Received: 6 May 2020 | Revised: 2 September 2020 | Accepted: 29 September 2020

DOI: 10.1002/jcc.26448

FULL PAPER

Journal of COMPUTATIONAL CHEMISTRY WILEY

Protein secondary structure motifs: A kinematic construction

Mosavverul Hassan[1] | Evangelos A. Coutsias[1,2]

[1]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York

[2]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York

Correspondence
Mosavverul Hassan, Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794.
Email: mosavverul.hassan@stonybrook.edu

Funding information
National Institutes of Health, Grant/Award Number: GM107104

Abstract

The kinematic geometry of protein backbone structures, constrained by either single or multiple hydrogen bonds (H-bonds), possibly in a periodic array, is discussed. These structures include regular secondary structure elements α-helices and β-sheets but also include other short H-bond stabilized irregular structural elements like β-turns. The work here shows that the variations observed in such structures have simple geometrical correlations consistent with constrained motion kinematics. A new classification of the ideal helices is given, in terms of the parameter α, the angle at a C_α atom to its two neighboring C_α's along the helix, and shown how it can be generalized to include nonideal helices. Specifically, we derive an analytical expression of the backbone dihedrals, (ϕ, ψ), in terms of the parameter α subject to the constraint that the peptide planes are parallel to the helical axis. Helices constructed in this way exhibit near-vertical alignment of the C = O and N – H units and are the canonical objects of this study. These expressions are easily modifiable to include perturbations of parameters relevant to nonplanar peptide units and noncanonical angles. The addition of a second parameter, ε_0, inclination of successive peptide planes along a helix with respect to the helical axis leads to a generalization of the previous expression and provides an efficient parametrization of such structures in terms of coordinates consistent with H-bond parameters. An analogs parametrization of β-turns, using inverse kinematic methods, is also given. Besides offering a unifying viewpoint, our results may find useful applications to protein and peptide design.

KEYWORDS

helices, kinematics, secondary structure elements, β-strands, β-turns

1 | INTRODUCTION

A regular *helical polygon* is a polygonal line in $\mathbb{R}^3$ space in which all segments have equal lengths, make equal angles with each other, and successive triplets form equal dihedral angles (see Figure 1). By extension, a helical structure results naturally by considering a chain of identical units (see the definition for *unit-SSE's* in Section 2.1) connected in such a way that the successive connections form a regular helical polygon.[1] Helical structures abound in bio-macromolecules such as proteins, nucleic acids, and other polymers, combining structural stability with packing efficiency. Corey et al., in their seminal work on stable polypeptide configurations, deduced the characteristics of the two *regular* structural motifs in proteins, α-helices (see Table 1[2]), and β-sheets.[3] In a series of articles, on the stereochemical requirements of the polypeptide chains, Ramachandran et al.,[4] provide a detailed description of all possible helical conformations. The helical region in the Ramachandran plot shows its precise extent in (ϕ, ψ) space.

In this work we discuss aspects of the inverse problem, namely determining (ϕ, ψ) pairs consistent with given helical parameters or other secondary structure descriptors, with focus on the role played by the intramolecular H-bonds in the secondary structure elements, regular and irregular, of a protein. While any sterically allowed pair of backbone dihedrals (ϕ, ψ) results in a helical structure if repeated

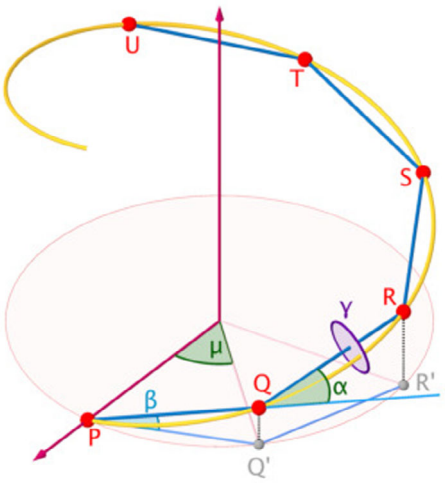

FIGURE 1 A schematic representation of a regular helix defined by a polygonal line in space. The red dots represent the pattern of C_α atoms. Intrinsic parameters (α, μ, β, γ) required for such a construction are as labeled. The Equations (5) and (6) define the relation among the intrinsic parameters assuming the repetitive pattern of C_α atoms

along an ideal polypeptide chain, it is only a narrow range of these values that result in biologically relevant helices.[4] The stability of these arrangements depends on their ability to form intramolecular H-bonds between the successive levels of the helical ladder. Although mathematical descriptions for a 2-D search over the backbone dihedrals (ϕ, ψ) for compatible values by construction and testing exist, analytical derivations of the compatible pairs are currently missing. By using ideas from kinematics, we give analytical expressions for such pairs that lead to helices with backbone geometries favorable for H-bonds. Unlike previous works, where an H-bond is the result of a search, our approach incorporates H-bond formation implicitly.

One of the earliest pieces of evidence regarding the crucial role of H-bonds in biomolecules comes from the crystal analysis of cell macromolecules. Subsequent efforts directed toward the prediction of stable tertiary protein structures using experimentally derived interatomic distances and angles led to the suggestion of patterned H-bond structures, α/γ-helices, the parallel and anti-parallel β-sheet configurations.[5] The two main assumptions under which the constructed α-helical structures were shown to exist, in the stable conformations of fibrous proteins and synthetic polypeptide chains, are (1) the trans-planarity of the amide bond and (2) the stereochemical requirement for H-bond formation, N···O ≈ 2.72 Å, ∡HNO ≤ 30° (see Figure 2).[6,7] The orientations assumed by the functional groups involved in the intramolecular H-bond formation within the α-helical structure in these stable conformations were such that the H-bonds were near parallel to the helical axis. The orientation of the participating functional groups with respect to the helical axis is an important geometrical construct in the sense that it is responsible for the H-bond orientation, parallel to the helical axis or away from it.

