Dihedral angle preferences of amino acid residues forming various non-local interactions in proteins

Konda Mani Saravanan¹ · Samuel Selvaraj^{1,2}

Received: 13 July 2016 / Accepted: 13 April 2017 / Published online: 2 June 2017
© Springer Science+Business Media Dordrecht 2017

Abstract In theory, a polypeptide chain can adopt a vast number of conformations, each corresponding to a set of backbone rotation angles. Many of these conformations are excluded due to steric overlaps. Ramachandran and coworkers were the first to look into this problem by plotting backbone dihedral angles in a two-dimensional plot. The conformational space in the Ramachandran map is further refined by considering the energetic contributions of various non-bonded interactions. Alternatively, the conformation adopted by a polypeptide chain may also be examined by investigating interactions between the residues. Since the Ramachandran map essentially focuses on local interactions (residues closer in sequence), out of interest, we have analyzed the dihedral angle preferences of residues that make non-local interactions (residues far away in sequence and closer in space) in the folded structures of proteins. The non-local interactions have been grouped into different types such as hydrogen bond, van der Waals interactions between hydrophobic groups, ion pairs (salt bridges), and $\pi\pi$ -stacking interactions. The results show the propensity of amino acid residues in proteins forming local and non-local interactions. Our results point to the vital role of different types of non-local interactions and their effect on dihedral angles in forming secondary and tertiary structural elements to adopt their native fold.

Keywords Protein folding · Local and non-local interactions · Conformational space · Inter-residue interactions · Dihedral angle preferences

✉ Samuel Selvaraj
selvarajsamuel@gmail.com

¹ Centre of Advanced Study in Crystallography & Biophysics, University of Madras, Guindy Campus, Chennai, Tamil Nadu 600 025, India

² Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli, Tamil Nadu 620024, India

1 Introduction

An understanding of protein structure is essential in solving the protein folding problem, which is one of the most important problems in computational molecular biology. The Ramachandran map [1, 2] is a powerful tool for the conformational analysis of polypeptide chain backbones in protein structures, where each residue of the protein is structurally characterized by a set of φ and ψ angles. The Ramachandran map is widely used to judge the stereo-chemical quality of the structures solved by experimental techniques as well as protein models built by predictive methods by analyzing dihedral angles in allowed or disallowed regions [3–6]. Extensive studies have been carried out on the backbone and side-chain conformations of amino acids in folded native proteins, unfolded non-native proteins, and peptides [7–16], with the help of the Ramachandran map. The folded native state of proteins is studied by looking at the inter-residue interactions between the residues in a protein. For the past four decades, this concept has been one of the main focuses of protein scientists to understand the mechanism of protein folding and stability.

Tanaka and Scheraga [17] proposed a three-step mechanism for the protein-folding problem based on inter-residue interactions: (i) formation of ordered backbone structures by local interactions (ii) formation of small contact regions by medium-range interactions, and (iii) association of the small contact regions into native structure by non-local interactions. Following Scheraga's work, Gromiha and Selvaraj classified these inter-residue interactions based on sequential and spatial separation of amino acids residues present in the protein chain and made an extensive study on globular proteins [18, 19]. Further, the authors brought out the importance of local and non-local interactions in four major structural classes such as all- α , all- β , $\alpha + \beta$, and α/β , respectively [20, 21] and they stated that the average number of residue contacts due to local and non-local interactions seems to balance each other in such a way to adopt their fold and stability. Also, they developed the measure “long-range order” (LRO) based on non-local interactions to predict folding and unfolding rates of proteins [22–24]. Many protein-engineering groups have shown that the balance between local and non-local interactions could alter a protein's conformation [25–29].

A careful analysis of the literature suggests that the conformation of a residue in a protein is largely determined by local interactions [30–33] and the dihedral angles of the residue are insensitive to its neighbor [34]. For example, in their classical textbook, Cantor and Schimmel stated that the residues in a protein are “sterically independent of rotations within neighboring residues” and hence the Ramachandran map is examined mostly based on local interactions [35]. Since the backbone conformation of a residue in a protein is fully determined by its corresponding dihedral angles, many analyses were carried out using these dihedral angles for native and denatured proteins [36, 37]. Porter and Rose showed that along with steric and hydrogen bonding constraints, the conformational space of a residue in a protein expands under folding conditions [38].

