Received: 13 May 2022 | Revised: 26 October 2022 | Accepted: 28 October 2022

DOI: 10.1002/pro.4491

TOOLS FOR PROTEIN SCIENCE

A backbone-dependent rotamer library with high (ϕ, ψ) coverage using metadynamics simulations

Jennifer C. Mortensen | **Jovan Damjanovic** | **Jiayuan Miao** | **Tiffani Hui** | **Yu-Shan Lin**

Department of Chemistry, Tufts University, Medford, Massachusetts, USA

Correspondence
Yu-Shan Lin, Department of Chemistry, Tufts University, Medford, MA 02155, USA.
Email: yu-shan.lin@tufts.edu

Funding information
National Institute of General Medical Sciences, Grant/Award Numbers: K12GM133314, R01GM124160

Review Editor: Nir Ben-Tal

Abstract

Backbone-dependent rotamer libraries are commonly used to assign the side chain dihedral angles of amino acids when modeling protein structures. Most rotamer libraries are created by curating protein crystal structure data and using various methods to extrapolate the existing data to cover all possible backbone conformations. However, these rotamer libraries may not be suitable for modeling the structures of cyclic peptides and other constrained peptides because these molecules frequently sample backbone conformations rarely seen in the crystal structures of linear proteins. To provide backbone-dependent side chain information beyond the α-helix, β-sheet, and PPII regions, we used explicit-solvent metadynamics simulations of model dipeptides to create a new rotamer library that has high coverage in the (ϕ, ψ) space. Furthermore, this approach can be applied to build high-coverage rotamer libraries for noncanonical amino acids. The resulting Metadynamics of Dipeptides for Rotamer Distribution (MEDFORD) rotamer library predicts the side chain conformations of high-resolution protein crystal structures with similar accuracy (~80%) to a state-of-the-art rotamer library. Our ability to test the accuracy of MEDFORD at predicting the side chain dihedral angles of amino acids in noncanonical backbone conformation is restricted by the limited structural data available for cyclic peptides. For the cyclic peptide data that are currently available, MEDFORD and the state-of-the-art rotamer library perform comparably. However, the two rotamer libraries indeed make different rotamer predictions in noncanonical (ϕ, ψ) regions. For noncanonical amino acids, the MEDFORD rotamer library predicts the χ_1 values with approximately 75% accuracy.

KEYWORDS
cyclic peptides, metadynamics simulations, protein, rotamer library, side chain rotamers

Jennifer C. Mortensen and Jovan Damjanovic contributed equally to this work.

1 | INTRODUCTION

Since its development in 2010, the Dunbrack backbone-dependent rotamer library[1] has been cited in over 600 papers and is used by common protein modeling software packages including Rosetta[2] and PyMOL.[3] Derived from protein structural data, this rotamer library provides the probability of discrete conformations of side chain dihedral angles (i.e., rotamers) based on the values of the backbone dihedral angles. While the Dunbrack rotamer library is highly used, it has limited data for many regions of the backbone space owing to the natural (ϕ, ψ) preferences of proteins. This limitation may lower its accuracy when predicting the side chain conformations for cyclic peptides, where the ring strain can force the backbone to sample dihedrals rarely seen in linear proteins.[4–6] Furthermore, while noncanonical amino acids are of high interest in peptide and protein therapeutic development,[7,8] little structural information is currently available for noncanonical amino acids, making it difficult to build rotamer libraries for them.[9–13]

To access uncommon regions of backbone space, molecular dynamics simulations can be used.[14–22] For example, the Dynameomics backbone-dependent rotamer library was created by running molecular dynamics simulations of proteins at room temperature.[14–16] The resulting rotamer library contains data across most of backbone space, producing a more thorough picture of the relationship between the backbone and side chain. But, to the best of our knowledge, the Dynameomics rotamer library has not been used in any side chain prediction algorithms and its accuracy in predicting side chain dihedral angles in proteins has not been studied.

In this work, we propose a new approach to develop backbone-dependent rotamer libraries with high backbone coverage for both canonical and noncanonical amino acids using explicit-solvent metadynamics simulations[23,24] of model dipeptides. By biasing the metadynamics simulations along (ϕ, ψ) of the dipeptides, we can obtain high coverage of backbone space with minimal impact on the side chain energetics. As this approach does not rely on experimental protein structural data, it can be applied to noncanonical amino acids as well. For the dipeptide simulations of canonical amino acids as well as norleucine (NLE; Figure S1), we use the RSFF2 force field which was recently parameterized to recapitulate the backbone and side chain preferences of amino acids in a coil library and has been shown to have improved accuracy over other molecular dynamics force fields.[21,22,25] In addition, we simulate dipeptides of five other noncanonical amino acids (Figure S1): α-Aminobutyric acid (ABA), citrulline (CIR), β-(2-naphthyl)-alanine (NAL), *N*,*N*-pyrrolidinylglutamine (PYE),[26] and 3-(1,3-thiazol-4-yl)-alanine (TZA). For these noncanonical amino acids (except NLE), we use the AMBER ff99SB force field[27] in conjunction with the generalized AMBER force field (GAFF) for missing parameters.[28]

The resulting rotamer library, Metadynamics of Dipeptides for Rotamer Distribution (MEDFORD), produces side chain rotamer probabilities similar to the Dunbrack backbone-dependent rotamer library in regions of backbone space that are highly sampled in protein crystal structures. Both the MEDFORD and the Dunbrack rotamer libraries predict the χ_1 values of the canonical amino acids in high-resolution protein structures with approximately 80% accuracy. The Dynameomics rotamer library has low prediction accuracy for some amino acids, possibly due to the force field used to generate the Dynameomics dataset.[14,15] In regions of backbone space where the Dunbrack rotamer library has limited data, the MEDFORD rotamer library provides additional data about the relationship between the backbone and side chain conformations. For noncanonical amino acids, the MEDFORD rotamer library also predicts the χ_1 values of NLE, CIR, and ABA with approximately 75% accuracy. We propose the use of the MEDFORD rotamer library when studying amino acids in regions of backbone space not commonly observed in the crystal structures of linear proteins and when studying noncanonical amino acids.

2 | ROTAMER LIBRARY DEVELOPMENT

2.1 | Metadynamics simulations

To create a rotamer library with high coverage of the backbone (ϕ, ψ) space, we performed bias-exchange metadynamics (BE-META) simulations[23,24] of model dipeptides. Dipeptides of the form Ac-Xaa-NMe were prepared in PyMOL[3] (Figure 1a). Xaa here was Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Trp, Tyr, Val, ABA, CIR, NAL, NLE, PYE, or TZA. To verify simulation convergence, for each amino acid, two initial structures, S1 and S2, were prepared, one with the backbone in the α-helical region and one in the β-sheet region. Each structure was energy minimized and then solvated with water. Minimal Na^+ or Cl^- ions were added to neutralize charges on Arg, Asp, Glu, and Lys. Each system was then further equilibrated before starting production runs. An independent simulation was performed for each initial structure, S1 and S2. Simulations for dipeptides of the canonical amino acids and NLE were performed with the RSFF2 force field with TIP3P water in the GROMACS software suite,[21,29,30] while

