PLOS | ONE

RESEARCH ARTICLE

The Ramachandran Number: An Order Parameter for Protein Geometry

Ranjan V. Mannige*, **Joyjit Kundu**, **Stephen Whitelam***

Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, United States of America

* rvmannige@lbl.gov (RVM); swhitelam@lbl.gov (SW)

CrossMark
click for updates

OPEN ACCESS

Citation: Mannige RV, Kundu J, Whitelam S (2016) The Ramachandran Number: An Order Parameter for Protein Geometry. PLoS ONE 11(8): e0160023. doi:10.1371/journal.pone.0160023

Editor: Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, RUSSIAN FEDERATION

Received: January 11, 2016

Accepted: July 12, 2016

Published: August 4, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability Statement: Figures presented in this paper were derived from the analysis of previously published and freely available databases (SCOP and DSSP) listed in the Methods section.

Funding: RVM and SW were supported by the Defense Threat Reduction Agency under contract no. IACRO-B0845281. This work was done at the Molecular Foundry at Lawrence Berkeley National Laboratory (LBNL), supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Abstract

Three-dimensional protein structures usually contain regions of local order, called secondary structure, such as α-helices and β-sheets. Secondary structure is characterized by the local rotational state of the protein backbone, quantified by two dihedral angles called ϕ and ψ. Particular types of secondary structure can generally be described by a single (diffuse) location on a two-dimensional plot drawn in the space of the angles ϕ and ψ, called a Ramachandran plot. By contrast, a recently-discovered nanomaterial made from peptoids, structural isomers of peptides, displays a secondary-structure motif corresponding to *two* regions on the Ramachandran plot [Mannige et al., *Nature* 526, 415 (2015)]. In order to describe such 'higher-order' secondary structure in a compact way we introduce here a means of describing regions on the Ramachandran plot in terms of a single *Ramachandran number*, $\mathcal{R}$, which is a structurally meaningful combination of ϕ and ψ. We show that the potential applications of $\mathcal{R}$ are numerous: it can be used to describe the geometric content of protein structures, and can be used to draw diagrams that reveal, at a glance, the frequency of occurrence of regular secondary structures and disordered regions in large protein datasets. We propose that $\mathcal{R}$ might be used as an order parameter for protein geometry for a wide range of applications.

1 Introduction

Many three-dimensional protein structures consist of regions of local order called secondary structure [1]. Consequently, the study of secondary structure has occupied a crucial role in structural biology [1–9]. A key insight from this study is the recognition that the conformation of a protein backbone near a given amino acid residue can be specified largely by two dihedral angles, called ϕ and ψ, as shown in Fig 1(a) (a third angle, ω, usually takes one of two values, defining *trans* and *cis* conformations [5, 6]). Ramachandran and co-workers deduced that peptide backbones inhabit only certain regions of dihedral angle (ϕ, ψ) configuration space [10]. Plots drawn in terms of this configuration space are called Ramachandran plots [1, 9, 11, 12], and they are among the most important innovations in structural biology, enabling immediate assessment of the geometric nature of protein structures [13–15].

Competing Interests: The authors have declared that no competing interests exist.

In general, residues that comprise particular protein secondary structures, such as α-helices and β-sheets, correspond to distinct, localized regions on the Ramachandran plot; see Fig 1(b). However, the possibility of secondary structure built from more than one rotational state, i.e. more than one region on the Ramachandran plot, was introduced in 1951 by Pauling and Corey. They proposed a 'pleated sheet' motif in which protein residues alternate between right- and left-handed forms of the α-helix. While not yet seen in nature, simulations indicate that α-pleated sheets can form as kinetic intermediates in unfolding processes [17–20]. More generally, a broad range of protein structures could in principle be built from polypeptide motifs possessing two rotational states [21, 22]. In the non-natural world, protein-mimetic polymers do form large-scale stable structures that simulation indicates harbor a secondary-structure motif built from more than one rotational state. The peptoid nanosheet [23] is a molecular bilayer that possesses macroscopic extent in two dimensions. It is made from peptoids, structural isomers of peptides. The nanosheet is flat because its constituent peptoid polymers are linear and untwisted, properties that result from the fact that backbone residues along each polymer alternate between two twist-opposed rotational states [16]. This secondary-structure motif, called a Σ-strand, corresponds to two regions on the Ramachandran plot, as shown in Fig 1(c).

To describe this structure and its possible generalizations it is convenient to be able to describe regions on the Ramachandran plot with a single number, so that the state of each residue along a polymer backbone can be compactly described. The desire for such a description is the motivation for this paper. We introduce in Section 2.1 a structurally meaningful combination of ϕ and ψ that we call the Ramachandran number, $\mathcal{R}$. Given a way of describing regions of the Ramachandran plot in terms of one number instead of two, one can then draw diagrams that give insight into protein geometry that is difficult to obtain by other means. In Section 2.2 we show that $\mathcal{R}$ can be used to assess in a compact manner the geometric content of protein structures, and can be used to draw diagrams that reveal at a glance the frequency of occurrence of regular secondary structures and disordered regions in large protein datasets. We also suggest that $\mathcal{R}$ may be useful in the analysis of intrinsically-disordered proteins, whose three-dimensional structures are less well understood than those of globular proteins [9, 24–26]. We conclude in Section 3.

