω backbone dihedral angles and the side-chain dihedral angles χ_i of the nine residues considered in this work are provided in Table S2 in *Supplementary Material*. All dihedral angles listed in the Table S2 in *Supplementary Material* range from 0° to 360°, except χ_2 for the aromatic residues Tyr, Trp, and Phe, which range from 0° to 180°. We show in the *Supplementary Material* that the side-chain dihedral angle distributions for Tyr, Trp, and Phe are similar for $0°<\chi_2<180°$ and $180°<\chi_2<360°$ because of the approximate ring inversion symmetry.

We constructed hard-sphere representations of dipeptide mimetics and α-helical segments for each of the nine residues. The structure of the dipeptide mimetics is N-acetyl-X-N'-methylamide, where X is one of the side chains of the nine residues we studied (Fig. 1). Particular combinations of the bond lengths, bond angles, and ω backbone dihedral angles for each dipeptide mimetic were obtained from the ≤ 1.0 Å subset of protein crystal structures. The atomic radii were set to 1.05 Å for Hydrogen, 1.5 Å for sp^3 Carbon, 1.4 Å for sp^2 Carbon, 1.4 Å for Nitrogen, 1.45 Å for Oxygen, based on our previous studies [14, 16, 20]. We fixed the radius of Sulfur to be 1.8 Å consistent with many prior studies [21–29] and the mean bond lengths μ=1.804±0.015 and 1.792±0.015 for C_γ-S_δ and C_ε-S_δ in Met from the 1.0 Å database. Hydrogen atoms were added to the structures using the REDUCE software package [30].

As shown in Fig. 5, the α-helical segments contain nine residues with the target residue positioned in the center at location i, and the side-chains of the other eight residues in the segment have been removed. The α-helical segments extracted from the 1.0Å set of structures satisfy the following criteria: 1) no missing atoms in the backbone of the segment, 2) no Pro residues, and 3) all backbone dihedral angle combinations possess $-80°<\varphi<-20°$ and $-65°<\psi<-20°$. In contrast to our previous studies [14, 16, 33], we do not change the backbone dihedral angles of the dipeptide mimetics and α-helical segments that were extracted from the high-resolution database.

Calculation of the probability distribution of sterically allowed side-chain dihedral angles

Given a particular dipeptide mimetic or α-helical segment, we rotate the side chain of the target residue to a particular conformation specified by $\{\chi_1, \chi_2, \ldots\}$ in small increments $\Delta\chi=5°$. For each side-chain conformation, we determine the separation r_{ij} between the centers of all pairs of nonbonded atoms i and j (with both atoms located on the side chain or with one on the side chain and the other on the backbone). If the separation r_{ij} between all nonbonded atom pairs for a conformation satisfies $r_{ij}\geq\sigma_{ij}$, where σ_{ij} is the sum of the radii of atoms i and j, this conformation is sterically allowed. We then perform this calculation for all possible side chain conformations for each dipeptide mimetic or α-helical segment.

To calculate the probability distributions of sterically allowed side-chain dihedral angles, $P(\chi_1)$ for Val, Thr, Ser, and Cys, and $P(\chi_1, \chi_2)$ for Leu, Ile, Phe, Tyr, and Trp in a dipeptide mimetic or α-helical segment, we first count the number of sterically allowed side-chain dihedral angle combinations in each 5° bin or 5° × 5° box, and then sum over all residues selected from the high-resolution Dunbrack database. Summing over structures in the Dunbrack database allows us to average over random bond length, bond angle, and ω dihedral angle fluctuations.

To investigate the dependence of the side-chain dihedral angle distributions on the backbone conformation, we identified dipeptide mimetics with φ and ψ within ±10° of canonical α-helix (φ=−57°, ψ=−47°) or β-sheet (φ=−119°, ψ=113°) values, and calculated the probability distributions for the side-chain dihedral angles for α-helical and β-sheet backbone conformations separately. We normalize the distributions so that $\int P(\chi_1)\, d\chi_1=1$ or $\int P(\chi_1, \chi_2)\, d\chi_1\, d\chi_2=1$.

Acknowledgments

We thank R.L. Dunbrack, Jr., for stimulating discussions and providing high-resolution sets of structures from the PDB. We also thank J.S. and D.C. Richardson for discussions and their valuable insights. This work was supported by the National Science Foundation (grants DMR-1006537, DMR-1307712 and PHY-1019147) and the Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences. Alice Qinhua Zhou is a Howard Hughes Medical Institute International Research Fellow. The computational studies were supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center and by the National Science Foundation under grant No. CNS 08-21132 that partially funded acquisition of the facilities. We thank the Regan and O'Hern labmates for helpful comments on the manuscript.

References

[1] Ramakrishnan GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 1963;7:95–99.

[2] Ramakrishnan C, and Ramachandran GN. Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* 1965;55:909–933.

[3] Laskowski RA, MacArthur MW, Moss DS and Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 1993;26:283–291.

[4] Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: Structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res*. 2004;32:W615–9.

[5] Zhu X, Lopes PEM, Shim J, & MacKerell AD. Intrinsic energy landscapes of amino acid side-chains. *J. Chem. Info. and Modeling.* 2012;52:1559–1572.

[6] Cornell WD; Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins. *J. Am. Chem. Soc.* 1995;117:5179–5197.

[7] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S and Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 1983;4:187–217.

[8] Van Der Spoel D, Lindahl E., Hess B., Groenhof G., Mark AE and Berendsen HJ. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 2005;26:1701–18.

[9] Kaminski GA, Friesner RA, Tirado-Rives J and Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B.* 2001;105:6474–6487

[10] MacKerell Jr. AD, Feig M, Brooks III CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 2004;25:14001415.

[11] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, and MacKerell AD. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and Side-Chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 2012;8:3257–3273

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

[13] Fleishman SJ, Whitehead TA, Ekiert DC, reyfus C, Corn JE, Strauch EM, Wilson IA, and Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science.* 2011 332:816–821.

[14] Zhou AQ, OHern CS, Regan L. The power of hard-sphere models: Explaining side-chain dihedral angle distributions of Thr and Val. *Biophys. J.* 2012;102:2345–2352.

[15] Richardson JS. The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* 1981;34:167–339.

