

Research Article

Dissection of Factors Affecting the Variability of the Peptide Bond Geometry and Planarity

Nicole Balasco,¹ Luciana Esposito,¹ Amarinder Singh Thind,²
Mario Rosario Guarracino,² and Luigi Vitagliano¹

¹Istituto di Biostrutture e Bioimmagini, CNR, Napoli, Italy

²Istituto di Calcolo e Reti ad Alte Prestazioni, CNR, Napoli, Italy

Correspondence should be addressed to Nicole Balasco; nicole.balasco@unicampania.it
and Luigi Vitagliano; luigi.vitagliano@unina.it

Received 3 July 2017; Accepted 5 September 2017; Published 15 October 2017

Academic Editor: Rita Casadio

Copyright © 2017 Nicole Balasco et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Proteins frequently assume complex three-dimensional structures characterized by marginal thermodynamic stabilities. In this scenario, deciphering the folding code of these molecular giants with clay feet is a cumbersome task. Studies performed in last years have shown that the interplay between backbone geometry and local conformation has an important impact on protein structures. Although the variability of several geometrical parameters of protein backbone has been established, the role of the structural context in determining these effects has been hitherto limited to the valence bond angle τ ($NC^\alpha C$). We here investigated the impact of different factors on the observed variability of backbone geometry and peptide bond planarity. These analyses corroborate the notion that the local conformation expressed in terms of (ϕ, ψ) dihedrals plays a predominant role in dictating the variability of these parameters. The impact of secondary structure is limited to bond angles which involve atoms that are usually engaged in H-bonds and, therefore, more susceptible to the structural context. Present data also show that the nature of the side chain has a significant impact on angles such as $NC^\alpha C^\beta$ and $C^\beta C^\alpha C$. In conclusion, our analyses strongly support the use of variability of protein backbone geometry in structure refinement, validation, and prediction.

1. Introduction

Proteins are large macromolecules that play a primary role in all biological processes. It is commonly assumed that their functions are strictly related to the three-dimensional structural organization of the constituent atoms. With the exception of rather few highly stable proteins, protein structured states, even when nontransient, are marginally stable. The simultaneous complexity and fragility of these structures make proteins a sort of giants with clay feet. These considerations clearly explain the difficulties encountered in the last decades to decipher the protein folding code [1–3]. Indeed, an appropriate description of protein structures should properly account for a huge number of different energetic factors, some of which have been identified only recently. Indeed, protein folding is the delicate balance of several distinct factors which include well known (H-bonds, electrostatic,

and hydrophobic) and more recently discovered ($n \rightarrow \pi^*$ interactions, intraresidue H-bonds) determinants [4, 5]. Studies carried out in the last two decades have shown that the interplay between local conformation and protein backbone geometry has important structural consequences [6–22], even in highly restrained contexts [6]. In this framework, it has been shown that local geometry has a crucial impact on allowing/disallowing specific conformations [23, 24]. Moreover, we have shown that the optimization of backbone geometry and local conformation provides an important contribution to the protein stability [7, 13]. Indeed, quantum mechanics calculations have shown that swapping geometrical parameters between different accessible conformations has an energetic cost of 1–2 kcal/mol *per* residue [7]. It has also been highlighted that the optimization of protein geometry may be important for improving protein structure prediction [25, 26].

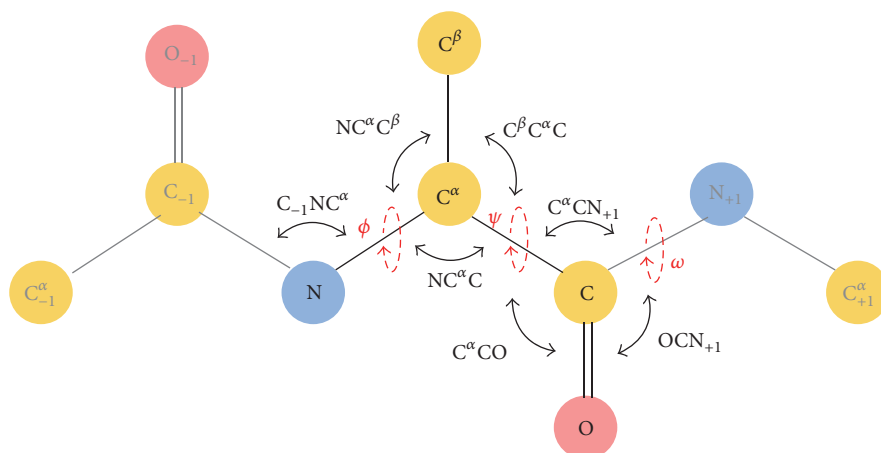


FIGURE 1: Schematic representation that shows the standard backbone torsion angles ϕ , ψ , and ω and the bond angles $NC^{\alpha}C^{\beta}$, $C^{\beta}C^{\alpha}C$, $NC^{\alpha}C^{\beta}$, $C^{\alpha}CO$, $C^{\alpha}CN^{+1}$, OCN^{+1} , and $C^{-1}NC^{\alpha}$. The dihedral angle ω_3 , which is used for the definition of the carbon carbonyl pyramidalization (see Methods), is defined by the atoms $OCN^{+1}C^{\alpha+1}$.

The variability of protein backbone geometry involves different parameters. These include bond distances, bond angles, and dihedral angles. Initial investigations have highlighted the conformational-dependent variability of the τ ($NC^{\alpha}C$) angle [9, 14]. Subsequent studies have extended this concept to the other backbone valence bond angles, to the peptide bond planarity, and, more recently, to bond distances [7, 8].

Many efforts have been made to unravel the factors that, besides conformation, may have an impact on τ angle. These investigations led to the conclusions that other factors such as secondary structure and residue type may affect the value of this angle, though at lower extent compared to conformation [9, 27–29]. Using statistical analyses of a recent ensemble of structures retrieved from the Protein Data Bank (PDB), we here extended these analyses to the other protein backbone parameters. Moreover, we also evaluated the dependence of peptide bond distortion (in terms of variations of the omega angle from planarity) and carbon carbonyl pyramidalization from the local structural context.

