Everyone Is a Protagonist: Residue Conformational Preferences in High-Resolution Protein Structures

RODRIGO LIGABUE-BRAUN[1], BRUNO BORGUESAN[2], HUGO VERLI[1], MATHIAS J. KRAUSE[3], and MÁRCIO DORN[1,2]

ABSTRACT

In many structural bioinformatics problems, there is a broad range of unanswered questions about protein dynamics and amino acid properties. Proteins are not strictly static objects, but rather populate ensembles of conformations. One way to understand these particularities is to analyze the information available in experimental databases. The Ramachandran plot, despite being more than half a century old, remains an utterly useful tool in the study of protein conformation. Based on its assumptions, we inspected a large data set (11,130 protein structures, amounting to 5,255,768 residues) and discriminated the conformational preferences of each residue type regarding their secondary structure participation. These data were studied for phi (ϕ), psi (ψ), and side chain chi (χ) angles, being presented in non-Ramachandranian plots. In the largest analysis of protein conformation made so far, we propose an original plot to depict conformational preferences in relation to different secondary structure elements. Despite confirming previous observations, our results strongly support a unique character for each residue type, whereas also reinforcing the observation that side chains have a major contribution to secondary structure and, by consequence, on protein conformation. This information can be further used in the development of more robust methods and computational strategies for structural bioinformatics problems.

Keywords: conformational preferences, structural biology, structure-based conformational preferences of amino acids, three-dimensional structure of proteins.

1. INTRODUCTION

G. N. RAMACHANDRAN (8 OCTOBER 1922 7 APRIL 2001) is considered one of the greatest scientists of the 20th century, often praised for making a lasting impression in the field of structural biology despite his somewhat far-off, obscure scientific origins (Ramakrishnan, 2001; Sarma, 2001). Among his many research topics in molecular biophysics, Ramachandran's probably greatest contribution to structural biology was the development of a diagram that now bears his name, the Ramachandran plot. Working on the (then) hot topic

[1]Center for Biotechnology, PPGBCM, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.
[2]Institute of Informatics, PPGC, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.
[3]Institute for Mechanical Process Engineering and Mechanics (MVM), Institute for Applied and Numerical Mathematics (IANM), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

of the collagen structure, Ramachandran proposed the triple, coiled-coil arrangement for the fibrous protein (Ramachandran et al., 1963). Even though this structure passed the test of time, he got little credit for it, especially due to controversies over details (Ramakrishnan, 2001; Sarma, 2001). These involved mainly the minimum possible distance between two nonbonded atoms, and this topic drew him and his team to an effort to prove that the distances he proposed for collagen were real and observed in other crystallographic structures. This research eventually led to the Ramachandran plot (Ramachandran et al., 1963), later expanded in a work that defined many key aspects of protein structure analysis (Ramachandran and Sasisekharan, 1968).

The Ramachandran plot defines limits that are imposed on the polypeptide chain conformation by the steric hindrance caused by nonbonded atoms. By plotting the angles that are allowed to rotate in a system of two linked peptide units, the plot is able to predict (or define) energetically forbidden regions for any dipeptide in a protein. The torsion angles (sometimes called *Ramachandran angles*) are defined as phi (ϕ) and psi (ψ), $C_{i-1}-N-C_{\alpha}-C$ and $N-C_{\alpha}-C-N_{i+1}$, respectively. The omega (ω) angle, set by the peptide bond, is practically fixed at 180°, since its partial double-bond character keeps it planar (Pauling et al., 1951). Similar to the polypeptide backbone, side chains have also dihedral angles, and the number of the so called chi (χ) angles of each side chain depends on the amino acid type.

The Ramachandran plot has been originally proposed based on interatomic distances and angles calculations (Carugo and Carugo, 2013). Afterward, many researchers confirmed its predictions by inspecting the range of values of ϕ/ψ torsion angle pairs achieved in real crystallographic structures (Kleywegt and Jones, 1996; Carugo and Carugo, 2013). These works also observed some ''blurring'' of the plot boundaries, chiefly explained by case-specific stabilization forces that allow specific plot violation (Carugo and Carugo, 2013). Still, Ramachandran's ingenious plot has been under continuous inspection. The most recent re-evaluation of Ramachandran graphs in light of current crystallographic data (Hollingsworth and Karplus, 2010) inspected 63, 149 residues and proposed new ways of presenting $\phi \times \psi$ data. In this context, and considering the continuous expansion of protein data bank (PDB) information in quantity and quality, in this work we revisited the Ramachandran plot by inspecting >5 million amino acid residues, expanding its analysis to a unparalleled level, which, in turn, supported the proposal of a new way of presenting amino acid conformational preferences in relation to different secondary structure elements.

Structural bioinformatics deals with problems in which the rules that govern the biochemical processes and relations are only partially known, making it hard to design efficient computational strategies to tackle them (de Lima Corrêa and Dorn, 2017). Especially in *Molecular Docking* and *Protein Structure Prediction* problems, there is a broad range of unanswered biological questions related to the protein folding process and dynamics. Predicting, for example, the three-dimensional (3D) structure of proteins requires an understanding of how tertiary structure depends on the primary amino acid sequence. Researchers are interested in the rules that determine the protein structure for a given sequence, which is influenced by the intrinsic conformational preference of each amino acid residue and also the interactions of neighbor amino acid residues (Dorn et al., 2014). An approach (knowledge-based prediction methods) to that problem is to explore the conformational particularities of local segments of amino acids in experimental structures solved at atomic resolution and stored in PDB (Borguesan et al., 2015).

Structural data are stored in databases or data banks whose number increases every year, the amount of data produced urged the necessity for fast and reliable ways of accessing, retrieving, researching, and understanding these data. In this article, we use high-quality conformational information from the PDB and explore novel approaches to present some of the fundamental aspects and conformational particularities of amino acids in proteins that cannot be easily explored by using traditional Ramachandran plots. Understanding the factors responsible for this behavior is necessary for protein studies and for the development of computational methods to tackle problems in structural biology.

2. MATERIALS AND METHODS

The database used in this study was built from a set of protein structures experimentally determined by X-ray diffraction with resolution $\leq$2.5Å and stored in PDB until December 2016. Only 3D protein structures with *R-factor* of ≤20% were considered. For proteins with sequence identity >30%, only one of them was considered and only amino acids with occupancy equal to 1 were used for further analysis. For each amino acid residue, the dihedral angles (ϕ, ψ, and χ's) and secondary structure information were assigned using secondary STRuctural IDEntification (STRIDE) (Frishman and Argos, 1995; Heinig and Frishman, 2004). This tool was chosen over more traditional tools [such as Dictionary of Secondary

Structure of Proteins (DSSP) (Kabsch and Sander, 1983)] due to its inclusion of dihedral angle potentials besides the hydrogen-bonding potentials. The data set analyzed in this work was produced using the Neighbors Influence of Amino acids and Secondary structures (NIAS) (Borguesan et al., 2017) Web-Server (http://sbcb.inf.ufrgs.br/nias). With this information, we have built a matrix $H_{aa,\,ss}$ of [−180.0°, +180.0° × −180.0°, +180.0°] cells for each amino acid residue (*aa*) and secondary structure (*ss*). Each cell contained the relative frequency of the amino acids and its secondary structure is observed in experimentally determined protein structures. This approach provides a way to identify the existence of conformational preferences for [*aa, ss*] pairs. Each cell (*i, j*) from *matrix H* has the number of times that a given amino acid residue *aa* in secondary structure *ss* has a pair of torsion angles ($i \leq \phi < i+1,\ j \leq \psi < j+1$).