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

[17] Mcgregor MJ, Islam SA, Sternberg MJE. Analysis of the relationship between side-chain conformation and secondary structure in globular-proteins. *J Mol Biol.* 1987;198:295-310

[18] Wang G, Dunbrack Jr RL. PISCES: A protein sequence culling server. *Bioinformatics* 2003;19:1589-1591.

[19] Wang G, Dunbrack Jr RL. PISCES: Recent improvements to a PDB sequence culling server. *Nucleic Acids Res.* 2005;33:W94–W98.

[20] ZhouAQ, O'Hern CS, ReganL. Revisiting the Ramachandran plot from a new angle. *Protein Sci.* 2011;20:1166–1171.

[21] Bondi A. van der Waals volumes and radii. *J. Phys. Chem.* 1964;68:441–452.

[22] Element data and radii, Cambridge Crystallographic Data Centre, http: //www.ccdc.cam.ac.uk/products/csd/radii; Accessed December 4, 2013.

[23] Seeliger D, de Groot BL. Atomic contacts in protein structures. A detailed analysis of atomic radii, packing, and overlaps. *Proteins.* 2007;68:595–601.

[24] Pauling L. The Nature of the Chemical Bond, 2nd ed.; Cornell University Press: Ithaca, NY 1948;187–193

[25] Chothia C. Structural invariants in protein folding. *Nature.* 1975;254:304–308.

[26] Richards FM. The interpretation of protein structures: Total volume, group volume distributions and packing density. *J. Mol. Biol.* 1974;82:1–14.

[27] Li AJ, Nussinov R. A set of van der Waals and Coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins: Struct. Funct. Genet.* 1998;32:111–127.

[28] Momany FA, Carruthers LM, Scheraga HA. 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.* 1974;78:1595–1630.

[29] Allinger NL, Yuh YH. Quantum Chemistry Program Exchange, Indiana University, 1980;Program 395

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

[31] McGregor MJ, Islam SA, Sternberg MJE. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J. Mol. Biol.* 1987;198:295–310.

[32] Gray TM, Matthews BW. Intrahelical hydrogen bonding of serine, threonine, and cysteine residues within α-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* 1984;175:75–81.

[33] Caballero D, Maatta J, Zhou AQ, Sammalkorpi M, O'Hern CS, Regan L. 2014. The intrinsic α-helical and β-sheet preferences: A computational case study of Alanine. Prot. Sci. [in press]

Tables and Figures

Figure 1: **Stick representation of dipeptide mimetics** Stick representation of dipeptide mimetics for nine residues: Val, Thr, Ser, Cys, Leu, Ile, Phe, Tyr, and Trp. The side-chain dihedral angles χ_1 and χ_2 are highlighted, with positive angles indicated by the arrows. Carbon, nitrogen, oxygen, hydrogen, and sulfur atoms are shaded pink, blue, red, white, and yellow, respectively. We also include the Ala dipeptide mimetic to label the backbone atoms and define the backbone dihedral angles (φ and ψ). Atoms with superscripts i-1 and i+1 refer to the residue order relative to the central residue i.

Figure 2: **Side-chain dihedral angle distributions for Ser and Cys dipeptide mimetics:** Comparison of the observed (*dotted red lines*) and calculated (*solid blue lines*) probability distributions $P(\chi_1)$ of the side-chain dihedral angle χ_1 for Ser and Cys in dipeptide mimetics with backbone dihedral angles φ and ψ within ±10° of canonical α-helix (φ=−57°, ψ=−47°) and β-sheet (φ=−119°, ψ=113°) values. The probabilities are normalized such that $\int P(\chi_1)d\chi_1=1$.

Figure 3: **Side-chain dihedral angle distributions for Val and Thr in dipeptide mimetics and α-helical segments:** Comparison of the observed (*dotted red lines*) and calculated (*solid blue lines*) probability distributions $P(\chi_1)$ for the side chain dihedral angle χ_1 for Val and Thr in dipeptide mimetics with backbone dihedral angles φ and ψ within ±10° of either the canonical α-helix (φ=−57°, ψ=−47°) or β-sheet (φ=−119°, ψ=113°) values (a, b, d, and e) and in α-helical segments (c and f). The probabilities are normalized such that $\int P(\chi_1)d\chi_1=1$.

Figure4: $P(\chi_1, \chi_2)$ **for Leu and Ile in dipeptide mimetics and α-helical segments:** Comparison of the observed and calculated probability distributions of side-chain dihedral angles $P(\chi_1, \chi_2)$ for Leu (*Left column*) and Ile (*Right column*). **Row (a):** Observed $P(\chi_1, \chi_2)$ for Leu and Ile in α-helical backbone conformations. **Row (b):** Calculated $P(\chi_1, \chi_2)$ for Leu and Ile dipeptide mimetics in α-helical backbone conformations. **Row (c):** Calculated $P(\chi_1, \chi_2)$ for Leu and Ile in α-helical segments. **Row (d):** Observed $P(\chi_1, \chi_2)$ for Leu and Ile in β-sheet backbone conformations. **Row (e):** Calculated $P(\chi_1, \chi_2)$ for Leu and Ile dipeptide mimetics in β-sheet backbone conformations. The probability distributions are normalized such that $\int P(\chi_1, \chi_2)\, d\chi_1\, d\chi_2=1$. The probability values, expressed as percentages, within each of the nine (χ_1, χ_2) regions defined by the dotted lines are labeled. The probabilities increase from white to yellow to orange to black.

Figure 5: **Stick representation of Ile in an α-helical segment: Top row:** Stick representation of an α-helical segment with Leu at position i and its side-chain dihedral angles set to (*Left*) $\chi_1=180°$ and $\chi_2=60°$ and (*Right*) $\chi_1=300°$ and $\chi_2=60°$. The C^{i-4} atoms are labeled and the atomic clashes between the carbonyl group of residue i-4 and the side-chain of Leu are indicated by a pink dashed circle. **Bottom row:** Stick representation of an α-helical segment with Ile at position I and its side-chain dihedral angles set to (*Left*) $\chi_1=180°$ and $\chi_2=180°$ and (*Right*) $\chi_1=300°$ and $\chi_2=180°$. The C^{i-4} and C^{i-3} atoms are labeled and the atomic clashes between the carbonyl group of residue i-4 and the side chain of Ile are indicated by a pink dashed circle. Carbon, nitrogen, oxygen, and hydrogen atoms are colored pink, blue, red, and white, respectively.