2. Methods

Statistical surveys of peptide bond geometrical parameters (bond and dihedral angles) were performed on ensembles of protein structures reported in the PDB (release of March 2016). These structures were selected using the PISCES culling server (<http://dunbrack.fccc.edu/PISCES.php>) applying specific criteria: resolution better than 1.6 Å for bond angles (Data 1.6) or 1.2 Å for dihedral angles (Data 1.2), sequence identity $\leq 25\%$, and R -factor ≤ 0.20 [30]. Additional selections of the structures of these datasets were carried out at residue level. In particular, in order to reduce local inaccuracies, we excluded the residues for which the ratio between the average backbone B -factor (atomic displacement parameter) of the residue and the same parameter calculated considering the entire chain was higher than 1.3. Data 1.6 and Data 1.2 datasets contain 3291 and 799 nonredundant protein chains, respectively.

The analyses dealt with all six bond angles involving non-H atoms of the protein backbone ($C^{\beta}C^{\alpha}C$, $NC^{\alpha}C^{\beta}$, $C^{\alpha}CO$, $C^{\alpha}CN^{+1}$, OCN^{+1} , and $C^{-1}NC^{\alpha}$) and two parameters that describe the peptide bond distortions ($\Delta\omega$ and θ_C) (Figure 1). In particular, $\Delta\omega$ defined as $(\omega - 180^\circ) \bmod 360^\circ$ represents the peptide bond deviations from planarity [12], whereas θ_C measured as $(\omega - \omega_3 + 180^\circ) \bmod 360^\circ$ (with ω_3 being the dihedral angle defined by the atoms $OCN^{+1}C^{\alpha+1}$) describes the displacement of the carbonyl carbon atom from the plane defined by its three bonded atoms (C^{α} , O, and N^{+1}) known as carbonyl carbon pyramidalization [10]. Some of the analyses were performed by computing average values of the geometrical parameters in specific (ϕ, ψ) boxes of the Ramachandran plot. In order to avoid the mixing of heterogeneous residues in terms of conformation, we minimized the size of these areas as much as possible while ensuring, at the same time, a significant number of observations.

The DSSP program [31] was used for the assignment of secondary structure elements as α -helix (H), 3(10)-helix (G), and β -sheet (E). Residues with a different notation (all but H, G, and E) were classified as coil (C). The statistical significance of the differences between the average values of pairs of angle distributions was evaluated assuming the so-called null hypothesis (no difference between the mean values) in a two-sample t -test analysis.

3. Results and Discussion

We initially evaluated the variability of both valence bond geometry and peptide bond planarity in the Ramachandran space using a recent database of protein structure (see Methods for details). In line with previous analyses [7, 14], we considered the bond angles formed by nonhydrogen atoms of the protein backbone ($NC^{\alpha}C$, $C^{\beta}C^{\alpha}C$, $NC^{\alpha}C^{\beta}$, $C^{\alpha}CO$, $C^{\alpha}CN^{+1}$, OCN^{+1} , and $C^{-1}NC^{\alpha}$) and $\Delta\omega$ and θ_C as indicators of the peptide bond distortions from planarity (Figure 1). Initial analyses were conducted by considering all non-Gly/non-Pro residues in all types of structures in recent protein

structure ensembles (Data 1.6 and Data 1.2 for bond and dihedral angles, resp.; see Methods for details), as Pro and Gly frequently display peculiar structural properties at geometry level [9, 29]. As shown in Figure S1 (in Supplementary Material available online at <https://doi.org/10.1155/2017/2617629>), all of the considered parameters display significant variability in the Ramachandran space. The comparison of these figures with those previously reported in the literature [7, 11–14] indicates a very close agreement. This observation clearly indicates that the increased size of the current databases does not have an impact on literature trends. Nevertheless, its larger content of structural information allows a more appropriate dissection of the possible factors influencing these variabilities.

As detailed in the following sections, for both backbone geometry and peptide bond planarity distortions, we evaluated the impact of the local (ϕ, ψ) conformation and of the structural context (occurrence of a specific secondary structure motif). For the backbone geometry, we also monitored the impact of the residue type on the observed variability.

3.1. Backbone Variability: Conformation versus Structural Context. These analyses were conducted by dissecting the Ramachandran space in (ϕ, ψ) boxes and considering collectively all eighteen non-Pro/non-Gly residues. In those boxes that were sufficiently populated (at least 50 residues per box), we separately evaluated the average values for each parameter for either residues belonging to secondary structure elements or residues embedded in nonregular regions. The correlation between the values computed in the same (ϕ, ψ) box is reported for the different parameters (NC^αC , $\text{NC}^\alpha\text{C}^\beta$, $\text{C}^\beta\text{C}^\alpha\text{C}$, C^αCO , $\text{C}^\alpha\text{CN}^{+1}$, OCN^{+1} , $\text{C}^{-1}\text{NC}^\alpha$, $\Delta\omega$, and θ_C) in Figures 2(a)–2(i). The values of the correlation coefficients and regression line parameters are reported in Table 1. An overview of these figures and of the correlation coefficients suggests that all these parameters tend to adopt similar values in different structural contexts (secondary structure or coil). More specifically, as for NC^αC [29], the valence bond angles C^αCO , $\text{C}^\alpha\text{CN}^{+1}$, $\text{C}^\beta\text{C}^\alpha\text{C}$, and $\text{C}^{-1}\text{NC}^\alpha$ exhibit very good agreements (correlation coefficient > 0.83) between the two ensembles. Indeed, the continuous fitting lines reported in Figures 2(a)–2(i) suggest that the variability of these parameters follows the same trends in the two distinct contexts. Moreover, the dashed-dotted diagonal line ($y = x$) indicates that also the absolute values of these bond angles are rather similar.

The correlation observed for $\text{NC}^\alpha\text{C}^\beta$ and OCN^{+1} , though highly significant, is less optimal. It is worth noting, however, that these latter angles display a limited overall variability ($\sim 3^\circ$ for $\text{NC}^\alpha\text{C}^\beta$ and $\sim 1.7^\circ$ for OCN^{+1} ; see Figure S1). Moreover, they also involve nitrogen and/or oxygen atoms whose position may be influenced by the local structural context being H-bond formers.

The analysis of the parameters that measure the deviations from planarity of the peptide bond indicates that for both $\Delta\omega$ and θ_C the structural environment plays a marginal role.

TABLE 1: Statistical parameters of the linear fitting of the graphs reported in Figure 2. The significance in terms of p value is less than 10^{-50} for all parameters.