Then, for each amino acid residue and secondary structure, the $APL_{aa,\,ss}$ (Eq. 1) was computed, representing the normalized frequency of each pair of angles. A higher rate associated with a pair ϕ and ψ indicates that this combination is more common in nature for the amino acid *aa* on the secondary structure *ss*.

$$APL_{aa,\,ss}(i,j) = \frac{H_{aa,\,ss}(i,j)}{\sum (H_{aa,\,ss})}. \tag{1}$$

3. RESULTS

By applying the described filtering settings on the PDB data, 11, 130 structures were selected, amounting to 5, 255, 768 residues (Tables 1 and 2). This curated data set was employed in all analyses described here. The traditional Ramachandran plots obtained did not diverge from the current knowledge on protein conformation. There are broader limits for each region in comparison with the original Ramachandran plot. Nonetheless, these regions are in agreement with more recent, large-scale surveys (Kleywegt and Jones, 1996; Hovmöller et al., 2002). To further explore amino acid conformational features, we plotted ϕ and ψ preferences for each residue, discriminating secondary structure preferences (Fig. 1). Side chain dihedral angles were also inspected in our work. The preferred conformations for χ angles (axes Ca-Cb, Cb-Cg, Cg-Cd, and Cd-Ne/Cd-Ce) were inspected and plotted in columns (Figs. 2 and 3) (angle description for all χ angles is presented in online Supplementary Material; available at http://sbcb.inf.ufrgs.br/chi). We observed a small number of experimental data, <1%, with amino acid residues in I (π-helix) and b (isolated bridge) conformational states (Tables 1 and 2, columns 7 and 9, respectively). Thus, we only consider six conformational states for further analysis: H (α-helix), G (3^{10}-helix), E (β-sheet), B (β-bridge), T (turn), and C (coil).

To facilitate interpretation, the conformational space populated by each residue type is commented hereunder. Unless otherwise noted, these observations regard the most populated zones for each angle. For these inspections, expected values for ϕ and ψ angles were in the interval of $\phi = (-70.0°, -60.0°)$, and $\psi = (-45.0°, -39.0°)$ for α-helices, $\phi = (-75.0°, -74.0°)$ and $\psi = (-5.0°, -4.0°)$ for 3^{10} helices; for β-sheets, there is a wider range of reference values ($-139.0°,\ -119.0°;\ -113.0°,\ -135.0°$), since parallel and antiparallel sheets, despite being very limited in their preferred conformations, have different distributions ($-119.0°,\ -113.0°$ and $-139.0°,\ -135.0°$, respectively) (Richardson, 1981; Bywater and Veryazov, 2015).

Ala (A): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -65.0°$, $-30.0° < \psi < -50.0°$. β sheet with $\phi \simeq -65.0°, 140.0° < \psi < 150.0°$.

Arg (R): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -68.0°$, $\psi \simeq -28.0°$. β sheet with traditional distribution. Side chain with four χ angles. χ_1 most abundant in the $-90.0° ¡ \chi_1 ¡ -60°$ range. χ_2 abundant in 180.0°. χ_3 abundant in 180.0°, with secondary, less populated regions in 60.0° and −60.0°. χ_4 has a wide distribution (when compared with the only other χ_4, from Lys), with almost forbidden occurrence in the $-60.0° \leq \chi_4 \leq 60.0°$ range, and most populated in 180.0° and secondarily abundant in $\simeq -90.0°$.

Asn (N): α-helix with traditional distribution. 3^{10}-helix with traditional, but broader, distribution. β-sheet with deviated ϕ values ($-90.0° \leq \phi \leq -130.0°$). Side chain with three χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 most abundant in the $-60.0° \leq \chi_2 \leq -30.0°$. χ_3 abundant in 180.0°.

Asp (D): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -70.0°$, $\psi \simeq -30.0°$. β-sheet with deviated ϕ values ($\phi \simeq -100.0°$). Side chain with two χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 populates the $-90.0° \leq \chi_2 \leq 90.0°$ range, being most abundant in $\simeq -30.0°$.

Table 1. STRIDE Secondary Structure Preferences of Amino Acid Residues in Proteins (Percentage Indicates the Preference of the Residue Type Toward Each Type of Secondary Structure)

	β-Bridge		*Coil*		*β-Sheet*		3_{10}-*Helix*		*α-Helix*		*π-Helix*		*Turn*		*β-Bridge*		*Total*
Amino	*B*		*C*		*E*		*G*		*H*		*I*		*T*		*b*		
ALA	3436	0.79%	5313	12.14%	80,338	18.36%	19,934	4.56%	21,145	48.33%	17	0.00%	69,096	15.79%	139	0.03%	437,540
ARG	3682	1.40%	41,014	15.62%	55,811	21.26%	10,891	4.15%	105,699	40.25%	32	0.01%	45,358	17.27%	89	0.03%	262,576
ASN	3015	1.35%	44,909	20.12%	3352	15.01%	9627	4.31%	58,138	26.04%	18	0.01%	73,949	33.12%	69	0.03%	223,245
ASP	3652	1.17%	61,138	19.63%	41,871	13.45%	16,753	5.38%	84,123	27.01%	54	0.02%	103,675	33.29%	141	0.05%	311,407
CYS	1248	1.81%	10,232	14.87%	22,306	32.42%	2342	3.40%	19,306	28.06%	0	0.00%	13,344	19.39%	24	0.03%	68,802
GLN	1883	0.97%	27,974	14.39%	38,143	19.63%	8849	4.55%	83,852	43.15%	12	0.01%	3351	17.24%	111	0.06%	194,334
GLU	2354	0.68%	42,913	12.49%	58,331	16.97%	2022	5.88%	155,325	45.20%	38	0.01%	64,332	18.72%	157	0.05%	343,670
GLY	3637	0.92%	110,208	27.99%	62,801	15.95%	11,028	2.80%	60,344	15.33%	79	0.02%	145,319	36.91%	271	0.07%	393,687
HIS	2172	1.70%	22,218	17.41%	31,062	24.34%	5518	4.32%	38,138	29.88%	25	0.02%	28,457	22.30%	37	0.03%	127,627
ILE	4717	1.56%	36,219	11.97%	119,789	39.60%	6706	2.22%	106,485	35.20%	32	0.01%	28,462	9.41%	79	0.03%	302,489
LEU	5689	1.16%	59,431	12.10%	131,631	26.80%	19,663	4.00%	21,836	44.45%	99	0.02%	56,134	11.43%	216	0.04%	491,223
LYS	3158	1.09%	46,123	15.94%	56,787	19.63%	12,804	4.43%	112,979	39.06%	44	0.02%	57,332	19.82%	39	0.01%	289,266
MET	1082	1.25%	11,112	12.84%	22,011	25.43%	3088	3.57%	38,154	44.09%	6	0.01%	1107	12.79%	20	0.02%	86,543
PHE	3652	1.68%	28,423	13.06%	74,044	34.02%	904	4.15%	70,425	32.36%	37	0.02%	31,992	14.70%	49	0.02%	217,662
PRO	2337	0.94%	81,714	32.97%	2599	10.49%	14,124	5.70%	34,623	13.97%	10	0.00%	88,996	35.91%	52	0.02%	247,846
SER	4952	1.61%	65,044	21.10%	6652	21.57%	15,277	4.95%	79,057	25.64%	40	0.01%	773	25.07%	131	0.04%	308,321
THR	483	1.63%	61,396	20.68%	87,669	29.53%	8247	2.78%	76,278	25.69%	43	0.01%	58,337	19.65%	128	0.04%	296,928
TRP	1065	1.32%	9959	12.32%	25,289	31.27%	4421	5.47%	27,067	33.47%	8	0.01%	1303	16.11%	27	0.03%	80,866
TYR	3392	1.73%	25,474	12.98%	65,833	33.54%	8151	4.15%	62,855	32.03%	46	0.02%	30,456	15.52%	54	0.03%	196,261
VAL	5806	1.55%	45,913	12.23%	166,292	44.29%	7349	1.96%	110,485	29.43%	51	0.01%	39,391	10.49%	188	0.05%	375,475
Total	65,759		884,544		1,266,038		214,032		1,753,143		691		1,069,540		2,021		5,255,768