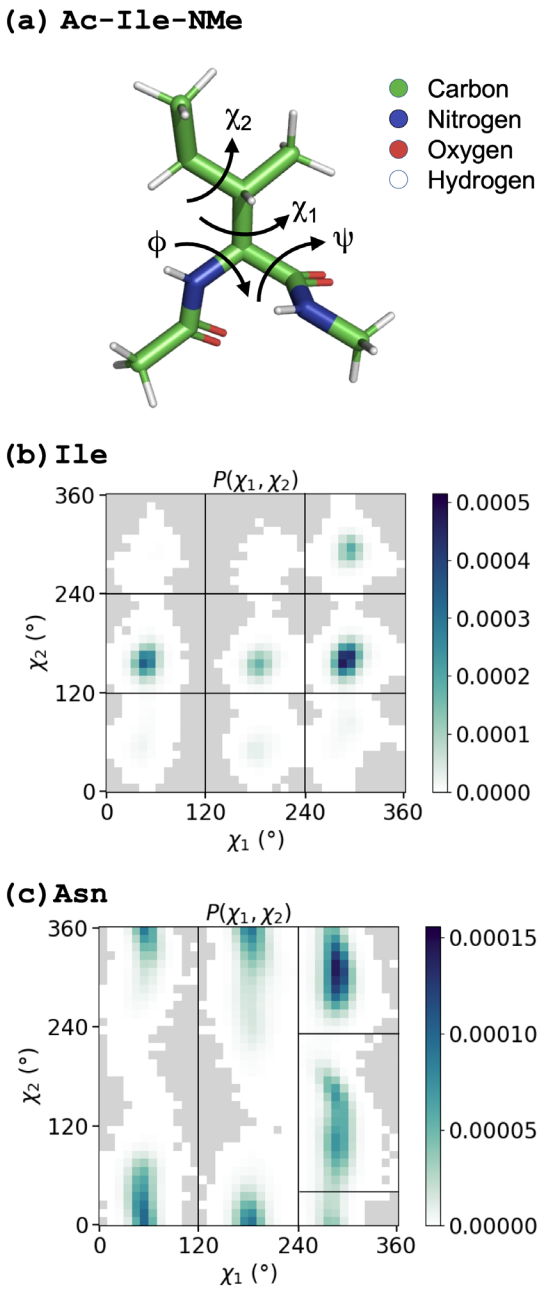

FIGURE 1 (a) Stick representations of Ac-Ile-NMe. The dihedral angles ϕ, ψ, χ_1, and χ_2 are indicated. (b) Side chain dihedral distributions (χ_1, χ_2) for Ac-Ile-NMe. (c) Side chain dihedral distributions (χ_1, χ_2) for Ac-Asn-NMe. Black lines indicate the edge of each rotamer definition.

simulations for dipeptides of the remaining noncanonical amino acids were performed with the AMBER ff99SB force field[27] in conjunction with GAFF for missing parameters.[28] More details on the simulation set up can be found in the Supporting Information.

BE-META runs with the equilibrated structures were performed at 300 K and 1 bar for 200 ns with 2 fs time steps using the PLUMED 2 plugin for the GROMACS suite.[31] Data were recorded every 500 steps (i.e., 1 ps). For each system, one biased and five neutral replicas were used. In the biased replica, a two-dimensional (ϕ, ψ) bias was applied to enable efficient sampling of the backbone dihedral space without modifying the side chain energy landscape. Gaussian hills of 0.1 kJ/mol with a width of 0.314 rad were added every 4 ps. We also aimed to obtain the equilibrium structural ensemble and verify that the resulting Ramachandran plot of each dipeptide system was consistent with that previously reported.[21,25] In BE-META simulations, the equilibrium ensemble can be obtained by adding a neutral replica with no bias.[24] To improve the statistics, multiple neutral replicas can be used.[32,33] In this study, we used five neutral replicas. Exchanges between replicas were attempted every 5 ps.

The backbone and side chain dihedral angles were calculated for every data point in the last 150 ns of each BE-META run. To test for simulation convergence, we calculated the equilibrium (ϕ, ψ) probability distributions of the neutral replicas of the BE-META runs initiated from S1 and S2, respectively, and computed the normalized integrated product (NIP) between the two distributions.[34] The NIP values between S1 and S2 for all the dipeptide systems were greater than 0.99, indicating high similarity between the two distributions and simulation convergence. The data from the five neutral and one biased replicas of the BE-META runs of S1 and S2 were then combined to create one large dataset for each dipeptide, resulting in 1.8×10^6 data points from a total of 1,800 ns of simulation (Figure 2b).

2.2 | Creation of the MEDFORD rotamer library

To create the MEDFORD rotamer library, we analyzed the data produced by our BE-META simulations of each dipeptide system. The backbone-dependent probability of each side chain rotamer combination, $P(\chi_{all} \mid \phi, \psi)$, was calculated by dividing the number of data points with the specific rotamer combination, χ_{all}, in the backbone (ϕ, ψ) bin by the total number of data points in the backbone (ϕ, ψ) bin. The average χ value of each dihedral angle for each rotamer combination was also calculated. For rotameric side chain dihedrals, three rotamers, r60, r180, and r300, for each dihedral angle were defined as (0°, 120°], (120°, 240°], and (240°, 360°], respectively (Figure 1b). Nonrotameric side chain dihedral angles were split into rotamers based on the minima of the dihedral angle distribution (Figure 1c; Figures S2 and S3). A detailed description of rotameric and nonrotameric side chain dihedral angles can be found in the Supporting Information. The total number of rotamer states for each amino acid is summarized in Table S1. To study the relationship

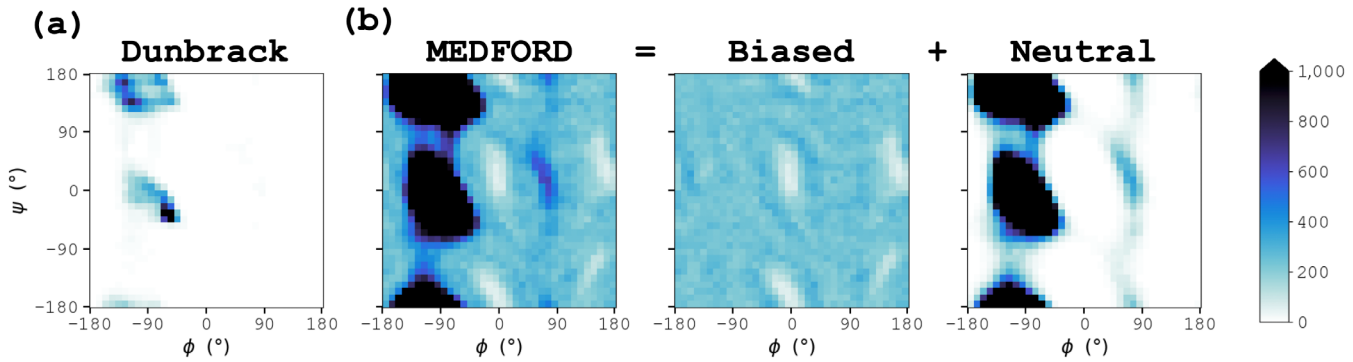

FIGURE 2 Number of data points in each of the 10° × 10° (ϕ, ψ) bins in the (a) Dunbrack and (b) MEDFORD rotamer libraries for Thr. The data used for the MEDFORD rotamer library come from both the biased and neutral replicas of BE-META simulations of Thr dipeptide. The biased replicas show fairly even coverage throughout the whole (ϕ, ψ) space, while the neutral replicas sample mostly the α-helix, β-sheet, and PPII regions.

between the backbone and χ_1, we also computed the backbone-dependent χ_1 probabilities, $P(\chi_1 \mid \phi, \psi)$, without considering the dihedral angle values of the rest of the side chain. (See Section 5 for additional details.)

The MEDFORD rotamer library is formatted to match the Dunbrack rotamer library and can be downloaded from https://github.com/jgaines42/MEDFORD-rotamer-library.

2.3 | Comparison to existing rotamer libraries

For the canonical amino acids, we compared the MEDFORD rotamer library to two existing rotamer libraries, Dunbrack[1] and Dynameomics.[15] The Dunbrack 2010 backbone-dependent rotamer library (hereafter referred to as the Dunbrack rotamer library) was generated by analyzing protein crystal structures and is considered a state-of-the-art rotamer library. The Dynameomics backbone-dependent rotamer library (hereafter referred to as the Dynameomics rotamer library) was created from molecular dynamics simulations of proteins. In each rotamer library, the probability of each side chain rotamer is reported along with the side chain dihedral angle values (see Section 5). Because proteins do not sample the full (ϕ, ψ) space, we investigated the backbone coverage of the raw data used to generate the Dunbrack rotamer library. We binned the raw amino acid data used to create the Dunbrack rotamer library into 10° × 10° (ϕ, ψ) bins centered on the (ϕ, ψ) values used to denote the bin. This information was used to define the "conventional low-coverage (ϕ, ψ) regions" for each amino acid as backbone regions with <25 data points in the (ϕ, ψ) bin and the "conventional high-coverage (ϕ, ψ) regions" as those with ≥25 data points in the (ϕ, ψ) bin.

To compare the $P(\chi_{\text{all}} \mid \phi, \psi)$ distributions between the Dunbrack, Dynameomics, and MEDFORD rotamer libraries, we needed to treat the nonrotameric side chains on the same footing. To do so, we combined the binned data of the nonrotameric side chains in each previously published rotamer library into rotamers that matched the dihedral cutoffs of the MEDFORD rotamer library (Tables S2–S5). We then calculated the NIP[34] values between the $P(\chi_{\text{all}} \mid \phi, \psi)$ distributions from the three rotamer libraries in the conventional high- and low-coverage regions of the backbone space, respectively.