Parameter	Correlation coefficient R	Regression line
NC^αC	0.96	$y = 0.85x + 16.9$
$\text{NC}^\alpha\text{C}^\beta$	0.71	$y = 0.68x + 35.3$
$\text{C}^\beta\text{C}^\alpha\text{C}$	0.88	$y = 0.78x + 24.1$
C^αCO	0.93	$y = 0.77x + 27.1$
$\text{C}^\alpha\text{CN}^{+1}$	0.93	$y = 0.64x + 41.8$
OCN^{+1}	0.72	$y = 0.41x + 72.2$
$\text{C}^{-1}\text{NC}^\alpha$	0.83	$y = 0.72x + 34.5$
$\Delta\omega$	0.89	$y = 0.75x + 0.68$
θ_C	0.89	$y = 0.85x + 0.04$

These observations clearly demonstrate that the local conformation is the predominant factor in determining the values of these geometrical parameters as residues in boxes with the same (ϕ, ψ) values but embodied in different structural contexts display rather similar values. This indicates that the general variability of peptide bond planarity is an intrinsic feature of the local conformation of the polypeptide chain.

Regression lines and correlation coefficients were also calculated separately for α -helix and β -sheet structures. As reported in Table S1, β -structures present very high correlation coefficients for all parameters. Highly significant correlation coefficients are generally exhibited also by α -helical residues. The two exceptions are the angles $\text{NC}^\alpha\text{C}^\beta$ and OCN^{+1} whose coefficients present either a limited ($\text{NC}^\alpha\text{C}^\beta$) or no (OCN^{+1}) statistical significance. This finding is not surprising taking into account the very limited variability of these two parameters in the helical regions.

To assess the role (if any) of secondary structure and to dissect the relative impact of structure and conformation, we performed additional analyses by comparing the mean values of each geometrical parameter of non-Gly/non-Pro residues in specific boxes of the Ramachandran space. The impact of the secondary structure was evaluated by comparing the average values of these parameters for residues either in coil region or in secondary structure elements (α -helix and β -structure). To maximize the significance of these analyses, we selected the most populated regions of the plot. In particular, we considered the boxes $3^\circ \times 3^\circ$ centered at $(\phi, \psi) = (-63^\circ, -43^\circ)$ and $15^\circ \times 15^\circ$ centered at $(\phi, \psi) = (-120^\circ, 130^\circ)$ corresponding to helical and extended states, respectively. It is worth mentioning that the standard deviations observed for these parameters in each box (Tables 2 and 3) are significantly lower than those associated in the Engh and Huber parameters [32], which are commonly used in protein refinement protocols (Table S2). This discrepancy is not surprising since the Engh and Huber analysis did not consider the overall variability of these angles in the Ramachandran space.

As shown in Tables 2 and 3 and Figures S2–S3, the differences are very limited. The p -test analysis indicates that the mean values are not significantly different for the angles C^αCO and $\text{C}^{-1}\text{NC}^\alpha$. On the other hand, the local structural context has a significant impact on the angles

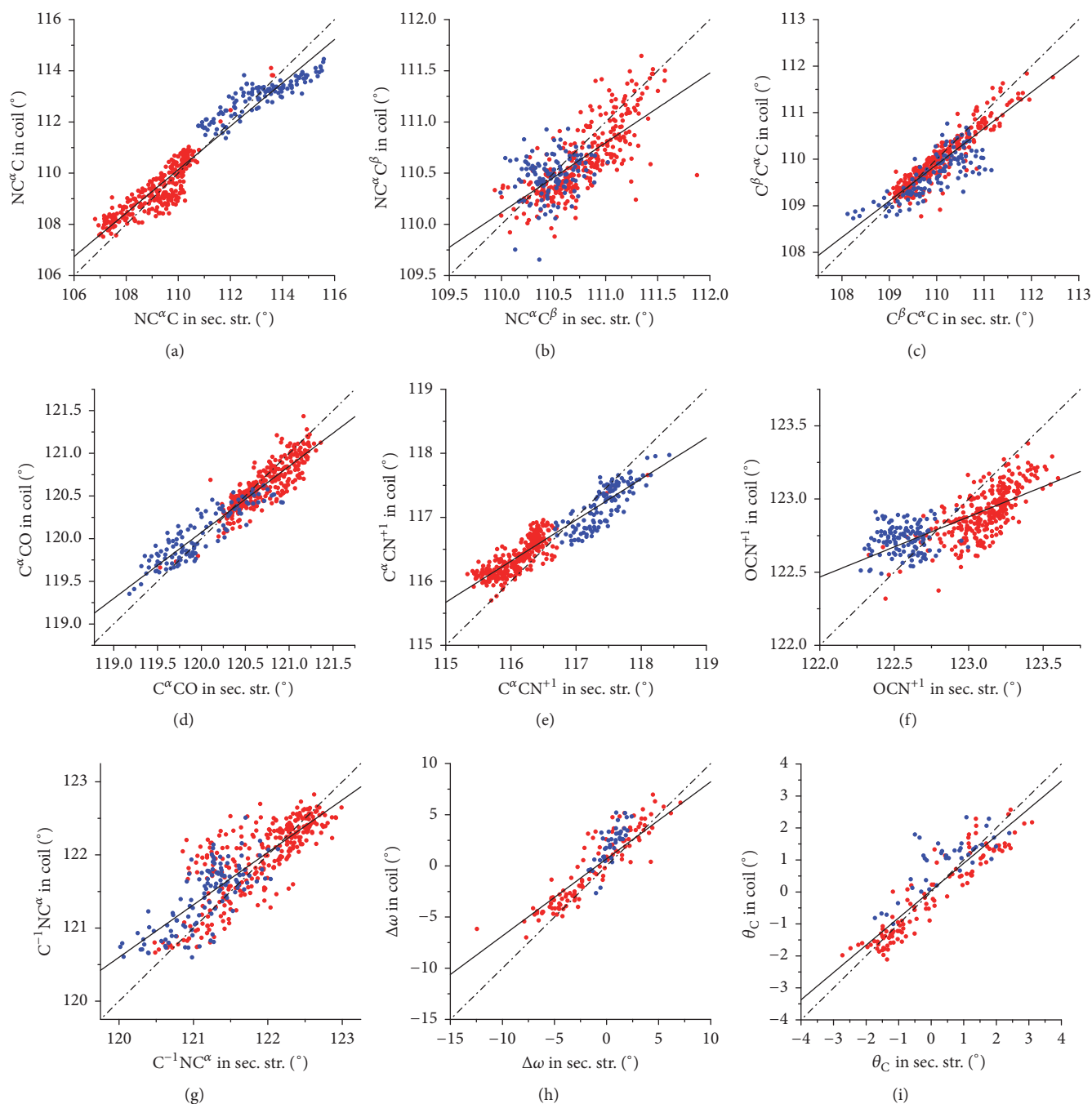


FIGURE 2: Correlation between the average values of each geometrical parameter computed in the same (ϕ, ψ) boxes of residues belonging to secondary structure elements (β -structure in red and α -helices in blue) with those of residues of coil structures: $NC^\alpha C$ (a), $NC^\alpha C^\beta$ (b), $C^\beta C^\alpha C$ (c), $C^\alpha CO$ (d), $C^\alpha CN^{+1}$ (e), OCN^{+1} (f), $C^{-1}NC^\alpha$ (g), $\Delta\omega$ (h), and θ_C (i). Continuous and dashed-dotted lines are used for the regression lines and diagonals ($y = x$), respectively.

