THE PROTEIN SOCIETY

FOR THE RECORD

Geometry motivated alternative view on local protein backbone structures

Jan Zacharias and Ernst Walter Knapp*

Institute of Chemistry and Biochemistry, Freie Universität Berlin, D-14195 Berlin, Germany

Received 19 July 2013; Revised 28 August 2013; Accepted 29 August 2013
DOI: 10.1002/pro.2364
Published online 3 September 2013 proteinscience.org

Abstract: We present an alternative to the classical Ramachandran plot (R-plot) to display local protein backbone structure. Instead of the (φ, ψ)-backbone angles relating to the chemical architecture of polypeptides generic helical parameters are used. These are the rotation or twist angle ϑ and the helical rise parameter *d*. Plots with these parameters provide a different view on the nature of local protein backbone structures. It allows to display the local structures in polar (*d*, ϑ)-coordinates, which is not possible for an R-plot, where structural regimes connected by periodicity appear disconnected. But there are other advantages, like a clear discrimination of the handedness of a local structure, a larger spread of the different local structure domains—the latter can yield a better separation of different local secondary structure motives—and many more. Compared to the R-plot we are not aware of any major disadvantage to classify local polypeptide structures with the (*d*, ϑ)-plot, except that it requires some elementary computations. To facilitate usage of the new (*d*, ϑ)-plot for protein structures we provide a web application (http://agknapp.chemie.fu-berlin.de/secsass), which shows the (*d*, ϑ)-plot side-by-side with the R-plot.

Keywords: alternative Ramachandran plot; protein secondary structure; (*d*, ϑ)-plot; generic helix parameters; backbone conformation plot; protein geometry

Statement for Broader Audience

In the Ramachandran plot (R-plot) backbone torsion angle pairs (φ, ψ) are plotted for each residue in a two-dimensional quadratic (φ, ψ)-diagram. Since its introduction in 1963[1] the R-plot has entered the introductory chapters on structural biology in practically all biochemical textbooks.[2,3] It has become a widespread and versatile tool in research to characterize and analyze local protein structures and to make approximate conclusions about the distribution of secondary structure motives in proteins. To visualize protein secondary structures also quaternion-based diagrams[4] or even three-dimensional representations on a (φ, ψ)-torus[5] were suggested. But, despite some early attempts,[6] no systematic studies were made to find alternative and potentially better suited representations to identify local protein secondary structures by means of generic helical parameters.

The torsion angles φ and ψ used in the R-plot are connected with and depend on the specific local chemical architecture of the polypeptide backbone. From the values of these torsion angles, one can

Abbreviations: DSSP, dictionary of protein secondary structure: pattern; R-plot, Ramachandran plot; VMD, visual molecular dynamics

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: E. W. Knapp, Institute of Chemistry and Biochemistry, Freie Universität Berlin, D-14195 Berlin, Germany. E-mail: knapp@chemie.fu-berlin.de