STRIDE, secondary STRuctural IDEntification

TABLE 2. STRIDE SECONDARY STRUCTURE PREFERENCES OF AMINO ACID RESIDUES IN PROTEINS (PERCENTAGE INDICATES THE CONTRIBUTION OF EACH RESIDUE TYPE TO EACH SECONDARY STRUCTURE)

	β-Bridge		*Coil*		*β-Sheet*		3_{10}-*Helix*		*α-Helix*		*π-Helix*		*Turn*		*β-Bridge*		
Amino	*B*		*C*		*E*		*G*		*H*		*I*		*T*		*b*		*Total*
ALA	3436	5.23%	5313	6.01%	80,338	6.35%	19,934	9.31%	21,145	12.06%	17	2.46%	69,096	6.46%	139	6.88%	437,540
ARG	3682	5.60%	41,014	4.64%	55,811	4.41%	10,891	5.09%	105,699	6.03%	32	4.63%	45,358	4.24%	89	4.40%	262,576
ASN	3015	4.58%	44,909	5.08%	3352	2.65%	9627	4.50%	58,138	3.32%	18	2.60%	73,949	6.91%	69	3.41%	223,245
ASP	3652	5.55%	61,138	6.91%	41,871	3.31%	16,753	7.83%	84,123	4.80%	54	7.81%	103,675	9.69%	141	6.98%	311,407
CYS	1248	1.90%	10,232	1.16%	22,306	1.76%	2342	1.09%	19,306	1.10%	0	0.00%	13,344	1.25%	24	1.19%	68,802
GLN	1883	2.86%	27,974	3.16%	38,143	3.01%	8849	4.13%	83,852	4.78%	12	1.74%	3351	3.13%	111	5.49%	194,334
GLU	2354	3.58%	42,913	4.85%	58,331	4.61%	2022	9.45%	155,325	8.86%	38	5.50%	64,332	6.01%	157	7.77%	343,670
GLY	3637	5.53%	110,208	12.46%	62,801	4.96%	11,028	5.15%	60,344	3.44%	79	11.43%	145,319	13.59%	271	13.41%	393,687
HIS	2172	3.30%	22,218	2.51%	31,062	2.45%	5518	2.58%	38,138	2.18%	25	3.62%	28,457	2.66%	37	1.83%	127,627
ILE	4717	7.17%	36,219	4.09%	119,789	9.46%	6706	3.13%	106,485	6.07%	32	4.63%	28,462	2.66%	79	3.91%	302,489
LEU	5689	8.65%	59,431	6.72%	131,631	10.40%	19,663	9.19%	21,836	12.46%	99	14.33%	56,134	5.25%	216	10.69%	491,223
LYS	3158	4.80%	46,123	5.21%	56,787	4.49%	12,804	5.98%	112,979	6.44%	44	6.37%	57,332	5.36%	39	1.93%	289,266
MET	1082	1.65%	11,112	1.26%	22,011	1.74%	3088	1.44%	38,154	2.18%	6	0.87%	1107	1.04%	20	0.99%	86,543
PHE	3652	5.55%	28,423	3.21%	74,044	5.85%	904	4.22%	70,425	4.02%	37	5.35%	31,992	2.99%	49	2.42%	217,662
PRO	2337	3.55%	81,714	9.24%	2599	2.05%	14,124	6.60%	34,623	1.97%	10	1.45%	88,996	8.32%	52	2.57%	247,846
SER	4952	7.53%	65,044	7.35%	6652	5.25%	15,277	7.14%	79,057	4.51%	40	5.79%	773	7.23%	131	6.48%	308,321
THR	483	7.35%	61,396	6.94%	87,669	6.92%	8247	3.85%	76,278	4.35%	43	6.22%	58,337	5.45%	128	6.33%	296,928
TRP	1065	1.62%	9959	1.13%	25,289	2.00%	4421	2.07%	27,067	1.54%	8	1.16%	1303	1.22%	27	1.34%	80,866
TYR	3392	5.16%	25,474	2.88%	65,833	5.20%	8151	3.81%	62,855	3.59%	46	6.66%	30,456	2.85%	54	2.67%	196,261
VAL	5806	8.83%	45,913	5.19%	166,292	13.13%	7349	3.43%	110,485	6.30%	51	7.38%	39,391	3.68%	188	9.30%	375,475
Total	65,759		884,544		1,266,038		214,032		1,753,143		691		1,069,540		2,021		5,255,768

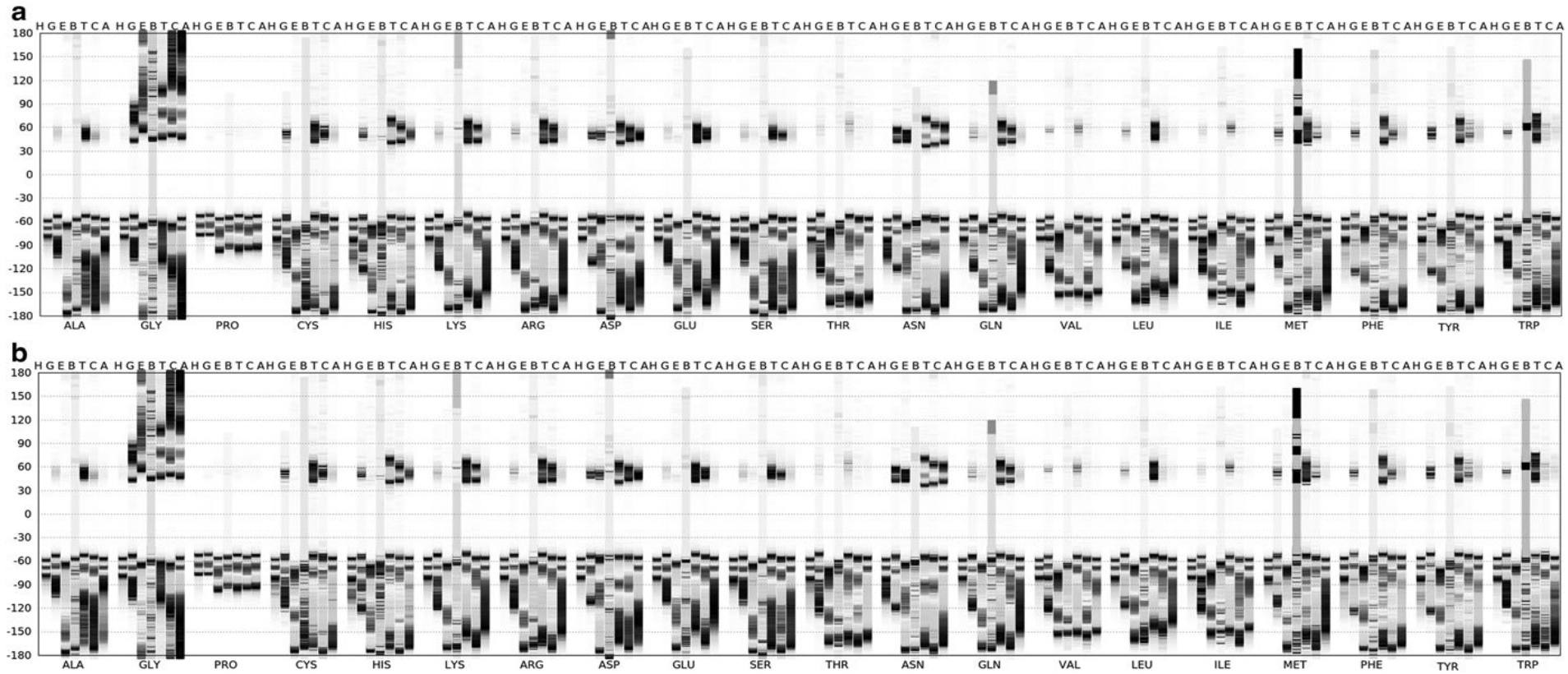

FIG. 1. Bar charts for ϕ (**a**) and ψ (**b**) angle preferences in the analyzed data set for each residue type, separated by secondary structure (nomenclature according to Table 1. ''A'' indicates the overall/total preference).