Several studies based on inter-residue interactions were undertaken to understand protein folding and stability, and no study was reported on the dihedral angle preferences of the amino acid residues contributing to non-local interactions. Although hydrogen bonding between pairs of residues in helices or strands would be responsible for forming the compact globular structure, it is also hypothesized that the residues that are far away in the sequence come closer through non-local interactions during the process of chain folding. Since protein structures are the end point of the protein folding pathway [39], it is of interest to analyze the dihedral angle preferences of various non-local interactions like van der Waals interactions

between hydrophobic residues, ion pair and $\pi\pi$ -stacking interactions, respectively. Considering the above facts, and the availability of a large number of structures in the Protein Data Bank, prompted us to conduct a comprehensive analysis of residues involved in local and non-local interactions.

2 Materials and methods

2.1 Dataset

The pre-compiled dataset derived from the latest release of Protein Data Bank (PDB) is available in the ccPDB database [40]. The dataset contains 5877 non-redundant protein chains and each residue in the dataset is assigned to regular secondary structure (Helix or Sheet) and solved experimentally at high resolution (less than 3 Å resolution). DSSP and PDB_Select were used to derive the dataset and the protein chains in the data set share less than 30% sequence identity. A brief description of the dataset is given at the ccPDB web page. The co-ordinates of the 5877 protein chains are obtained and kept in a suitable format by using in-house PERL programs to do computations.

2.2 Computation of non-local interactions

The non-local interactions in a protein are computed by using the procedure described in our earlier works [18, 41]. Representing each residue in the protein molecule by its C α atom and fixing the first residue as the center, the distances between this atom and the rest of the C α atoms in the protein molecules are computed. The residues appearing around the reference residue within a sphere of 8 Å radius are determined. From the computed surrounding residues, those contributing to non-local interactions (C α atom $> \pm 4$ residues) were computed [19, 42]. We also computed the number of non-local interactions for each and every residue in the 5877 non-redundant protein chains. The residues are grouped into two types, those with one non-local interaction and those with ten or more non-local interactions. Since the number of residues with 11 non-local interactions is too small to make a comparative analysis, we considered the residues with ten or more non-local interactions in our study.

2.3 Computation of different types of non-local interactions

In order to study the interaction specific dihedral angle preferences of amino acid residues contributing to different types of non-local interactions, we computed hydrogen bonding, van der Waals interactions between hydrophobic groups, salt bridges (ion pairs), $\pi\pi$ -stacking interactions. The van der Waals interactions between hydrophobic groups have been computed following a similar procedure described above to compute non-local interactions by using a distance cut off as 5 Å instead of 8 Å [43]. The amino acid residues ala, val, leu, ile, met, phe, trp, pro, and tyr, respectively, are considered as hydrophobic amino acid residues [44]. To compute ion pair interactions, ionic amino acid residue pairs (arg, lys, his, asp, and glu) falling within 6 Å are computed [44]. In the case of computing $\pi\pi$ -stacking interactions, a pair of phenyl ring centroids separated by a distance between 4.5 Å and 7 Å is used [45]. The distance criteria used to compute hydrogen bond between donor and acceptor atoms (oxygen and nitrogen) is 3.5 Å, whereas a 4 Å distance cutoff is used to compute hydrogen bonds between

sulfur atoms [46]. All of the above distance cutoffs used in the present work have been followed from the Protein Interaction Calculator [44].

2.4 Computation of dihedral angles

The dihedral angles (φ , ψ) are computed by using the executable file obtained from Dr. Andrew CR Martin's group via the web interface at www.bioinf.org.uk. A PERL program is coded to compute dihedral angles for all 5877 protein chains in a single run. Since the program uses a value of “-9999” to indicate that no angle is found for some residues, we neglected such residues from the data set.

2.5 Construction of the Ramachandran map

The dihedral angles of residues that have one non-local interaction, ten or more non-local interactions, and four different types of non-local interactions are used to plot the Ramachandran map. The percentage distribution of dihedral angles of amino acid residues at each 30° bins from -180° to +180° is computed. Grouping of dihedral angles of amino acid residues with local and non-local interactions and grouping of residues with different types of non-local interactions are done by using in-house PERL scripts. The Ramachandran map is plotted by using the open-source code of GNUPLOT available from <http://www.gnuplot.info/>. Further, the propensity of amino acid residues contributing to four different types of non-local interactions in predefined secondary structure space such as alpha helix, beta sheet, polyproline II, and turns in the Ramachandran map is computed. The definition of the above four secondary structural elements in the Ramachandran map is adopted from existing literature [47–51].