$NC^\alpha C$ and OCN^{+1} . The influence of the local structure on the angles $C^\alpha CN^{+1}$, $NC^\alpha C^\beta$, and $C^\beta C^\alpha C$ shows a nonsystematic dependence on the type of secondary structure.

To further investigate the role of the conformation versus local structure, we also compared the values of these parameters for residues adopting the same structural motif (β -sheet)

but in boxes characterized by significantly different (ϕ, ψ) angles. As shown in Table 4, differences are remarkable and statistically significant in all cases. A collective analysis of the data reported in Tables 2–4 corroborates the notion that the contribution of (ϕ, ψ) angles overcomes the impact of the local structural motif. A significant contribution of secondary

TABLE 2: Average values of bond angles with standard deviations and occurrences (in bracket) of non-Gly/non-Pro residues in α -helix or coil in the $3^\circ \times 3^\circ$ box centered at $(\varphi, \psi) = (-63^\circ, -43^\circ)$. The differences (coil minus α -helix) between the average values of the distributions and the statistical significance of these differences (p value) are reported.

Bond angle	$\langle \text{angle} \rangle \pm \sigma$		Difference	p value
	α -Helix (11480)	Coil (181)		
NC^αC	110.9 ± 1.2	111.6 ± 1.6	0.7	$2.9 * 10^{-14}$
C^αCO	120.6 ± 0.9	120.4 ± 0.9	-0.2	0.0051
$\text{C}^\alpha\text{CN}^{+1}$	116.9 ± 0.9	116.7 ± 1.1	-0.2	0.019
OCN^{+1}	122.5 ± 0.9	122.8 ± 0.9	0.3	$9.3 * 10^{-7}$
$\text{C}^{-1}\text{NC}^\alpha$	120.6 ± 1.1	120.7 ± 1.2	0.1	0.87
$\text{NC}^\alpha\text{C}^\beta$	110.6 ± 1.1	110.6 ± 1.1	0.0	0.81
$\text{C}^\beta\text{C}^\alpha\text{C}$	110.6 ± 1.3	110.2 ± 1.4	-0.4	$8.1 * 10^{-4}$

TABLE 3: Average values of bond angles with standard deviations and occurrences (in bracket) of non-Gly/non-Pro residues in β -structure or coil in the $15^\circ \times 15^\circ$ box centered at $(\varphi, \psi) = (-120^\circ, 130^\circ)$. The differences (coil minus β -sheet) between the average values of the distributions and the statistical significance of these differences (p value) are reported.

Bond angle	$\langle \text{angle} \rangle \pm \sigma$		Difference	p value
	β -Sheet (14076)	Coil (1968)		
NC^αC	108.8 ± 1.9	109.0 ± 2.2	0.2	$2.8 * 10^{-7}$
C^αCO	120.4 ± 0.9	120.4 ± 0.9	0.0	0.58
$\text{C}^\alpha\text{CN}^{+1}$	116.4 ± 1.1	116.6 ± 1.1	0.2	$8.4 * 10^{-16}$
OCN^{+1}	123.2 ± 0.9	123.0 ± 1.0	-0.2	$1.3 * 10^{-22}$
$\text{C}^{-1}\text{NC}^\alpha$	122.5 ± 1.3	122.6 ± 1.4	0.1	0.15
$\text{NC}^\alpha\text{C}^\beta$	110.9 ± 1.4	110.7 ± 1.4	-0.2	$3.1 * 10^{-13}$
$\text{C}^\beta\text{C}^\alpha\text{C}$	109.7 ± 1.5	109.7 ± 1.6	0.0	0.27

TABLE 4: Average values of bond angles with standard deviations and occurrences (in bracket) of non-Gly/non-Pro residues in β -structure in the $15^\circ \times 15^\circ$ boxes centered at $(\varphi, \psi) = (-120^\circ, 130^\circ)$ and $(\varphi, \psi) = (-60^\circ, 150^\circ)$. The differences between the average values of the distributions and the statistical significance of these differences (p value) are reported.

Bond angle	$\langle \text{angle} \rangle \pm \sigma$		Difference	p value
	$(\varphi, \psi) =$ $(-120^\circ, 130^\circ)$ (14076)	$(\varphi, \psi) =$ $(-60^\circ, 150^\circ)$ (680)		
NC^αC	108.8 ± 1.9	110.3 ± 1.8	1.5	$8.6 * 10^{-94}$
C^αCO	120.4 ± 0.9	121.1 ± 0.9	0.7	$2.0 * 10^{-86}$
$\text{C}^\alpha\text{CN}^{+1}$	116.4 ± 1.1	115.9 ± 1.1	-0.5	$2.0 * 10^{-23}$
OCN^{+1}	123.2 ± 0.9	122.9 ± 1.0	-0.3	$3.8 * 10^{-15}$
$\text{C}^{-1}\text{NC}^\alpha$	122.5 ± 1.3	120.6 ± 1.4	-1.9	$3.7 * 10^{-306}$
$\text{NC}^\alpha\text{C}^\beta$	110.9 ± 1.4	110.4 ± 1.4	-0.5	$3.1 * 10^{-26}$
$\text{C}^\beta\text{C}^\alpha\text{C}$	109.7 ± 1.5	109.5 ± 1.7	-0.2	0.0025

structure is limited to angles which involve atoms that are usually engaged in H-bonding interactions and, therefore, more susceptible to the structural context.