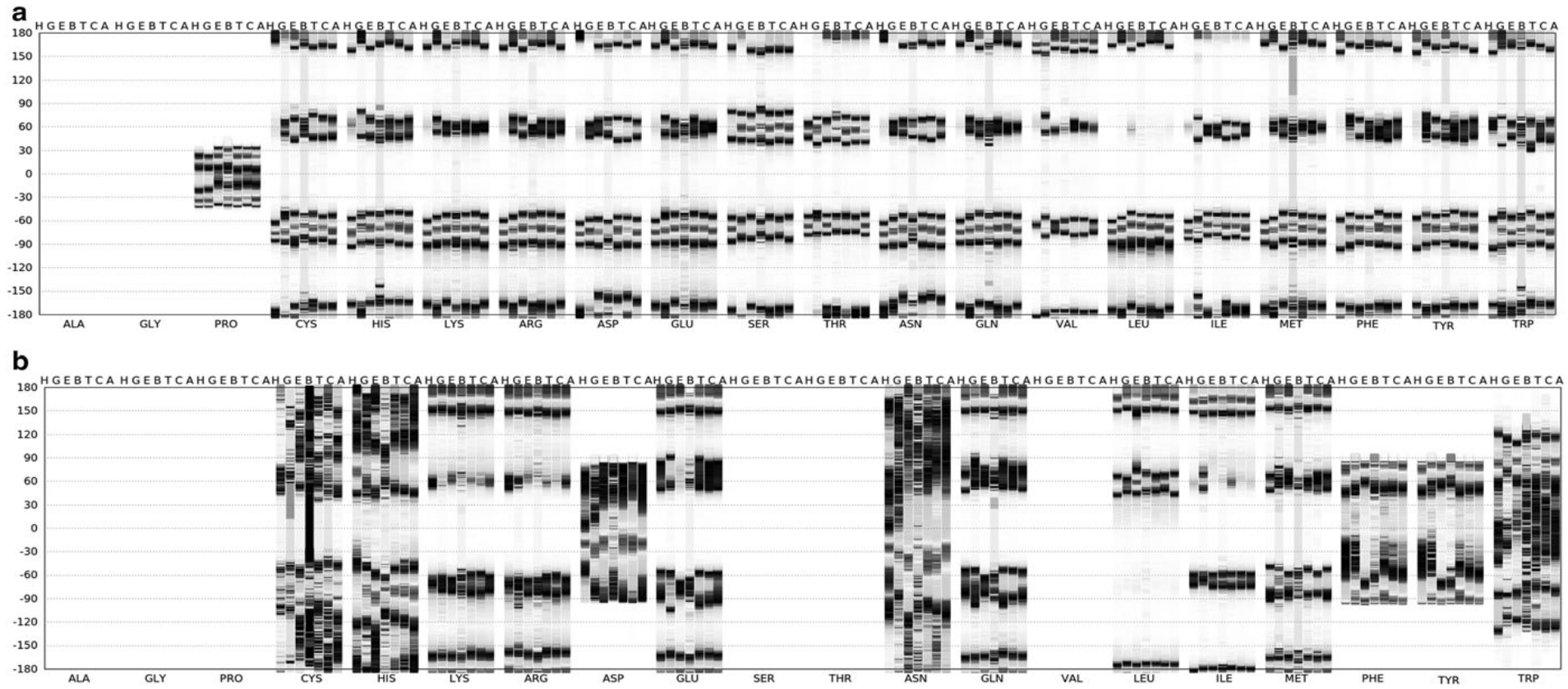

FIG. 2. Bar charts for χ_1 (**a**) and χ_2 (**b**) angle preferences in the analyzed data set for each residue type, separated by secondary structure (nomenclature according to Table 1. ''A'' indicates the overall/total preference).

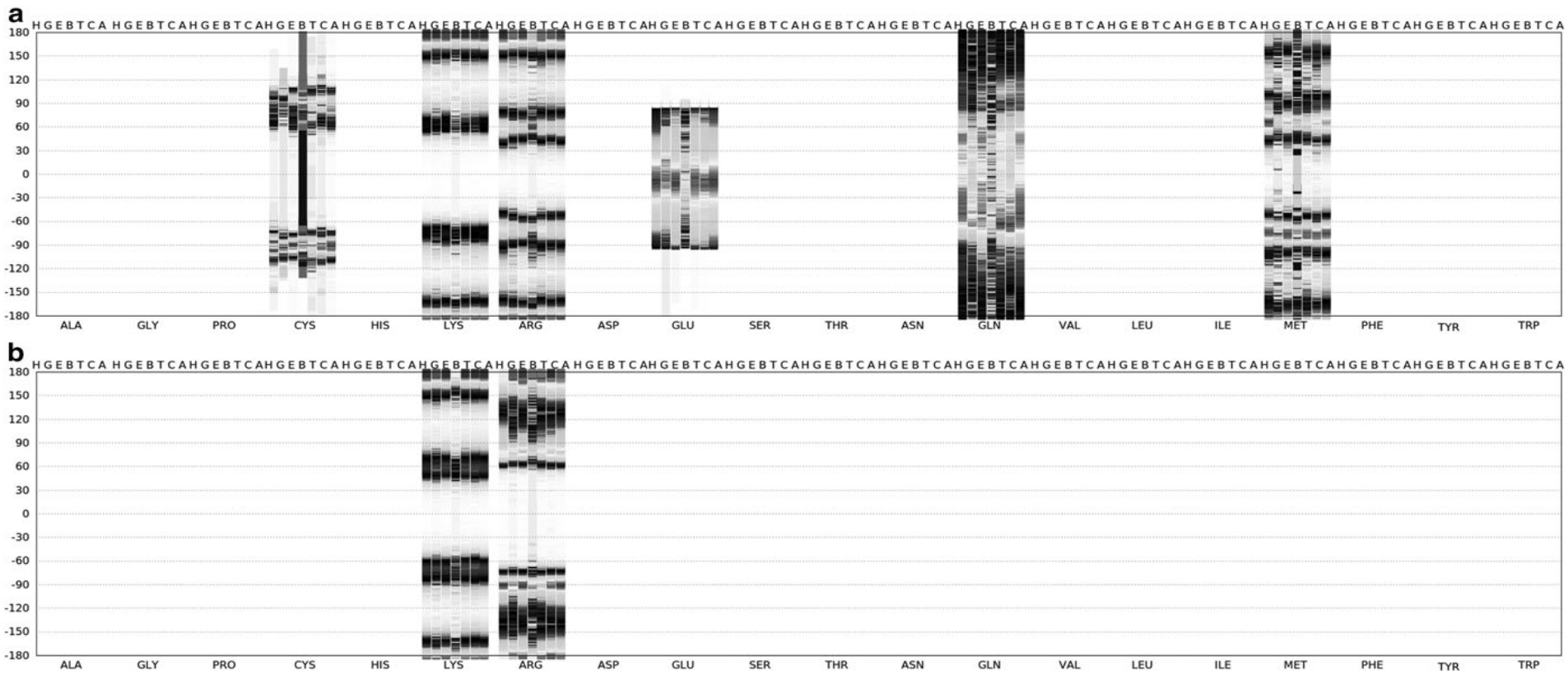

FIG. 3. Bar charts for χ_3 (**a**) and χ_4 (**b**) angle preferences in the analyzed data set for each residue type, separated by secondary structure (nomenclature according to Table 1. ''A'' indicates the overall/total preference).

Cys (C): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -70.0°$, $\psi \simeq -30.0°$. β-sheet with deviated ϕ values ($\phi \simeq -65.0°$). Side chain with three χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 most abundant in the $-90.0° \leq \chi_2 \leq -60.0°$ range. χ_3 most abundant in $\simeq -90.0°$ and secondarily in $\simeq 80.0°$.