3 Results

3.1 Dihedral angle preferences of amino acid residues with minimal and maximum non-local interactions

The Ramachandran maps of amino acid residues with one non-local interaction (A) and with ten or more non-local interactions (B) are shown in Fig. 1. The ranges of allowed regions are shown as red color lines. Most of the amino acid residues (92.03%) in the non-redundant dataset with minimal and maximum non-local interactions lie in the allowed region of the Ramachandran map. Some residues are scattered in the disallowed regions of the Ramachandran map and these residues could be found at the edges of helices or strands and involve in weak stabilizing interactions [52]. The dihedral angles of amino acid residues with one non-local interaction are mostly populated as a dense cluster in the conformational space of helix regions (-60°, -40°) and relatively sparse in the conformational space of extended regions of the Ramachandran map (Fig. 1a). In contrast, the dihedral angles of residues with ten or more non-local interactions form a dense cluster in the extended region of the Ramachandran map (Fig. 1a) and are mainly populated in the region of φ and ψ near -120°, 140°.

The approximate dihedral angles for regular structures like different types of helices and strands are given in a table of the textbook by Cantor and Schimmel [35]. The residues with

Fig. 1 Ramachandran map of residues with one non-local interaction (**a**) and with ten or more non-local interactions (**b**)

one non-local interaction occupy a conformational space as a small scattered cluster in the extended region of the Ramachandran map at φ , ψ near -70° , $+150^\circ$, where those with maximum non-local interactions occupy a conformational space as a small scattered cluster in the right-handed alpha helical region of the Ramachandran map near -60° , -40° . Some amino acid residues with one non-local interaction and ten or more non-local interactions are also populated in the conformational space of left-handed alpha helices ($+57^\circ$, $+47^\circ$) in the Ramachandran map. Two small clusters fall in the allowed region (not in the favored region) of the Ramachandran map near 30° , -30° and -60° , 30° , respectively, whereas no such small clusters are found in the Ramachandran map of amino acid residues with maximum non-local interactions. Other residues form a small cluster at the helical region near -60° , -50° . We believe that these small clusters of residues may appear in a specific context of secondary structure elements (like residues immediately before/after a certain secondary structural element, or connecting two secondary structures in a specific way).

The percentage occurrences of φ angles of residues with minimal and maximum non-local interactions at each 30° interval of -180° to 180° are shown in Fig. 2a. φ angles of residues with one non-local interaction lie mostly in -89° to -60° (53.81%) where those with ten or more non-local interactions lie mostly in three intervals such as -149° to -120° (30.01%), -119° to -90° (33.01%) and -89° to -60° (23.90%), respectively. The percentage of occurrences of residues with ψ angles at each 30° interval of -180° to 180° is shown in Fig. 2b. ψ angles of residues with minimal non-local interactions lie mostly in -59° to -30° (40.87% of residues) where those with ten or more non-local interactions (about 45% of residues) lie mostly in 121° to 150° . From our observation, we found that the φ angles of the residues with maximum non-local interactions shift from $(-89^\circ$ to $-60^\circ)$ to $(-119^\circ$ to -90° and -149° to $-120^\circ)$ as that of residues with minimal non-local interactions where ψ angles of the residues with maximum non-local interactions shift from $(-59^\circ$ to $30^\circ)$ to $(121^\circ$ to $150^\circ)$. From Fig. 2, φ mostly prefers negative angular values where ψ prefer to lie in positive angular values on the Ramachandran map.

3.2 Dihedral angle preferences of amino acid residues contributing to different types of non-local interactions

A review of the relevant literature reveals that the backbone interactions determine the secondary structure formation and backbone dihedral angles are important structural

Fig. 2 Percentage of occurrences of phi (a) and psi (b) angles in different bins

parameters which represent the overall topology of proteins [7, 31, 35]. However, we believe other interactions are also crucial for protein folding and hence we have grouped the amino acid residues contributing different types of non-local interactions such as hydrogen bonding, van der Waals between hydrophobic groups, ion pair interactions, and $\pi\pi$ -stacking interactions. The percentage of occurrences of (φ , ψ) angles of amino acid residues contributing to four different major non-local interactions in each 30° interval are shown in Fig. 3. φ angles of amino acid residues contributing to van der Waals interactions between hydrophobic residues and $\pi\pi$ -stacking interactions prefer a higher percentage of occurrences in -119° to -90° whereas φ angles of amino acid residues contributing to ion pair interactions and hydrogen bonding interactions prefer a higher percentage of occurrences in -89° to -60° interval. Interestingly, the ψ angles of amino acid residues contributing to hydrogen bonding interactions prefer a higher percentage of occurrences in -59° to -30° (negative bin) where ψ angles of amino acid residues contributing to the other three non-local interactions prefer a higher percentage of occurrences in 121° to 150° (positive bin). An opposite trend was observed between φ and ψ angles of amino acid residues contributing to non-local interactions, except for non-local hydrogen bonding interactions, which is shown in Fig. 3.