For the noncanonical amino acids, we compared the MEDFORD rotamer library to other available rotamer libraries. As the NLE side chain is very similar to that of Met, the rotamer library for Met is often used for NLE. Therefore, we compared the performance of MEDFORD on NLE to that of the Dunbrack rotamer library for Met. In addition, Renfrew et al. developed a protocol, called MakeRotLib, which can create backbone-dependent rotamer libraries for noncanonical amino acids.[9] Renfrew et al. have reported a backbone-dependent rotamer library for ABA, and we constructed a rotamer library using MakeRotLib for NLE and CIR to compare the performance of the rotamer libraries constructed using MakeRotLib to that of MEDFORD for these three noncanonical amino acids.

3 | RESULTS AND DISCUSSION

3.1 | Metadynamics simulations increase coverage of backbone space

The protein data used to create the Dunbrack rotamer library have low coverage in most regions of the (ϕ, ψ) space. For example, in Figure 2a, we see that the (ϕ, ψ) of Thr is frequently found in the α-helix, β-sheet, and PPII

regions but rarely found in other regions of the backbone space. In fact, 88% of the (ϕ, ψ) bins for Thr contain fewer than 25 data points (see Table S6 for statistics for the other amino acids). Looking at all amino acids in the Dunbrack dataset, we find that most of the residues belong to the high-coverage regions, with only 4% of residues found in the low-coverage regions (Figure 3a, blue bar). Looking instead at cyclic peptides, we found that the noncanonical (ϕ, ψ) regions are more frequently sampled, with 20% of cyclic peptide residues found in the low-coverage regions (Figure 3a, red bar, and Figure 3b, red × 's).

Including the biased replicas from the BE-META simulation of Ac-Xaa-NMe dipeptides in our analysis allowed us to obtain high coverage of the full (ϕ, ψ) space. The added Gaussian biasing potentials removed the energy barriers between backbone conformations, allowing efficient sampling. As an example, in Figure 2b, the BE-META simulations of Thr dipeptide sample the full (ϕ, ψ) space with more than 25 data points in every (ϕ, ψ) bin and more than 100 data points in 97% of the (ϕ, ψ) bins (Table S6). Therefore, BE-META simulations allowed us to create a new rotamer library with improved statistics for (ϕ, ψ) regions that are rarely sampled by linear proteins.

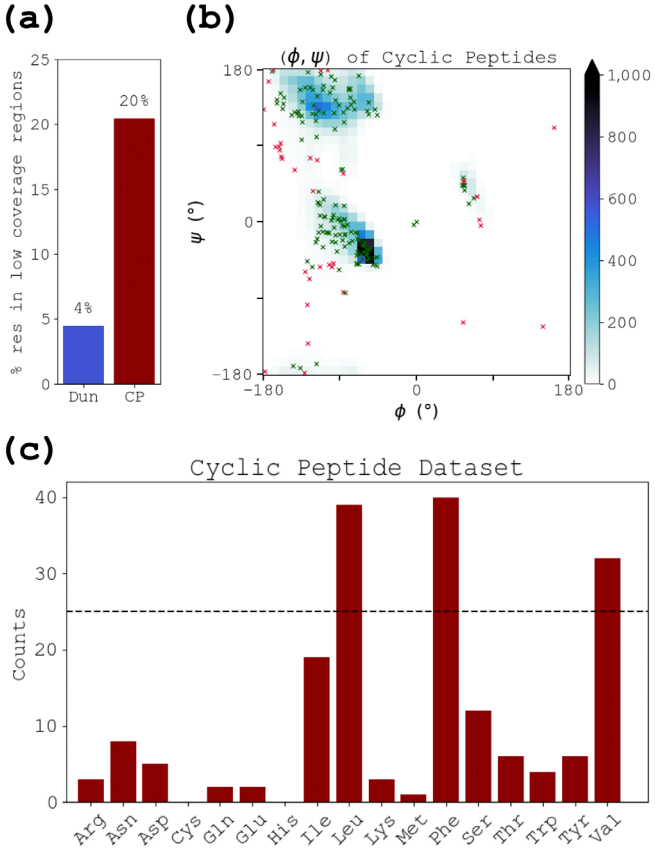

FIGURE 3 (a) Percentage of residues in (ϕ, ψ) bins with "low coverage" in the Dunbrack rotamer library (Dun), and in a dataset of cyclic peptides (CP). Low coverage is defined as a 10° × 10° (ϕ, ψ) bin containing <25 data points in the dataset used to create the Dunbrack rotamer library. (b) Location of the 181 (ϕ, ψ) values observed in the cyclic peptide dataset overlaid on the average frequency of backbone conformations from the Dunbrack rotamer library. Points in the conventional "high-coverage" (ϕ, ψ) bins are marked in green, "low-coverage" in red. (c) Frequency of each amino acid in the cyclic peptide dataset

3.2 | MEDFORD contains different rotamer probability distributions than existing rotamer libraries, particularly in the protein low-coverage regions

We next sought to compare the side chain dihedral distributions in the Dunbrack, Dynameomics, and MEDFORD rotamer libraries. Visual inspection of the $P(\chi_1 \mid \phi, \psi)$ distributions for various amino acids between the three rotamer libraries reveals differences in the probability distributions. As an example, Figure 4a–c show the backbone-dependent probability of the three χ_1 rotamers of Thr, and Figure 4d shows the probability of the most likely rotamer in each (ϕ, ψ) bin. The probability distribution from Dynameomics is quite different from the other two rotamer libraries. The Dunbrack and MEDFORD rotamer libraries also differ when $\phi > 0°$, corresponding to regions where the Dunbrack rotamer library has low coverage (Figure 2a).

We quantified the similarity between the three rotamer libraries by calculating the NIP[34] between their side chain probability distributions, $P(\chi_{all} \mid \phi, \psi)$. Figure 5 shows that the three rotamer libraries have rather similar $P(\chi_{all} \mid \phi, \psi)$ distributions for Leu, Phe, and Tyr, especially for regions with high coverage (NIP > 0.9). When comparing the Dunbrack and MEDFORD rotamer libraries, for regions with high coverage in the Dunbrack dataset, the two rotamer libraries have highly similar $P(\chi_{all} \mid \phi, \psi)$ distributions with NIP generally >0.9 (Figure 5b, left column). This result confirms that our simulation methods create side chain distributions that are consistent with those observed in protein crystal structures. A noticeable exception is Trp, where the NIP is only 0.80 between the MEDFORD and the Dunbrack rotamer libraries in the high-coverage region. The strong similarity between the MEDFORD and Dunbrack rotamer libraries in the high-coverage areas is consistent with previous studies that have shown that the full protein environment is not needed to produce accurate χ_1 side chain rotamer distributions.[35–40] Instead, the conformation of χ_1 is based on short-range interactions with the backbone and the backbone-dependent χ_1 rotamer preferences are fully present at the peptide level.[41]