Gln (Q): α-helix with traditional distribution. 3^{10}-helix with deviated ψ values ($\psi \simeq -32.0°$). β-sheet with traditional distribution. Side chain with three χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 abundant in 180.0°. χ_3 with broad distribution, being abundant in the $-60.0° \leq \chi_3 \leq 60.0°$ range, and most populated in $-60.0° \leq \chi_3 \leq -30.0°$.

Glu (E): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -65.0°$, $\psi \simeq -30.0°$. β-sheet with traditional distribution, but ϕ has a broader highly populated zone. Side chain with three χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 abundant in 180.0°. χ_3 populates only the $-90.0° \leq \chi_3 \leq 90.0°$ range, being most abundant in $\simeq 0°$.

Gly (G): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -70.0°$, $\psi \simeq -55.0°$. β-sheet with broad distribution (forbidden zones: $-60.0° \leq \phi \leq 45.0°$; $-80.0° \leq \psi \leq -30.0°$). Helical segments are bound by very strict limits, whereas other structures, in special sheets, have broader distributions than expected.

His (H): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -72.0°$, $\psi \simeq -55.0°$. β-sheet with traditional distribution. Side chain with two χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 most abundant around $-75.0°$, with the exception of the occurrences in α-helices, which populate mostly $\simeq 75.0°$.

Ile (I): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -65.0°$, $\psi \simeq -30.0°$. β-sheet with deviated values, $-130.0° \leq \phi \leq -120.0°$, $118.0° \leq \psi \leq 122.0°$. Side chain with two χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 abundant in $\simeq 150.0°$.

Leu (L): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -70.0°$, $-20.0° \leq \psi \leq -10.0°$. β-sheet with deviated values, $-122.0° \leq \phi \leq -118.0°$, $118.0° \leq \psi \leq 122.0°$. Side chain with two χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 abundant in $\simeq 180.0°$.

Lys (K): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -70.0°$, $\psi \simeq -25.0°$. β-sheet with traditional distribution. Side chain with four χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 and χ_3 abundant in $\simeq 180.0°$. χ_4 abundant in $\simeq 180.0°$, with secondary, less populated regions in $\simeq 60.0°$ and $\simeq -60.0°$.

Met (M): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -72.0°$, $\psi \simeq -30.0°$. β-sheet with traditional distribution. Side chain with three χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range. χ_2 abundant in $\simeq 180.0°$. χ_3 abundant in $\simeq -75.0°$ and secondarily in $\simeq 60.0°$.

Phe (F): α-helix with deviated ψ value, $\psi \simeq -60.0°$. 3^{10}-helix deviates from expected with $\phi \simeq -50.0°$, and broader ψ values, $-50.0° \leq \psi \leq 0.0°$. β-sheet with deviated values, $-140.0° \leq \phi \leq -120.0°$, $120.0° \leq \psi \leq 150.0°$. Side chain with two χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range, with the exception of the occurrences in α-helices, which populate mostly $\simeq 180.0°$. χ_2 populates the $-90.0° \leq \chi_2 \leq 90.0°$ range, being most abundant in $\simeq -90.0°$ and $\simeq 90.0°$.

Pro (P): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -50.0°$, $\psi \simeq -40.0°$. β-sheet with narrow distribution, $-80.0° \leq \phi \leq -75.0°$, $140.0° \leq \psi \leq 150.0°$. Side chain with a single χ angle. χ_1 populates the $\simeq -30.0°$ region when in α-helices and β-sheets, and the $\simeq 30.0°$ region when in other structures.

Ser (S): α-helix with narrow distribution. 3^{10}-helix deviates from expected with $\phi \simeq -50.0°$, $\psi \simeq -20.0°$. β-sheet with deviated ψ values ($145.0° \leq \psi \leq 152.0°$). Side chain with a single χ angle. χ_1 populates the $\simeq 60.0°$ region, and the $\simeq -60.0°$ region when in α-helices and β-sheets.

Thr (T): α-helix with deviated ψ value, $\psi \simeq -55.0°$. 3^{10}-helix deviates from expected with $\phi \simeq -70.0°$, $\psi \simeq -20.0°$. β-sheet with deviated ψ values ($120.0° \leq \psi \leq 130.0°$). Side chain with a single χ angle. χ_1 populates the $\simeq 60.0°$ region, and the $\simeq -60.0°$ region when in α-helices and β-sheets.

Trp (W): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -125.0°$, $-35.0° \leq \psi \leq -25.0°$. β-sheet with deviated ϕ values ($115.0° \leq \phi \leq 125.0°$). Side chain with two χ angles. χ_1 most abundant in the $-90.0° \leq \chi_1 \leq -60.0°$ range, with the exception of the occurrences in α-helices, which populate mostly $\simeq 180.0°$. χ_2 populates the $-120.0° \leq \chi_2 \leq 120.0°$ range, being most abundant in $\simeq 90.0°$, with the exception of the occurrences in 3^{10}-helices, which populate mostly $\simeq -90.0°$.

Tyr (Y): α-helix with deviated ψ value, $\psi \simeq -50.0^\circ$. 3^{10}-helix deviates from expected with $\phi \simeq -70.0^\circ$, $-40.0^\circ \leq \psi \leq 0.0^\circ$. β-sheet with traditional distribution. Side chain with two χ angles. χ_1 most abundant in the $-90.0^\circ \leq \chi_1 \leq -60.0^\circ$ range, with the exception of the occurrences in α-helices, which populate mostly $\simeq 180.0^\circ$. χ_2 populates the $-90.0^\circ \leq \chi_2 \leq 90.0^\circ$ range, being most abundant in $\simeq -90.0^\circ$ and $\simeq 90.0^\circ$.

Val (V): α-helix with traditional distribution. 3^{10}-helix deviates from expected with $\phi \simeq -70.0^\circ$, $-35.0^\circ \leq \psi \leq -30.0^\circ$. β-sheet with deviated ϕ values ($115^\circ \leq \phi \leq 125.0^\circ$). Side chain with a single χ angle. χ_1 populates the $160^\circ \leq \chi_1 \leq 180.0^\circ$ range, with helices more restricted to $\simeq 160.0^\circ$.

4. DISCUSSION

In this work, we performed the most encompassing scan of the PDB to date, aiming to analyze and update the Ramachandran plot with experimental data. This led to a data set that is larger than similar studies (Table 3). There have been many efforts to redesign the Ramachandran plot, that is, plot the same or similar data onto different planes, spaces, or forms. These attempts aim to convey more information using the same rationale from the original plot, either with only $\phi \times \psi$ information or with the addition of extra data, such as hydrogen bonding constraints. These alternative graphs include mirrored, wrapped, and geo-style Ramachandran plots [as reviewed in Hollingsworth (Hollingsworth and Karplus, 2010), and Carugo (Carugo and Carugo, 2013)]. Current efforts apply the Ramachandran strategy to larger structures, such as in the proteomic Ramachandran plot (Carugo and Carugo, 2013), or refine the original Ramachandran model by adding hydrogen bonding constraints (Porter and Rose, 2011). Despite being extremely useful, most of these attempts to redraw the Ramachandran plot lack residue-specific resolution, something that impairs more in-depth description of protein structural variations.

The pervasive textbook statement that secondary structure does not depend on amino acid sequence (due to hydrogen bonds being formed by the backbone only) overtly simplifies our understanding of protein conformational preferences. Since the late 1970s (at least), it is recognized that different side chains make residues behave differently regarding their ϕ and ψ angle preferences (Chou and Fasman, 1974; Williams et al., 1987; Street and Mayo, 1999). Apart from single-residue plots, these conformational preferences seemed to have remained largely overlooked. Here, we present stripe charts for ϕ and ψ angles, colored by population, separated by secondary structure types for each amino acid. We also add side chain torsional angles (chi-χ), which contribute to depict conformational preferences.