Figure6: $P(\chi_1, \chi_2)$ **for a dipeptide mimetic versus α-helical segment for aromatic residues:** Comparison of the observed and calculated probability distributions of side-chain dihedral angles $P(\chi_1, \chi_2)$ for Phe (*Left column*), Tyr (*Middle column*), and Trp (*Right column*). **Row (a):** Observed $P(\chi_1, \chi_2)$ for Phe, Tyr, and Trp residues in an α-helical backbone conformation. **Row (b):** Calculated $P(\chi_1, \chi_2)$ for Phe, Tyr, and Trp dipeptide mimetics with φ and ψ angles within ±10° of the canonical α-helix values (φ=−57°, ψ=−47°). **Row (c):** Calculated $P(\chi_1, \chi_2)$ for Phe, Tyr, and Trp in α-helical segments. **Row (d):** Observed $P(\chi_1, \chi_2)$ for Phe, Tyr, and Trp residues in a β-sheet backbone conformation. **Row (e):** Calculated $P(\chi_1, \chi_2)$ for Phe, Tyr, and Trp dipeptide mimetics with φ and ψ angles within ±10° of the canonical β-sheet values (φ=−119°, ψ=113°). The probabilities are normalized such that $\int P(\chi_1, \chi_2)\, d\chi_1\, d\chi_2=1$. The probability values, expressed as percentages, are defined by regions between the two vertical dotted lines. The probabilities increase from white to yellow to orange to black.

Figure 7: **Stick representation of Phe within an α-helical segment** Stick representation of Phe at the central position i within a nine-residue α-helical segment and its side-chain dihedral angle χ_1 set to 180° (*Left*) and 300° (*Right*). The C^{i+4} and C^{i-4} atoms are labeled and the atomic clashes that occur between the carbonyl group of residue i-4 the side-chain of Phe when $\chi_1=300°$ are

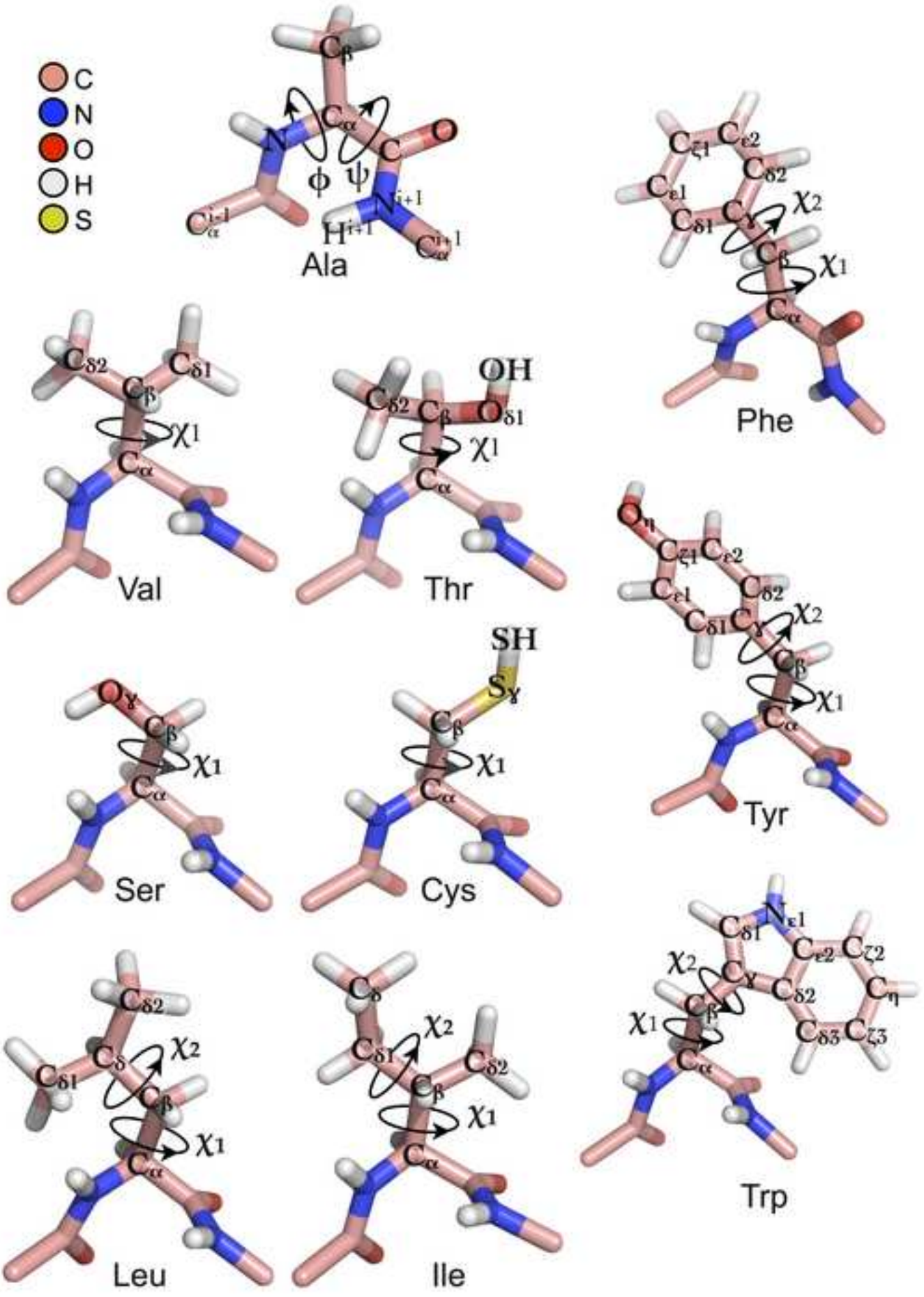
indicated by the pink dashed circle. Carbon, nitrogen, oxygen, and hydrogen atoms are colored pink, blue, red, and white, respectively.

Figure 8: **Stick representation of the Met and Nle dipeptide mimetics** The side-chain dihedral angles χ_1, χ_2, and χ_3 are defined with positive angles indicated by the arrows. Carbon, nitrogen, oxygen, hydrogen, and sulfur atoms are shaded pink, blue, red, white, and yellow, respectively.