The Ramachandran map of amino acid residues in proteins contributing to four different non-local interactions are shown in Fig. 4. In Fig. 4, the Ramachandran map at the center represents dihedral angles of all the residues in the dataset (i.e., dihedral angles of amino acid residues in proteins contributing to both local and non-local

Fig. 3 Percentage of occurrences of phi and psi angles of amino acid residues contributing to different types of non-local interactions in different bins

interactions). Favored regions of the Ramachandran map are shown in red lines, whereas partially allowed regions are shown in blue lines. The amino acid residues contributing to the four different types of non-local interactions prefer to occupy the extended region (beta sheet) than the right-handed helical region in the Ramachandran map. The dihedral angle preferences of amino acid residues contributing to hydrogen bonding (Ramachandran map in the bottom of Fig. 4) and amino acid residues contributing to both local and non-local interactions (Ramachandran map in the center of Fig. 4) seem to be similar. This is due to the fact that among the four types of non-local interactions responsible for folding and stability of proteins, hydrogen bonds are distinct in having their origin in electron delocalization, which is clearly revealed in a paper by Bartlett et al. [53]. A small cluster in the region (30° , -45°) is absent in the Ramachandran map of non-local $\pi\pi$ -stacking interactions, whereas it is present in the other three Ramachandran maps. Another cluster in the region (30° , -150°) is present in the Ramachandran map of hydrogen bonding where it is absent in the other three Ramachandran maps in Fig. 4.

The amino acid residues contributing to hydrogen bonds prefer maximum to occupy the helical region (0.66) of Ramachandran map, whereas those residues contributing to van der Waals hydrophobic (0.30), ion pairs (0.08), and $\pi\pi$ -stacking interactions (0.0143), respectively, prefer to occupy beta-sheet regions of the Ramachandran map. The increase in beta-sheet

Fig. 4 Ramachandran maps of amino acid residues contributing to different types of non-local interactions. The central panel of the figure represents the Ramachandran map of all residues in the dataset contributing to local and non-local interactions

propensities caused by non-local interactions is not surprising because it is reported as a local order promoting effect in intrinsically disordered proteins [54, 55].

Interestingly, the amino acid residues contributing to van der Waals hydrophobic interactions prefer to occupy polyproline II space in the Ramachandran map than alpha helical space [56, 57]. The statistical propensities evaluated for amino acid residues contributing to hydrogen bonding and van der Waals hydrophobic interactions occupy polyproline II space on the Ramachandran maps were compared and a minimal propensity difference (0.17) was observed. Similarly, the amino acid residues contributing to ion pair interactions have low propensity difference (0.02) in adopting beta-sheet and polyproline II conformation. This implies that polyproline II helices and beta sheets may be able to convert into one another because polyproline II occupy a conformational space in the Ramachandran map, which is contiguous to that of the beta sheets [50]. Also, polyproline II is not an overall conformation of unfolded peptides and proteins, whereas it is one of many local conformations [58].

Turn regions of the proteins help a polypeptide chain to fold back into a globular manner rather than linear. About 50% (propensity is 0.5) of amino acid residues contributing to hydrogen bonding interactions prefer to occupy a conformational space defined as turns in the Ramachandran map (Fig. 5). The amino acid residues contributing to van der Waals hydrophobic, ion pairs, and $\pi\pi$ -stacking interactions have propensities such as 0.10, 0.04, and 0.0035. The role of non-local interactions in forming certain “native-like topology” in the denatured state of proteins results in turns and unstable helical and strand structures and these amino acid residues display different conformational preferences [59, 60]. The high propensity of conformational preferences of amino acid residues in turn regions are found and it is because conformational ensembles of amino acids in an unfolded state are varied, unique, and unexplainable in the context of pure steric interaction [61].