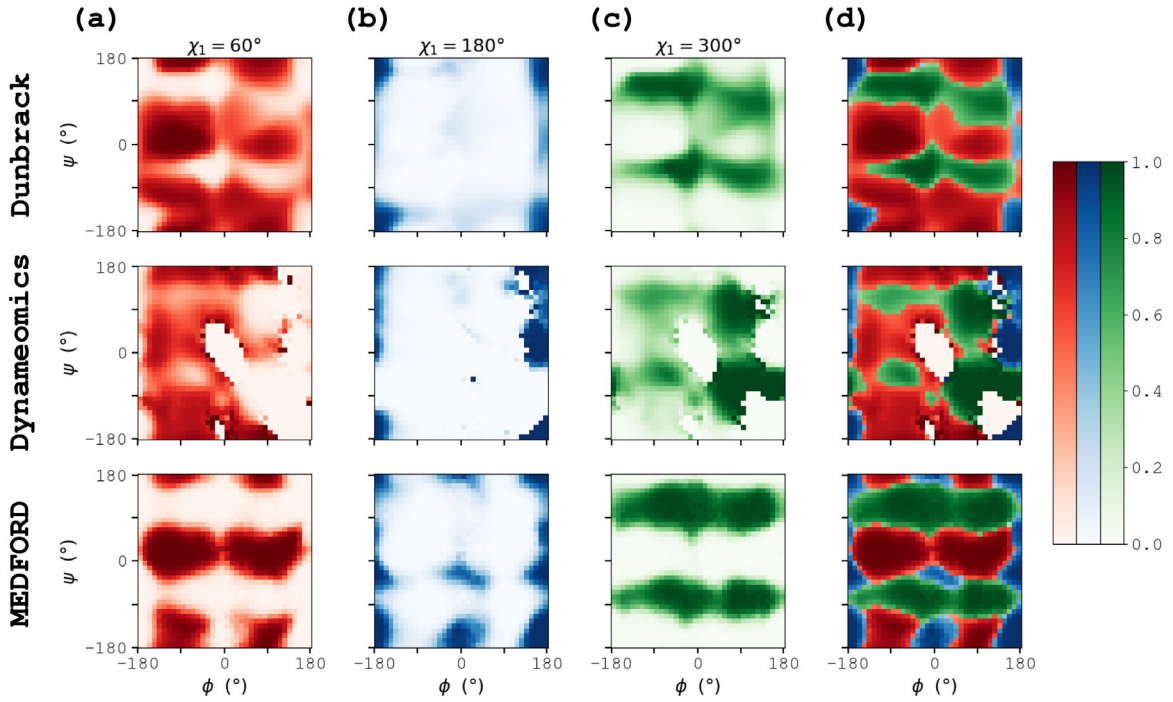

FIGURE 4 (a–c) Backbone-dependent probability of each χ_1 rotamer for Thr and (d) the most probable χ_1 rotamer colored by probability for $\chi_1 = 60°$ (red), $\chi_1 = 180°$ (blue), and $\chi_1 = 300°$ (green) in the Dunbrack (top), Dynameomics (middle), and MEDFORD (bottom) rotamer libraries

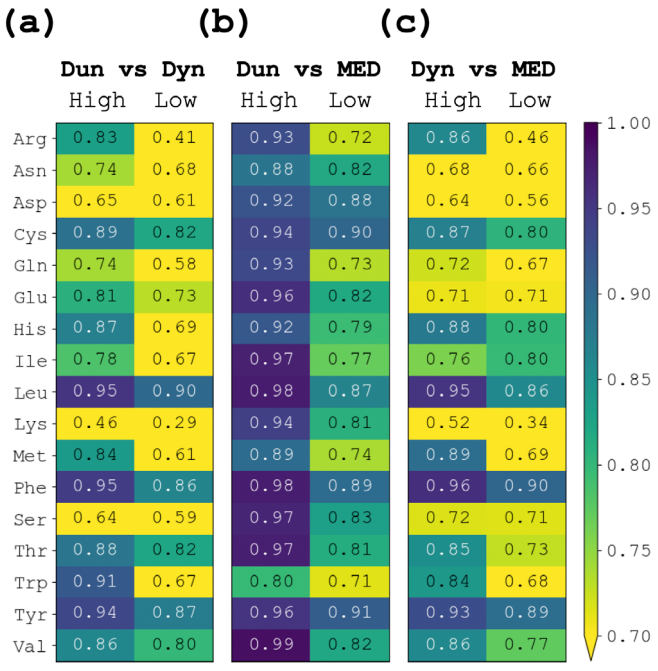

FIGURE 5 Similarity of $P(\chi_{all} \mid \phi, \psi)$ distributions between (a) the Dunbrack and Dynameomics rotamer libraries, (b) the Dunbrack and MEDFORD rotamer libraries, and (c) the Dynameomics and MEDFORD rotamer libraries, for regions of (ϕ, ψ) with high coverage in Dunbrack (left columns) and regions with low coverage (right columns)

Looking instead at low-coverage regions, the $P(\chi_{all} \mid \phi, \psi)$ distributions of the MEDFORD and Dunbrack rotamer libraries are less similar, with NIP values as low as 0.71 (Figure 5b, right column). Lower NIP values indicate larger differences between the Dunbrack and MEDFORD rotamer libraries. These differences could be due to inaccuracy in the extrapolation method used to generate rotamer probabilities for low-coverage regions for the Dunbrack rotamer library or inaccuracy in the RSFF2 force field used to create the MEDFORD rotamer library.

For most amino acids, comparison of the Dynameomics rotamer library to either the Dunbrack (Figure 5a) or the MEDFORD rotamer library (Figure 5c) show different $P(\chi_{all} \mid \phi, \psi)$ distributions with NIP values generally <0.9 in the high-coverage regions (left column), and even lower values for regions with low coverage (right column). The significant differences between the Dunbrack and Dynameomics rotamer libraries suggest that the simulations used to create the Dynameomics rotamer library did not reproduce the side chain dihedral distributions found in natural protein structures. Dynameomics was developed using the ENCAD force field[14,42–44] which may have resulted in atypical interactions between the backbone and side chains. Alternatively, the Dynameomics simulations were run in water, intending to have better predictions for proteins in aqueous solutions.[14] It is possible that some amino acids—for example, Asp, Lys, and Ser, which have very low NIP between the Dunbrack and Dynameomics rotamer libraries—have particularly pronounced differences between the side chain preferences in a protein in water and in crystal structures.

3.3 | $P(\chi_1 \mid \phi, \psi)$ depends strongly on local interactions between the side chain and backbone

Further analysis of the MEDFORD rotamer library revealed similarities between $P(\chi_1 \mid \phi, \psi)$ distributions of different amino acids. In Figure 6, we show the NIP values between the $P(\chi_1 \mid \phi, \psi)$ distributions of all pairs of amino acids, clustered using a nearest-neighbor search. The amino acids form six clusters: (1) Leu, His, Trp, Tyr, Phe, Gln, Lys, Met, Arg, and Glu, (2) Cys and Asn, (3) Ser, (4) Asp, (5) Val and Ile, and (6) Thr. These results are similar to those previously reported by Jiang et al. by analyzing the χ_1, ϕ, and ψ distributions in a coil library[45] and highlight the influence of the polarity and branching of the side chain close to the backbone on the $P(\chi_1 \mid \phi, \psi)$ distribution (Figure S4). By grouping amino acids by their $P(\chi_1 \mid \phi, \psi)$ distributions, we can assess the side chain prediction accuracy of each rotamer library based on the type of side chain present.

3.4 | The MEDFORD rotamer library has comparable prediction accuracy to the Dunbrack rotamer library

We assessed the ability of the Dunbrack rotamer library, the Dynameomics rotamer library, and our new MEDFORD rotamer library to predict experimental side chain dihedral angles using two methods ("Top ranked rotamer" and "Rosetta repacking"). In the first method, we compared the top ranked rotamer in each rotamer library to the experimental data without the use of a repacking algorithm; in the second method, we implemented the rotamer libraries in Rosetta[46] and evaluated the accuracy of the rotamer libraries after repacking using Rosetta (see Sections 5.3–5.4).

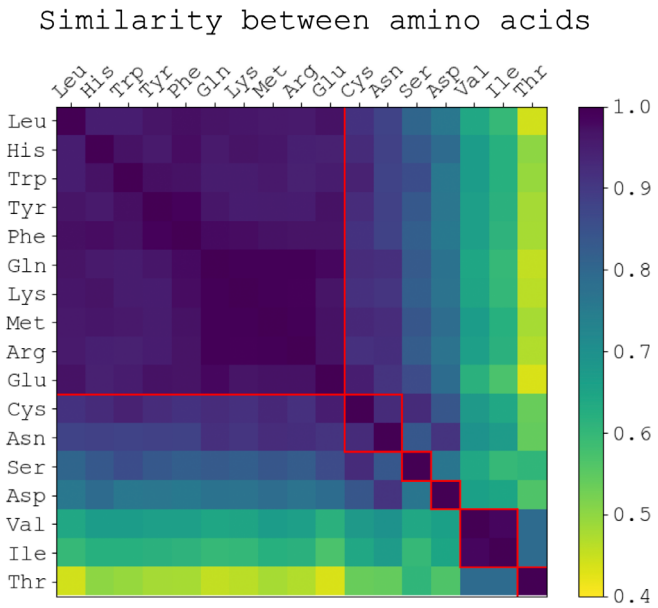

FIGURE 6 Similarity between $P(\chi_1 \mid \phi, \psi)$ distributions from the MEDFORD rotamer library. For this analysis, the rotamer definition of χ_1 in Val was swapped to match the assignments of χ_1 in Ile.