Figure 9: **Side-chain dihedral angle distributions for Met and Nle dipeptide mimetics:** Comparison of the observed (*dotted red lines*) and calculated (*solid blue lines*) probability distributions $P(\chi_1)$ (*top*), $P(\chi_2)$ (*middle*), and $P(\chi_3)$ (*bottom*) for the side-chain dihedral angles χ_1, χ_2, and χ_3 for Met (*left*, binned by 5°) and Nle (*right*, binned by 20°) in dipeptide mimetics with α-helix backbone conformations. The probabilities are normalized such that $\int P(\chi_n)d\chi_n=1$, with n=1, 2, or 3.

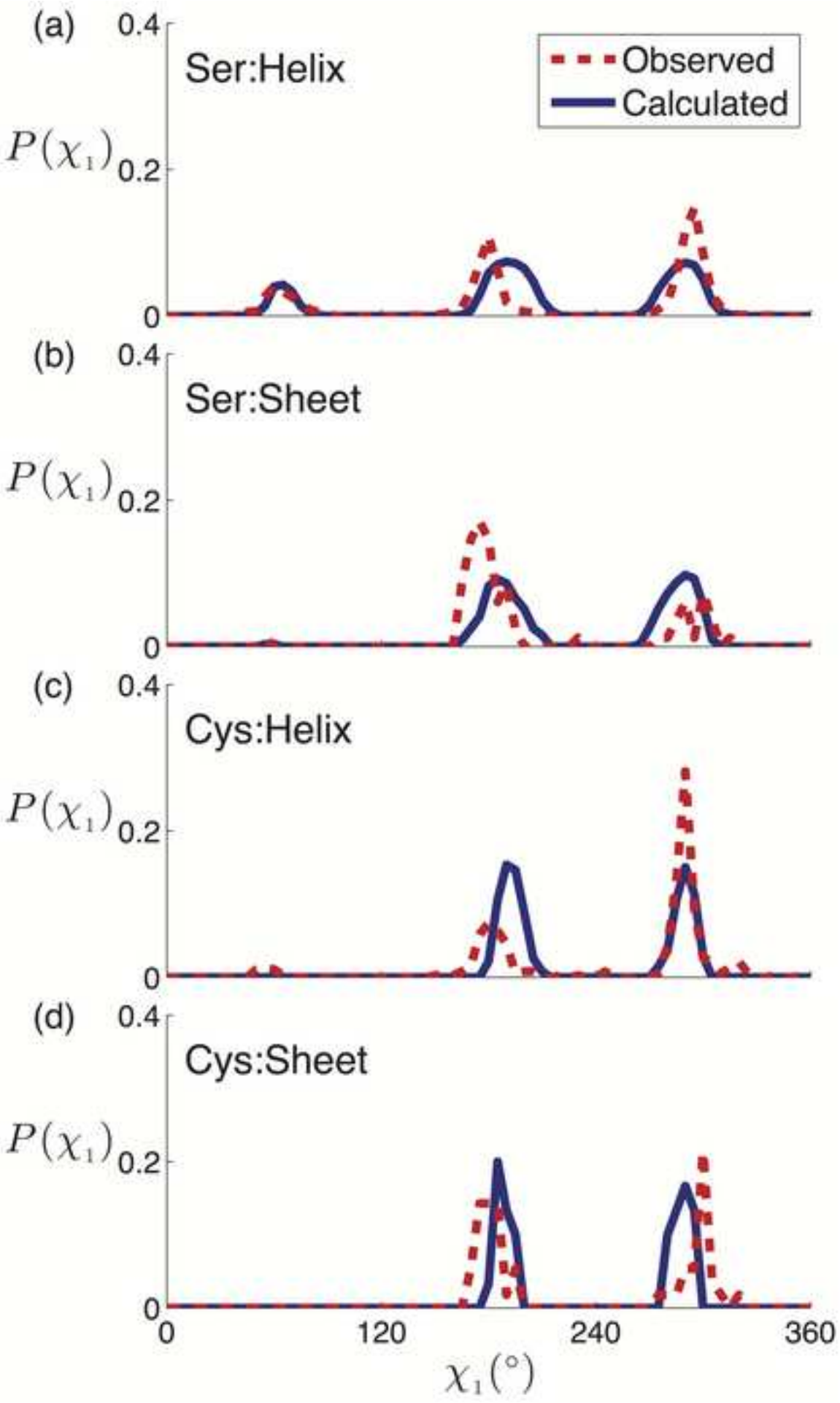
Figure

Click here to download high resolution image