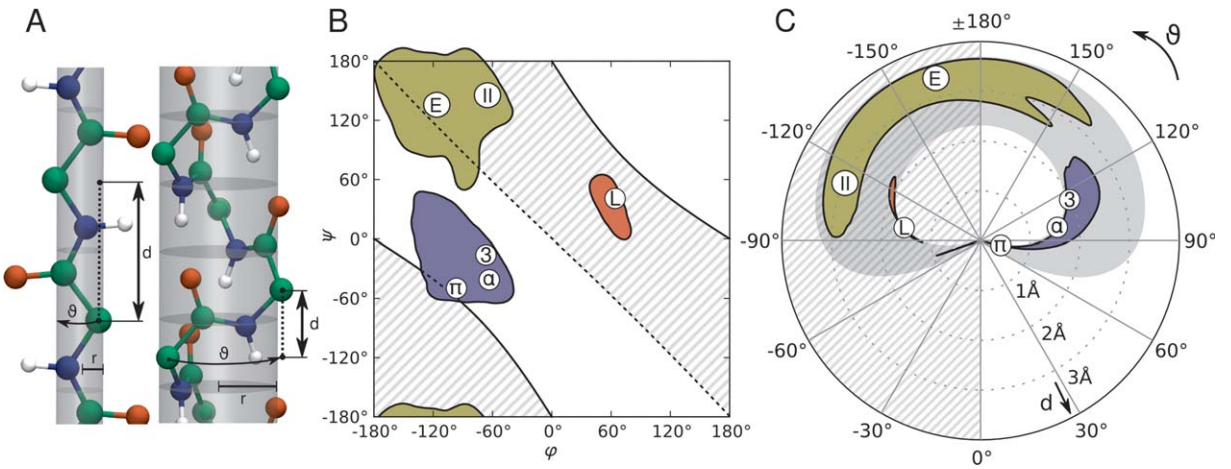

Figure 1. A: Definition of the three parameters (d, ϑ, r) characterizing helical geometry shown for an idealized strand (d = 3.7, ϑ = 180°, r = 0.5) and 3_{10}-helix (d = 2.0, ϑ= 115°, r = 2.0) structure, left and right part, respectively. B: Ramachandran (φ, ψ)-plot. Three disjoint easily accessible domains[8] are highlighted (green: strands; blue: helices; red: regime of left-handed conformations). The two disjoint hatched areas mark regimes of left-handed conformations. Separation lines between left and right-handed conformations are defined by $\vartheta = \pm 180°$ (dashed line, close to the diagonal) and d = 0 (solid lines), see text. Average locations of secondary structures are indicated: E: β-strand; II: polyproline II helix, 3, α, π: 3_{10}-, α-, π-helix; L: left-handed helical conformations. C: Polar (d, ϑ)-plot. The formally accessible region involving all (φ, ψ)-values are highlighted in gray. The hatched area marks left-handed conformations as in part B. Same notation as in part B is used for the easily accessible domains.

deduce secondary structure features of larger polypeptide patches assuming that the (φ, ψ) angles are approximately invariant for several consecutive residues. The latter is often the case for native protein structures. Interestingly, it is generally valid that a polypeptide backbone structure with repeated occurrence of the same (φ, ψ) angles is helical.[7] This is even true for the fully stretched polypeptide backbone conformation with (φ, ψ) = (180°, 180°), since this structure can be interpreted as a helix with rotation or twist angle of ϑ = 180° per residue and two residues per full turn (n = 2). The advantage to use (φ, ψ) torsion angles is that they can easily be derived from three-dimensional protein structure files. The disadvantage is that they do not provide an intuitive understanding of the precise nature of helical structures resulting from repeatedly occurring specific values of (φ, ψ) angles.

In the present contribution, we explore alternatives to the R-plot to identify and discriminate local protein structures. They make use of geometric parameters that are derived from the (φ, ψ) angles by a nonlinear transform. These plots can directly be used to characterize the geometries of the various helix-like structures arising for repeated values of the (φ, ψ) angles. The most useful plot scheme that we suggest in the present contribution uses the following two parameters measured per residue: d, rise along helix axis and ϑ, angle of rotation around helix axis. The advantage of the (d, ϑ)-plot is a simple separation between left and right turning helical structures, a clearer separation of different local secondary structure motifs, and a more direct understanding of the different local secondary structures occurring in native proteins.

The Ramachandran Plot

The R-plot in the space of the (φ, ψ) backbone torsion angles is shown in Figure 1(B) using the data about contours for the easily accessible regions.[8] These are three main domains (green: β-strands, blue: right-handed helices, red: left-handed helical conformations) stretched in a diagonal orientation from the upper left to the lower right, that is along straight lines fulfilling the relation $\varphi + \psi$ = const. This behavior was already mentioned by Ho *et al.*[9] and explained by the nature of H-bond interactions between backbone CO and NH groups and the avoidance of steric clashes between atoms of adjacent residues. Due to these steric clashes and clashes between backbone and side chain atoms, a large part of the right side of the R-plot is only rarely occupied under normal conditions[1] [Fig. 1(B)]. Hence, not all (φ, ψ) angles are energetically and sterically allowed and even if they are allowed, they do not necessarily appear repeatedly[10] to generate a visible helix-like structure.