A similar approach coupled with the cis–trans orientations of atoms about the single bonds adjacent to the C_α atoms along the backbone, resulted in two stable pleated β-sheets type structures.

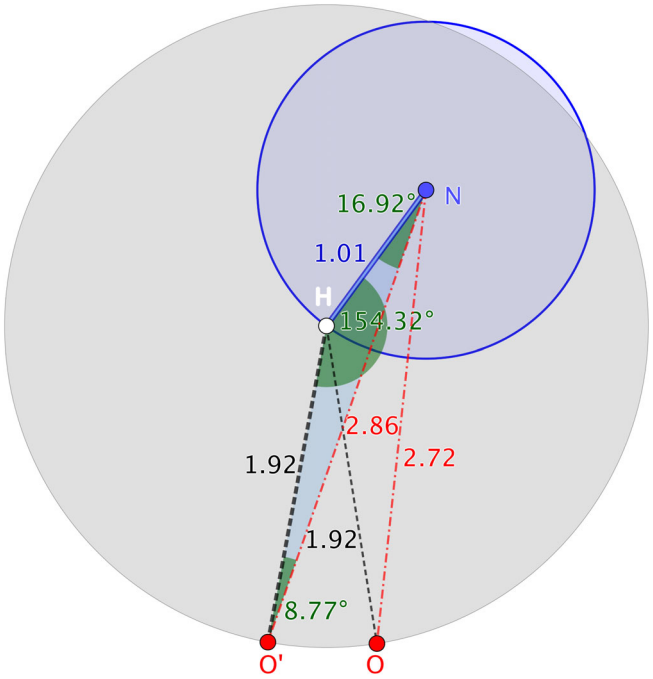

FIGURE 2 H-bond distance-related geometry with fixed ∡HNO = 30°. The blue circle shows the locus of the hydrogen atom (H) covalently bonded to the nitrogen (N). The gray circle shows the locus of oxygen (O) atom bonded noncovalently to the hydrogen (H). The d(N···O) = f(d(N − H), d(H···O), ∡NHO). The position of the H-atom causes the d(H···O) distance to vary from 1.71 Å, collinear position ∡NHO = 180° to 1.92 Å, current position ∡NHO′ ≈ 154°

Among these two configurations, one was disfavored by all residues except glycine due to side-chain sterics. The favored pleated β-sheet configuration demonstrated a lateral arrangement of H-bonds in which the consecutive residue C_α – C_α distance is ≈ 3.34 Å, possibly due to the distortions in the trans-planarity of the peptide bonds, lateral C_α – C_α distances ≈ 4.75 and 4.85 Å for antiparallel and parallel, respectively, and ∡HNO ≈ 6° − 10°. The lateral H-bond distance as a consequence of these metrical properties falls within the range of those proposed for helical structures, for example, 2.75 Å.[8]

The construction of the helix and pleated sheet structure, a result of an exhaustive exploration for H-bond structures within a protein, may be understood as an optimal arrangement of the planar peptide units, joined at the C_α atoms, that maximizes the formation of the H-bonds with prescribed stereochemical properties. This construction process is pivotal to the determination of protein tertiary structures, the arrangement of the helices and sheets, for two reasons. First, it underlies most of the structural parameters which may be used in the construction of patterned H-bond structures and second, broaches the study of the conformational flexibility of the polypeptide chains under the energetically favored trans-planarity of the amide bond geometry, that is, the Pauling–Corey peptide unit.[3]

Ramakrishnan et al., recognized that questions related to the conformational flexibility of the polypeptide chain are of much more general nature than those undertaken earlier since they must describe, with reason, not only those conformations which are sterically and

energetically feasible, but also provide necessary and sufficient conditions to distinguish them from those that are not. The former objective depends on the latter. For example, if the necessary conditions for unfeasible conformations are known they could be used to describe sterically feasible conformations. If these conformations also show a close agreement with experimentally derived data, the requirements for sufficiency and energetic feasibility are automatically satisfied. This details the general approach followed by Ramakrishnan et al. for the description of feasible conformations.[9,10] These descriptions require parameters that affect the relative orientations of the peptide planes in the polypeptide chains. The rotatable bonds adjacent to the C_α atoms, the ϕ, ψ dihedrals, serve this purpose.

The helical parameters H_p, the unit twist, unit translation and the direction cosines of the helical axis, expressed as a function of the dihedral angle parameters, were used to determine the general nature of the helical conformations. This formulation, $H_p = F(\phi, \psi)$ does not provide a priori bounds on the input dihedrals and was thus coupled with the distance contact restraints to serve as a guide in the search for feasible conformations.[4] Specifically, the procedure adopted requires the calculation of the atomic coordinates based on different discrete combinations of the dihedrals ϕ, ψ. Nonbonded atomic-distances calculated from the atomic coordinates were in turn evaluated against a predefined, van der Waals radii dependent, minimum contact atomic-distance list. Conformations that violate the minimum contact distance criteria indicate a steric clash and were subsequently discarded. Contact distances that fall within the favorable H-bond forming range determine the relevant dihedral ranges. The resulting $\phi - \psi$ conformational map outlines the allowed flexibility, in a helical formation, of a peptide chain. The associated dihedrals of the final ensemble specify the feasible, sterically and energetically, range of the helical parameters H_p.