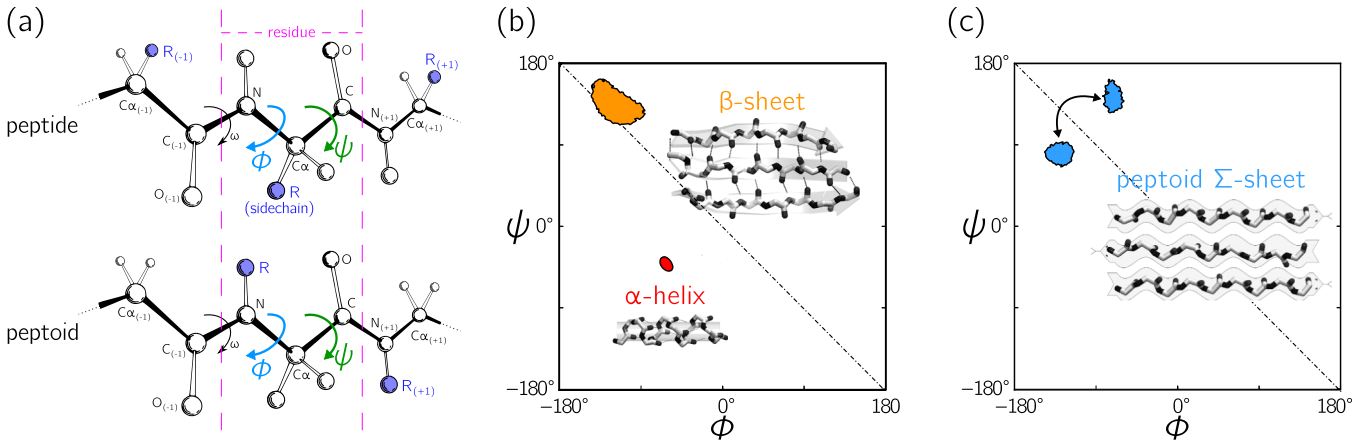

Fig 1. The Ramachandran plot is an important way of describing protein secondary structure. (a) The state of a residue within a peptide (top) and a peptoid (bottom) can be largely specified by the two dihedral angles ϕ and ψ. (b) Regular protein secondary structures, such as α-helices and β-sheets, correspond to single diffuse regions on a plot drawn in terms of ϕ and ψ, called a Ramachandran plot (see Methods). (c) Peptoid Σ-sheets [16] harbor a secondary-structure motif in which backbone residues alternate between two regions on the Ramachandran plot. In order to describe each region in terms of a single number, so that the state of each residue in a backbone can be compactly indicated, we describe in this paper the development and properties of a structurally meaningful combination of ϕ and ψ that we call the Ramachandran number, $\mathcal{R}$. [Panel (a) was adapted from an image found on Wikimedia Commons (link) by Dcrjsr (CC BY 3.0 (link)). The contours in (b) and (c) represent regions within which 70% of a secondary structure resides; see Section 4.1.].

doi:10.1371/journal.pone.0160023.g001

2 Results and Discussion

2.1 One possible Ramachandran number

One physical factor that suggests a compact way of describing the Ramachandran plot is the sense of residue twist implicit in the plot, which changes sign as one crosses the negative-sloping diagonal; see Fig 2(a). Structures whose backbones occupy the bottom left-hand triangle of the Ramachandran plot have a right-handed (dextrorotatory or D) sense of twist, while structures in the top right-hand triangle have a left-handed (levorotatory or L) sense of twist [1, 16] (note that the terms 'L' and 'D' are also independently used to distinguish between distinct amino acid enantiomers [27–30]). This observation suggests an indexing system that proceeds from the bottom left of the plot to the top right of the plot. To gain insight into how this should be done, we built protein backbones with dihedral angles chosen from designated regions of the Ramachandran plot. We examined the behavior of the end-to-end distance R_e of polymers built in this way (the polymer radius of gyration behaves similarly). This behavior is shown in Fig 2(a) and 2(b). The shapes of the contours of R_e suggest an indexing system that proceeds in a sweeping fashion across the Ramachandran plot, as shown in Fig 2(c), so that R_e changes as slowly as possible. Proceeding in this manner one moves from structures having one sense of twist to structures having the opposite sense of twist, with the degree of compactness of the backbone varying only in a gradual fashion. Thus the indexing system suggested graphically in Fig 2(c) is sensitive both to the twist state and the degree of compactness of the polymer backbone, allowing one to distinguish, for example, compact α-helices from extended β-sheets, or nearly twist-free β sheets from twisted loop regions.

To construct such an indexing system we take the Ramachandran plot axes to have the range $[-\lambda/2, \lambda/2)$ where $\lambda = 360°$ [1, 11, 13, 15]. We divide the plot into a square grid of $(360° \sigma)^2$ sites, where σ is a scaling factor that is measured in reciprocal degrees. We shall show that it is straightforward to make σ large enough that the error incurred upon converting angles from structures in the protein databank to Ramachandran numbers and back again is less than the characteristic error associated with the coordinates of structures in that database.

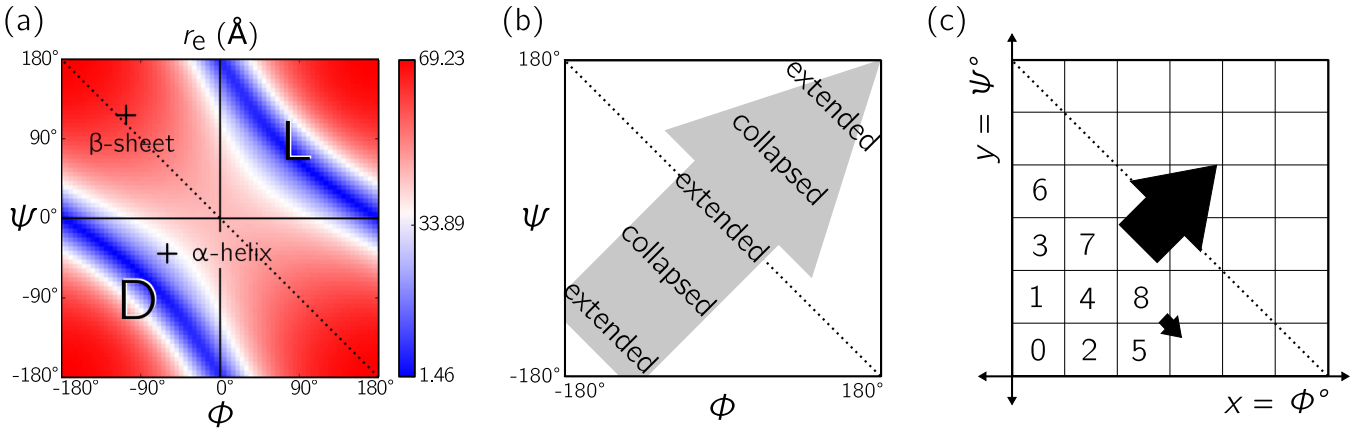

Fig 2. Physical trends within the Ramachandran plot suggest a way of describing regions of it with a single number. (a) First, the sense of residue twist changes from right-handed ('D') to left-handed ('L') as one moves from the bottom left of the Ramachandran plot to the top right. Second, contours (colored) of end-to-end polymer distance R_e (here calculated for a 20-residue glycine) have a negative slope, resulting in the general trend shown in panel (b). Panel (c) indicates one method of indexing the Ramachandran plot so as to move from the region of right-handed twist to the region of left-handed twist with R_e changing as slowly as possible. This method provides the basis for the construction of the Ramachandran number, $\mathcal{R}$.

doi:10.1371/journal.pone.0160023.g002

Given a choice of grid resolution σ, we define the integer-valued Ramachandran number

$$R_{\mathbb{Z}}(\phi,\psi) \equiv \lfloor \phi' \rceil + \lambda' \lfloor \psi' \rceil, \tag{1}$$

in which the coordinates

$$\phi' \equiv (\phi - \psi + \lambda)\sigma/\sqrt{2} \tag{2}$$

and

$$\psi' \equiv (\phi + \psi + \lambda)\sigma/\sqrt{2} \tag{3}$$

correspond to a clockwise rotation by 45°, a shift, and a rescaling of the original coordinates ϕ and ψ (see Section 4.2), and the parameter

$$\lambda' \equiv \left\lfloor \sqrt{2}\lambda\sigma \right\rceil. \tag{4}$$

In these relations the symbol $\lfloor x \rceil$ means the integer closest to the real number x, i.e. $\lfloor 2.49 \rceil = 2$ and $\lfloor 2.51 \rceil = 3$. Combined as a single expression, Eqs (1)–(4) read

$$R_{\mathbb{Z}}(\phi,\psi) \equiv \left\lfloor (\phi - \psi + \lambda)\sigma/\sqrt{2} \right\rceil + \left\lfloor \sqrt{2}\lambda\sigma \right\rceil \left\lfloor (\phi + \psi + \lambda)\sigma/\sqrt{2} \right\rceil. \tag{5}$$