3.4.1 | HQ54

Figure 7 shows that both the Dunbrack and MEDFORD rotamer libraries have high χ_1 prediction accuracy across all amino acids in the HQ54 protein dataset. When we

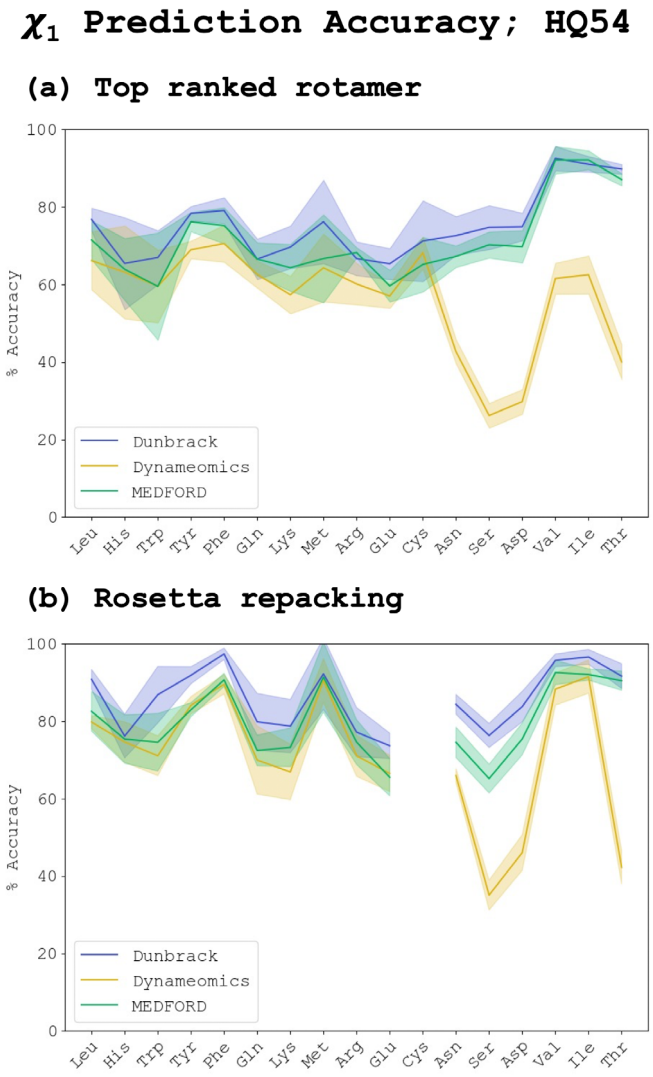

FIGURE 7 Average χ_1 prediction accuracy for amino acids in the HQ54 dataset of protein structures using each rotamer library (a) by selecting the rotamer with the highest probability in the rotamer libraries or (b) after Rosetta repacking. A prediction is considered accurate when χ_1 is predicted within 20° of the crystal structure value. Shaded regions show the standard deviation found using five-fold cross validation.

select the highest-probability rotamer as the prediction, the Dunbrack and MEDFORD rotamer libraries predict the χ_1 dihedral angles for different types of amino acids in the HQ54 dataset with an average accuracy of 75% and 73%, respectively (Figure 7a). After repacking in Rosetta, the accuracies further increase (Figure 7b). Surprisingly, the Dynameomics rotamer library performs poorly for Asn, Asp, Ser, Thr, Ile, and Val, that is, amino acids with polar C_γ or C_δ atoms and β-branched amino acids when selecting the top ranked rotamers as the predictions (Figure 7a, yellow curve); after Rosetta repacking, the accuracies for Asn, Val, and Ile greatly improve, while the accuracies for Ser, Asp, and Thr remain relatively low.

Predictions for the combined side chain dihedral angles for amino acids with χ_2, χ_3, and χ_4 have lower accuracy for all three rotamer libraries (Figure S5a) due to the weaker relationship between these dihedrals and the backbone. This issue is often dealt with by using an additional energy function that incorporates both the backbone and surrounding protein environment to predict the full side chain structure of long amino acids.[1,47,48] Indeed, after repacking using Rosetta, we see that both accuracies for predicting χ_1 (Figure 7b) and all χ (Figure S5b) generally increase. However, it remains difficult to predict all the χ angles for long side chains, such as those of Gln, Lys, Arg, and Glu. It should be noted that the predictions are compared against X-ray structures. It is possible that in solution, these side chains might sample a different conformation or multiple conformations.

3.4.2 | Cyclic peptide dataset

Next, we assessed the ability of the Dunbrack, Dynameomics, and MEDFORD rotamer libraries to predict the side chain dihedral angles of cyclic peptides. Unfortunately, there are currently very few data points in the cyclic peptide dataset (Figure 3c). In Figure S6, we show the prediction accuracy for those amino acids that have 25 or more instances in the cyclic peptide dataset, that is, Leu, Phe, and Val. The MEDFORD rotamer library has similar performance to the Dunbrack rotamer library for these three amino acids. As seen with the prediction for the HQ54 residues, the Dynameomics rotamer library in general performs worse than the other two rotamer libraries.

3.4.3 | High- versus low-coverage regions

We expected that sufficient sampling of the full backbone space in the MEDFORD rotamer library would improve prediction accuracy of amino acids in the regions that have low coverage in the Dunbrack rotamer library. However, the MEDFORD rotamer library performs very similarly to the Dunbrack rotamer library in both the conventional high- and low-coverage backbone regions for the HQ54 and cyclic peptide datasets (Figure 8). To further investigate this unexpected result, we identified the (ϕ, ψ) bins where the two rotamer libraries make different χ_1 rotamer predictions (dotted areas in Figure 9 and Figure S7), and overlaid them with the (ϕ, ψ) free energy landscape calculated from the neutral replicas from the BE-META simulations for each dipeptide (Figure 9 and Figure S7). The regions of backbone space with different χ_1 rotamer predictions span a wide range of the free energy landscape. Amino acids in the low-coverage regions in the HQ54 and cyclic peptide datasets (circles and triangles in Figure 9 and Figure S7) are typically found near the edges of the canonical regions, in areas with the same rotamer prediction from the Dunbrack and MEDFORD rotamer libraries. The similarity between the predictions made by the Dunbrack and MEDFORD rotamer libraries near the canonical regions suggests that the extrapolation method used to create the Dunbrack rotamer library was sufficient to produce accurate rotamer predictions in these regions.

It is difficult to evaluate and distinguish the performances of the Dunbrack and MEDFORD rotamer libraries in the conventional low-coverage regions due to the

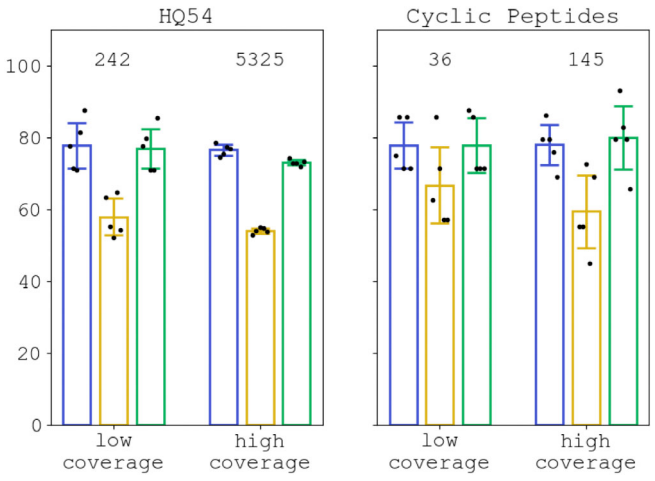

FIGURE 8 Prediction accuracy of χ_1 for amino acids in the HQ54 dataset (left) and cyclic peptide dataset (right) by the Dunbrack (blue), Dynameomics (yellow), and MEDFORD (green) rotamer libraries split by low-coverage and high-coverage regions of (ϕ, ψ) space by selecting the rotamer with the highest probability in the rotamer libraries. A prediction is accurate when each dihedral angle is predicted within 20° of the crystal structure value. Coverage is defined based on the Dunbrack dataset. Error bars show the standard deviation from five-fold cross validation.

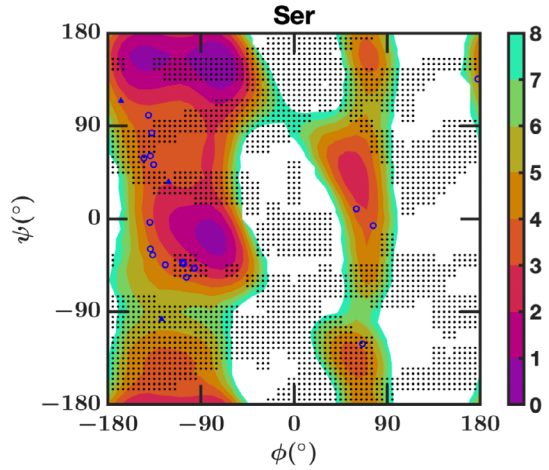

FIGURE 9 The (ϕ, ψ) locations of Ser in the low-coverage backbone regions from HQ54 (blue circles) and the cyclic peptide dataset (blue triangles) overlaid with the (ϕ, ψ) free energy profile calculated using the neutral replicas from the Ser dipeptide BE-META simulations (shown in k_BT). The black dots mark the (ϕ, ψ) bins for which the Dunbrack and MEDFORD rotamer libraries make different χ_1 rotamer predictions.