The polypeptide conformations, generated over the feasible dihedral angle range, were analyzed for two different types of H-bonds, forward and backward. Forward H-bonds, associate the oxygen of the first residue to the nitrogen of the latter, while the backward H-bonds, associate the oxygen of the latter residue to the nitrogen of the first. The results obtained in the context of the forward and backward H-bonds are of particular significance since they rule out the presence of backward H-bond conformations especially the γ-helices, of the kind $H_n \cdots O_{n+4}$, $n \geq 2$, proposed by Pauling et al. The α–helical formation was also noted as sensitive to the range of the $\measuredangle NC_\alpha C$, and the allowed perturbation for this angle is ±2° from its canonical value ($\measuredangle NC_\alpha C \approx 109.5°$). Besides α–helices, 3_{10} and π helices were also identified (see Table 1).

Irregular secondary structure elements, sections of the loop regions which allow peptide chains to make abrupt directional changes, are known connectives of the *regular* secondary structure elements in protein structures and form an integral part of cyclic peptides.[11] Such units are nonhelical in nature and exhibit a conserved, but variable, H-bond geometry. Venkatachalam et al. characterized such peptide conformations, dihedral angle (ϕ, ψ) based descriptive approach extended to three linked peptide units at the C_α atom, to three different types with H-bonds between residues (n, $n+3$).[12]

TABLE 1 Types of helices in biological molecules

Helix type	H-bond residue	Avg helix length	Observed frequency[a]
3_{10}	(n, $n+3$)	3 – 5 res	83%
α	(n, $n+4$)	10 – 12 res	78.4%
π	(n, $n+5$)	5 – 8 res	1.8% (this percentage is assumed to be much higher ≈15%)[2]

[a]Secondary structure method DSSP is used here to determine the helix type content in each protein structure. The values correspond to a percentage of 161757 RCSB structures.

Amino-acid propensity analysis based on feasible conformations of the linked peptide units suggest a maximum likelihood of glycine at the $n+2$ residue position (see Section SI.6). The availability of high resolution protein structures has enabled further classification of protein turns into 18 types based on residue contacts and H-bonds.[13]

It is apparent that the descriptive approach complements the earlier construction methodology proposed by Pauling et al. by providing a detailed observation of the geometrical nature of polypeptide chains and its helical formation using two linked peptide groups at the C_α atom. However, both methods use the H-bond criteria as a tool for either construction or characterization of the polypeptide chain conformation and neither attempts an ab-initio investigation of the H-bond geometry itself. Such an investigation must adhere to a construction method where the H-bond geometry is not a subject for search but a consequence of the polypeptide chain conformation. Finding a more natural devise that allows control over H-bond formation is one of the main motives of the work undertaken here.

Studies that pertain specifically to the analysis of H-bonds, both intra, and inter molecular, are forked in two main directions. The first is concerned with the recognition of H-bonds and statistical analysis of its associated geometry while the second concerns itself with the effects of H-bond formation on the molecular geometry of macromolecules, leading to a thorough understanding of its structure–function relation. Both experimental and theoretical techniques are applied to these two types of studies wherein the validation process requires the ability of the theoretical methods to substantiate the experimentally determined qualitative data. Within this setting, an H-bond formation is defined as an attractive force based on the electronegativities of the donor-acceptor atoms giving rise to the widely accepted, although debatable, electrostatic nature of the H-bond in biomolecules.[14,15] There is, however, an associated partial covalent character to this bond which institutes a *directionality*, (see Figure 2) the direction of approach of the proton donor toward the proton acceptor in relation to the bond(s) of the acceptor. The nature of the acceptor atoms is responsible for this partial covalent character and directionality is a consequence of an anisotropic electron density distribution toward the hydrogen atom.[16] Statistics from crystallographic data on the varied nature of bimolecular H-bonds suggest that the directional requirement of the acceptor atom, usually the oxygen atom, is (1) not strong and (2) affected by the polarizability of the

electron density, possibly induced by the configuration of the functional group to which it is attached.[17] In general, the H-bond directionality strengthens with a heightened covalent character and is observed geometrically as its tendency toward a collinear arrangement of the donor, hydrogen, and acceptor atom.

The definition of an H-bond has evolved. It serves as a precursor to its recognition and is complemented by theoretical studies which provide a basis for its intrinsic descriptors.[18,19] The H-bond definition highlights two of its main characteristics, strength, and directionality, where the latter may not necessarily depend on the former.[20] Experimental studies clearly distinguish strong and weak H-bonds.[21,22] In contrast, there is no correlation between the strengths of the various types of weak H-bonds through their exhibited directionalities. Statistical studies which attempt to differentiate between weak H-bonds and the van der Waals attraction show that even the weakest H-bonds exhibit a range of directional attributes.[23] It is also deceptive to associate the strength of the H-bond with the distance of approach since such associations do not take into consideration the nature of the donor-acceptor atoms,[17] although environmental conditions which affect the bond strength have been considered.[24–26] Strength is associated with the total energetics of the H-bond formation and, it is well accepted that it has no single energetic contributor.[27–29] Quantum mechanical methods based on the molecular orbital theory approximate the energetic contributions as the sum of electrostatics, charge-transfer, polarization, exchange-repulsion, and dispersion. The first three terms are attractive and the last two terms repulsive in nature. Although theoretical studies target the exact nature of the H-bond formation, which includes the coupling of the different contributing factors, they are limited in their applicability to systems with many degrees of freedom. Ground state molecular geometry predictions based on theoretical studies for biomolecular systems, where the competing nature of inter and intramolecular H-bonds is significant, must also be used with caution. Bias toward the latter may result in unfeasible collapsed conformations.[30,31] In general, there is a disconnect between the geometrical descriptors of the H-bond, referenced in prior discussions, and its strength. Theoretical studies that attempt to rationalize directionality thus seek to quantify the energetic difference of the different H-bond contributors as a function of this characteristic.[32–34]