Since glycine carries no side chain, it is more flexible and can appear also in the right half of the R-plot. On the other hand, the pyrrolidine ring of proline locks the φ backbone angle to approximately −60° and constrains the accessible (φ, ψ)-space of the neighbor residue toward the N-terminus [symbol II in Fig. 1(B,C)].

Repeated occurrences of approximately the same (φ, ψ) angles give rise to specific regular

polypeptide structures, that is secondary structures. The well-known α-helix structure predicted by Pauling *et al.*[11] as early as 1951 represents the largest uniform secondary structure motif, located around (φ, ψ) = (−63°, −41°) in the R-plot. A small cluster overlapping with the α-helices involves 3_{10}-helix motifs. The π-helix occurs only in very small patches mostly as a distortion of an α-helix with a relatively large variation in (φ, ψ) angles.[10,12–14] Therefore, its occurrence was generally underestimated[15,16] and occasionally even its existence was denied.[17] Residues in β-strand conformation form the second largest cluster of regular secondary structures (by number of residues), located around (φ, ψ) = (−119°, +136°).

Two disjoint hatched areas in Figure 1(B) mark the regime of left-handed conformations, while the non-hatched areas belong to the regime of right-handed conformations. However, on the surface of a torus in three-dimensions the areas belonging to the same handedness are connected. The computation of the boundaries between these regimes is explained below. The small cluster [red area in Fig. 1(B)] on the right side of the R-plot contains exclusively left-handed helix-like conformations. β-Strand structures can be either left- or right-handed [green area in Fig. 1(B)]. However, also a small part of the helix regime [blue area in Fig. 1(B)] involves left-handed conformations. Some of them are actually π-helix-like conformations. These three regions (helices, strands, left-handed helical conformations) correspond to secondary structure motifs that are usually defined by hydrogen-bond pattern, which are used by secondary structure assignment tools.[15,18] However, there are also a larger number of irregular local structures with practically no repeat in (φ, ψ) angles, which are spread also outside of the easily accessible regions.

Geometric Parameters Suitable to Characterize Helical Structures

Regular helix structures of molecules are generally described by three geometric parameters. These are rise per monomer (residue) along helix axis: d, radius of helix: r, and rotation angle per residue with respect to helix axis: ϑ. For a definition see Figure 1(A) and Supporting Information Figure S1. Alternatively to the angle ϑ, the number of residues per full turn n can be used, $n = 360°/\vartheta$ but n or d could also be replaced by $D = dn$, which is the rise per full turn. However, to characterize a helical conformation only two of the three parameters (d, ϑ, r) are necessary. Note that even the fully stretched all-*trans* conformation with (φ, ψ) = (180°, 180°) can be interpreted as a helix with rotation angle per residue of $\vartheta = \pm 180°$ and two residues per full turn ($n = 2$).

The helix parameters (d, ϑ, r) can be computed from the (φ, ψ) angles accounting for bond lengths and bond angles and the amide torsion angle ω of the polypeptide backbone. How one can convert the (φ, ψ) angles to the helix parameters (d, ϑ, r) was already demonstrated by Miyazawa[19] and Ramachandran and Sasisekharan.[20] Assuming idealized values of the bond lengths and bond angles taken from Ref. 21 and a perfectly planar amide plane (ω = 180°), the parameters d and ϑ can be computed from the following relations

$$\cos\left(\frac{\vartheta}{2}\right) = -0.8235\,\sin\left(\frac{\psi+\phi}{2}\right) - 0.0222\,\sin\left(\frac{\psi-\phi}{2}\right) \quad (1)$$

$$d\,\sin\left(\frac{\vartheta}{2}\right) = 2.999\,\cos\left(\frac{\psi+\phi}{2}\right) - 0.657\,\cos\left(\frac{\psi-\phi}{2}\right) \quad (2)$$