The quantitative comparison of dihedral angle distributions of amino acid residues affected by non-local interactions is shown in Table 1. Table 1 shows the percentage of amino acid residues contributing to four different types of important non-local interactions. About 35% of amino acid residues contribute to the non-local hydrogen bonding interactions, whereas 7.65,

Fig. 5 Secondary structural propensity of amino acid residues contributing to four important non-local interactions

Table 1 Percentage of amino acid residues participating in local and non-local interactions

Type	Interactions	Percentage of amino acid residues
Non-local	Hydrogen bond	35.03
	van der Waals hydrophobic	7.65
	Ion pairs	2.71
	III stacking	0.36
	Total	45.75
Local interactions/other		54.25

2.75, and 0.36% of amino acid residues contribute to van der Waals hydrophobic, ion pair, and $\pi\pi$ -stacking interactions, respectively. It is observed that 45.75% amino acid residues participate in four important stabilizing non-local interactions and 54.25% of amino acid residues participate in other/local interactions, which are clearly revealed in Table 1. The balance between the contribution of amino acids to local and nonlocal interactions is observed from the present analysis.

4 Discussion

The conformation adopted by a residue in a protein structure depends upon various factors such as secondary and tertiary interactions [62, 63]. The relative contribution of local and non-local interactions plays an important role in folding and stability of proteins [64] and a more detailed explanation about origin of studies on local and non-local interactions were reviewed in the literature [41, 65]. The conformational properties of residues in a protein provide important clues about their fold and topology. Dihedral angles of amino acid residues are one of the important conformational properties to define the complex three-dimensional structure of protein. However, the dihedral angles of residues in a protein mainly depend on local interactions [35]; these dihedral angles are very important in protein tertiary prediction [66, 67]. Also, the dihedral angles of a residue define the backbone of a protein; they almost represent the entire protein's topology. Although several studies on dihedral angle preferences of each amino acid type were presented, a comprehensive survey of dihedral angle preferences of residues based on non-local interactions on a large data set has not been done yet to the best of our knowledge. The present study has addressed the relationship between amino acids contributing to non-local interactions and their preferred dihedral angles.

The axial distance between adjacent residues of the beta strand is 3.5 Å, whereas the axial distance between adjacent residues of the alpha helix is only 1.5 Å. Hence, the extended regions can accommodate more non-local interacting residues compared to the alpha helices which are clearly reflected in the Ramachandran map (Figs. 1 and 4). A clear distinction in the ϕ, ψ regions adopted by two groups of residues (amino acid residues contributing local and non-local interactions) are observed from our results. The amino acid residues contributing to non-local interactions favor β -sheet conformation because the more extended the local chain is the more other residues can get close to a central residue. The amino acid residues contributing to local interaction favor bent-chain conformations (especially the broad alpha region of the Ramachandran map) because a bent chain putting a residue out on the surface would decrease the chance of having distant neighbors packing near it. Also, we have observed a cluster of residues near ϕ, ψ ($-60, +30$), ($30^\circ, -150^\circ$), and ($+30, -45$), which are defined as disallowed

regions. Also, we have observed a cluster of residues near φ, ψ ($-60, +30$), ($30^\circ, -150^\circ$), and ($+30, -45$), which are defined as disallowed regions. The cluster in the disallowed region is due to the location of residues at short loops or at the junction of two secondary structural elements [52].

A clear discrimination is observed from Ramachandran maps (Fig. 4) of four different types of non-local interactions responsible for protein folding and stability. Since the original Ramachandran maps were derived from a dipeptide model, the question arises whether amino acid residues contributing to non-local interactions in a polypeptide chain has an influence on Ramachandran maps or it is just a geometric necessity for a large number of proteins. The results show that the 55% of amino acid residues in the dataset (Table 1) that are not involved in non-local interactions make the protein polymer flexible. The amino acid residues exhibiting flexibility in proteins as random coil structures have no significant intra-protein or non-local interactions [61]. The amino acid residues exhibiting random coil-like structures occupy various conformational spaces of the Ramachandran map instead of well-defined standard conformational space of secondary structural elements such as helices, sheets, polyproline II, and turns [61, 68–70].

The conformational propensity of amino acid residues forming non-local interactions (Fig. 5) were compared with the deviations of statistical coil-like behavior of amino acid residues in the ht40 protein (an intrinsically disordered protein with 441 amino acid residues) [71]. However, while the deviations in statistical coil-like behavior of amino acid residues in intrinsic disordered proteins are interpreted as indicating the transient formation of a secondary structure [72], our results suggest that nearest-neighbor interactions shift conformational propensities of amino acid residues in the protein rather than facilitating formation of regular secondary structures like right-handed helices, β -strands, and polyproline II-like helices. Through the present work, we highlight the relevance of four major types of non-local interactions on protein secondary and tertiary structures, which has been emphasized over the past 60 years [73–76] and the formation of non-local interactions is a crucial step in the protein-folding process and these interactions might be key drivers to force the residues to adopt certain dihedral angles.