An accurate representation of the H-bond geometry is one of the major bottlenecks in biomolecular simulations of proteins and nucleic acids employing widely used force fields or their variants. These force fields, empirically or theoretically parameterized, treat the noncovalent interactions in terms of the isotropic van der Waals and electrostatic interaction between fixed dipoles. The putative argument based on the optimization of the parameters and charge refinement procedures in the molecular mechanic's force field seems to suggest a balance sought between the spectroscopic properties of the H-bond and the overall, computationally efficient, conformational accuracy of the biomolecules.[17] A ramification of this balance often results in H-bond geometries that differ significantly from those predicted theoretically. The atom centered charge assignment in these force fields tends to bias the H-bond geometries toward a collinear arrangement of the participating atoms.[35,36] As a remedy, off center charge assignment methods, placing fixed or iteratively obtained partial atomic charges at defined sites away from the atom centers, or multipole (MTP) electrostatic interactions are used in force fields to replicate the geometrical changes observed as a consequence of the lone-pair (LP) effect.[37–40] Curative measures for polarizability induced geometric changes, for a consistent transferability among the different bimolecular dielectrics, are incorporated in polarizable force fields.[41] Force fields which attempt to explicitly account for quantum mechanically derived energetics, charge transfer (CT) effects which cause an elongation of the H-bonds, of the observed H-bond geometries do so by including specific angular terms, $\measuredangle O: \measuredangle 17-4-3$, $\measuredangle H: \measuredangle 13-17-4$, or $\measuredangle N: 4-13-17$ or their combinations (see Section 3.6) in an additive manner.[42,43]

Knowledge-based potentials provide a simple, computationally efficient, alternative to account for the H-bond geometries in crystallized structures. These potentials, developed for both protein structure prediction methods and H-bond geometrical variability, rely on the central idea that the empirically observed frequency of certain *structural properties* (S_p) correspond to the Boltzmann like probability density function.[44,45] Owing to the natural caveats of such an assumption, that is, the ensemble of high-resolution structures are not in thermal equilibrium, Grzybowski et al., provide a statistical framework for such an assumption wherein it is assumed true provided S_p is a function of some intrinsic property of the molecular structures in the ensemble.[44,46] This rationalization provides for an additive H-bond potential, based on the *orientation* of the H-bond geometry, designed to encompass both the anisotropic charge distribution toward the hydrogen atom and the electrostatic effects captured by the point charge models.[47,48] The additive nature of the H-bond potential developed by Kortemme and Morozov is an attempt to decouple the parameter set interdependence, in absence of a single intrinsic parameter, and serves as a simplification of the H-bonds complex energy landscape. The orientation-dependent geometric descriptors, scrutinized in a high-resolution protein database, are the bond length $\delta_{HA,KM}: 17-4$, bond angles $\measuredangle\Theta_{KM}: \measuredangle 17-4-3$, and $\measuredangle\Psi_{KM}: \measuredangle 13-17-4$, and dihedral $X_{KM}: 17-4-3-5^1$ and their associated energies are obtained as the negative logarithm of their frequency distribution.[35]

In general, there are six parameters involved in the formation of the H-bond, two in addition to those used by Kortemme et al. These are torsions about $N-H: 11-13-17-4$ and the $O\cdots H: 13-17-4-3$. The inverse kinematic solution to the H-bond geometrical problem developed here allows an exploration of the full six parameter set, including its implicit interdependence, as a function of a single parameter α (see Section 4). It is also possible to explore the inverse parameter set interdependence problem using the *Builder for Recursive Inverse Kinematic Assembly and Ring Design* (BRIKARD) application[49] as implemented in this work for H-bonds in β-turns (for the parameter α range see Section 3.5). The parameter α thus serves as the intrinsic parameter and may supplement or replace the current

parameter set in H-bond potentials. Such an H-bond potential, although not explicitly presented here, would simplify substantially the dependence on relevant empirical data and the strain on parametrization to correctly reproduce the different test results. The kinematic-geometric analysis of the secondary structure elements demonstrates the viability of such a program.

We focus exclusively on the kinematics and geometric analysis of these building blocks of the biomolecular machinery, and show that (1) the helical parameters H_p, are functions of the intrinsic parameters, (2) the construction of the ideal helical structures depends solely on the parameter α, (3) the H-bond geometries exhibited for ideal helices, represented by the six parameter set (see Section 3.6), are a functions of the parameter α, (4) with set peptide plane orientation, controlled by the value of ε_0, or with set parameter α values (see Section 3.4) it is possible to explore the (ϕ, ψ) space and construct nonideal helices (see Section 4),[50] and (5) the inverse kinematic construction, adept to exhaustively explore β-turn backbone torsions, can implicitly explore the inverse H-bond parameter relations and the resulting β-turn *Bricard Curves* (2-D figures, Section 3.5) present a simple and efficient way of exploring torsional interdependence.

This article is organized as follows: In Section 2, we begin with the geometrical underpinning for the SSE motifs and provide formulas which relate H_p to the intrinsic parameters μ, β, γ, and α. A mathematical expression for the peptide plane orientation parallel to the helical axis Z, is included in the *Tetrahedral Equation* (see Equation (21)) and analytical expressions are presented to convert the solutions of the *Tetrahedral Equation* to the backbone dihedrals (ϕ, ψ). Next, the general theoretical foundation for all possible peptide plane orientations is developed. Results of the kinematic equations developed in Section 2 are presented in Section 3. Step-by-step construction protocols are provided here and the implications of the results obtained are discussed in relation to the experimentally derived data. Our findings are summarized in Section 4. The Supporting Information contains mainly the derivations of the various relations obtained in Section 2. Figures related to the peptide plane perturbations, additional β-turn *Bricard Curves* and other details are also presented in this section for a complete overview of the results in Section 3.