The numerical values in Eqs. (1) and (2) differ to some extent from the original values of Miyazawa[19] due to the more precise parameters of polypeptide bond geometry from Engh and Huber.[21,22] With the same assumptions also the helix radius r can be calculated from the rise per residue d and rotation angle per residue ϑ by simple geometric considerations (see Supporting Information Fig. S1) yielding

$$r = \sqrt{\frac{d_0^2 - d^2}{2 - 2\cos\vartheta}} \quad (3)$$

where d_0 = 3.8 Å is the distance of two consecutive C_α-atoms in a polypeptide chain. The precise dependence on bond lengths and bond angles of the Eqs. (1) and (2) is given in the Supporting Information.

In the Eqs. (1) and (2) the first term depending on $\varphi + \psi$ dominates. Hence, large variations in the (φ, ψ) torsion angles will have no significant effect on the values of d and ϑ, if the variations are anti-correlated keeping $\varphi + \psi$ = const. As a consequence, the larger variations of (φ, ψ) along the diagonal ($\varphi + \psi$ = const), observed in the R-plot mainly for the two helical regimes [blue and red areas in Fig. 1(B)], will have no significant effect on the nature of helix-type conformations. This correlates with the persistence of local helical structures observed along the diagonal $\varphi + \psi$ = const in the R-plot.

Inverting the non-linear transform, Eqs. (1) and (2), back to the (φ, ψ) torsion angles is a more difficult task, albeit it is not necessary, since the (φ, ψ) values are generally known as the primary data. On the other hand, not for all (d, ϑ) values corresponding (φ, ψ) values do exist [see Fig. 1(C)]. But, if (φ, ψ) values exist, there is always a pair of such values except for singular points. It is interesting to realize that due to this two-fold degeneracy, local α-helix structures with the same (d, ϑ) value can formally be generated by two different but very similar values of the (φ, ψ) backbone torsion angles (see Supporting Information Fig. S2). For strands, only