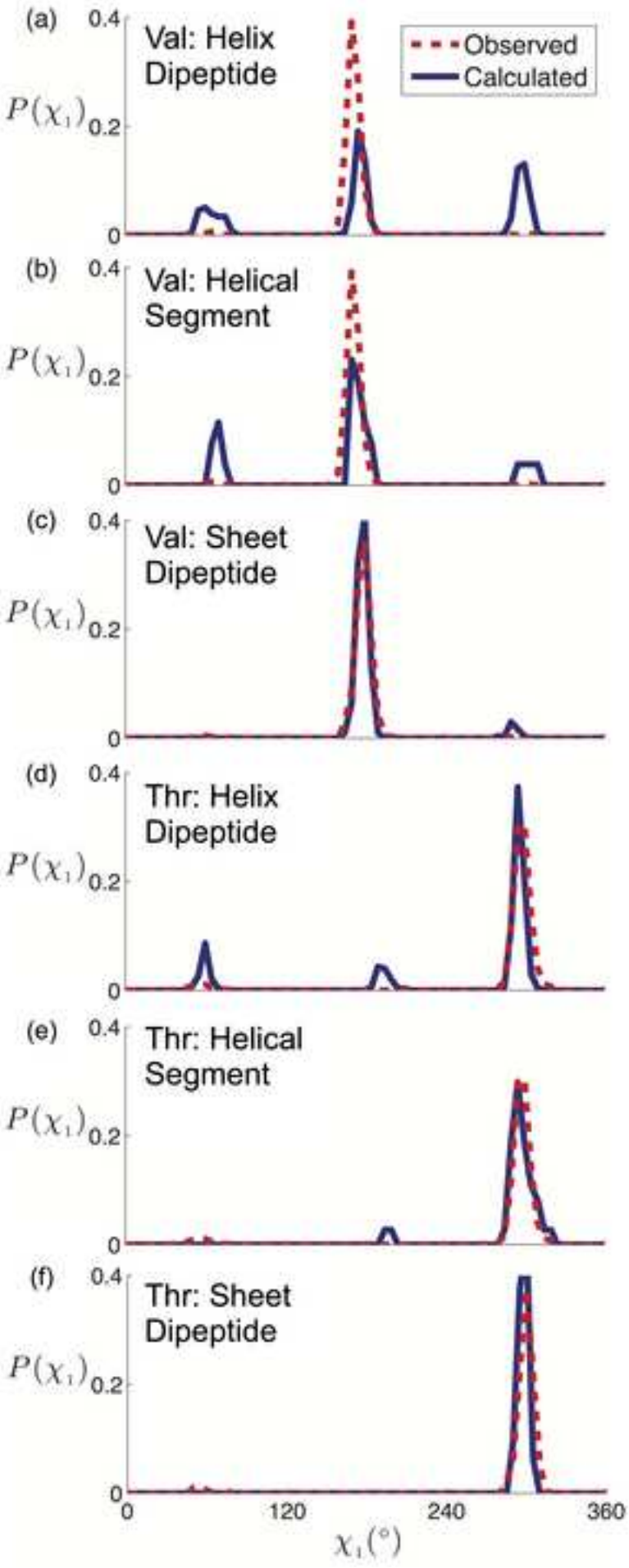
Figure

Click here to download high resolution image



Figure

Click here to download high resolution image



Figure

Click here to download high resolution image

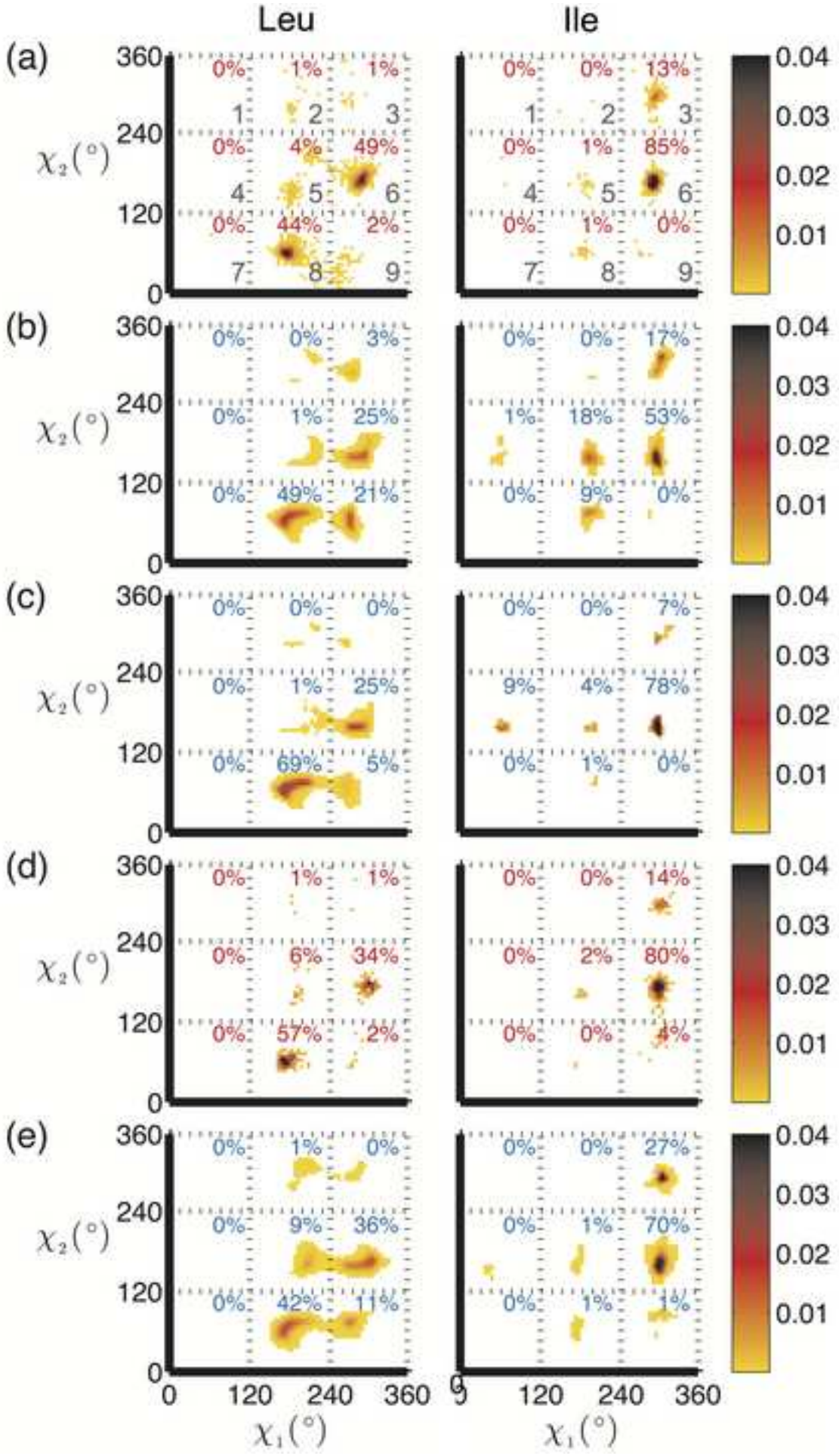
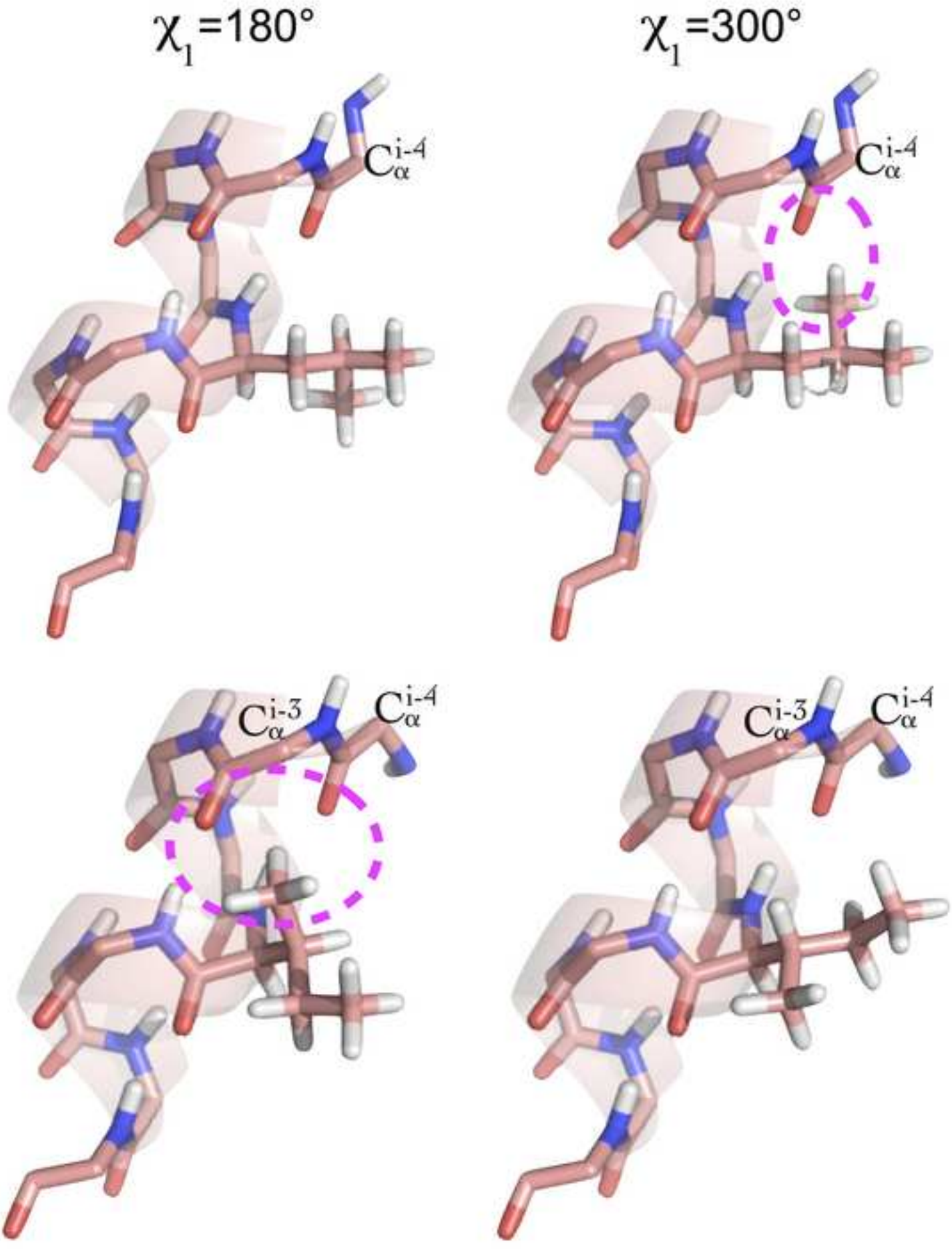
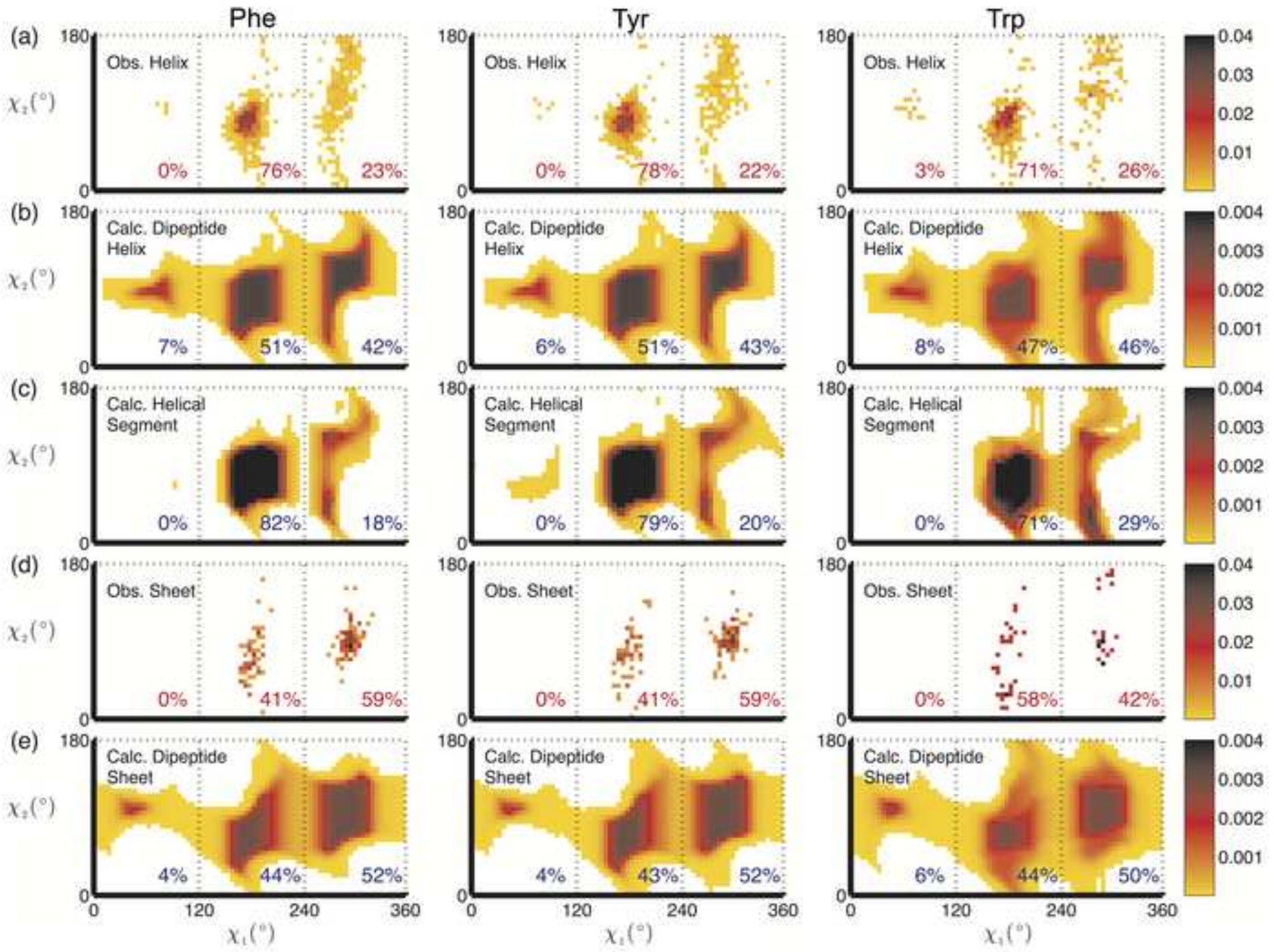