2 | METHOD

2.1 | Geometrical construction for helices and β-strands

Consider a regular right handed helix and introduce a collection of right handed co-ordinate systems with Z axis along the helical axis. A projection of the helix onto the X – Y plane is therefore a circle $\mathcal{C}$ of some radius (say *r*).

Arrange the backbone of a polypeptide chain so that the n^{th} C_α atom is placed on the helix at a distance *r* from the origin, along the positive X-axis. For simplicity in the sequel, we define the length unit as the canonical distance between two successive C_α atoms in a *trans* conformation

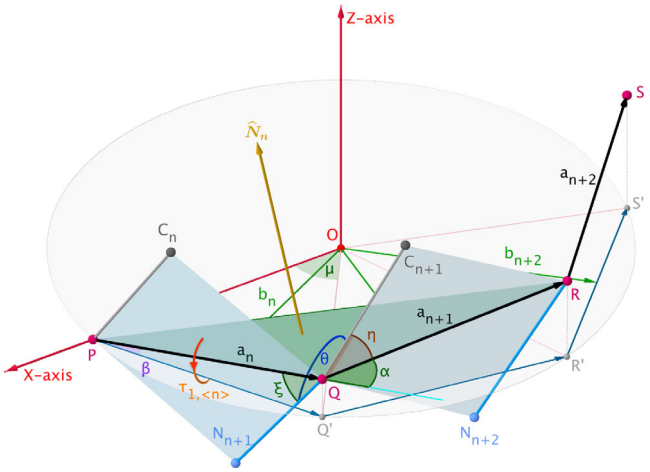

FIGURE 3 Geometrical skeletal of a helix. P, Q, R and S shown as red dots represent the C_α atoms forming a helical polygon

(C_α – C_α distance is ≈ 3.8 Å). Let the unit vector along $\{C_{\alpha,n}, C_{\alpha,n+1}\}$, $\hat{a}_n$, make an angle β with the X – Y plane and let the angle between the two consecutive vectors $\hat{a}_n$ and $\hat{a}_{n+1}, n = \{1,2,\ldots\}$ be denoted by the parameter α. The projection of $\{C_{\alpha,n}, C_{\alpha,n+1}\}$, n = {1, 2, …} onto the *X* – *Y* plane subtends an angle μ at the origin O. The unit normal to the plane <*n*> formed by the triad $\{C_{\alpha,n}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$ is denoted as $\hat{N}_n$. The dihedral γ defined by the tetrad $\{C_{\alpha,n}, C_{\alpha,n+1}, C_{\alpha,n+2}, C_{\alpha,n+3}\}$ is therefore the angle between the unit vectors $\hat{N}_n$ and $\hat{N}_{n+1}$ (not shown in Figure 3). In Figure 3, P, Q, R, and S represent the consecutive C_α atoms. PQ' represents the projection of $\hat{a}_n$ onto the X – Y plane. We denote the unit vector ⊥ to the plane containing $\Delta PQQ'$ by $\hat{b}_n$. The angles ξ, θ, and η are as defined in Figure 3.[51] A tetrahedral is defined as, *an arrangement of four rotors in a cyclic sequence such that each makes a given angle with its predecessor and successor*.[52] We define a *unit-SSE* as such a tetrahedral.[53] In general, the angles ξ, θ, η, and α are variable across the different planes <*n*> formed by the triad $\{C_{\alpha,n}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$. However, these values are held constant under the assumption of identical repeated *unit-SSE's*. The numerical values of ξ, θ, and η used in all calculations are consistent with the trans-planarity of the peptide bonds which require $C_{\alpha,n} - C_{\alpha,n+1}$, n = {1, 2, …} distance as ≈ 3.8 Å (see Section 2.6.2 for the values of the tetrahedral variables used).

2.2 | Vector definitions for Whitworth's relation

Considering $\Delta OPQ'$, the radius *r* of $\mathcal{C}$, using from the law of sines, is obtained as a function of the intrinsic parameters (see Figure 1) that is,

$$r = \frac{\cos\beta}{2\sin\frac{\mu}{2}}. \tag{1}$$

From $\Delta PQQ'$, we define the rise $h = \bar{Q'Q}$, that is, $h = \sin\beta$. With the defined co-ordinate system we may write the unit vector

$$\hat{a}_1 = \cos\beta\left(-\sin\frac{\mu}{2}\right)\hat{i} + \cos\beta\cos\frac{\mu}{2}\hat{j} + \sin\beta\,\hat{k}. \quad (2)$$

The consecutive vector $\hat{a}_2$ can be obtained by rotating about the Z axis by an angle μ, $\hat{a}_2 = \mathcal{R}_z(\mu)\hat{a}_1$, where $\mathcal{R}_z(\mu) = \begin{pmatrix} \cos\mu & -\sin\mu & 0 \\ \sin\mu & \cos\mu & 0 \\ 0 & 0 & 1 \end{pmatrix}$, so that

$$\hat{a}_2 = \cos\beta\left(-\sin\frac{3\mu}{2}\right)\hat{i} + \cos\beta\cos\frac{3\mu}{2}\hat{j} + \sin\beta\,\hat{k}.$$

In general, with $\hat{a}_n = \mathcal{R}_z(\mu)\hat{a}_{n-1}$, we have

$$\hat{a}_n = -\cos\beta\sin\mu_n\,\hat{i} + \cos\beta\cos\mu_n\,\hat{j} + \sin\beta\,\hat{k}; \mu_n = \frac{(2n-1)\mu}{2} \& n = \{1,2,\ldots\}. \quad (3)$$