The integer-valued Ramachandran number runs between $R_{\mathbb{Z},\min} = R_{\mathbb{Z}}(-180,-180) = \lfloor \lambda\sigma/\sqrt{2} \rceil$ and $R_{\mathbb{Z},\max} = R_{\mathbb{Z}}(180,180) = R_{\mathbb{Z},\min} + \lfloor \sqrt{2}\lambda\sigma \rceil^2$. The range of these values are dependent on σ, which makes $R_{\mathbb{Z}}$ a difficult value to intuitively grasp. Therefore, for ease of plotting, we define the real-valued, normalized, Ramachandran number

$$\mathcal{R}(\phi,\psi) \equiv \frac{R_{\mathbb{Z}}(\phi,\psi) - R_{\mathbb{Z},\min}}{R_{\mathbb{Z},\max} - R_{\mathbb{Z},\min}}, \tag{6}$$

which is a value that is practically invariant of σ. Given that $\phi, \psi \in [-\lambda/2, \lambda/2)$ or $[-180, 180)$, the ranges of $R_{\mathbb{Z}}$ and $\mathcal{R}$ are respectively $[R_{\mathbb{Z},\min}, R_{\mathbb{Z},\max})$ and $[0, 1)$.

The closest approximations to the original coordinates ϕ and ψ that may be retrieved from $R_{\mathbb{Z}}$ are [31]

$$\tilde{\phi} = \frac{1}{\sigma\sqrt{2}}(\lfloor R_{\mathbb{Z}}/\lambda' \rfloor + R_{\mathbb{Z}} \,\%\, \lambda' - \lambda'), \tag{7}$$

$$\tilde{\psi} = \frac{1}{\sigma\sqrt{2}}(\lfloor R_{\mathbb{Z}}/\lambda' \rfloor - R_{\mathbb{Z}} \,\%\, \lambda'), \tag{8}$$

where $\lfloor x \rceil$ is the largest integer smaller than the real number x, and $\alpha\%\beta$ is the remainder obtained upon dividing the integer α by the integer β. Eqs (6) to (8) define our mapping of the dihedral angles to the Ramachandran number, i.e. $(\phi,\psi) \to R_{\mathbb{Z}} \to \mathcal{R}$, and the subsequent approximate recovery of those angles, $R_{\mathbb{Z}} \to (\tilde{\phi},\tilde{\psi})$. We show in Section 4.3 that this recovery can be done to within the characteristic precision of the protein databank.

By 'slicing' across the Ramachandran plot we group together structures that might be relatively distant in dihedral angle space, the more so as we approach the negative-sloping diagonal (near $\mathcal{R} = 0.5$). One consequence of this grouping is that the set of structures described by a small interval of $\mathcal{R}$ displays a distribution of properties, such as end-to-end distance, as shown in Fig 3(a). The mean of this distribution gives rise to a smoothly-varying trend, but the variance of this distribution is nonzero, and is largest near $\mathcal{R} = 0.5$. Some unavoidable structural

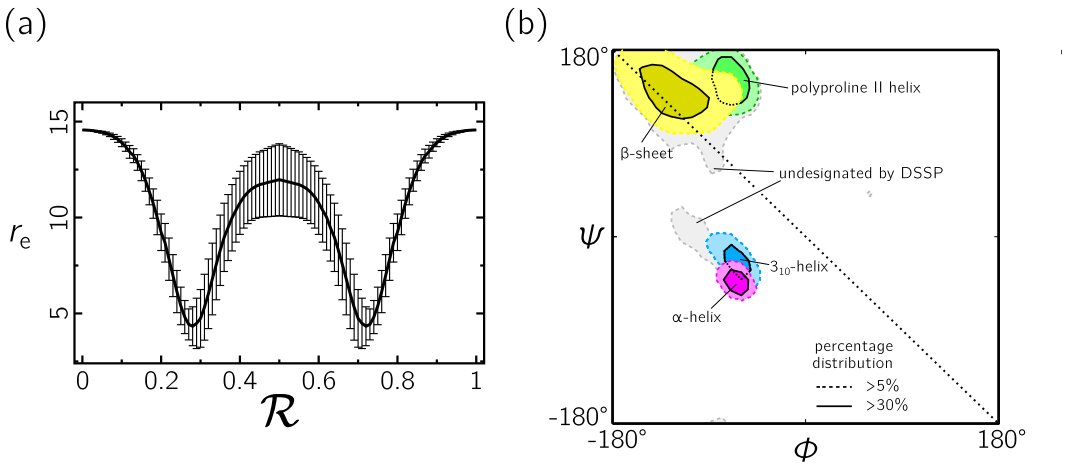

Fig 3. Potential pathologies of $\mathcal{R}$ are avoided by the sparse occupancy of the Ramachandran plot. (a) We construct $\mathcal{R}$ by slicing across the Ramachandran plot, which can cause points distant in dihedral angle space to be grouped together, the more so as we approach the negative-sloping diagonal (near $\mathcal{R} = 0.5$). This grouping can be inferred by superposing the standard deviation (error bars) in polymer end-to-end distance on top of the mean value (smooth line) for hypothetical structures built from the relevant part of the Ramachandran diagram. (b) However, many structures distant in dihedral angle space but close in $\mathcal{R}$ do not arise in proteins; the Ramachandran diagram is in general relatively sparsely occupied. Consequently, $\mathcal{R}$ can resolve the major types of protein secondary structure, which can be inferred from the fact that lines parallel to the negative-sloping diagonal (marked), along which $\mathcal{R}$ varies only slowly, can touch each region of known secondary structure (colored) individually. This sensitivity allows $\mathcal{R}$ to function as an order parameter for protein geometry. [Data in (a) were calculated for a 5-residue peptoid; $\mathcal{R}$ values are shown at discrete intervals of 0.01.].

doi:10.1371/journal.pone.0160023.g003

coarse-graining therefore occurs upon going from the Ramachrandran plot to the Ramachandran number. Despite this drawback, we shall show that $\mathcal{R}$ can function as an order parameter for protein geometry, in large part because the Ramachandran plot is in general relatively sparsely occupied: many hypothetical structures that possess distinctly different structural properties but that would be assigned similar Ramachandran numbers simply do not arise in the protein world. Consequently, $\mathcal{R}$ can resolve the major classes of protein secondary structure, such as the α and β motifs; see Figs 3(b) and 4.

2.2 Properties and uses of $\mathcal{R}$

The indexing system defined by Eqs (6) and (1) collapses the Ramachandran plot into a single line, the Ramachandran number $\mathcal{R}$. As shown in Fig 4, this number can act as an order parameter for types of polymer secondary structure. Given such an order parameter, we can then draw diagrams that reveal the abundance and spatial connectivity of different forms of secondary structure within polymers.