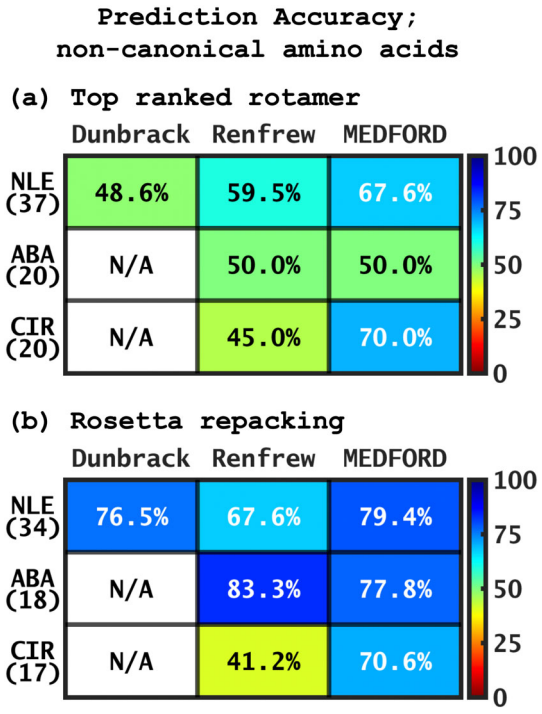

FIGURE 10 Prediction accuracy of χ_1 for amino acids in the noncanonical amino acid dataset (a) by selecting the rotamer with the highest probability in the rotamer libraries or (b) after Rosetta repacking. A prediction is accurate when χ_1 is predicted within 20° of the crystal structure value.

limited protein and cyclic peptide data available in these regions (note the large error bars for the low-coverage datasets in Figure 8). Backbone conformations with different predictions from Dunbrack and MEDFORD are often found in the noncanonical backbone regions, which tend to have higher free energy. However, some of these (ϕ, ψ) conformations have free energy values that are accessible by proteins and cyclic peptides (Figure 9 and Figure S7). A larger library of cyclic peptides or proteins with backbone conformations in noncanonical regions would allow for a stronger comparison of the Dunbrack and MEDFORD rotamer libraries. The MEDFORD rotamer library has substantially higher sampling than the Dunbrack rotamer library in the noncanonical (Φ, ψ) regions, and the two rotamer libraries behave less similarly in these regions. Although we expect the MEDFORD rotamer library may perform better than the Dunbrack rotamer library in these regions, unfortunately, there are currently insufficient data to draw such conclusions on the prediction accuracy, and the two rotamer libraries have rather similar performance.

3.5 | The MEDFORD rotamer library predicts the side chains of noncanonical amino acids with approximately 75% accuracy

Using explicit-solvent BE-META simulations, we constructed rotamer libraries for six noncanonical amino acids, ABA, CIR, NAL, NLE, PYE,[26] and TZA (Figure S1). We were able to find 20, 20, 5, 37, 3, and 1 examples of ABA, CIR, NAL, NLE, PYE, and TZA, respectively, in the PDB. We tested the performance of the MEDFORD rotamer library on the amino acids with more than 15 instances in the PDB: NLE (37), ABA (20), and CIR (20). We found that the MEDFORD rotamer library predicts the side chains of NLE, ABA, and CIR with approximately 60% accuracy when using the top ranked rotamers and 75% accuracy after repacking in Rosetta (Figure 10). The MEDFORD rotamer library also shows comparable or better performance compared to the previously reported rotamer libraries for the three noncanonical amino acids. However, it should be noted the amount of data made it difficult to establish the uncertainties of these results.

4 | CONCLUSIONS

Using data from explicit-solvent metadynamics simulations of dipeptides, we created a backbone-dependent rotamer library, MEDFORD, that has extensive data in all (ϕ, ψ) regions. We found that the MEDFORD rotamer library has similar $P(\chi_{all} \mid \phi, \psi)$ distributions to the Dunbrack backbone-dependent rotamer library for (ϕ, ψ) regions where the Dunbrack dataset has high coverage. It

also predicts side chain conformations for proteins with similar accuracy. However, the probability distributions from the two rotamer libraries differ in many (ϕ, ψ) areas where the Dunbrack dataset has low coverage. The approach of using explicit-solvent metadynamics simulations of dipeptides to generate high-coverage rotamer libraries can also be applied to noncanonical amino acids. We find that the MEDFORD rotamer library can predict χ_1 of noncanonical amino acids with an accuracy of around 75%.

5 | MATERIALS AND METHODS

5.1 | Creating the MEDFORD rotamer library

Following the example of the Dunbrack rotamer library, the backbone space was divided into a $10^\circ \times 10^\circ$ grid where the bin name identifies the center of that bin. For example, ($\phi = 10^\circ$, $\psi = 10^\circ$) contains data with $5^\circ \leq \phi < 15^\circ$ and $5^\circ \leq \psi < 15^\circ$. Each data point was assigned to a $10^\circ \times 10^\circ$ (ϕ, ψ) bin and a side chain rotamer combination (e.g., χ_1 = r60, χ_2 = r180). For rotameric side chain dihedrals, three rotamers, r60, r180, and r300, for each dihedral angle were defined as (0°, 120°], (120°, 240°], and (240°, 360°], respectively (Figure 1b). For nonrotameric side chain dihedrals, we divided the dihedral space into rotamers based on the minima of the dihedral angle distribution (Figure 1c). To define the edges of the rotamer states for nonrotameric amino acids with two dihedral angles (Asn, Asp, His, Phe, Trp, and Tyr), three backbone-independent $P(\chi_2 \mid \chi_1 = Y)$ plots were made where Y = r60, r180, and r300. The minimum value(s) for each $P(\chi_2 \mid \chi_1 = Y)$ plot were used as the rotamer cutoffs, resulting in the rotamer states as defined in Table S1. For nonrotameric amino acids with three dihedral angles (Gln and Glu), nine $P(\chi_3 \mid \chi_1 = Y_1, \chi_2 = Y_2)$ plots were made where Y_i = r60, r180, and r300 and the minimum value(s) for each $P(\chi_3 \mid \chi_1 = Y_1, \chi_2 = Y_2)$ were saved as the rotamer cutoffs. The resulting χ cutoffs are presented in Figures S2 and S3.

5.2 | Description of two previously reported rotamer libraries: The Dunbrack rotamer library and the Dynameomics rotamer library

5.2.1 | The Dunbrack rotamer library

The Dunbrack 2010 backbone-dependent rotamer library was generated by analyzing 3,985 protein chains from 3,845 protein crystal structures with resolution ≤ 1.8 Å, R factor ≤ 0.22, and mutual sequence identity ≤50%.[1] Adaptive kernel density estimates were used to calculate the density of (ϕ, ψ) for each side chain rotamer rX,$\rho(\phi, \psi \mid rX)$. Bayes' rule was then used to invert these densities to $P(rX \mid \phi, \psi)$ values. The resulting rotamer library reports the rotamer probabilities of rotameric side chains in each $10^\circ \times 10^\circ$ (ϕ, ψ) bin. Nonrotameric side chain dihedrals are reported in both 10° and 30° bins. In this paper, we used the version of the rotamer library with 5% kernel smoothing, which provides a balance between rotamer details and smoothed probabilities, and 30° bins for nonrotameric side chains. To evaluate the coverage of the Dunbrack rotamer library, we binned the raw amino acid data into $10^\circ \times 10^\circ$ (ϕ, ψ) bins centered on the (ϕ, ψ) values used to denote the bin. This information was used to define conventional low-coverage (ϕ, ψ) regions for each amino acid as backbone regions with <25 data points in the (ϕ, ψ) bin and conventional high-coverage (ϕ, ψ) regions as those with ≥25 data points in the (ϕ, ψ) bin.