By definition, $\hat{N}_n = \kappa\hat{a}_n \times \hat{a}_{n+1}$ where $\kappa = 1/\sin\alpha$, so that

$$\hat{N}_n = \kappa(\sin\beta\cos\beta\{\cos\mu_n - \cos\mu_{n+1}\}\hat{i} + \sin\beta\cos\beta\{\sin\mu_n - \sin\mu_{n+1}\}\hat{j} + \cos^2\beta\sin(\mu_{n+1} - \mu_n)\hat{k}), \quad \text{where } \mu_{n+1} - \mu_n = \mu. \quad (4)$$

From the definition for $\hat{N}_n$, we have $\sin\alpha = |\hat{a}_n \times \hat{a}_{n+1}|$. Using Equation (4) we can show that

$$\sin\frac{\mu}{2}\cos\beta = \sin\frac{\alpha}{2}. \quad (5)$$

Since, γ is the angle between planes <n> and <n + 1>, the relation $\cos(2\pi - \gamma) = \hat{N}_n \cdot \hat{N}_{n+1}$ holds. Again, using Equation (4) we can show that

$$\cos\frac{\gamma}{2}\cos\frac{\alpha}{2} = \cos\frac{\mu}{2}. \quad (6)$$

A formal derivation of the Whitworth's relation, Equations (5) and (6),[54] consistent with the above described geometry is provided in the Supporting Information (see Section SI.1).

2.3 | Geometry based torsional relation for SSE's

Here, we derive an expression relating torsions $\tau_{1,<n>}$ and $\tau_{2,<n>}$, $n = \{1, 2, \ldots\}$ where $\tau_{1,<n>}$ and $\tau_{2,<n>}$ are the angles between the planes formed by atoms $\{C_n, C_{\alpha,n}, C_{\alpha,n+1}\}$ and $\{C_{n+1}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$ with <n>, respectively. For a regular helix $\hat{b}_n$ represents the surface normal of the plane containing $\Delta PQQ'$ and hence the plane formed by $\{C_n, C_{\alpha,n}, C_{\alpha,n+1}\}$. This gives us

$$\cos\tau_{1,<n>} = \hat{b}_n \cdot \hat{N}_n, \quad \cos\tau_{2,<n>} = \hat{b}_{n+1} \cdot \hat{N}_n \text{ where } \hat{b}_n = \cos\mu_n\,\hat{i} + \sin\mu_n\,\hat{j}, \quad (7)$$

which on simplification using Equation (4) results in

$$\cos\tau_{1,<n>} = \kappa\sin\beta\cos\beta\{1 - \cos\mu\} = \sec\frac{\alpha}{2}\sin\beta\sin\frac{\mu}{2}. \quad (8)$$

$$\cos\tau_{2,<n>} = -\kappa\sin\beta\cos\beta\{1 - \cos\mu\} = -\cos\tau_{1,<n>}. \quad (9)$$

For a complete derivation of the relation obtained in Equation (9) we refer the reader to the Supporting Information (see Section SI.2).

2.3.1 | An approximate formula from Miyazawa's expression

An intricate geometrical approach was developed by Mizushima et al., to express H_p as a function of a repetitive pattern of single or multiple atom type positions defining a helical polygon in $\mathbb{R}^3$ and is isolated from the backbone dihedral related work, $H_p = F(\phi, \psi)$ of Ramachandran et al. We supply a complete set of relations among the intrinsic parameters μ, β, γ, and α missing in this treatment through our geometrical approach and show that parameter interdependence is indeed dissociable. Well known approximations originating from the intricate treatment are also discussed and its limitations are highlighted in this section.

The mathematical treatment of the helical conformation as a function of the internal coordinates, bond lengths, bond angles and torsions about the covalent rotatable bonds, attempted earlier by Mizushima et al., expresses μ (angle subtended at the center, see Figure 1), $h = \sin\beta$ (assuming $C_\alpha - C_\alpha$ distance as unit) and radius r (assuming all atoms are of the same type) as a function of the consecutive $C_\alpha - C_\alpha$ distance, $\pi - \alpha$ and γ.[55] There are two main points we make here regarding this rigorous mathematical treatment. First, the complicated formulas arrived at, resulting from a relation between two right handed Cartesian coordinate transformations, are essentially the Whitworth's relation. These are easily arrived at from our fairly simple geometrical treatment using the intrinsic parameters α, μ, β, and γ. To see this, we direct our attention to a simplification of these results under different assumptions (1) All atoms considered are of the same type that is, a repetitive pattern of C_α atoms and (2) All atoms considered are of multiple types that is, a repetitive pattern of two or more atom types. For example, $\{N_n, C_{\alpha,n}\}$ atoms or $\{C_{\alpha,n}, C_n\}$ atoms or $\{N_n, C_n\}$ atoms or a repetitive pattern of backbone atoms $\{C_{\alpha,n}, C_n, N_{n+1}\}$.[56] Second, the construction of a helical string passing through a polygonal line or frame in space requires only a single atom type. In our case, this is the repetitive pattern of C_α atoms forming the polygonal line and is afforded to us through the parameter α search. The dihedrals (ϕ, ψ) are calculated based on this pattern (see Table 2).