In Fig 5 we show four different molecular structures described in terms of (a) spatial configurations, (b) the Ramachandran plot, (c) a histogram (bar code or '$\mathcal{R}$-code') of $\mathcal{R}$-values, and (d) a plot of $\mathcal{R}$ versus residue number (the structure of the coiled coil was deduced by a number of authors, Ramachandran among them [32–35]). The $\mathcal{R}$-code of panel (c) can be regarded as a way of assaying the residues of a protein by geometry, much as gel electrophoresis, which results in similar-looking pictures, is used to tell apart macromolecules by their size and charge. The plot of $\mathcal{R}$ versus residue number in panel (d) reveals the spatial connectivity of distinct ordered domains. It shows the distinct segments of secondary structure (α-helix and β-sheet) and loop regions in the proteins, and shows that the peptoid Σ-strand's residues alternate between twist-opposed rotational states. This representation makes clear that the two

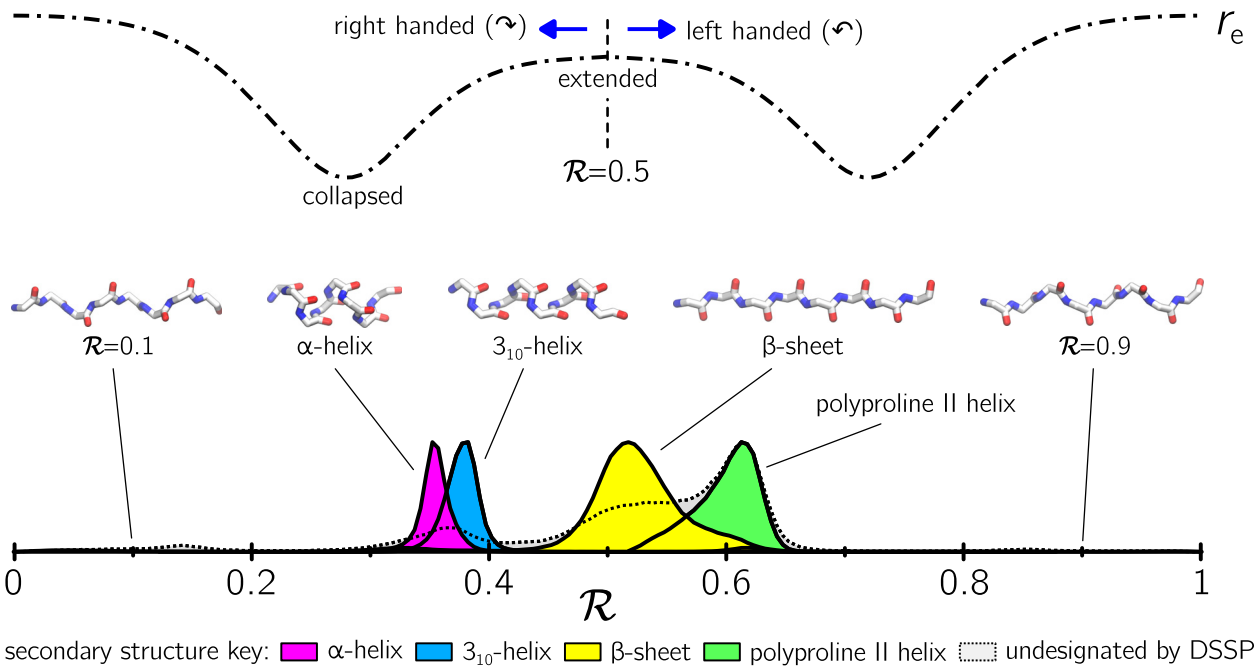

Fig 4. The indexing system defined by Eqs (6) and (1) collapses the Ramachandran plot into a single line, the Ramachandran number $\mathcal{R}$. This number can act as an order parameter to distinguish secondary structures of different geometry, as shown (the overlap between distributions exists in the original Ramachandran plot representation; see Fig 3(b)). Top: $\mathcal{R}$ interpolates between regions of right-handed and left-handed twist, with polymer extension R_e varying smoothly throughout.

doi:10.1371/journal.pone.0160023.g004

rotational states of the Σ-strand motif are incorporated within a single type of secondary structure; the Ramachandran plot alone does not distinguish between that outcome and the alternative, that the two rotational states exist within two distinct types of secondary structure.

$\mathcal{R}$ can be used to compactly describe the abundance of secondary structure types with large protein datasets, as shown in Fig 6. There we show histograms of $\mathcal{R}$ ('$\mathcal{R}$-codes') for proteins belonging to distinct SCOP classes [36]. These diagrams identify a number of trends within this dataset. As expected, proteins belonging to classes 'a' and 'b' are rich in α-helical ($\mathcal{R}$ ≈0.36) and β-sheet ($\mathcal{R}$ ≈0.52) regions, respectively. More surprisingly, α-helical regions are abundant in all protein classes, even in the 'all-β' class 'b'. Loop regions ($\mathcal{R}$ ≈0.62) are also prominent; loops connect regions of ordered secondary structure. The $\mathcal{R}$-code also highlights the symmetry of the peptoid backbone about the twist-free region $\mathcal{R}$ ≈0.5 (panel (b)).

$\mathcal{R}$ can also be used in a time- and space-resolved way, as shown in Fig 7. Here we show the results of molecular dynamics simulations of the peptoid nanosheet [16, 23], which reveal the existence of the Σ-strand secondary structure motif in which residues possess two distinct rotational states. A time series of the $\mathcal{R}$-code of the bilayer (panel (b)) shows the emergence of the Σ-strand motif within molecular dynamics simulations via a breaking of the initially-imposed molecular symmetry. In panel (c), we show the geometric state of each residue in one peptoid as a function of time, revealing the emergence of the Σ-strand structure and the subsequent fluctuations of individual residues on a nanosecond timescale.

Such time- and space-resolved analysis of polymer backbones may be useful for the analysis of intrinsically disordered proteins (IDPs), whose backbone conformations are heterogeneous in space and time [9, 24–26]. Insight into the behavior of IDPs can likely be obtained by tracking the state of every residue of a protein as a function of time, which can be done with $\mathcal{R}$ in