3.2. Backbone Variability: The Impact of Residues Type. The role of specific properties of residue side chains on the variability has been initially demonstrated by Touw and Vriend [28] and later confirmed by us [29] for the prototypical NC^αC angle. We here extended these analyses to the other valence bond angles of protein backbone. In this framework, to achieve statistically significant results, we considered the highly populated (φ, ψ) boxes for the helical and the extended regions described in the previous section. It is worth mentioning that this choice ensured the occurrence of at least 100 residues of each type in the two selected boxes. The inspection of Tables 5-6 and Figures 3(a)-3(g) clearly indicates that the valence bond angles characterized by a central atom endowed with sp^2 hybridization display a very limited dependence on the residue type. For proline, a specific value is observed for the angle $\text{C}^{-1}\text{NC}^\alpha$ in the helical box due to the cyclic nature of this residue (Table 5). A significantly lower value is also displayed by the same angle of Gly residues in the extended context (Table 6). A more significant impact of the residue type is occasionally observed for the valence bond angles centered at the C^α atom that is spatially close to the side chain. One clear trend is observed for the angle $\text{NC}^\alpha\text{C}^\beta$ which is systematically higher for the β -branched residues Val and Ile independently of the structural contexts (Tables 5 and 6). These latter residues also exhibit high values of the $\text{C}^\beta\text{C}^\alpha\text{C}$ angle, although the effect is evident only in the helical state. Since for these residues a decrease of the related NC^αC angle is observed due to steric effects of the branched side chain (Tables 5-6 and [28, 29]), the enlargement of the $\text{NC}^\alpha\text{C}^\beta$ and $\text{C}^\beta\text{C}^\alpha\text{C}$ may be a consequence of the NC^αC variability. It is worth mentioning that, even in the most sterically allowed and populated rotamer of β -branched residues (*trans* for Val and *gauche* for Ile), the two C^γ atoms are *gauche* to both N and C atoms of their own backbone (1-4 interactions).

The repulsive interactions between these groups likely produced slight displacement of the side chain with respect to the main chain. This causes an enlargement of the angles involving the C^β atom ($\text{NC}^\alpha\text{C}^\beta$ and $\text{C}^\beta\text{C}^\alpha\text{C}$) (Figure 4(a)). In addition to these interactions which are independent of backbone conformation, there are possible interactions displayed by the C^γ atoms due to the fact that they are in a five-atom chain (1-5 interactions) with heavy atoms whose position is determined by backbone conformation [33, 34] (Figures 4(b) and 4(c)). These (φ, ψ) -dependent interactions produce a slight repulsion between the C^γ atom and the O atom of the same residue in the preferred *trans* rotamer (experimental population 89%) of the α -helical conformation. This causes a further enlargement of the $\text{C}^\beta\text{C}^\alpha\text{C}$ angle in the helical conformation (Tables 5-6 and Figure 4(b)). This proximity between C^γ and O atoms does not occur in the extended conformation in the most preferred (*trans*) rotameric state (Figure 4(c)).

Other significant peculiarities are observed for the angles $\text{NC}^\alpha\text{C}^\beta$ and $\text{C}^\beta\text{C}^\alpha\text{C}$ of proline residues (Table 5), again

TABLE 5: Average values ($^{\circ}$) of the bond angles $\text{NC}^{\alpha}\text{C}$, $\text{NC}^{\alpha}\text{C}^{\beta}$, $\text{C}^{\beta}\text{C}^{\alpha}\text{C}$, $\text{C}^{\alpha}\text{CO}$, $\text{C}^{\alpha}\text{CN}^{+1}$, OCN^{+1} , and $\text{C}^{-1}\text{NC}^{\alpha}$ for each amino acid residue type in α -helix in the $3^{\circ} \times 3^{\circ}$ box centered at $(\phi, \psi) = (-63^{\circ}, -43^{\circ})$.

Residue	Occurrence	$\text{NC}^{\alpha}\text{C}$	$\text{NC}^{\alpha}\text{C}^{\beta}$	$\text{C}^{\beta}\text{C}^{\alpha}\text{C}$	$\text{C}^{\alpha}\text{CO}$	$\text{C}^{\alpha}\text{CN}^{+1}$	OCN^{+1}	$\text{C}^{-1}\text{NC}^{\alpha}$
Ala	2036	111.0	110.4	110.6	120.6	116.8	122.5	120.8
Leu	1434	111.1	110.6	110.5	120.5	117.0	122.4	120.6
Met	272	111.1	110.7	110.5	120.5	117.0	122.5	120.6
Ile	830	110.5	110.9	111.3	120.6	117.0	122.4	120.7
Val	851	110.4	111.1	111.2	120.6	117.1	122.3	120.6
Cys	122	111.0	110.6	110.3	120.4	117.0	122.5	120.6
Asn	294	111.3	110.6	110.3	120.7	116.8	122.5	120.4
Gln	650	111.1	110.6	110.4	120.5	116.9	122.5	120.6
Ser	429	111.2	110.3	110.3	120.6	116.8	122.5	120.7
Thr	491	110.8	110.4	110.9	120.6	116.9	122.5	120.8
Phe	350	110.8	110.5	110.4	120.7	116.9	122.4	120.4
Trp	183	111.0	110.3	110.8	120.7	117.0	122.3	120.7
Tyr	299	110.9	110.3	110.4	120.5	116.9	122.5	120.7
Asp	559	110.9	110.8	109.8	120.7	116.8	122.5	120.6
Glu	1002	111.1	110.5	110.4	120.6	116.8	122.5	120.7
His	208	111.3	110.4	110.3	120.7	116.8	122.5	120.4
Lys	798	110.8	110.6	110.5	120.6	116.9	122.4	120.7
Arg	746	111.0	110.7	110.3	120.6	116.9	122.5	120.5
Gly	459	112.1	—	—	120.8	116.6	122.5	120.7
Pro*	28	113.5	103.3	111.7	119.8	117.8	122.3	117.7

*The number of observations for Pro residues is very limited.

TABLE 6: Average values ($^{\circ}$) of the bond angles $\text{NC}^{\alpha}\text{C}$, $\text{NC}^{\alpha}\text{C}^{\beta}$, $\text{C}^{\beta}\text{C}^{\alpha}\text{C}$, $\text{C}^{\alpha}\text{CO}$, $\text{C}^{\alpha}\text{CN}^{+1}$, OCN^{+1} , and $\text{C}^{-1}\text{NC}^{\alpha}$ for each amino acid residue type in β -sheet in the $15^{\circ} \times 15^{\circ}$ box centered at $(\phi, \psi) = (-120^{\circ}, 130^{\circ})$. No Pro residues have been found in Data 1.6 for this box.