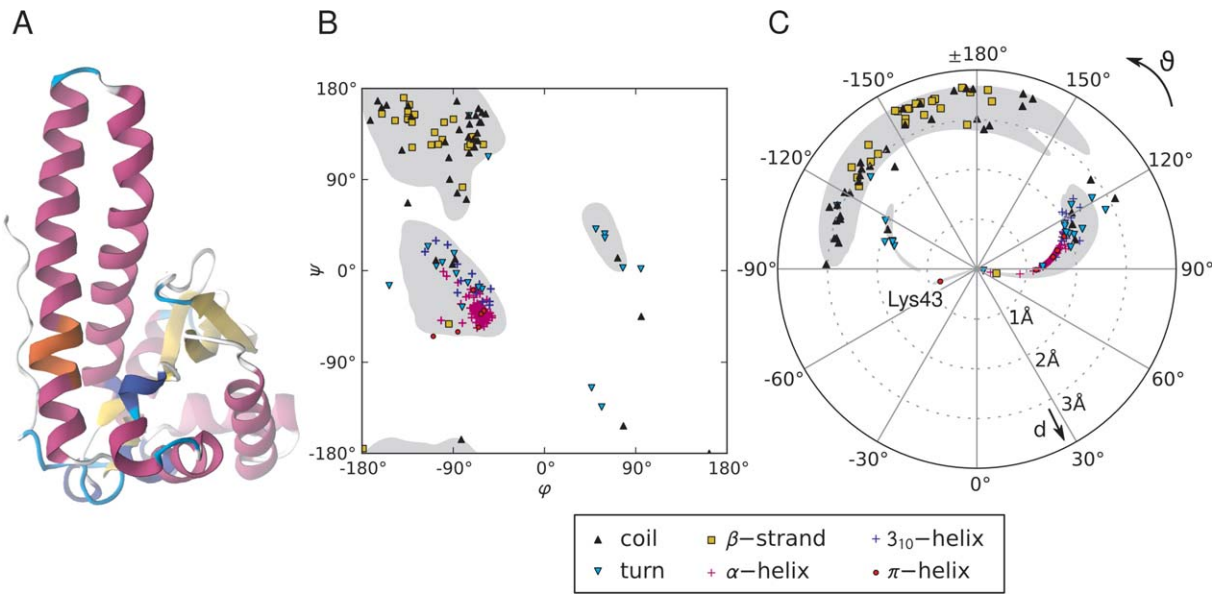

Figure 2. A: Crystal structure of chain A of human superoxide dismutase (PDB id: 1KKC:A). Structure was generated with VMD[26] with default colors: blue: 3_{10}-helix, purple: α-helix, red: π-helix, cyan: turn, light gray: coil. Secondary structure was assigned by DSSP[15] with the following changes: Residues belonging to the bend class were assigned to coil class; the β-bridge class was merged with the β-strand class; residues were assigned to the π-helix class when two or more π-helix-like H-bonds occurred successively. B: R-plot. C: (d, ϑ)-plot of the protein 1KKC:A.

one of these (φ, ψ) values belongs to the easily accessible regimes [Fig. 1(B), Supporting Information Fig. S2].

The Eqs. (1) and (2) have been used to draw contour lines in the R-plot either with constant d or ϑ.[23–25] The contour lines with $\vartheta = \pm 180°$ and $d = 0$, shown in [Fig. 1(B)], are particularly interesting, since they separate right- from left-handed helical conformations. Note that practically in all R-plots where contour lines are displayed they differ slightly from the present one, since they are based on old values of polypeptide bond geometries.[19] An R-plot with updated contour lines is shown in Supporting Information Figure S3.

Polar Plots with Generic Helix Parameters

The main focus of the present contribution is to advocate an alternative two-dimensional plot based on generic helix parameters that will complement the R-plot in the (φ, ψ)-space of backbone torsion angles. Since we need only two of the three helix parameters (d, ϑ, r) there are formally three possibilities: (d, ϑ), (r, ϑ), (d, r). Discriminating left- and right-handed structures can be achieved with ϑ [right-handed with $\vartheta \in (-180°, 0°)$; left-handed with $\vartheta \in (0°, +180°)$] or alternatively with d [right-handed with $d > 0$; left-handed with $d < 0$]. The (d, r)-plot shows a disconnected area for the β-strand region and requires to use also negative values of the rise parameter d (Supporting Information Fig. S4).

The other two possible combinations (d, ϑ) and (r, ϑ) involve an angle and a distance parameter, such that a two-dimensional polar plot is suitable to account for periodicity. In both cases, we discriminate between right- and left-handed helical conformations with the angle parameter ϑ. To account for periodicity in (φ, ψ)-space, polar coordinates would require to map the data on the surface of a torus in three-dimensional space, which is impractical.

Due to the intrinsic constraints of bond geometry only part of the areas in the polar plots are formally accessible. This is shown by the gray shaded area in the polar (d, ϑ)-plot, Figure 1(C). A Cartesian (d, ϑ)-plot where the helix regime is disconnected is shown in Supporting Information Figure S5. Note that mainly the upper half of the polar (d, ϑ)-plot belongs to the formally accessible area. In the polar (r, ϑ)-plot, the formally accessible area is smaller and the important strand structures have a small radius. They are therefore located close to the origin, where they occupy a small area only (Supporting Information Fig. S6). Hence, the most suitable set of helix parameters to identify local protein structures is (d, ϑ). An analogous polar plot was introduced by Perez and Vergelati[6] though they focused on its application to polysaccharides.