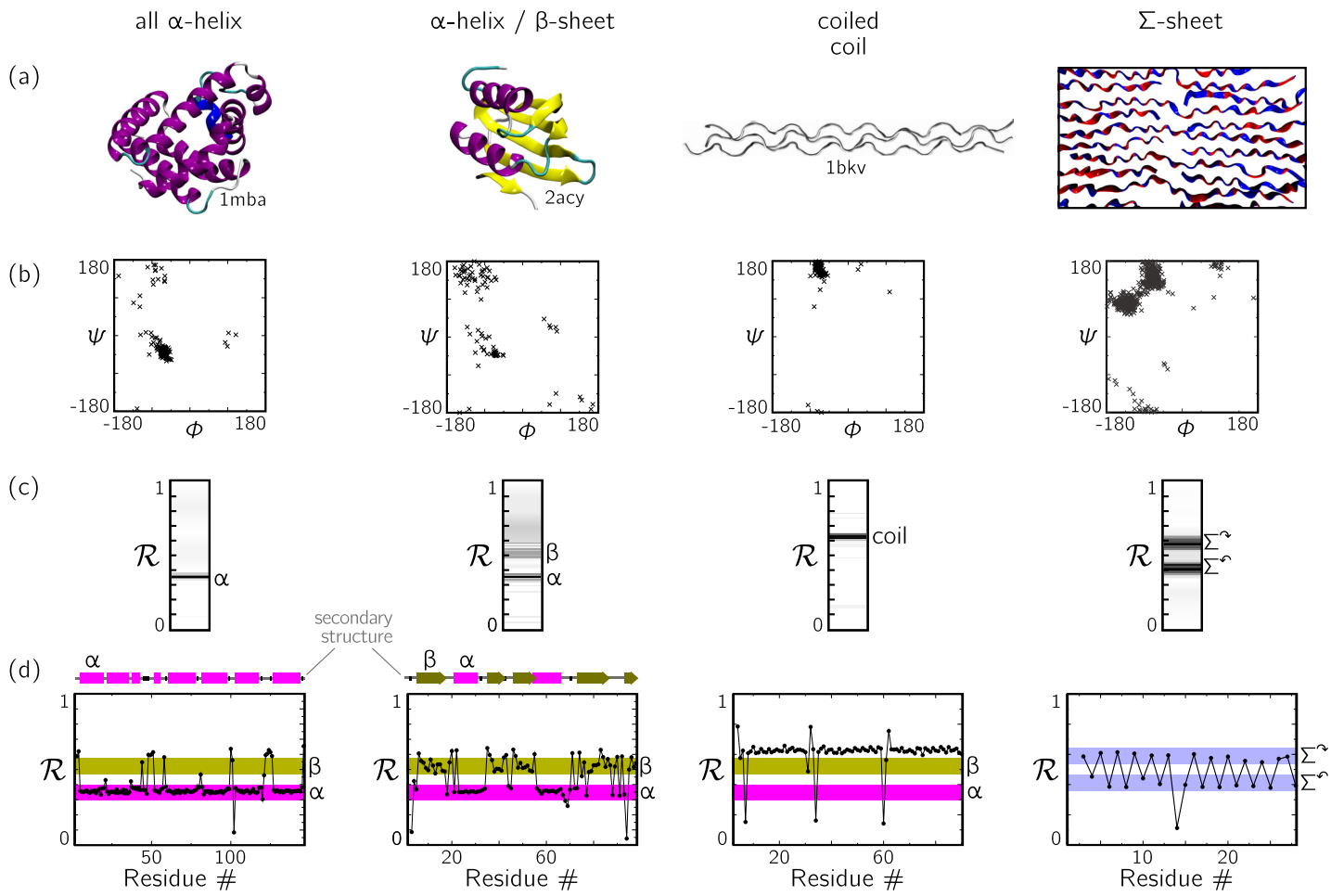

Fig 5. Four ways of looking at secondary structure: a) Molecular configurations; b) Ramachandran plot; c) Histogram ($\mathcal{R}$-code) of Ramachandran numbers; and (d) $\mathcal{R}$ as a function of residue number (for the Σ-sheet we have chosen a single polymer). Panel (c) provides a compact assay-by-geometry of the residues within molecular structures, while panel (d) shows that one can use $\mathcal{R}$ to identify the spatial connectivity of domains of secondary structure within a polymer.

doi:10.1371/journal.pone.0160023.g005

the simple manner shown in Fig 7(c). Presenting similar data using the conventional Ramachandran plot would require a movie or a complex three-dimensional figure.

3 Conclusions

The Ramachandran plot is a central element of structural biology. We have introduced here a way of describing regions of the Ramachandran plot in terms of a single *Ramachandran number*, $\mathcal{R}$, which is a structurally meaningful combination of ϕ and ψ. The are many possible ways of constructing such a number, and the one we have chosen is sensitive to the local twist state and degree of compactness of a polymer backbone. Given the ability to describe a two-dimensional space with a single number, one can draw diagrams that furnish insight into polymer structure that is difficult to obtain through other means. For instance, we have shown that $\mathcal{R}$ can be used to describe the geometric content of protein and protein-inspired structures, in a space- and time-resolved way, and can be used to draw diagrams that reveal at a glance the frequency of occurrence of regular secondary structures and disordered regions in large protein datasets. We speculate that $\mathcal{R}$ may also be useful in analyzing the behavior and evolution of

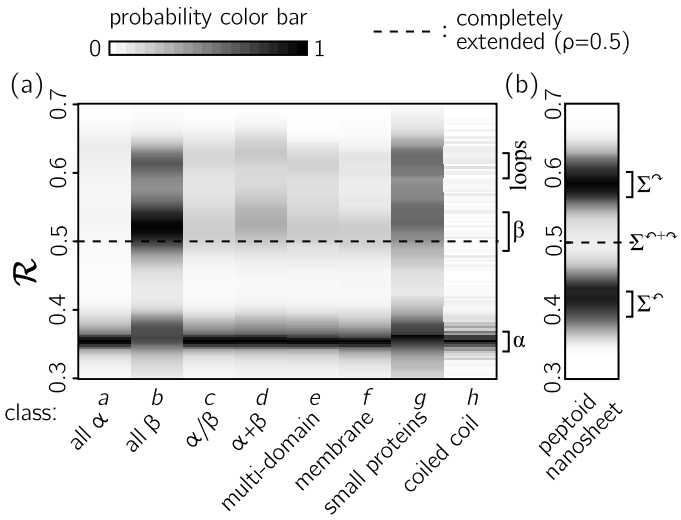

Fig 6. (a) $\mathcal{R}$-codes for the SCOP protein dataset reveal at a glance several geometric properties of the set. Each column represents a histogram of the indicated protein class, normalized so that the largest value is unity. A feature common to all classes is the prominence of α-helices ($\mathcal{R} \approx 0.36$). Another common feature is the presence of loops that connect ordered secondary structure ($\mathcal{R} \approx 0.62$). Moreover, α-helical regions are prominently visible in 'all-β' proteins. (b) The $\mathcal{R}$-code for a peptoid nanosheet shows two dominant rotational states, which coexist within a single secondary structure (see Figs 5 and 7).