5.2.2 | The Dynameomics rotamer library

The Dynameomics backbone-dependent rotamer library was created from molecular dynamics simulations of 807 proteins.[15] Simulations of the native state of each protein were run at 298 K using the *in lucem* molecular mechanics package (*il*mm) with the ENCAD force field.[7,30–32] The resulting dataset contains over 51,000 data points for each amino acid residue in each protein. Combined across all 807 proteins, the dataset covers over 97% of the (ϕ, ψ) space.[9] Backbone-dependent rotamer probabilities were calculated as the percentage of data points of a rotamer in each $10^\circ \times 10^\circ$ (ϕ, ψ) bin. Because the simulations used to create this dataset sampled most of the (ϕ, ψ) space, no smoothing or extrapolation was done to produce the rotamer library. Nonrotameric side chain dihedrals were reported with bin widths of 60° in Asn, Asp, Gln, Glu, and His, 90° for those in Phe and Tyr, and 120° for Trp.

5.3 | Test datasets for determining prediction accuracy

5.3.1 | HQ54

To assess the accuracy of the Dunbrack, Dynameomics, and MEDFORD rotamer libraries, we used the three rotamer libraries to predict the side chain dihedrals in a dataset of high-resolution proteins called HQ54.[49] This

dataset was chosen to represent the highest quality of protein structures and contains 54 high-resolution, non-redundant, monomeric protein crystal structures. From the 54 proteins, we created a dataset of amino acids with one or more side chain dihedral angle, that is, excluding Ala, Gly, and Pro. Amino acids with missing heavy atoms were removed and amide flips were corrected using the Reduce software package.[50] The resulting dataset contains a total of 5,567 data points.

5.3.2 | Cyclic peptide dataset

To assess the accuracy of each rotamer library on regions of backbone space less commonly found in proteins, we also created a dataset of amino acids from head-to-tail cyclic peptides. Cyclic peptides were curated from collections of cyclic peptides reported in three recent papers.[6,51,52] From the cyclic peptide datasets in these three papers, we obtained a list of head-to-tail cyclic peptides with at least one natural, side chain containing, L-amino acid. We further filtered this list by removing equal-length cyclic peptides with more than 75% sequence identity. The resulting list contained 52 cyclic peptides with lengths from 4 to 13 amino acids. From these cyclic peptide structures, we extracted amino acid residues with one or more side chain dihedral angle, that is, excluding Ala, Gly, and Pro. Amino acid residues were included in the dataset if they had L-chirality at C_α, a *trans* peptide bond preceding and succeeding the amino acid, and were unmethylated. The resulting dataset contains 181 L-amino acids from 52 cyclic peptides with a composition as shown in Figure 3c. The full composition of the cyclic peptide dataset can be found in Tables S7 and S8. However, as seen in Figure 3c, the cyclic peptide dataset has very limited data: All amino acids have ≤40 data points, and for most of the amino acids except Leu, Phe, and Val, there are fewer than 25 examples.

5.3.3 | Noncanonical amino acid dataset

To assess the accuracy of the rotamer libraries for noncanonical amino acids, we searched the Protein Data Bank for high resolution structures (≤ 2.5 Å) containing ABA, CIR, NAL, NLE, PYE, and TZA. Each resulting structure was manually inspected to ensure that the noncanonical amino acid was indeed present. Amino acids were removed from the list if they had D-chirality or a *cis* peptide bond preceding or succeeding the amino acid. We also filtered the dataset for sequence similarity, removing the structure with poorer resolution if the same chain was present in a second structure in our dataset. The resulting datasets contained 20 ABA, 20 CIR, 5 NAL, 37 NLE, 3 PYE, and 1 TZA amino acids. Because there are very few NAL (5), PYE (3), and TZA (1) examples, sidechain predictions and analysis were only performed on NLE (37), ABA (20), and CIR (20).

5.4 | Methods for evaluating prediction accuracy

Each applicable rotamer library was used to predict the side chain dihedrals of each amino acid in the HQ54, the cyclic peptide, and the noncanonical amino acid datasets using two methods. In the first method (top ranked rotamer), for each amino acid, the rotamer χ_{all} with the highest backbone-dependent probability in the rotamer library was selected. The corresponding dihedral values of that rotamer (χ_1, χ_2, etc.) were then compared to the experimental data.

For the HQ54 and the noncanonical amino acid datasets, we also used a second method (Rosetta repacking) to evaluate prediction accuracy of the three rotamer libraries. We used Rosetta to allow for the use of rotamer libraries in conjunction with a built-in scoring function for protein side chain packing.[46] By default, natural amino acids are handled by Rosetta's implementation of the Dunbrack rotamer library. Additional rotamer libraries may only be included as MakeRotLib-format libraries used for noncanonical amino acids.[9] With this in mind, we created copies of the parameter files for the natural amino acids used in this work, and edited them to include new three-letter codes and pointers to the rotamer library to be tested, converted to MakeRotLib format. Since Rosetta distinguishes between free sulfhydryl and disulfide bonded cysteine outside of the parameters file, we could not easily replicate this functionality and thus elected to omit cysteine from this benchmark. In the parameter files for the remaining amino acids, we also removed the ROTAMER_AA parameter line, as well as the annotation describing the amino acid as canonical, from the parameter files. Since several terms of the Rosetta ref2015 scoring function can only be applied to amino acids recognized by Rosetta as natural, instead of using the default implementation of the Dunbrack rotamer library, we converted the Dunbrack rotamer library to the MakeRotLib format as well and tested it alongside MEDFORD and Dynameomics. PDB files were edited to include the new three-letter codes, and packing was performed using Rosetta's fixbb utility. The NATAA command was included in the packer resfile to prevent mutations. In addition, the -ignore_zero_occupancy false flag was passed to fixbb in order to ensure that all atoms from the provided structure files would be loaded, instead

of poorly resolved atoms being omitted altogether. It should be noted that several PDB's (1JBE and 2DFB in the HQ54 dataset; 2WFJ, 4KTY, 4OKF, 5OCK, 5OCX, 5UHR, and 5W89 in the noncanonical amino acid dataset) were excluded from the analysis using Rosetta repacking because the protein chains contain residues such as pyroglutamic acid, (2 S)-2-amino-7-methoxy-7-oxoheptanoic acid, and 2,3-didehydroalanine that we were unable to treat properly in Rosetta.

All calculations were performed on the full PDB structures. However, for the accuracy analysis, if multiple chains were present in the crystal structure, we only considered the first chain in our analysis. The prediction was considered accurate if all dihedrals were within 20° of the experimental values. If a crystal structure has multiple conformations present for an amino acid residue, the prediction was compared to all of the conformations and considered accurate if all dihedral angles were within 20° in any of the experimental conformations. We also assessed the ability of each rotamer library to predict χ_1 alone by selecting the rotamer conformation, χ_{all}, with the highest probability but only comparing its χ_1 prediction to the crystal structure value. To assess the variability of our results, we performed a five-fold cross validation, randomly splitting each dataset into five groups, and calculating the percentage of amino acids predicted correctly in each group.

AUTHOR CONTRIBUTIONS

Jennifer C. Mortensen: Conceptualization (lead); data curation (supporting); formal analysis (lead); investigation (lead); methodology (equal); visualization (lead); writing – original draft (equal); writing – review and editing (equal). **Jovan Damjanovic:** Data curation (equal); formal analysis (equal); methodology (equal); writing – original draft (supporting); writing – review and editing (equal). **Yu-Shan Lin:** Conceptualization (equal); funding acquisition (lead); resources (lead); supervision (lead); writing – original draft (equal); writing – review and editing (equal). **Jiayuan Miao:** Investigation (equal). **Tiffani Hui:** Investigation (equal).

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM124160 (PI: Y.-S.L.) and Award Number K12GM133314 (the Tufts IRACDA Program). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. We thank Dr. Maxim Shapovalov and Dr. Matthew Childers for helpful discussions. We thank Dr. Parisa Hosseinzadeh for help with running calculations using MEDFORD in Rosetta. We are grateful for the support from the Tufts Technology Services and for computing resources at the Tufts Research Cluster. We thank the YSL group members for their invaluable feedback on the manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The MEDFORD rotamer library can be downloaded from https://github.com/jgaines42/MEDFORD-rotamer-library.