Acknowledgments We thank the anonymous reviewers for constructive comments. KMS is supported by a National Post-Doctoral Fellowship (File No: PDF/2015/000276) by Science Engineering Research Board (SERB), Government of India under the mentorship of Prof. P. Karthe at University of Madras. SS thanks UGC for the award of Emeritus Fellowship (Grant No: F.6-6/2014-15/EMERITUS-2014-15-GEN-4545/(SA-II)).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V.: Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963)
2. Ramachandran, G.N., Sasisekharan, V.: Conformation of polypeptides and proteins. *Adv. Prot. Chem.* **23**, 283–437 (1968)

3. Brocchieri, L., Karlin, S.: Geometry of interplanar residue contacts in protein structures. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9297–9301 (1994)
4. Gunasekaran, K., Ramakrishnan, C., Balaram, P.: Disallowed Ramachandran conformations of amino acid residues in protein structures. *J. Mol. Biol.* **264**, 191–198 (1996)
5. Walthers, D., Cohen, F.E.: Conformational attractors on the Ramachandran map. *Acta. Cryst. D* **55**, 506–517 (1999)
6. Ho, B.K., Brasseur, R.: The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* **5**, 14 (2005)
7. Sasisekharan, V., Ponnuswamy, P.K.: Backbone and side chain conformations of amino acids and amino acid residues in peptides. *Biopolymers* **9**, 1249–1252 (1970)
8. Rose, G.D., Gierasch, L.M., Smith, J.A.: Turns in peptides and proteins. *Adv. Prot. Chem.* **37**, 1–109 (1985)
9. Venkatachalam, C.M.: Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* **6**, 1425–1436 (1968)
10. ElKettani, M.A.E.C., Zakrzewska, K., Durup, J., Lavery, R.: An analysis of the conformational paths of citrate synthase. *Proteins* **16**, 393–407 (1993)
11. Vijayakumar, S., Vishveshwara, S., Ravishanker, G., Beveridge, D.L.: Differential stability of beta-sheets and alpha-helices in beta-lactamase: a high-temperature molecular dynamics study of unfolding intermediates. *Biophys. J.* **65**, 2304–2312 (1993)
12. Betts, M.J., Sternberg, M.J.E.: An analysis of conformational changes on protein–protein association: implications for predictive docking. *Protein Eng.* **12**, 271–283 (1999)
13. Hagler, A.T., Honig, B.: On the formation of protein tertiary structure on a computer. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 554–558 (1978)
14. Neha, V.K., Ramakrishnan, C., Balaram, P.: Sparsely populated residue conformations in protein structures: revisiting experimental Ramachandran maps. *Proteins* **82**, 1101–1112 (2014)
15. Lakshmi, B., Sinduja, C., Archunan, G., Srinivasan, N.: Ramachandran analysis of conserved glycol residues in homologous proteins of known structure. *Protein Sci.* **23**, 843–850 (2014)
16. Saravanan, K.M., Krishnaswamy, S.: Analysis of dihedral angle preferences for alanine and glycine residues in alpha and beta transmembrane regions. *J. Biomol. Struct. Dyn.* **33**, 552–562 (2014)
17. Tanaka, S., Scheraga, H.A.: Model of protein folding: inclusion of short, medium, and long-range interactions. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 3802–3806 (1975)
18. Gromiha, M.M., Selvaraj, S.: Influence of medium and long range interactions in different structural classes of globular proteins. *J. Biol. Phys.* **23**, 209–217 (1997)
19. Gromiha, M.M., Selvaraj, S.: Importance of long-range interactions in protein folding. *Biophys. Chem.* **77**, 49–68 (1999)
20. Selvaraj, S., Gromiha, M.M.: Importance of long-range interactions in (α/β)₈ barrel fold. *J. Prot. Chem.* **17**, 691–697 (1998)
21. Gromiha, M.M., Selvaraj, S.: Role of medium–and long-range interactions in discriminating globular and membrane proteins. *Int. J. Biol. Macromol.* **29**, 25–34 (2001)
22. Gromiha, M.M., Selvaraj, S.: Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J. Mol. Biol.* **310**, 27–32 (2001)
23. Harihar, B., Selvaraj, S.: Refinement of the long-range order parameter in predicting folding rates of two-state proteins. *Biopolymers* **91**, 928–935 (2009)
24. Harihar, B., Selvaraj, S.: Application of long-range order to predict unfolding rates of two-state proteins. *Proteins* **79**, 880–887 (2011)
25. Kabsch, W., Sander, C.: On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 1075–1078 (1984)
26. Minor, D.L., Kim, P.S.: Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730–734 (1996)
27. Alexander, P.A., He, Y., Chen, Y., Orban, J., Bryan, P.N.: A minimal sequence code for switching protein structure and function. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21149–21154 (2009)
28. Saravanan, K.M., Balasubramanian, H., Nallusamy, S., Selvaraj, S.: Sequence and structural analysis of two designed proteins with 88% identity adopting different folds. *Protein Eng. Des. Sel.* **23**, 911–918 (2010)
29. Saravanan, K.M., Selvaraj, S.: Search for identical octapeptides in unrelated proteins: structural plasticity revisited. *Biopolymers* **98**, 11–26 (2012)
30. Pace, C.N.: Conformational stability of globular proteins. *Trends Biochem. Sci.* **15**, 14–17 (1990)
31. Ponnuswamy, P.K., Gromiha, M.M.: On the conformational stability of folded proteins. *J. Theor. Biol.* **166**, 63–74 (1994)
32. Engin, O., Sayar, M., Erman, B.: The introduction of hydrogen bond and hydrophobicity effects into the rotational isomeric states model for conformational analysis of unfolded peptides. *Phys. Biol.* **6**, 016001 (2009)