doi:10.1371/journal.pone.0160023.g006

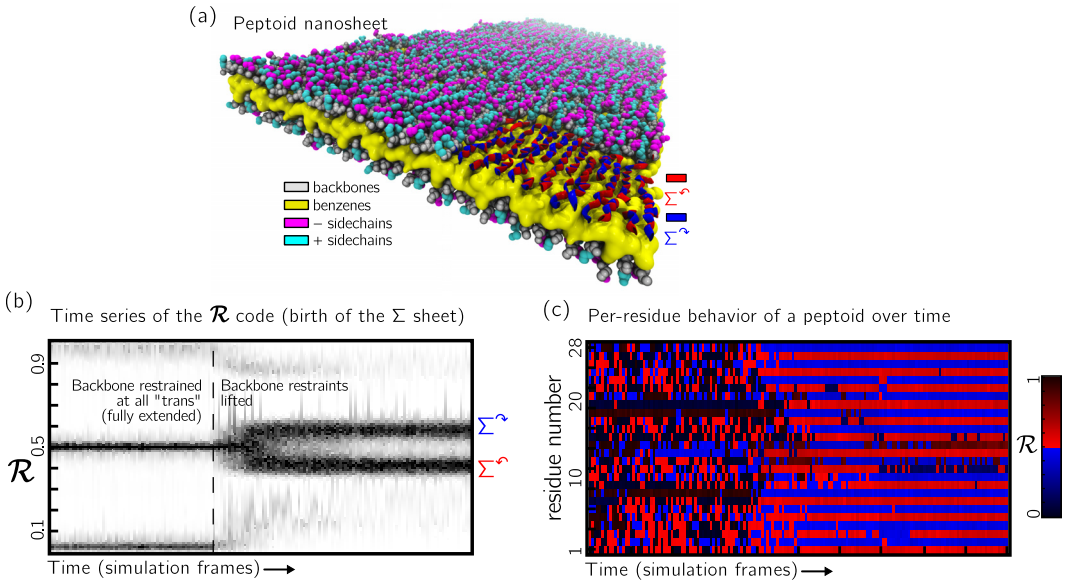

Fig 7. (a) Molecular dynamics simulations of the peptoid nanosheet [16, 23] show the existence of the Σ-strand secondary structure motif, within which residues possess two distinct rotational states (colored red and blue in the bottom-right-hand cutaway). (b) A time series of the $\mathcal{R}$-code of the bilayer shows the emergence (to the right of the vertical dotted line) of the Σ-strand motif within molecular dynamics simulations. Polymers in these simulations were initially fully extended, and adopted the Σ-strand motif upon relaxation of their backbone constraints. (c) Geometric state of each residue in one peptoid as a function of time, revealing the emergence of the Σ-strand structure and the subsequent fluctuations of individual residues on a nanosecond timescale.

doi:10.1371/journal.pone.0160023.g007

intrinsically-disordered proteins (IDPs), important to e.g. the study of diseases [37]. Such proteins are less well-characterized than globular proteins [1]. IDPs spend substantial amounts of time in unfolded or disordered conformations [9, 24–26], but may harbor local or transient regions of structure such as α-helices [38]. The Ramachandran number described here may be a useful complement to existing bioinformatics metrics for IDP sequences [39] for understanding the behavior of these proteins in simulations [40–44] and experiments [45–48]. More generally, $\mathcal{R}$ may be useful as an order parameter for polymer geometry for a wide range of applications.

4 Methods

4.1 Obtaining polymer (protein/peptide/peptoid) statistics

The contours in Figs 1b and 3b describe the distribution of secondary structures in a Ramachandran plot, while Fig 4 represent histogram distributions of secondary structures on the Ramachandran line. To obtain statistics on all secondary structures (excepting the polyproline II helix; see below), a protein structure database was obtained from the Structural Classification of Proteins or SCOPe (Release 2.03) [36] that contains proteins with no more than 40% sequence identity (downloaded from http://scop.berkeley.edu/downloads/pdbstyle/pdbstyle-sel-gs-bib-40-2.03.tgz). Secondary structural elements such as α-helices, 3_{10}-helices and β-sheets were identified using the DSSP algorithm [49–51].

The statistics for polyproline II helices (used to generate the green distributions in Figs 3b and 4) were obtained from segments within 16,535 proteins annotated by PolyprOnline [52] to contain three or more residues of the secondary structure.

Fig 6a represents $\mathcal{R}$-codes for entire classes of proteins. We utilized the SCOP classification system and the individual proteins from each class were extracted from the SCOP dataset described above. Altogether, there were 8560 proteins that were amenable to analysis from this database.

The distribution for the Σ-sheet on a Ramachandran Plot (Fig 1c) and in an $\mathcal{R}$-code (Figs 5 and 6b) were obtained from a 50 nanosecond interval of a molecular dynamics trajectory [16]. Fig 7(b) and 7(c) describes a trajectory of the same system before and after the symmetry-enforcing backbone restraints were lifted. This process was part of a molecular dynamics equilibration step used in Ref. [16].

The end-to-end distances for glycine peptides of 20 residues (Fig 2(a)) and 5 residues (Fig 3 (a)) were generated using the PeptideBuilder library [53] and analyzed using BioPython [54].

4.2 Coordinate transformation used to obtain $\mathcal{R}$

Eqs (6)–(8) were obtained by rotating the Ramachandran plot so that contours of constant polymer extension R_e (see Fig 2) lie roughly horizontal in the new coordinate representation. To this end we define

$$\begin{pmatrix} \phi' \\ \psi' \end{pmatrix} \equiv \frac{\sigma}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi \\ \psi \end{pmatrix} + \frac{\lambda\sigma}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{9}$$

so that ϕ' and ψ' are obtained by rotating the original coordinate system (ϕ, ψ) clockwise by 45°, shifting the resulting coordinates linearly (so that the new coordinates are non-negative), and rescaling the result by the grid resolution σ. Fig 8(a) shows graphically this transformation. Indexing the new coordinate system according to Eq (1) corresponds to the counting system

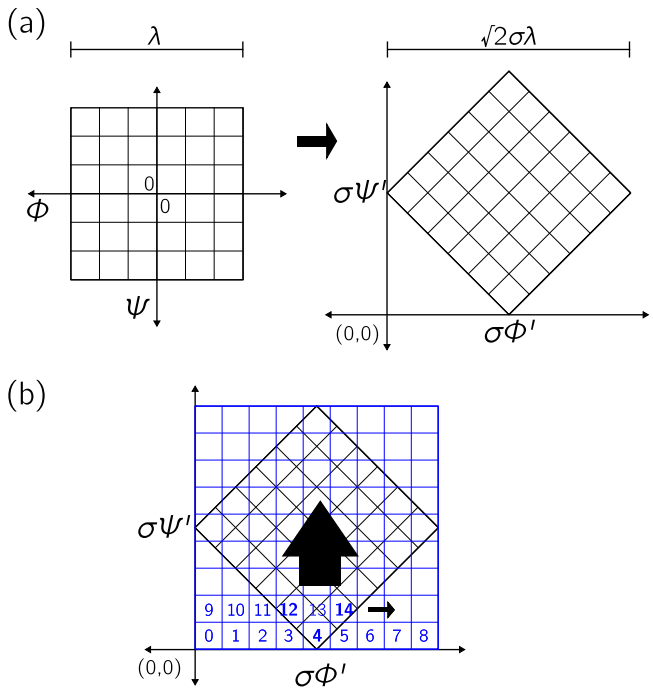

Fig 8. (a) Coordinate transformation applied to the Ramachandran plot in order to compute the Ramachandran number $\mathcal{R}$. A rotation, shift, and rescaling of ϕ and ψ results in a representation in which horizontal cuts run roughly along contours of polymer extension; see Fig 2. (b) The indexing system defined by Eq (1), where bold numbers are those that fall on the original Ramachandran plot.

doi:10.1371/journal.pone.0160023.g008

shown in Fig 8(b). To undo the transformation Eq (1) approximately we compute

$$\tilde{\phi}' = R_{\mathbb{Z}} \,\%\, \lambda', \tag{10}$$

$$\tilde{\psi}' = \lfloor R_{\mathbb{Z}}/\lambda' \rfloor. \tag{11}$$

We then insert Eqs (10) and (11) into the equations $\tilde{\phi}' \equiv (\tilde{\phi} - \tilde{\psi} + \lambda)\sigma/\sqrt{2}$ and