Table 3. Details for Data Sets Employed in Previous, Similar Works Exploring the Ramachandran Plot

References	*Year*	*N °proteins*	*N.°residues*	*Resolution (Å)*	*Type*	*R factor*
Beck et al.	2008	188	989, 001	NA	X ray, NMR	NA
Carugo and Carugo	2013	2725	NA	NA	NA	NA
Dahl et al.	2008	6702	NA	≤ 2.50	X ray	< 1.00
Gromiha et al.	2002	162	NA	NA	NA	NA
Hintze et al.	2016	8000	983, 574	≤ 2.00	X ray	
Ho et al.	2003	500	97, 368	< 1.8	X ray	NA
Hollingsworth and Karplus	2010	NA	72, 376	< -1.20	X ray	NA
Hooft et al.	1997	295	60, 000	< 2.8	X ray	< 0.25
Hovmöller et al.	2002	1042	137, 384	≤ 2.00	X ray	≤ 0.20
Hooft et al.	2005	2020	NA	≤ 2.0	X ray	≤ 0.30
Keskin et al.	2004	1646	NA	$0.54 \leq res \leq 3.00$	X ray, NMR	NA
Kleywegt and Jones	1996	403	81, 782	≤ 2.00	X ray	NA
Kolaskar and Sawant	1996	221	51, 998	NA	X ray	NA
Oldfield and Hubbard	1994	83	18, 503	≤ 2.00	X ray	NA
Pal and Chakrabarti	2002	353	63, 949	≤ 2.00	X ray	≤ 0.20
Tanty and Delsuc	2011	500	110, 018	NA	X ray	NA
Ting et al.	2010	3038	180, 847	≤ 1.70	X ray	≤ 0.25
Xia and Xie	2002	7342	NA	NA	X ray	NA
Zhou et al.	2011	850	86, 229	≤ 1.70	X ray	≤ 0.25

NA, not available; NMR, nuclear magnetic resonance.

Classical exceptions to the general ''Ramachandranian'' behavior are confirmed here, with Gly populating wide ranges of torsional angles and Pro being much more restricted. There is a clear trend for ψ values of nonhelical structures to populate a similar, more restrict angle range, whereas for ϕ, β-sheets have the most dissimilar distributions. Of all secondary structures, we confirm that helices have the narrowest ranges of ϕ and ψ values. β-sheets, in contrast, populate a broader range of values than originally expected. Ribosome environment simulations (using density functional theory) support these observations, showing that helical energy wells are deep and well defined, whereas β-sheets are energetically unfavorable, having shallower, wider energy wells (Bywater and Veryazov, 2015). In addition, when compared with helices, β-sheets are more complex, more flexible, and especially more prone to discordant and inaccurate observations (Parisien and Major, 2005; Fujiwara et al., 2012). Variations of the same kind in helices are much more restricted (Richardson and Richardson, 1988; Ho et al., 2003; Engel and DeGrado, 2004).

Side chain conformational preferences are less frequently explored in the literature than ''traditional'' $\phi \times \psi$ plots. Most notably, side chains have been studied with regard to their effect on the protein backbone and ways to classify their variation (Dahl et al., 2008). Here, we observed some preferences that may group some amino acids, but these clusters are not as clearly separated as observed for previous, theoretical inspections (such as Dahl et al. 2008; Carrascoza et al. 2014). For the χ_1 angle, Ser adopts a *gauche+* orientation when in α-helix, whereas Thr adopts *gauche+* for α-helix and β-sheet, probably due to an additional -CH_3 group in the latter when compared with the former. It is interesting to note that Ser is the most likely residue to adopt a nonallowed conformation when taking part in a protein structure, whereas Thr is the fourth more likely to do so (Pal and Chakrabarti, 2002). The χ_1 behavior observed here may be able to justify some of this propensity. In Val, χ_1 is near-*trans* for helices and *trans* for all other secondary structures. χ_1 in Phe, Tyr, and Trp is *trans* when in α-helix, a configuration that may be favored due to differences in hydrophobicity for main chain and side chain in these residues, which leads to repulsion.

Phe and Tyr have χ_2 in $+90.0^\circ$ when in α-helices, whereas all other secondary structures are found in $+90.0^\circ$ and -90.0°. A similar behavior is observed for His, despite having a broader distribution. For Trp, α-helix and β-sheet are in $+90.0$, with this restriction reflecting the increased rigidity of the residue double-ring when compared with single-ringed side chains. Asp and Asn have similar χ_2 and χ_3 distributions (although Asp distribution being narrower than Asn). Such similarity points to their charges, despite opposite, being the driving force for side chain orientation, since their molecular length is comparable. It is interesting to note that Asn was the sole residue to have a major deviation when comparing experimental and predicted preferences for β-sheet formation, something that was explained as a greater contribution of hydrogen bonding over other properties in this specific residue (Street and Mayo, 1999). Residue-level Ramachandran maps for Asn and Asp are considered more complex than those for other amino acids (Hovmöller et al., 2002), and here we also observe that χ_2 and χ_3 behave in a similar manner in Glu and Gln. χ_3 shows no other clear pattern. The differences in χ_4 between Lys and Arg (the only residues with this angle) are probably due to the difference in their chemical components. Although in Lys a C atom takes part in χ_4, in Arg an N atom takes part in this angle.

The complexity of protein conformational preferences becomes more evident when one considers the multidimensional representation of Ramachandran plots, adding residue preferences and including χ angles to the ϕ and ψ combinations (Fig. 4; Supplementary Material available at http://sbcb.inf.ufrgs.br/chi). These Ramachandran cubes are very different among themselves, pointing to each residue acting as an almost exclusive piece in the protein puzzle. Such Ramachandran plot variations have also been observed for single-residue or small peptide density functional theory calculations (Porter and Rose, 2011; Carrascoza et al., 2014). Comparing molecular dynamics simulations of pentapeptides, large proteins, and PDB structures, it has been shown that many factors overcome the ''intrinsic'' properties of the amino acids. Such factors include packing constraints, solvation, hydrogen bonding potential, and the tendency that amino acids have to form specific structures (Beck et al., 2008; Porter and Rose, 2011; Penner et al., 2014). With the current exception of selenocysteine and pyrrolysine (Rother and Krzycki, 2010), which are taxon-specific, the same 20 amino acids are found in all organisms. In fact, there seems to be a need for even more types of residues, as observable from post-translational modifications.

The amount of naturally occurring amino acids (i.e., amino acids for which there are corresponding tRNAs) is restrained by the genetic code, despite the fact that some amino acids have up to six codons coding for them, whereas some have only one, with no detrimental effect for the organism (Vella, 2003). Davis (1999) and Wong (2005) developed proposals in which the genetic code coevolved with the biochemical pathways needed to synthesize the different amino acids. These propositions place Asp, Asn, Glu, and Gln, which are residues that behave similarly in the analyses presented here, earlier in the evolution of