Residue	Occurrence	$\text{NC}^{\alpha}\text{C}$	$\text{NC}^{\alpha}\text{C}^{\beta}$	$\text{C}^{\beta}\text{C}^{\alpha}\text{C}$	$\text{C}^{\alpha}\text{CO}$	$\text{C}^{\alpha}\text{CN}^{+1}$	OCN^{+1}	$\text{C}^{-1}\text{NC}^{\alpha}$
Ala	559	108.9	110.7	110.0	120.4	116.3	123.3	122.4
Leu	1717	108.9	110.4	109.3	120.3	116.5	123.2	122.8
Met	252	109.0	110.6	109.9	120.4	116.3	123.3	122.4
Ile	2816	108.4	111.4	109.8	120.4	116.3	123.2	122.6
Val	3793	108.6	111.4	109.9	120.4	116.3	123.2	122.5
Cys	182	109.7	110.2	109.3	120.2	116.5	123.3	122.5
Asn	170	109.0	110.5	109.4	120.2	116.4	123.3	122.6
Gln	325	109.0	110.6	109.7	120.5	116.2	123.3	122.3
Ser	382	109.3	110.5	109.6	120.4	116.4	123.3	122.4
Thr	1369	109.0	110.9	109.8	120.4	116.2	123.3	122.4
Phe	670	109.0	110.5	109.3	120.3	116.4	123.2	122.8
Trp	202	108.7	110.6	109.1	120.4	116.4	123.2	123.0
Tyr	624	109.1	110.5	109.4	120.4	116.4	123.2	122.9
Asp	131	108.7	110.5	109.4	120.2	116.5	123.3	122.5
Glu	575	108.8	110.7	109.8	120.5	116.3	123.2	122.4
His	245	109.2	110.3	109.2	120.3	116.4	123.3	122.7
Lys	541	108.8	110.7	109.9	120.4	116.4	123.1	122.5
Arg	580	108.8	110.6	109.8	120.4	116.4	123.2	122.5
Gly	140	110.0	—	—	120.3	116.3	123.3	121.9

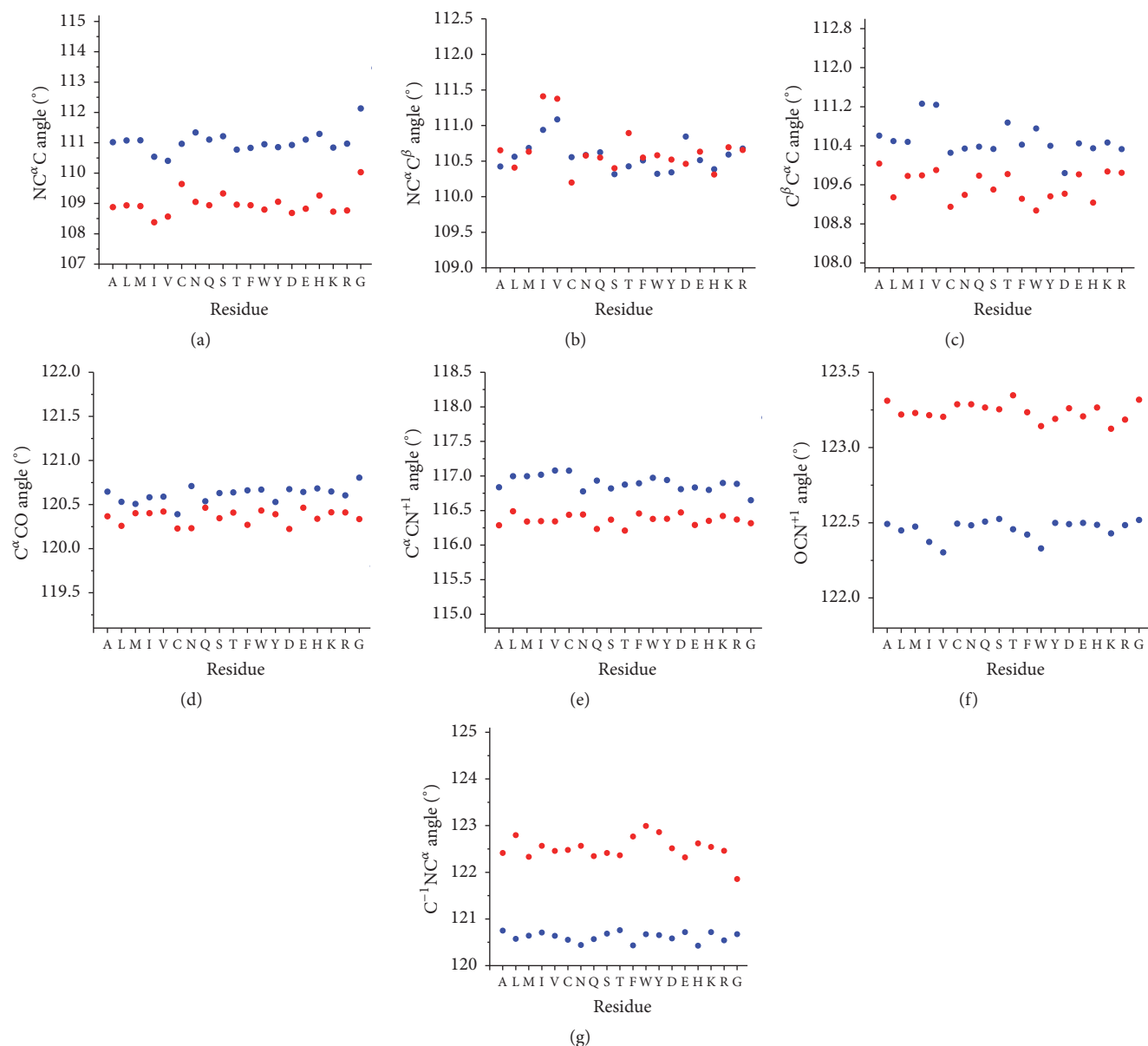


FIGURE 3: Average values of bond angles of protein residues in α -helix in the $3^\circ \times 3^\circ$ box centered at $(\varphi, \psi) = (-63^\circ, -43^\circ)$ (in blue) and in β -sheet in the $15^\circ \times 15^\circ$ box centered at $(\varphi, \psi) = (-120^\circ, 130^\circ)$ (in red) in the Ramachandran plot. Pro is not reported in the diagram since the observations for this residue are very limited.

ascrivable to its cyclic nature. Our analysis also highlights that the $C^\beta C^\alpha C$ tends to adopt low values for Asp residues in the helical context. The limited distortions may be due to the potential interaction that charged Asp side chains may form with the local backbone.