To make a direct comparison with the trigonometric relations of Miyazawa et al., for the single atom type, we substitute for the variables $\theta_M \equiv \mu$, $\tau_M \equiv \gamma$, $\phi_M \equiv \pi - \alpha$ in Miyazawa's eq. 9: $\cos\frac{\theta_M}{2} = \cos\frac{\tau_M}{2}\sin\frac{\phi_M}{2}$. This gives us,

TABLE 2 Key expressions

Input → output	Expression
$\{\xi, \theta, \eta, \alpha\} \rightarrow \tau_1$	$U = \frac{-(D\pm C+B)\pm\sqrt{(D\pm C+B)^2-4EA}}{2E}$, Equation (23)
	where $U = u^2 = \tan^2\frac{\tau_1}{2}$,
$\{\alpha, \tau_1\} \rightarrow \mu$	$\tan\frac{\mu}{2} = \pm\sqrt{\frac{1}{\sin^2\tau_1\cos^2\alpha/2}-1}$, Equation (26)
$\{\alpha, \mu\} \rightarrow \beta$	$\sin\frac{\mu}{2}\cos\beta = \sin\frac{\alpha}{2}$, Equation (5)
$\{\alpha, \mu\} \rightarrow \gamma$	$\cos\frac{\gamma}{2}\cos\frac{\alpha}{2} = \cos\frac{\mu}{2}$, Equation (6)
$\{\mu, \beta\} \rightarrow r$	$r = \frac{\cos\beta}{2\sin\frac{\mu}{2}}$, Equation (1)
$\{\tau_1, \xi, \theta, \eta, \alpha\} \rightarrow \phi$	$x = \frac{1}{C_{\phi,\tau_1}u}\left\{E_{\phi,\tau_1}u^2 + D_{\phi,\tau_1} - \left[\frac{G_2v^2+J_2}{F_2v^2+H_2}\right](B_{\phi,\tau_1}u^2 + A_{\phi,\tau_1})\right\}$, Equation (31)
$\{\tau_1, \xi, \theta, \eta, \alpha\} \rightarrow \psi$	$y = \frac{1}{C_{\tau_2,\psi}v}\left\{B_{\tau_2,\psi}v^2 + E_{\tau_2,\psi} - \left[\frac{G_1u^2+J_1}{F_1u^2+H_1}\right](A_{\tau_2,\psi}v^2 + D_{\tau_2,\psi})\right\}$, Equation (32)

$$\cos\frac{\mu}{2} = \cos\frac{\gamma}{2}\cos\frac{\alpha}{2},$$

which is Equation (6) (see Table 2).

Similarly, Miyazawa's eq. $10: d_M\sin\frac{\theta_M}{2} = r_M\sin\frac{\tau_M}{2}\sin\frac{\phi_M}{2}$ takes the form

$$\begin{aligned} &d_{C_\alpha-C_\alpha}\sin\beta\sin\frac{\mu}{2} = d_{C_\alpha-C_\alpha}\sin\frac{\gamma}{2}\sin\frac{\pi-\alpha}{2}, \text{where } \frac{\gamma}{2} = \frac{\pi}{2}+\tau_1, \\ &\Rightarrow \cos\tau_1 = \sec\frac{\alpha}{2}\sin\beta\sin\frac{\mu}{2}, \end{aligned} \tag{10}$$

which is Equation (8).

In particular, we note that Whitworth's relation Equation (5) is missing in this work, which results in the reliance on the two parameters μ and β to calculate γ, using Equations (6) and (10) together, and hence α using Equation (6) separately. Thus, with the internal coordinates, bond length $d_{C_\alpha-C_\alpha}$, bond angle $\pi-\alpha$ and torsion γ about the virtual $C_\alpha-C_\alpha$ bond now known, it is possible to calculate the Cartesian coordinates of the polygonal line. In contrast, our only reliance is on the parameter α to calculate the helical parameters (μ, β, γ) given the kinematic approach which results in solutions for τ_1. The path taken to calculate the backbone dihedrals (ϕ, ψ) from (α, τ_1) is self-contained and the a priori requirement of dihedrals for the calculation of the helical parameters and vice versa is redundant (see Figure 4).

The three atom type consideration, repeats of the kind $\{C_{\alpha,n}, C_n, N_{n+1}\}$, although not necessary to analyze in the context of helical construction, gives rise to a simplified formula for μ in terms of the dihedral angles (ϕ, ψ),[57] that is,

$$3\cos\mu = 1-4\cos^2\left(\frac{\phi+\psi}{2}\right). \tag{11}$$

We show here that this simplified formula is an approximation derived from Miyazawa's eq. (36) and not an exact formulation. Assuming the Pauling-Corey geometry for trans peptides, Miyazawa et al. simplify their eq. (36) to

$$\cos\frac{\theta_M}{2} = -0.817\sin\left(\frac{\tau_{12,M}+\tau_{31,M}}{2}\right) - 0.045\sin\left(\frac{\tau_{12,M}-\tau_{31,M}}{2}\right).$$

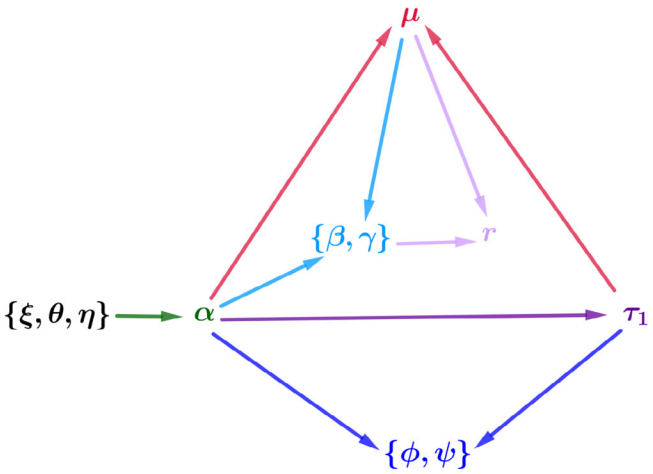

FIGURE 4 Graphical sketch of the variable dependence (notice that there are no arrows from the dihedrals (ϕ, ψ) toward the helical parameters and vice versa). The tetrahedral angular variables $\{\xi, \theta, \eta\}$ limit the range of α values which result in a viable solution for τ_1. It should *not* be construed that the tetrahedral variables are related to the parameter α through an explicit mathematical expression like the other variables in the graph