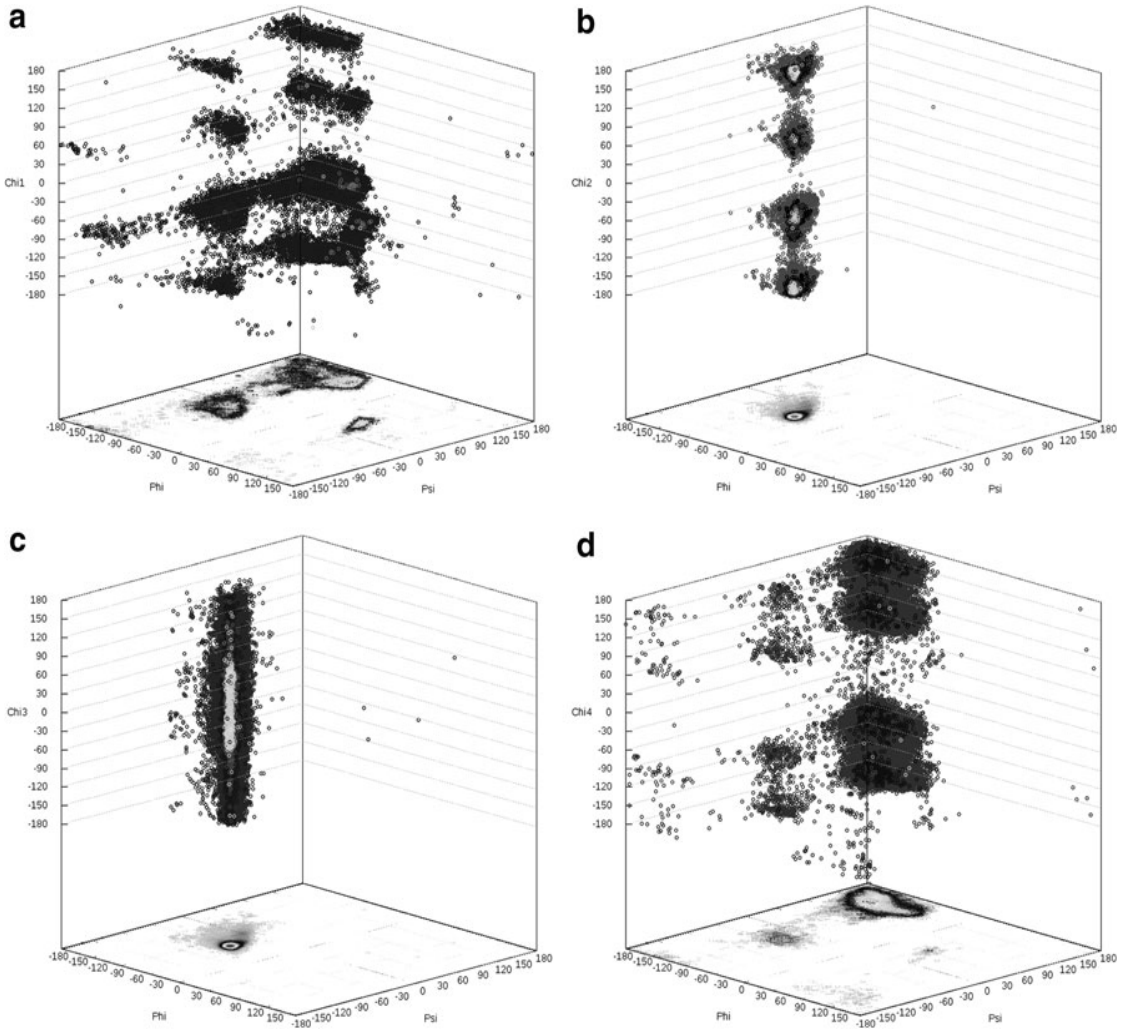

FIG. 4. Multidimensional Ramachandran cube plots. The dark red color represents the most densely occupied regions of the plot. Similar cubes for all residues in the combinations Phi × Psi × Chi_1 (**a**)/Chi_2 (**b**)/Chi_3 (**c**)/Chi_4 (**d**) for each secondary structure type are available at http://sbcb.inf.ufrgs.br/chi

the genetic code, whereas hydrophobic residues (Phe, Tyr, and Phe) would be the last to emerge (Griffiths, 2007). It is tempting to propose that, given the universality of amino acid composition in proteins, evolution had enough time to tinker with the system as to keep only neutral or beneficial residues in the making of proteins. Thus, redundant pieces in the puzzle would have been shunned earlier in evolutionary terms, being supplanted by unique, not totally interchangeable amino acids. This is in accordance with the observation that the genetic code is almost maximally optimized in its current state (Koonin and Novozhilovn, 2009). In this context, divergence in the canonical Ramachandran plot for each amino acid in each type of secondary structure is nothing but expected.

The contribution of residue neighborhood, that is, the residues that come before and after the residue position of interest, for the amino acid conformation has been extensively demonstrated (Kabat and Wu, 1973; Crasto and Feng, 2001; Hovmöller et al., 2002; Xia and Xie, 2002; Ting et al., 2010; Borguesan et al., 2015; Hollingsworth et al., 2016). In this work, we show that side chains also have clear preferences when taking part into different secondary structures, and not only the main chain, as previously observed (Hovmöller et al., 2002). A similar observation has been made recently (Hintze et al., 2016). Here, we expand this analysis, establishing a catalog of possible side chain angle preferences and their relation to secondary structure formation, explicitly depicting the information implicitly available in rotamer libraries. This information is especially interesting when one considers its application for structural modeling.

The construction of protein models is currently based on two strategies: the template-based modeling and the template-free modeling (Dorn et al., 2014). The knowledge about amino acid preferences for main chain and side chain dihedral angles constitute a requisite for proper template-based modeling. In this

context, use of side chain rotamer libraries (Shapovalov and Dunbrack, 2011; Towse et al., 2016; Hintze et al., 2016) may be greatly improved by taking the data presented here into account. The knowledge on such preferences, however, is not restricted to template-based modeling. Considering that the PDB may not harbor all possible templates anytime soon (Brylinski, 2015), the experimentally derived information presented here may help in selecting the most native-like models among multiple protein decoys generated by template-free modeling (Mishra et al., 2016). The application of structural models is not restricted to the acquisition of the models themselves, since these structures greatly favor function activity prediction pipelines, which can include docking of small ligands and protein protein docking. The proper prediction of protein backbone and side chain conformations also enhances our ability to design and propose more adequate protein mutants or synthetic peptides with biotechnological applications, even in large scale, reducing time and costs of valuable in vitro experimentation (Hakansson, 2002; Nazari-Robati et al., 2013).

5. CONCLUSIONS

The Ramachandran plot remains an utterly valuable tool to depict and understand protein structure. The current work improves upon previous observations, confirming in detail that residues behave very differently when taking part in diverse types of secondary structures. This variation is especially true for side chain behavior. Further developments in protein modeling and engineering must take into account the residue-level specificity of torsional angles to reliably depict natural proteins.

The observed protagonist, that is, the particularities of each amino acid residue when participating in different secondary structures, can be used to answer some important questions about protein dynamics. In protein structure prediction, this information can be employed, for example, to reduce the protein conformational search space and improve the robustness of prediction methods. In the field of computational protein design, this knowledge may be used to guide the design process, by offering reliable information on the statistical distribution of the amino acid residues, the natural preferences of amino acids to participate in segments of secondary structures, etc. Prediction of local structural properties of a protein molecule should be easier when computational protein design methods consider structural template information.

ACKNOWLEDGMENT

This research received funding from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (473692/2013-9 and 311022/2015-4); the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (473692/2013-9 and 311022/2015-4); Deutscher Akademischer Austauschdienst (DAAD), Alexander von Humboldt-Stiftung (AvH) Germany, and the Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS) (PRONUPEQ 16/2551-0000 520-6). This research is supported by Microsoft under a Microsoft Azure for Research Award.

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Beck, D., Alonso, D., Inoyama, D., et al. 2008. The intrinsic conformational propensities of the 20 naturally occurring amino acids and reflection of these propensities in proteins. *Proc. Natl. Acad. Sci. U. S. A.* 105, 12259 12264.

Borguesan, B., Barbachan e Silva, M., Grisci, B., et al. 2015. APL: An angle probability list to improve knowledge based metaheuristics for the three dimensional protein structure prediction. *Comput. Biol. Chem.* 59, 142 157.

Borguesan, B., Inostroza Ponta, M., and Dorn, M. 2017. Nias server: Neighbors influence of amino acids and secondary structures in proteins. *J. Comput. Biol.* 24, 255 265.

Brylinski, M. 2015. Is the growth rate of protein data bank sufficient to solve the protein structure prediction problem using template based modeling? *Bio Algorithms Med Syst.* 11, 1 7.

Bywater, R., and Veryazov, V. 2015. The dipeptide conformations of all twenty amino acid types in the context of biosynthesis. *SpringerPlus* 4, 1 12.

Carrascoza, F., Zaric, S., and Silaghi Dumitrescu, R. 2014. Computational study of protein secondary structure ele ments: Ramachandran plots revisited. *J. Mol. Graphics Modell.* 50, 125 133.