4. Conclusions

Proteins frequently assume complex three-dimensional structures characterized by marginal thermodynamic stabilities. Therefore, a full understanding of the principles underlying their folding requires a profound knowledge of all the aspects involved in this process. The variability

of several geometrical parameters of protein backbone has attracted much attention and it is believed to play a role in protein folding as well as in other contexts such as structure refinement and validation. Although the structural variability of several geometrical parameters of protein backbone has been well established, the role of the structural environment in determining/modulating these effects has been hitherto limited to the prototypical τ ($NC^\alpha C$) valence bond. We here extended the analysis of the peptide backbone geometry and planarity with the aim of gaining insights into the structural determinants of this variability. As expected, present statistical surveys confirm the remarkable variability of these parameters. Collectively, present findings

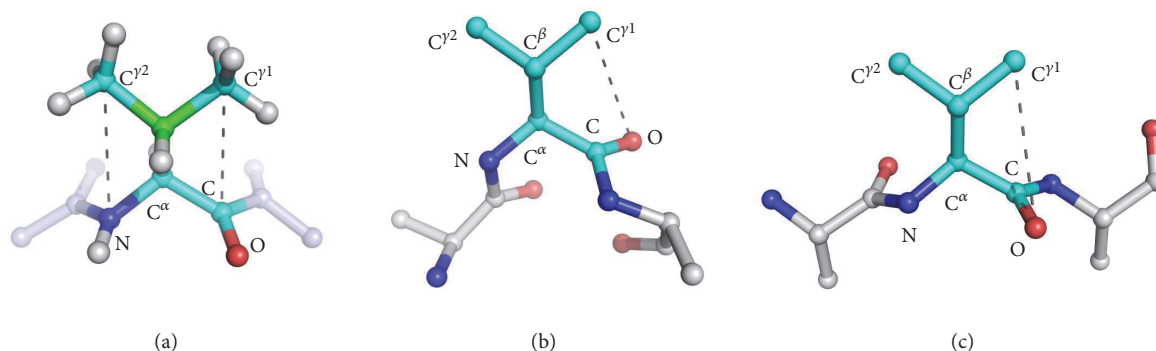


FIGURE 4: Local steric hindrance in β -branched residues. Ball-and-stick representations of the most populated rotamer of Val (used here as a representative model of a β -branched residue) are reported. (a) The backbone-independent repulsive interactions between the two heavy C^γ atoms and both backbone N and C atoms are shown as dashed gray lines; they can explain the widening of the $NC^\alpha C^\beta$ and $C^\beta C^\alpha C$ angles (the C^β atom is shown in green). (b) The backbone-dependent contact between $C^{\gamma 1}$ atom and the O backbone atom is shown (dashed gray lines) in the α -helical conformation ($\phi, \psi = (-63^\circ, -43^\circ)$). (c) In β -sheet conformation ($\phi, \psi = (-120^\circ, 130^\circ)$), the same atoms are on opposite sides. In both panels (b) and (c), the H atoms are omitted for clarity.

corroborate the notion that the contribution of (ϕ, ψ) angles overcomes the impact of the local structural motif. A significant contribution of secondary structure is limited to angles which involve atoms that are usually engaged in H-bonding interactions and, therefore, more susceptible to the structural context. In this scenario, it is not surprising that the highest dependence from the structural context is exhibited by the $OCN^{\alpha 1}$ angle.

Present data also show that the impact of the nature of the residues' side chain is marginal in most of the cases. However, we observe that, in addition to the impact of some side chains on $NC^\alpha C$ [28, 29], the values of angles such as $NC^\alpha C^\beta$ and $C^\beta C^\alpha C$ may depend on the nature of residue type. In particular, these angles tend to adopt larger values in the β -branched residues Val and Ile. This finding may be interpreted on the basis of steric effects generated by the simultaneous presence of the bulky groups that are linked to the C^β atom. It is worth mentioning that Thr, the other β -branched residue, has a distinct behavior. Evidently, the presence of an oxygen atom in the Thr side chain, which may establish H-bonding interactions with the local environment, has a significant impact on the geometry of this residue. Local H-bonding interactions likely cause the peculiar values observed for the $C^\beta C^\alpha C$ angle of Asp in helical contexts.

In conclusion, the rather tight association between conformation and geometry explains the high energetic costs associated with the swapping of geometries between different structural states. Moreover, our analysis further corroborates the necessity of considering the variability of protein backbone geometry in structure refinement, validation, and prediction.

Abbreviations

PDB: Protein Data Bank
 H: α -Helix
 G: 3(10)-Helix

E: β -Sheet
 τ : $NC^\alpha C$ bond angle.

Conflicts of Interest

The authors of the manuscript declare that there are no conflicts of interest.

Acknowledgments

Luigi Vitagliano would like to thank the Short Term Mobility (STM) program of CNR. The authors also thank Luca De Luca for skillful technical assistance.

References

- [1] K. A. Dill and J. L. MacCallum, "The protein-folding problem, 50 years on," *Science*, vol. 338, no. 6110, pp. 1042–1046, 2012.
- [2] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, "A backbone-based theory of protein folding," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 45, pp. 16623–16633, 2006.
- [3] R. Casadio, P. Fariselli, P. L. Martelli, and G. Tasco, "Thinking the impossible: How to solve the protein folding problem with and without homologous structures and more," *Methods in Molecular Biology*, vol. 350, pp. 305–320, 2006.
- [4] R. W. Newberry and R. T. Raines, "A prevalent intraresidue hydrogen bond stabilizes proteins," *Nature Chemical Biology*, 2016.
- [5] G. J. Bartlett, A. Choudhary, R. T. Raines, and D. N. Woolfson, " $n \rightarrow \pi^*$ interactions in proteins," *Nature Chemical Biology*, vol. 6, no. 8, pp. 615–620, 2010.
- [6] L. Esposito, N. Balasco, A. De Simone, R. Berisio, and L. Vitagliano, "Interplay between peptide bond geometrical parameters in nonglobular structural contexts," *BioMed Research International*, vol. 2013, Article ID 326914, 2013.
- [7] R. Improta, L. Vitagliano, and L. Esposito, "The determinants of bond angle variability in protein/peptide backbones: A comprehensive statistical/quantum mechanics analysis," *Proteins*:

- Structure, Function and Bioinformatics*, vol. 83, no. 11, pp. 1973–1986, 2015.
- [8] R. Improta, L. Vitagliano, and L. Esposito, “Bond distances in polypeptide backbones depend on the local conformation,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, pp. 1272–1283, 2015.
- [9] P. A. Karplus, “Experimentally observed conformation-dependent geometry and hidden strain in proteins,” *Protein Science*, vol. 5, no. 7, pp. 1406–1420, 1996.
- [10] L. Esposito, L. Vitagliano, A. Zagari, and L. Mazzarella, “Experimental evidence for the correlation of bond distances in peptide groups detected in ultrahigh-resolution protein structures,” *Protein Engineering*, vol. 13, no. 12, pp. 825–828, 2000.
- [11] L. Esposito, L. Vitagliano, A. Zagari, and L. Mazzarella, “Pyramidalization of backbone carbonyl carbon atoms in proteins,” *Protein Science*, vol. 9, no. 10, pp. 2038–2042, 2000.
- [12] L. Esposito, A. De Simone, A. Zagari, and L. Vitagliano, “Correlation between ω and ψ dihedral angles in protein structures,” *Journal of Molecular Biology*, vol. 347, no. 3, pp. 483–487, 2005.
- [13] R. Improta, L. Vitagliano, and L. Esposito, “Peptide bond distortions from planarity: New insights from quantum mechanical calculations and peptide/protein crystal structures,” *PLoS ONE*, vol. 6, no. 9, Article ID e24533, 2011.
- [14] D. S. Berkholz, M. V. Shapovalov, R. L. Dunbrack Jr., and P. A. Karplus, “Conformation Dependence of Backbone Geometry in Proteins,” *Structure*, vol. 17, no. 10, pp. 1316–1325, 2009.
- [15] M. W. MacArthur and J. M. Thornton, “Deviations from planarity of the peptide bond in peptides and proteins,” *Journal of Molecular Biology*, vol. 264, no. 5, pp. 1180–1195, 1996.
- [16] D. S. Berkholz, P. B. Krenesky, J. R. Davidson, and P. A. Karplus, “Protein Geometry Database: A flexible engine to explore backbone conformations and their relationships to covalent geometry,” *Nucleic Acids Research*, vol. 38, no. 1, Article ID gkp1013, pp. D320–D325, 2009.
- [17] D. E. Tronrud, D. S. Berkholz, and P. A. Karplus, “Using a conformation-dependent stereochemical library improves crystallographic refinement of proteins,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 7, pp. 834–842, 2010.
- [18] D. S. Berkholz, C. M. Driggers, M. V. Shapovalov, R. L. Dunbrack Jr., and P. A. Karplus, “Nonplanar peptide bonds in proteins are common and conserved but not biased toward active sites,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 2, pp. 449–453, 2012.
- [19] E. I. Howard, R. Sanishvili, R. E. Cachau et al., “Ultrahigh resolution drug design I: Details of interactions in human aldose reductase-inhibitor complex at 0.66 Å,” *Proteins: Structure, Function and Genetics*, vol. 55, no. 4, pp. 792–804, 2004.
- [20] P. A. Karplus, M. V. Shapovalov, R. L. Dunbrack, and D. S. Berkholz, “A forward-looking suggestion for resolving the stereochemical restraints debate: Ideal geometry functions,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 64, no. 3, pp. 335–336, 2008.
- [21] M. Jaskolski, M. Gilski, Z. Dauter, and A. Wlodawer, “Stereochemical restraints revisited: How accurate are refinement targets and how much should protein structures be allowed to deviate from them?” *Acta Crystallographica Section D: Biological Crystallography*, vol. 63, no. 5, Article ID wd5076, pp. 611–620, 2007.
- [22] I. J. Tickle, “Experimental determination of optimal root-mean-square deviations of macromolecular bond lengths and angles from their restrained ideal values,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 63, no. 12, pp. 1274–1281, 2007.
- [23] A. Q. Zhou, C. S. O’Hern, and L. Regan, “Revisiting the Ramachandran plot from a new angle,” *Protein Science*, vol. 20, no. 7, pp. 1166–1171, 2011.
- [24] D. Caballero, J. Määttä, A. Q. Zhou, M. Sammalkorpi, C. S. O’Hern, and L. Regan, “Intrinsic α -helical and β -sheet conformational preferences: A computational case study of Alanine,” *Protein Science*, vol. 23, no. 7, pp. 970–980, 2014.
- [25] P. Conway, M. D. Tyka, F. DiMaio, D. E. Konderding, and D. Baker, “Relaxation of backbone bond geometry improves protein energy landscape modeling,” *Protein Science*, vol. 23, no. 1, pp. 47–55, 2014.
- [26] P. Bradley, K. M. S. Misura, and D. Baker, “Toward high-resolution de novo structure prediction for small proteins,” *Science*, vol. 309, no. 5742, pp. 1868–1871, 2005.
- [27] S. M. Malathy Sony, K. Saraboji, N. Sukumar, and M. N. Ponnuswamy, “Role of amino acid properties to determine backbone τ (N-C α -C’) stretching angle in peptides and proteins,” *Biophysical Chemistry*, vol. 120, no. 1, pp. 24–31, 2006.
- [28] W. G. Touw and G. Vriend, “On the complexity of Engh and Huber refinement restraints: The angle τ as example,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 12, pp. 1341–1350, 2010.
- [29] N. Balasco, L. Esposito, and L. Vitagliano, “Factors affecting the amplitude of the τ angle in proteins: a revisitation,” *Acta Crystallographica*, vol. D73, pp. 618–625, 2017.
- [30] G. Wang and R. L. Dunbrack Jr., “PISCES: a protein sequence culling server,” *Bioinformatics*, vol. 19, no. 12, pp. 1589–1591, 2003.
- [31] W. Kabsch and C. Sander, “Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features,” *Biopolymers—Peptide Science Section*, vol. 22, no. 12, pp. 2577–2637, 1983.
- [32] R. A. Engh and R. Huber, “Accurate bond and angle parameters for X-ray protein structure refinement,” *Acta Crystallographica Section A*, vol. 47, no. 4, pp. 392–400, 1991.
- [33] R. L. Dunbrack and M. Karplus, “Conformational analysis of the backbone-dependent rotamer preferences of protein side-chains,” *Nature Structural Biology*, vol. 1, no. 5, pp. 334–340, 1994.
- [34] R. L. Dunbrack Jr., “Rotamer libraries in the 21st century,” *Current Opinion in Structural Biology*, vol. 12, no. 4, pp. 431–440, 2002.