33. Fitzkee, N.C., Rose, G.D.: Sterics and solvation winnow accessible conformational space for unfolded proteins. *J. Mol. Biol.* **353**, 873–887 (2005)
34. Ohkubo, Y.Z., Brooks, C.L.: Exploring Flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13916–13921 (2003)
35. Cantor, C., Schimmel, P.: *Biophysical Chemistry*. WH Freeman, New York (2011)
36. Park, B., Levitt, M.: Energy functions that discriminate X-ray and near native folds from well-constructed decoys. *J. Mol. Biol.* **258**, 367–392 (1996)
37. Keskin, O., Yuret, D., Gursoy, A., Turkay, M., Erman, B.: Relationships between amino acid sequence and backbone torsion angle preferences. *Proteins* **55**, 992–998 (2004)
38. Porter, L.L., Rose, G.D.: Redrawing the Ramachandran plot after inclusion of hydrogen-bonding constraints. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 109–113 (2011)
39. Thornton, J.M.: *Protein Folding*. WH Freeman, New York (1992)
40. Singh, H., Chauhan, J.S., Gromiha, M.M., Open Source Drug Discovery Consortium, Raghava, G.P.: ccPDB: compilation and creation of data sets from protein data Bank. *Nucleic Acids Res.* **40**, D486–D489 (2012)
41. Gromiha, M.M., Selvaraj, S.: Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol.* **86**, 235–277 (2004)
42. Manavalan, P., Ponnuswamy, P.K.: Hydrophobic character of amino acid residues in globular proteins. *Nature* **275**, 673–674 (1978)
43. Kyte, J., Doolittle, R.F.: A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982)
44. Tina, K.G., Bhadra, R., Srinivasan, N.: PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, W473–W476 (2007)
45. Burley, S.K., Petsko, G.A.: Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science* **229**, 23–28 (1985)
46. Overington, J.P., Johnson, M.S., Sali, A., Blundell, T.L.: Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. Roy. Soc. Biol. Sci.* **241**, 132–145 (1990)
47. Munoz, V., Serrano, L.: Intrinsic secondary structure propensities of the amino acids using statistical φ, ψ matrices: comparison with experimental scales. *Proteins* **20**, 301–311 (1994)
48. Srinivasan, R., Rose, G.D.: LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins* **22**, 81–89 (1995)
49. Street, A.G., Mayo, S.L.: Intrinsic β -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 (1999)
50. Berisio, R., Loguercio, S., De Simone, A., Zagari, A., Vitagliano, L.: Polyproline helices in protein structures: a statistical survey. *Prot. Pept. Lett.* **13**, 847–854 (2006)
51. Wilmot, C.M., Thornton, J.M.: β -turns and their distortions: a proposed new nomenclature. *Protein Eng.* **3**, 479–493 (1990)
52. Pal, D., Chakrabarti, P.: On residues in the disallowed region of the Ramachandran map. *Biopolymers* **63**, 195–206 (2002)
53. Bartlett, G.J., Choudhary, A., Raines, R.T., Woolfson, D.N.: N-II interactions in proteins. *Nat. Chem. Biol.* **6**, 615–620 (2010)
54. Kraqelj, J., Ozenne, V., Blackledge, M., Jensen, M.R.: Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *Chem. Phys. Chem.* **14**, 3034–3045 (2013)
55. Jensen, M.R., Zweckstetter, M., Huang, J., Blackledge, M.: Exploring free energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014)
56. Adzhubei, A.A., Sternberg, M.J., Makarov, A.A.: Polyproline II helix in proteins: structure and function. *J. Mol. Biol.* **425**, 2100–2132 (2013)
57. Cubellis, M.V., Caillez, F., Blundell, T.L., Lovell, S.C.: Properties of polyproline II, a secondary structure element implicated in protein-protein interactions. *Proteins* **58**, 880–892 (2005)
58. Makowska, J., Rodziewicz-Motowidlo, K., Makowski, M., Vila, J.A., Liwo, A., Chmurzynski, L., Scheraga, H.A.: Polyproline II conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1744–1749 (2006)
59. Hagarman, A., Mathieu, D., Toal, S., Measey, T.J., Schwalbe, H., Schweitzer-Stenner, R.: Amino acids with hydrogen-bonding side chains have an intrinsic tendency to sample various turn conformations in aqueous solutions. *Chemistry* **17**, 6789–6797 (2011)
60. Jha, A.K., Colubri, A., Zaman, M.H., Koide, S., Sosnick, T.R., Freed, K.F.: Helix, sheet and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* **44**, 9691–9702 (2005)