In the polar (d, ϑ)-plot, the left-handed conformations are found on the left side of the graph while right-handed conformations are on the right side [Fig. 1(C)]. This contrasts with the R-plot that shows two disconnected areas for left and right-handed conformations. In the (d, ϑ)-plot we see the large variation in the twist angle ϑ for strand-like conformations [green area in Fig. 1(C), where ϑ covers an interval of more than 130° mostly extending in

the left-handed part]. On the other hand, the right-handed helix area (blue) is more compact (ϑ covers an interval of only 45°), since it is structurally more constrained by intra-helix H-bond pattern. Interestingly, there is a narrow branch attached to the right-handed helix area pointing toward the origin of the (d, ϑ)-plot and continuing even in the regime of left-handed conformations. This branch involves short motifs of relatively flat spiral-like structures, which can be left- or right-handed. The right-handed part of this branch is populated by patches of π-helices (distributed over larger parts of the branch) and other spiral-like conformations with large radius r and small rise d.

Additional information is obtained from a scatter plot of a set of 50 representative protein crystal structures shown in the (d, ϑ)- and R-plot mode (Supporting Information Fig. S7). Based on secondary structure definition of DSSP,[15,16] this (d, ϑ)-plot shows that for $\vartheta \in (-150°, -180°) \cup (+180°, +165°)$ and $d \in (2.8, 3.6$ Å) the local conformations are mainly β-strands. For $\vartheta \in (-90°, -130°)$ regardless of the d value the conformations are mainly coil- or turn-like structures. In fact, backbone conformations with a twist angle ϑ deviating to much from $\pm 180°$ are not able to be part of a β-sheet stabilized by the typical H-bond pattern and are therefore classified as being coil-like by DSSP.[15,16] Also the left-handed helical conformations with $d \in (1.0, 2.2$ Å) and $\vartheta \in (-90°, -130°)$ are nearly exclusively assigned to turn structures [Supporting Information Fig. S7(B)]. Such a clear separation of different local structures is not so easily possible in the R-plot [Supporting Information Fig. S7(A)]. The helix regime shows a strong correlation between d and ϑ values and is located at the inner edge (small d values) of the formally allowed regime in the (d, ϑ)-space, which is due to the geometric constraints from the helical H-bond pattern.

To demonstrate the usefulness of the (d, ϑ)-plot for an individual protein, we consider the crystal structure of a manganese superoxide dismutase [PDB code 1KKC:A, Fig. 2(A)]. This protein involves all three helix types α-, 3_{10}-, and π-helix as well as a three-stranded β-sheet. The π-helix is a pronounced disturbance of a longer α-helix [marked in red in Fig. 2(A)]. Most residues are located in the easily accessible regions of the (d, ϑ)-plot [gray shaded area, in Fig. 2(B,C)]. Although Lys43 belongs to the π-helix, which is part of a larger right-handed α-helix, it surprisingly adopts locally a left-handed conformation (isolated red point in the lower left of Figure 2(C).

Summary

The proposed polar (d, ϑ)-plot displays the local secondary structure of proteins in terms of generic helical parameters measuring the rotation or twist angle ϑ and rise d per residue relative to the axis of the helical structure, which forms if the local structure with specific (φ, ψ) backbone torsion angles occurs repeatedly. The (d, ϑ)-plot offers several advantages: (1) The proximity of local structures lying on opposite border regions of the R-plot is no longer masked. (2) One obtains direct and therefore easier interpretable information of the nature of the local structures. (3) The handedness of a local structure can be read immediately from the (d, ϑ)-plot, which displays left-handed conformations on the left side and right handed conformations on the right side, while this discrimination is not so easily possible with the R-plot. (4) The (d, ϑ)-plot also provides a more realistic spread of the three different easily accessible structure domains (strand, helix, and left-handed helix regime). Only in the (d, ϑ)-plot one can observe that the strand domain shows a large variation in the twist angle ϑ, while the rise per residue d is large but varies in a rather narrow interval (2.8–3.6 Å) reflecting the relatively stretched nature of these conformations. The opposite is the case for the helix domain, which comprises a relatively small interval of the angle ϑ, while the rise parameter varies practically from 0.0 to 2.5 Å exhibiting a strong correlation between d and ϑ values. The low populated left-handed helix region consists of a narrow branch but is otherwise practically a mirror image to the right-handed normal helical region.