Carugo, O., and Carugo, K. 2013. Half a century of Ramachandran plots. *Acta Crystallogr.* 69, 1333 1341.

Chou, P., and Fasman, G. 1974. Conformational parameters for amino acids in helical, β sheet, and random coil regions calculated from proteins. *Biochemistry 13.* 211 222.

Crasto, C., and Feng, J. 2001. Sequence codes for extended conformation: A neighbor dependent sequence analysis of loops in proteins. *Proteins* 42, 399 413.

Dahl, D., Bohannan, Z., Mo, Q., et al. 2008. Assessing side chain perturbations of the protein backbone: A knowledge based classification of residue ramachandran space. *J. Mol. Biol.* 378, 749 758.

Davis, B.K. 1999. Evolution of the genetic code. *Prog. Biophys. Mol. Biol.* 72, 157 243.

de Lima Corrêa, L., and Dorn, M. 2017. Multi agent systems in three dimensional protein structure prediction, 241 278. *In Multi Agent Based Simulations Applied to Biological and Environmental Systems*. D. F. Adamitti (Ed). IGI Global, Hershey, PA.

Dorn, M., Barbachan e Silva, M., Buriol, L., et al. 2014. Three dimensional protein structure prediction: Methods and computational strategies. *Comput. Biol. Chem.* 53PB, 251 276.

Engel, D., and DeGrado, W. 2004. Amino acid propensities are position dependent throughout the length of helices. *J. Mol. Biol.* 337, 1195 1205.

Frishman, D., and Argos, P. 1995. Knowledge based protein secondary structure assignment. *Proteins* 23, 566 579.

Fujiwara, K., Toda, H., and Ikeguchi, M. 2012. Dependence of α helical and β sheet amino acid propensities on the overall protein fold type. *BMC Struct. Biol.* 12, 1 15.

Griffiths, G. 2007. Cell evolution and the problem of membrane topology. *Nat. Rev. Mol. Cell Biol.* 8, 1018 2024.

Gromiha, M., Oobatake, M., Kono, H., et al. 2002. Importance of mutant position in Ramachandran plot for predicting protein stability of surface mutations. *Biopolymers* 64, 210 220.

Hakansson, K. 2002. The strand helix motif is a recurring theme in biological hydrolysis. does the conformation of the Ramachandran outlier enhance its electrophilicity? *Int. J. Biol. Macromol.* 30, 273 277.

Heinig, M., and Frishman, D. 2004. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* 32, W500 W502.

Hintze, B., Lewis, S., Richardson, J., et al. 2016. Molprobity's ultimate rotamer library distributions for model vali dation. *Proteins* 84, 1177 1189.

Ho, B., Thomas, A., and Brasseur, R. 2003. Revisiting the Ramachandran plot: Hard sphere repulsion, electrostatics, and h bonding in the alpha helix. *Protein Sci.* 12, 2508 2522.

Hollingsworth, S., and Karplus, P. 2010. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol. Concepts* 1, 271 283.

Hollingsworth, S., Lewis, M., and Karplus, P. 2016. Beyond basins: ϕ,ψ preferences of a residue depend heavily on the ϕ,ψ values of its neighbors. *Protein Sci.* 25, 1757 1762.

Hooft, R., Sander, C., and Vriend, G. 1997. Objectively judging the quality of a protein structure from a Ramachandran plot. CABIOS Comput. *Appl. Biosci.* 13, 425 430.

Hooft, R., Sander, C., and Vriend, G. 2005. Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* 44, 9691 9702.

Hovmöller, S., Zhou, T., and Ohlson, T. 2002. Conformations of amino acids in proteins. *Acta Crystallogr.* 58, 768 776.

Kabat, E., and Wu, T. 1973. The influence of nearest neighbor amino acids on the conformation of the middle amino acid in proteins: Comparison of predicted and experimental determination of β sheets in concanavalin a. *Proc. Natl. Acad. Sci. U. S. A.* 70, 1473 1477.

Kabsch, W., and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 22, 2577 2637.

Keskin, O., Yuret, D., Gursoy, A., et al. 2004. Relationships between amino acid sequence and backbone torsion angle preferences. *Proteins* 55, 992 998.

Kleywegt, G., and Jones, T. 1996. Phi/psi chology: Ramachandran revisited. *Structure* 4, 1395 1400.

Kolaskar, A., and Sawant, S. 1996. Objectively judging the quality of a protein structure from a Ramachandran plot. *Int. J. Pept. Protein Res.* 47, 110 116.

Koonin, E.V., Novozhilovn, A.S. 2009. Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* 61, 99 111.

Mishra, A., Iqbal, S., and Hoque, M. 2016. Discriminate protein decoys from native by using a scoring function based on ubiquitous phi and psi angles computed for all atom. *J. Theor. Biol.* 398, 112 121.

Nazari Robati, M., Khajeh, K., Aminian, M., et al. 2013. Enhancement of thermal stability of chondroitinase ABC I by site directed mutagenesis: An insight from Ramachandran plot. *Biochim. Biophys. Acta* 1834, 479 486.

Oldfield, T., and Hubbard, R. 1994. Analysis of c_α geometry in protein structures. *Proteins* 18, 324 337.

Pal, D., and Chakrabarti, P. 2002. On residues in the disallowed region of the ramachandran map. *Biopolymers* 63, 195 206.
Parisien, M., and Major, F. 2005. A new catalog of protein beta sheets. *Proteins* 61, 545 558.
Pauling, L., Corey, R., and Branson, H. 1951. The structure of proteins: Two hydrogen bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* 37, 205 211.
Penner, R., Andersen, E., Jensen, J., et al. 2014. Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture. *Nat. Commun.* 5, 1 12.
Porter, L., and Rose, G. 2011. Redrawing the Ramachandran plot after inclusion of hydrogen bonding constraints. *Proc. Natl. Acad. Sci. U. S. A.* 108, 109 113.
Ramachandran, G., and Sasisekharan, V. 1968. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23, 283 438.
Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configu rations. *J. Mol. Biol.* 7, 95 99.
Ramakrishnan, C. 2001. Ramachandran and his map. *Resonance* 6, 48 56.
Richardson, J. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34, 167 339.
Richardson, J., and Richardson, D. 1988. Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240, 1648 1652.
Rother, M., Krzycki, J. 2010. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. *Archaea* 453642, 1 14.
Sarma, R. 2001. G.N. Ramachandran (1922 2001). *Structure* 9, 765 767.
Shapovalov, M., and Dunbrack, R. 2011. A smoothed backbone dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* 19, 844 858.
Street, A., and Mayo, S. 1999. Intrinsic beta sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9074 9076.
Tanty, M., and Delsuc, M. A. 2011. RamaDA: Complete and automated conformational overview of proteins. Ar Xiv.1111.5586 [q bio.BM] 1 7.
Ting, D., Wang, G., Shapovalov, M., et al. 2010. Neighbor dependent ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model. *PLoS Comput. Biol.* 6, 1 21.
Towse, C. L., Rysavy, S., Vulovic, I., et al. 2016. New dynamic rotamer libraries: Data driven analysis of side chain conformational propensities. *Structure* 24, 187 199.
Vella, F. 2003. Is there and answer? *Life* 55, 695 696.
Williams, R., Chang, A., Jureti, D., et al. 1987. Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta* 916, 200 204.
Wong, J.T. F. 2005. Coevolution theory of the genetic code at age thirty. *Bioessays* 27, 416 425.
Xia, X., and Xie, Z. 2002. Protein structure, neighbor effect, and a new index of amino acid dissimilarities. *Mol. Biol. Evol.* 19, 58 67.
Zhou, A., O'Hern, C., and Regan, L. 2011. Revisiting the Ramachandran plot from a new angle. *Protein Sci.* 20, 1166 1171.

Address correspondence to:
Dr. Márcio Dorn
Institute of Informatics
PPGC
Federal University of Rio Grande do Sul
Porto Alegre
Brazil

E-mail: mdorn@inf.ufrgs.br