61. Toal, S., Schweitzer-Stenner, R.: Local order in the unfolded state: conformational biases and nearest neighbor interactions. *Biomolecules* **4**, 725–773 (2014)
62. Swindells, M.B., MacArthur, M.W., Thornton, J.M.: Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. *Nat. Struct. Biol.* **2**, 596–603 (1995)
63. Thornton, J.M., Jones, D., MacArthur, M., Orengo, C., Swindells, M.: Protein folds: towards understanding folding from inspection of native structures. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **348**, 71–79 (1995)
64. Bellesia, G., Jewett, A.I., Shea, J.E.: Sequence periodicity and secondary structure propensity in model proteins. *Protein Sci.* **19**, 141–154 (2010)
65. Munoz, V., Seranno, L.: Local versus nonlocal interactions in protein folding and stability – an experimentalist's point of view. *Fold. Des.* **1**, 71–77 (1996)
66. Thukral, L., Shenoy, S.R., Bhushan, K., Jayaram, B.: ProRegIn: a regularity index for the selection of native-like tertiary structures of proteins. *J. Biosci.* **32**, 71–81 (2007)
67. Ponnuraj, K., Saravanan, K.M.: Dihedral angle preferences of DNA and RNA binding amino acid residues in proteins. *Int. J. Biol. Macromol.* **97**, 434–439 (2017)
68. Zaman, M.H., Shen, M.Y., Berry, R.S., Freed, K.F., Sosnick, T.R.: Investigations into sequence and conformational dependence of backbone entropy, inter chain dynamics and the Flory isolated-pair hypothesis for peptides. *J. Mol. Biol.* **331**, 693–711 (2003)
69. Ting, D., Wang, G., Shapovalov, M., Mitra, R., Jordan, M.I., Dunbrack Jr., R.L.: Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput. Biol.* **6**, e1000763 (2010)
70. Espinoza-Fonseca, L.M., Llizalitturi-Flores, I., Correa-Basurto, J.: Backbone conformational preferences of an intrinsically disordered protein in solution. *Mol. BioSyst.* **8**, 1798–1805 (2012)
71. Schwalbe, M., Ozenne, V., Bibow, S., Jaremko, M., Jaremko, L., Gajda, M., Jensen, M.R., Biernat, J., Becker, S., Mandelkow, E., Zweckstetter, M., Blackledge, M.: Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α -synuclein in solution from NMR and small angle scattering. *Structure* **22**, 238–249 (2014)
72. Schweitzer-Stenner, R., Toal, S.E.: Construction and comparison of the statistical coil states of unfolded and intrinsically disordered proteins from nearest neighbor corrected conformational propensities of short peptides. *Mol. BioSyst.* **12**, 3294 (2016)
73. Kauzmann, W.: Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63 (1959)
74. Thomas, P.D., Dill, K.A.: Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci.* **2**, 2050–2065 (1993)
75. Dill, K.A., Truskett, T.M., Vlachy, V., Hribar-Lee, B.: Modeling water, the hydrophobic effect, and ion salvation. *Ann. Rev. Biophys. Biomol. Struct.* **34**, 173–199 (2005)
76. Dasgupta, D., Kaushik, R., Jayaram, B.: From Ramachandran maps to tertiary structures of proteins. *J. Phys. Chem. B* **119**, 11136–11145 (2015)