To facilitate usage of (d, ϑ)-plots for protein structures, we have prepared an easy to use webapplication (http://agknapp.chemie.fu-berlin.de/secsass), where one can submit coordinates in PDB format or just the PDB id of a protein and obtains the polar (d, ϑ)-plot and the R-plot side-by-side for display and download. The web plots have additional functionalities. They allow for instance to load only the residues belonging to α-helices or strands. Pointing on individual spots of either the R- or the (d, ϑ)-plot one obtains specific information about the corresponding residue.

References

1. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99.
2. Lehninger AL (1978) Biochemistry: the molecular basis of cell structure and function, 2nd ed. New York: Worth Publishers.
3. Voet DV (2011) Biochemistry. Hoboken, NJ: Wiley.
4. Hanson RM, Kohler D, Braun SG (2011) Quaternion-based definition of protein secondary structure straightness and its relationship to Ramachandran angles. Proteins 79:2172–2180.
5. Boomsma W, Mardia KV, Taylor CC, Ferkinghoff-Borg J, Krogh A, Hamelryck T (2008) A generative, probabilistic model of local protein structure. Proc Natl Acad Sci USA 105:8932–8937.
6. Perez S, Vergelati C (1985) Unified representation of helical parameters: application to polysaccharides. Biopolymers 24:1809–1822.

7. Schellman JA, The conformation of polypeptide chains in proteins. In: Neurath H, Ed (1964) The proteins composition, structure, and function. New York: Academic Press, pp 1–137.
8. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50:437–450.
9. Ho BK, Thomas A, Brasseur R (2003) Revisiting the Ramachandran plot: hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. Protein Sci 12:2508–2522.
10. Hollingsworth SA, Berkholz DS, Karplus PA (2009) On the occurrence of linear groups in proteins. Protein Sci 18:1321–1325.
11. Pauling L, Corey RB, Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl Acad Sci USA 37: 205–211.
12. Fodje MN, Al-Karadaghi S (2002) Occurrence, conformational features and amino acid propensities for the pi-helix. Protein Eng 15:353–358.
13. Cooley RB, Arp DJ, Karplus PA (2010) Evolutionary origin of a secondary structure: π-helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. J Mol Biol 404: 232–246.
14. Riek RP, Graham RM (2011) The elusive π-helix. J Struct Biol 173:153–160.
15. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.
16. Weaver TM (2000) The pi-helix translates structure into function. Protein Sci 9:201–206.
17. Creighton TE (1992) Proteins: structures and molecular properties, 2nd ed. New York: Freeman WH.
18. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23:566–579.
19. Miyazawa T (1961) Molecular vibrations and structure of high polymers. II. Helical parameters of infinite polymer chains as functions of bond lengths, bond angles, and internal rotation angles. J Polym Sci 55:215–231.
20. Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Prot Chem 23: 283–438.
21. Engh RA, Huber R, Structure quality and target parameters. In: Rossmann MG, Arnold E, Eds (2001). International tables for crystallography, Volume F: Crystallography of biological macromolecules SE-42. The Netherlands: Springer. pp 382–392.
22. Engh RA, Huber R (1991) Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 47:392–400.
23. Dickerson RE, Geis I (1969) The structure and action of proteins. New York: Harper & Row.
24. Hollingsworth SA, Karplus PA (2010) A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol Concepts 1:271–283.
25. Garrett RH, Grisham CM (2012) Biochemistry, 5th ed. Belmont, CA: Brooks/Cole, Cengage Learning.
26. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graphics 14:33–38.