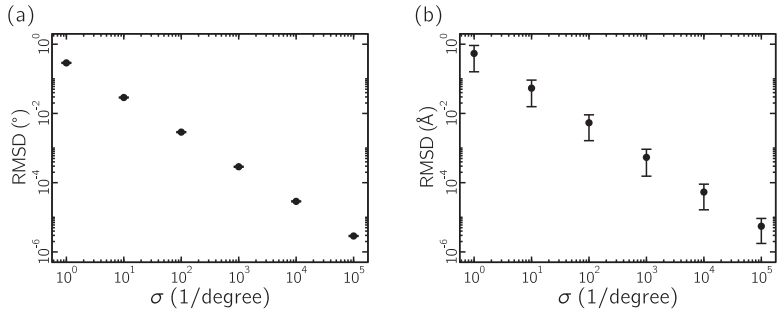

Fig 9. Dihedral angles converted to Ramachandran numbers can be recovered only approximately, but the error incurred during this back-mapping can be made much smaller than the standard error (typically 1Å) associated with structures in the protein databank. Here we show the root-mean-squared-deviation (RMSD) in dihedral angles (a) and in protein α-carbon spatial coordinates (b) generated upon taking 8560 protein structures obtained from SCOP [36], converting their dihedral angles to Ramachandran numbers, and recovering approximately those dihedral angles using Eqs (7) and (8). The parameter σ indicates the grid resolution used to calculate R; see Eq (1).

doi:10.1371/journal.pone.0160023.g009

$\tilde{\psi}' \equiv (\tilde{\phi} + \tilde{\psi} + \lambda)\sigma/\sqrt{2}$, and solve these for the closest approximations $\tilde{\phi}$ and $\tilde{\psi}$ to the original angles ϕ and ψ. The results are Eqs (7) and (8).

4.3 Recovering dihedral angle values from the Ramachandran number

It is convenient to be able to retrieve from the Ramachandran number a good approximation to the dihedral angles used to calculate it. This can be done for a range of choices of grid resolution σ. We took 8560 protein structures obtained from SCOP [36]; see Section 4.1. For a given protein we took the dihedral angles associated with each residue, and used these to compute the 3D protein structure (given values of the ω dihedral angle). We carried out the conversion from dihedral angles to Ramachandran number, defined by Eq (1), and from this used Eqs (7) and (8) to obtain an approximation to the original dihedral angles. We used these approximate angles (with the original values of the ω dihedral angle) to compute the 3D structure of the protein. We then calculated the root-mean-squared deviation (RMSD) between the original and recovered sets of angles and heavy-atom positions, shown in Fig 9. For a range of values of grid resolution σ we find these RMSD values to lie well within the 1Å characteristic of the protein databank. For the calculations done in this paper we took $\sigma = 10^5$ reciprocal degrees.

Acknowledgments

The authors thank Ronald N. Zuckermann for important discussions. RVM thanks Alana Canfield Mannige for input. This work, RVM and SW were supported by the Defense Threat Reduction Agency under contract no. IACRO-B0845281. This work was done at the Molecular Foundry at Lawrence Berkeley National Laboratory (LBNL), supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Author Contributions

Wrote the paper: RVM JK SW.

Conceived of the Ramachandran number, and developed it with assistance from JK and SW: RVM.

Designed the calculations described in this paper: RVM SW.

Carried out the calculations described in this paper: RVM.

References

1. Berg JM, Tymoczko JL, Stryer L. Biochemistry, International Edition. 7th ed. New York: WH Freeman & Co.; 2010.
2. Chothia C, Hubbard T, Brenner S, Barns H, Murzin A. Protein folds in the all-beta and all-alpha classes. Annu Rev Biophys Biomol Struct. 1997; 26:597–627. doi: 10.1146/annurev.biophys.26.1.597 PMID: 9241431
3. Cooper GM, Hausman RE. The Cell: A Molecular Approach. 6th ed. Sunderland, MA: Sinauer Associates, Inc.; 2013.
4. Linderstrøm-Lang KU. Lane Medical Lectures: proteins and enzymes. vol. 6. Stanford University Press; 1952.
5. Pauling L, Corey RB. The pleated sheet, a new layer configuration of polypeptide chains. Proceedings of the National Academy of Sciences of the United States of America. 1951; 37(5):251. doi: 10.1073/pnas.37.5.251 PMID: 14834147
6. Pauling L, Corey RB, Branson HR. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proceedings of the National Academy of Sciences. 1951; 37(4):205–211. doi: 10.1073/pnas.37.4.205

7. Bragg L, Kendrew JC, Perutz MF. Polypeptide chain configurations in crystalline proteins. Proceedings of the Royal Society of London Series A Mathematical and Physical Sciences. 1950; 203(1074):321–357. doi: 10.1098/rspa.1950.0142
8. Eisenberg D. The discovery of the α-helix and β-sheet, the principal structural features of proteins. Proceedings of the National Academy of Sciences. 2003; 100(20):11207–11210. doi: 10.1073/pnas.2034522100
9. Mannige RV. Dynamic New World: Refining Our View of Protein Structure, Function and Evolution. Proteomes. 2014; 2(1):128–153. doi: 10.3390/proteomes2010128
10. Ramachandran G, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. Journal of molecular biology. 1963; 7(1):95–99. doi: 10.1016/S0022-2836(63)80023-6 PMID: 13990617
11. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. New York: Garland Science; 2002. Classic textbook now in its 5th Edition. 2002;.
12. Subramanian E. GN Ramachandran. Nature Structural & Molecular Biology. 2001; 8(6):489–491. doi: 10.1038/88544
13. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography. 1993; 26(2):283–291. doi: 10.1107/S0021889892009944
14. Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Computer applications in the biosciences: CABIOS. 1997; 13(4):425–430. PMID: 9283757
15. Laskowski RA. Structural quality assurance. Structural Bioinformatics, Volume 44. 2003;p. 273–303.
16. Mannige RV, Haxton TK, Proulx C, Robertson EJ, Battigelli A, Butterfoss GL, et al. Peptoid nanosheets exhibit a new secondary structure motif. Nature. 2015; 526:415–420. doi: 10.1038/nature15363 PMID: 26444241
17. Armen RS, DeMarco ML, Alonso DO, Daggett V. Pauling and Corey's α-pleated sheet structure may define the prefibrillar amyloidogenic intermediate in amyloid disease. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(32):11622–11627. doi: 10.1073/pnas.0401781101 PMID: 15280548
18. Daggett V. α-Sheet: the toxic conformer in amyloid diseases? Accounts of chemical research. 2006; 39(9):594–602. doi: 10.1021/ar0500719 PMID: 16981675
19. Wu H, Canfield A, Adhikari J, Huo S. Quantum mechanical studies on model α-pleated sheets. Journal of computational chemistry. 2010; 31(6):1216–1223. doi: 10.1002/jcc.21408 PMID: 19827141
20. Babin V, Roland C, Sagui C. The α-sheet: A missing-in-action secondary structure? Proteins: Structure, Function, and Bioinformatics. 2011; 79(3):937–946. doi: 10.1002/prot.22935
21. Hayward S, Milner-White EJ. The geometry of α-sheet: Implications for its possible function as amyloid precursor in proteins. Proteins: Structure, Function, and Bioinformatics. 2008; 71(1):415–425. doi: 10.1002/prot.21717
22. Hayward S, Leader DP, Al-Shubailly F, Milner-White EJ. Rings and ribbons in protein structures: Characterization using helical parameters and Ramachandran plots for repeating dipeptides. Proteins: Structure, Function, and Bioinformatics. 2014; 82(2):230–239. doi: 10.1002/prot.24357
23. Nam KT, Shelby SA, Choi PH, Marciel AB, Chen R, Tan L, et al. Free-floating ultrathin two-dimensional crystals from sequence-specific peptoid polymers. Nat Mater. 2010 May; 9(5):454–460. doi: 10.1038/nmat2742 PMID: 20383129
24. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. J Mol Graph Model. 2001; 19(1):26–59. doi: 10.1016/S1093-3263(00)00138-8 PMID: 11381529
25. Dunker AK, Silman I, Uversky VN, Sussman JL. Function and structure of inherently disordered proteins. Curr Opin Struct Biol. 2008 Dec; 18(6):756–764. doi: 10.1016/j.sbi.2008.10.002 PMID: 18952168
26. Dunker AK, Babu M, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, et al. What's in a name? Why these proteins are intrinsically disordered? Intrinsically Disordered Proteins. 2013; 1:e24157. doi: 10.4161/idp.24157
27. Cintas P. Chirality of living systems: a helping hand from crystals and oligopeptides. Angewandte Chemie International Edition. 2002; 41(7):1139–1145. doi: 10.1002/1521-3773(20020402)41:7%3C1139::AID-ANIE1139%3E3.0.CO;2-9
28. Branden CI, et al. Introduction to protein structure. Garland Science; 1999.
29. Zawadzke LE, Berg JM. The structure of a centrosymmetric protein crystal. Proteins: Structure, Function, and Bioinformatics. 1993; 16(3):301–305. doi: 10.1002/prot.340160308