Figure
Click here to download high resolution image



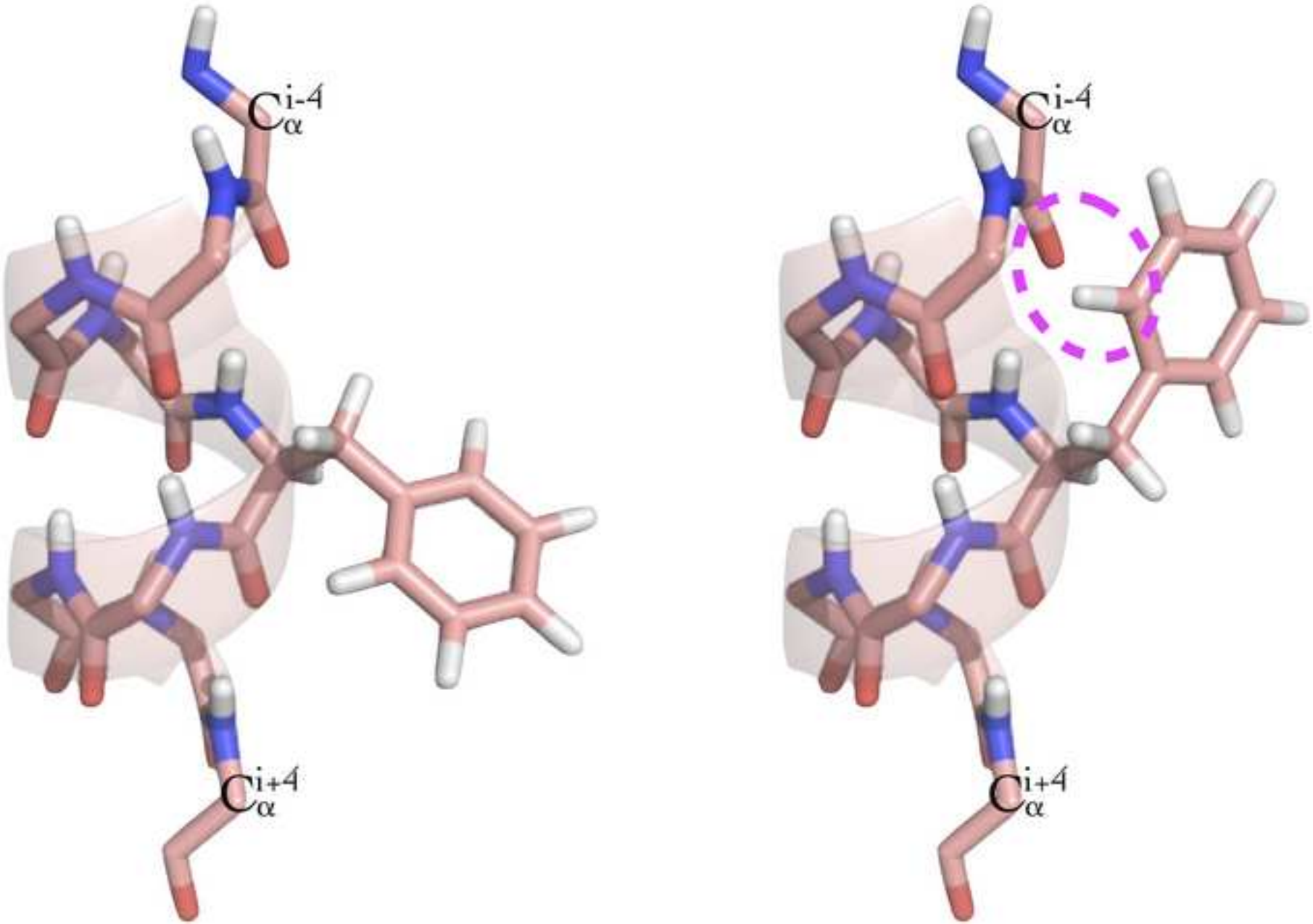
Figure

Click here to download high resolution image



Figure

Click here to download high resolution image



Figure

Click here to download high resolution image