ORCID

Jennifer C. Mortensen https://orcid.org/0000-0001-8080-337X
Jovan Damjanovic https://orcid.org/0000-0002-6019-4738
Jiayuan Miao https://orcid.org/0000-0003-1112-1927
Tiffani Hui https://orcid.org/0000-0002-1355-389X
Yu-Shan Lin https://orcid.org/0000-0001-6460-2877

REFERENCES

1. Shapovalov MV, Dunbrack RL Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure. 2011;19:844–858.
2. Kaufmann KW, Lemmon GH, Deluca SL, Sheehan JH, Meiler J. Practically useful: What the rosetta protein modeling suite can do for you. Biochemistry. 2010;49:2987–2998.
3. Schrödinger L. The pymol molecular graphics system, version 2.0. 2015. Available from: Https://pymol.Org/2/.
4. Geng H, Jiang F, Wu YD. Accurate structure prediction and conformational analysis of cyclic peptides with residue-specific force fields. J Phys Chem Lett. 2016;7:1805–1810.
5. Slough DP, McHugh SM, Lin Y-S. Understanding and designing head-to-tail cyclic peptides. Biopolymers. 2018;109:e23113.
6. Huang H, Damjanovic J, Miao J, Lin Y-S. Cyclic peptides: Backbone rigidification and capability of mimicking motifs at protein–protein interfaces. Phys Chem Chem Phys. 2021;23:607–616.
7. Zhang H, Chen S. Cyclic peptide drugs approved in the last two decades (2001–2021). RSC Chem Biol. 2022;3:18–31.
8. Rezhdo A, Islam M, Huang M, Van Deventer JA. Future prospects for noncanonical amino acids in biological therapeutics. Curr Opin Biotechnol. 2019;60:168–178.
9. Renfrew PD, Choi EJ, Bonneau R, Kuhlman B. Incorporation of noncanonical amino acids into rosetta and use in computational protein-peptide interface design. PLoS One. 2012;7:e32637.
10. Renfrew PD, Craven TW, Butterfoss GL, Kirshenbaum K, Bonneau R. A rotamer library to enable modeling and design of peptoid foldamers. J Am Chem Soc. 2014;136:8772–8782.
11. Holden JK, Pavlovicz R, Gobbi A, Song Y, Cunningham CN. Computational site saturation mutagenesis of canonical and

non-canonical amino acids to probe protein-peptide interactions. Front Mol Biosci. 2022;9:848689.
12. Gfeller D, Michielin O, Zoete V. Swisssidechain: A molecular and structural database of non-natural sidechains. Nucleic Acids Res. 2013;41:D327–D332.
13. Gfeller D, Michielin O, Zoete V. Expanding molecular modeling and design tools to non-natural sidechains. J Comput Chem. 2012;33:1525–1535.
14. Beck DA, Jonsson AL, Schaeffer RD, et al. Dynameomics: Mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. Protein Eng Des Sel. 2008;21:353–368.
15. Scouras AD, Daggett V. The dynameomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. Protein Sci. 2011;20:341–352.
16. Towse CL, Rysavy SJ, Vulovic IM, Daggett V. New dynamic rotamer libraries: Data-driven analysis of side-chain conformational propensities. Structure. 2016;24:187–199.
17. Joo H, Qu X, Swanson R, McCallum CM, Tsai J. Fine grained sampling of residue characteristics using molecular dynamics simulation. Comput Biol Chem. 2010;34:172–183.
18. Petrovic D, Wang X, Strodel B. How accurately do force fields represent protein side chain ensembles? Proteins. 2018;86:935–944.
19. Shim J, Zhu X, Best RB, MacKerell AD Jr. $(Ala)_4$-X-$(Ala)_4$ as a model system for the optimization of the χ_1 and χ_2 amino acid side-chain dihedral empirical force field parameters. J Comput Chem. 2013;34:593–603.
20. Xu C, Wang J, Liu H. A Hamiltonian replica exchange approach and its application to the study of side-chain type and neighbor effects on peptide backbone conformations. J Chem Theory Comput. 2008;4:1348–1359.
21. Zhou C-Y, Jiang F, Wu Y-D. Residue-specific force field based on protein coil library. RSFF2: Modification of AMBER ff99SB. J Phys Chem B. 2015;119:1035–1047.
22. Li S, Elcock AH. Residue-specific force field (RSFF2) improves the modeling of conformational behavior of peptides and proteins. J Phys Chem Lett. 2015;6:2127–2133.
23. Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci U S A. 2002;99:12562–12566.
24. Piana S, Laio A. A bias-exchange approach to protein folding. J Phys Chem B. 2007;111:4553–4559.
25. Jiang F, Wu HN, Kang W, Wu YD. Developments and applications of coil-library-based residue-specific force fields for molecular dynamics simulations of peptides and proteins. J Chem Theory Comput. 2019;15:2761–2773.
26. Taechalertpaisarn J, Ono S, Okada O, Johnstone TC, Lokey RS. A new amino acid for improving permeability and solubility in macrocyclic peptides through side chain-to-backbone hydrogen bonding. J Med Chem. 2022;65:5072–5084.
27. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65:712–725.
28. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general Amber force field. J Comput Chem. 2004;25:1157–1174.
29. Hess B, Kutzner C, van der Spoel D, Lindahl E. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008;4:435–447.
30. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79:926–935.
31. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Comput Phys Commun. 2014;185:604–613.
32. Mehrabian H, Trout BL. In silico engineering of hydrate anti-agglomerant molecules using bias-exchange metadynamics simulations. J Phys Chem C. 2020;124:18983–18992.
33. Bian Y, Tan C, Wang J, Sheng Y, Zhang J, Wang W. Atomistic picture for the folding pathway of a hybrid-1 type human telomeric DNA g-quadruplex. PLoS Comput Biol. 2014;10:e1003562.
34. Damas JM, Filipe LCS, Campos SRR, et al. Predicting the thermodynamics and kinetics of helix formation in a cyclic peptide model. J Chem Theory Comput. 2013;9:5148–5157.
35. Dunbrack RL Jr, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nat Struct Biol. 1994;1:334–340.
36. Marcus E, Keller DA, Shibata M, Ornstein RL, Rein R. Comparing theoretical and experimental backbone-dependent side-chain conformational preferences for linear, branched, aromatic and polar residues. Chem Phys. 1996;204:157–171.
37. Chakrabarti P, Pal D. Main-chain conformational features at different conformations of the side-chains in proteins. Protein Eng Des Sel. 1998;11:631–647.
38. Samudrala R, Moult J. Determinants of side chain conformational preferences in protein structures. Protein Eng Des Sel. 1998;11:991–997.
39. Jiang F, Han W, Wu YD. The intrinsic conformational features of amino acids from a protein coil library and their applications in force field development. Phys Chem Chem Phys. 2013;15:3413–3428.
40. Zhou AQ, O'Hern CS, Regan L. Predicting the side-chain dihedral angle distributions of nonpolar, aromatic, and polar amino acids using hard sphere models. Proteins. 2014;82:2574–2584.
41. Avbelj F, Grdadolnik SG, Grdadolnik J, Baldwin RL. Intrinsic backbone preferences are fully present in blocked amino acids. Proc Natl Acad Sci U S A. 2006;103:1272–1277.
42. Levitt M, Hirshberg M, Sharon R, Daggett V. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Comput Phys Commun. 1995;91:215–231.
43. Levitt M, Hirshberg M, Sharon R, Laidig KE, Daggett V. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. J Phys Chem B. 1997;101:5051–5061.
44. Beck DA, Daggett V. Methods for molecular dynamics simulations of protein folding/unfolding in solution. Methods. 2004;34:112–120.
45. Jiang F, Han W, Wu Y-D. Influence of side chain conformations on local conformational features of amino acids and implication for force field development. J Phys Chem B. 2010;114:5840–5850.

46. Leaver-Fay A, Tyka M, Lewis SM, et al. Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. 2011;487:545–574.
47. Krivov GG, Shapovalov MV, Dunbrack RL Jr. Improved prediction of protein side-chain conformations with scwrl4. Proteins. 2009;77:778–795.
48. Gaines JC, Virrueta A, Buch DA, Fleishman SJ, O'Hern CS, Regan L. Collective repacking reveals that the structures of protein cores are uniquely specified by steric repulsive interactions. Protein Eng Des Sel. 2017;30:387–394.
49. Leaver-Fay A, O'Meara MJ, Tyka M, et al. Scientific benchmarks for guiding macromolecular energy function improvement. Methods Enzymol. 2013;523:109–143.
50. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol. 1999;285:1735–1747.
51. Malde AK, Hill TA, Iyer A, Fairlie DP. Crystal structures of protein-bound cyclic peptides. Chem Rev. 2019;119:9861–9914.
52. Prasad VK, Otero-de-la-Roza A, DiLabio GA. Pepconf, a diverse data set of peptide conformational energies. Sci Data. 2019;6:180310.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mortensen JC, Damjanovic J, Miao J, Hui T, Lin Y-S. A backbone-dependent rotamer library with high (ϕ, ψ) coverage using metadynamics simulations. Protein Science. 2022;31(12):e4491. https://doi.org/10.1002/pro.4491