30. Hung LW, Kohmura M, Ariyoshi Y, Kim SH. Structure of an enantiomeric protein, d-monellin at 1.8 Å resolution. Acta Crystallographica Section D: Biological Crystallography. 1998; 54(4):494–500. doi: 10.1107/S0907444997012225

31. Anderson JM, Amarasinghe SP, Lam MS. Data and computation transformations for multiprocessors. In: ACM SIGPLAN Notices. vol. 30. ACM; 1995. p. 166–178.

32. Ramachandran G, Kartha G. Structure of collagen. Nature. 1954; 174:269–270. doi: 10.1038/174269c0 PMID: 13185286

33. Ramachandran G, Kartha G. Structure of collagen. Nature. 1955; 176:593–595. doi: 10.1038/176593a0 PMID: 13265783

34. Rich A, Crick F. The Structure of Collagen. Nature. 1955; 176:915–916. doi: 10.1038/176915a0 PMID: 13272717

35. Bella J, Eaton M, Brodsky B, Berman HM. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. Science. 1994; 266(5182):75–81. doi: 10.1126/science.7695699 PMID: 7695699

36. Fox NK, Brenner SE, Chandonia JM. SCOPe: Structural Classification of Proteins–extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Res. 2014 Jan; 42 (Database issue):D304–D309. doi: 10.1093/nar/gkt1240 PMID: 24304899

37. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. Current opinion in structural biology. 2011; 21(3):432–440. doi: 10.1016/j.sbi.2011.03.011 PMID: 21514144

38. Geist L, Henen MA, Haiderer S, Schwarz TC, Kurzbach D, Zawadzka-Kazimierczuk A, et al. Protonation-dependent conformational variability of intrinsically disordered proteins. Protein Science. 2013; 22(9):1196–1205. doi: 10.1002/pro.2304 PMID: 23821606

39. Dosztányi Z, Mészáros B, Simon I. Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Brief Bioinform. 2010 Mar; 11(2):225–243. doi: 10.1093/bib/bbp061 PMID: 20007729

40. Staneva I, Huang Y, Liu Z, Wallin S. Binding of two intrinsically disordered peptides to a multi-specific protein: a combined Monte Carlo and molecular dynamics study. 2012;.

41. Chong SH, Ham S. Conformational entropy of intrinsically disordered protein. The Journal of Physical Chemistry B. 2013; 117(18):5503–5509. doi: 10.1021/jp401049h PMID: 23531173

42. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences. 2013; 110(33):13392–13397. doi: 10.1073/pnas.1304749110

43. Mittal A, Lyle N, Harmon TS, Pappu RV. Hamiltonian Switch Metropolis Monte Carlo simulations for improved conformational sampling of intrinsically disordered regions tethered to ordered domains of proteins. Journal of chemical theory and computation. 2014; 10(8):3550–3562. doi: 10.1021/ct5002297 PMID: 25136274

44. Baruah A, Rani P, Biswas P. Conformational Entropy of Intrinsically Disordered Proteins from Amino Acid Triads. Scientific reports. 2015; 5. doi: 10.1038/srep11740

45. Kosol S, Contreras-Martos S, Cedeño C, Tompa P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. Molecules. 2013; 18(9):10802–10828. doi: 10.3390/molecules180910802 PMID: 24008243

46. Sibille N, Bernado P. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochemical society transactions. 2012; 40(5):955–962. doi: 10.1042/BST20120149 PMID: 22988847

47. Jensen MR, Ruigrok RW, Blackledge M. Describing intrinsically disordered proteins at atomic resolution by NMR. Current opinion in structural biology. 2013; 23(3):426–435. doi: 10.1016/j.sbi.2013.02.007 PMID: 23545493

48. Konrat R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. Journal of Magnetic Resonance. 2014; 241:74–85. doi: 10.1016/j.jmr.2013.11.011 PMID: 24656082

49. Zhao Y, Schultz NE, Truhlar D. Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. The Journal of chemical physics. 2005; 123(16):161103. doi: 10.1063/1.2126975 PMID: 16268672

50. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec; 22(12):2577–2637. doi: 10.1002/bip.360221211 PMID: 6667333

51. Joosten RP, te Beek TAH, Krieger E, Hekkelman ML, Hooft RWW, Schneider R, et al. A series of PDB related databases for everyday needs. Nucleic Acids Res. 2011 Jan; 39(Database issue):D411–D419. doi: 10.1093/nar/gkq1105 PMID: 21071423

52. Chebrek R, Leonard S, de Brevern AG, Gelly JC. PolyprOnline: polyproline helix II and secondary structure assignment database. Database. 2014; 2014:bau102. doi: 10.1093/database/bau102 PMID: 25380779

53. Matthew ZT, Sydykova DK, Meyer AG, Wilke CO. PeptideBuilder: A simple Python library to generate model peptides. PeerJ. 2013; 1:e80. doi: 10.7717/peerj.80

54. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25(11):1422–1423. doi: 10.1093/bioinformatics/btp163 PMID: 19304878