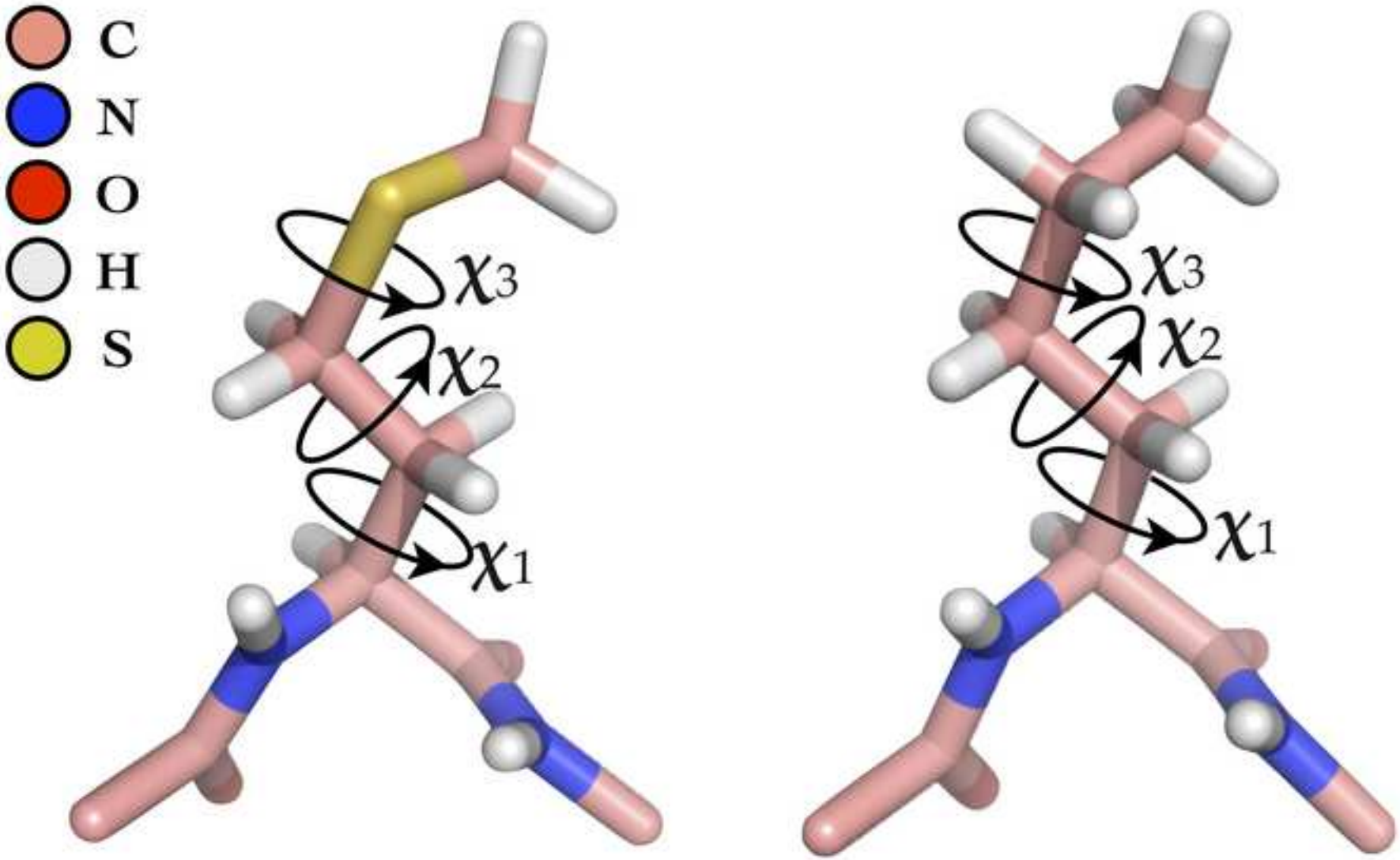
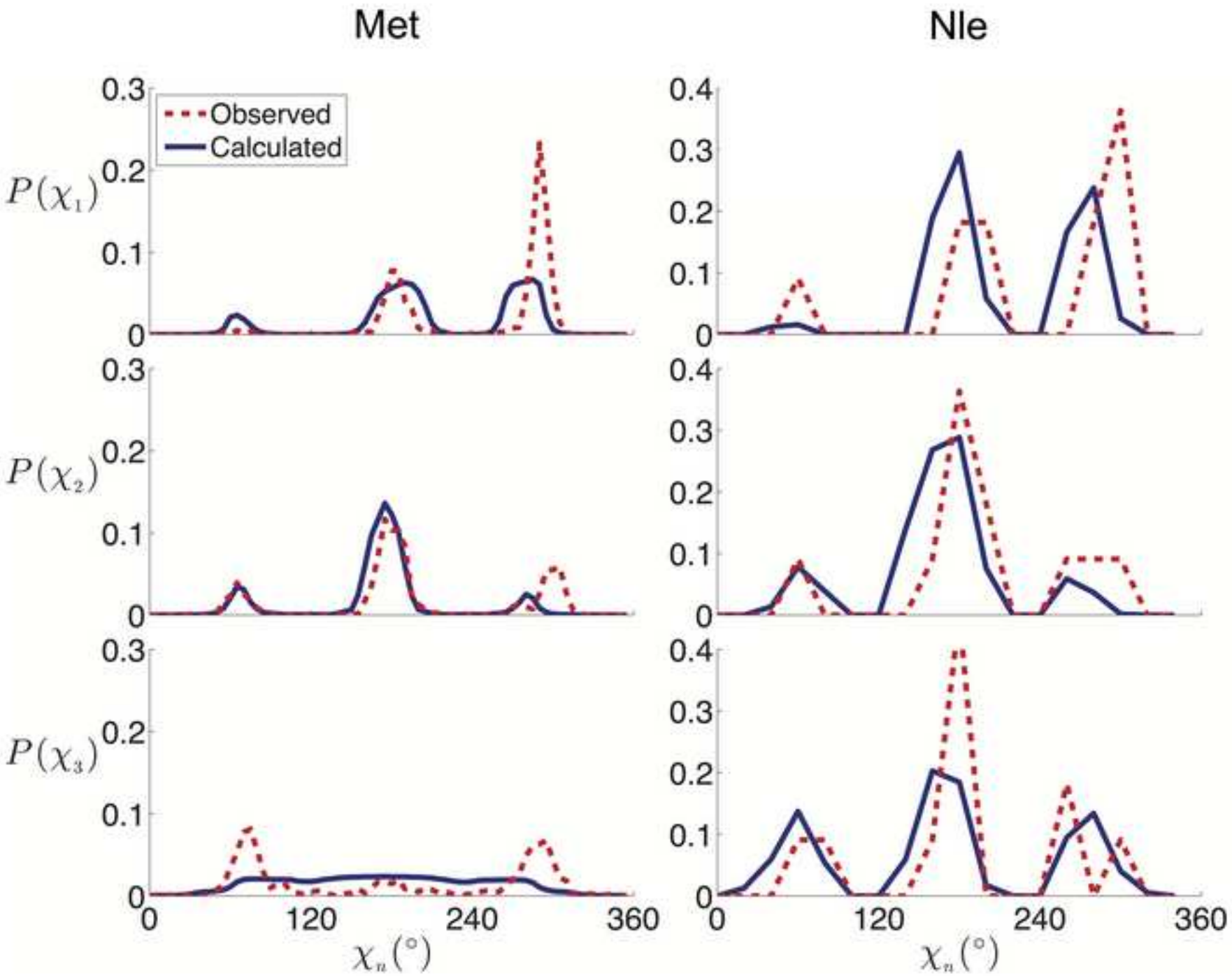


Figure
Click here to download high resolution image



Supplementary Material (To be Published)

Click here to download Supplementary Material (To be Published): Supplementary.docx