Substituting for the variables as above and maintaining the current convention of dihedral rotation we have

$$\begin{aligned} \cos\frac{\mu}{2} &= -0.817\sin\left(\frac{\psi+\phi}{2}\right) + 0.045\sin\left(\frac{-\psi+\phi}{2}\right) \\ \Rightarrow \cos\mu &= -0.330486 - 0.667489\cos(\psi+\phi) - 0.07353\cos\psi \\ &\quad + 0.07353\cos\phi \quad -0.002025\cos(-\psi+\phi) \\ \Rightarrow 3\cos\mu &= 1.011009 - 4.004934\cos^2\left(\frac{\psi+\phi}{2}\right) - 0.22059\cos\psi \\ &\quad + 0.22059\cos\phi - 0.006075\cos(-\psi+\phi). \end{aligned} \tag{12}$$

Discarding the last three terms and rounding up to the nearest integer we have the truncated formula Equation (11). The approximate formula works well for the experimentally observed helical region bounded by the range of $\phi \in \{0, -\pi\}$ and $\psi \in \{0, -\pi\}$. However, Equation (12) or its truncated

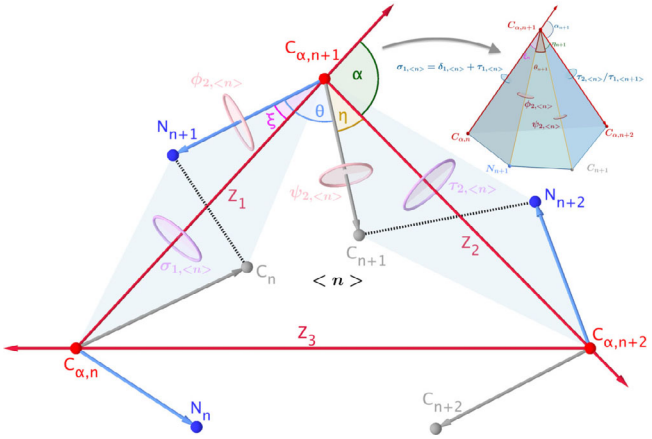

FIGURE 5 General description of a constrained motion tetrahedral in terms of its tetrahedral angles ξ, θ, η, and α. The tetrahedral angles are labeled as following. To each virtual bond $C_{\alpha,n} - C_{\alpha,n+1}$ we associate the angles η_n and ξ_n and to each $C_{\alpha,n}$ atom we associate the angles θ_n and α_n (see Table 3 for angular definitions). Dihedral angles ϕ, ψ, σ, and τ are shown along their respective axis of rotation. The inset figure on the upper right hand corner shows the bonds used to form the tetrahedral at the central $C_{\alpha,n+1}$ atom. The dihedral angles in this figure are labeled according to the plane definition. However, in our case all planes are identical and hence the angles will take on the same values across the planes <*n*> , $n = \{1, 2, \ldots,\}$

version Equation (11) require a priori knowledge of the dihedrals. Consequently, $(\phi,\psi) \rightleftharpoons_{\text{not possible}}^{\text{possible}} \mu$.

2.4 | General tetrahedral reconstruction

For consistency we will refer to the three consecutive C_α atoms which form the plane <*n*> as $C_{\alpha,n+k}^{(k+1)}; k = j \bmod 3$ where $j = \{0, 1, \ldots\}$. When the context of the plane of reference for the angles and their associated polynomial variables is clear, we will drop the suffix <*n*>. For a definition of the torsional angles please see Figure 5 and Table 3. The general tetrahedral equation with respect to the plane <*n*> and vertex at $C_{\alpha,n+1}^{(2)}$ is written in compact form as[58]

$$\mathcal{B}_n(\sigma_{1,<n>}; \tau_{2,<n>}; \xi_n, \theta_{n+1}, \eta_{n+1}, \alpha_{n+1}) = 0, \text{ where } \sigma_{1,<n>} = \tau_{1,<n>} + \delta_{1,<n>}. \tag{13}$$

Converting to polynomial form using the following half tangents,

$$u_{i,<n>} = \tan\frac{\tau_{i,<n>}}{2}, i = \{1,2\} \quad w_{1,<n>} = \tan\frac{\sigma_{1,<n>}}{2}, \tag{14}$$

we have

$$\mathcal{B}_n(w_1, u_2, A, B, C, D, E) = 0. \tag{15}$$

In view of the relation between σ_1 and $\tau_1 \; \forall$ <*n*> where $n \in \{1, 2, \ldots, N_{\text{res}} - 1\}$ we may eliminate w_1 using the following relations

TABLE 3 Notation

Notation	Description
$\hat{a}_n$	Unit vector from $C_{\alpha,n}$ to $C_{\alpha,n+1}$.
$\hat{b}_n$	Surface normal to the plane formed by $\{C_n, C_{\alpha,n}, C_{\alpha,n+1}\}$ atoms.
$\mathcal{C}$	Helical projection onto the X – Y plane.
$C_{\alpha,n}$	n^{th} C_α atom.
<*n*>	Plane formed by the triad $\{C_{\alpha,n}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$.
$\hat{N}_n$	Unit normal to <*n*>.
O	Center of $\mathcal{C}$
r	Radius of $\mathcal{C}$.
X, Y, Z	Fixed axes attached to the helix.
α	Angle between $\hat{a}_n$ and $\hat{a}_{n+1}$.
β	Angle $\hat{a}_n$ makes with the X – Y plane.
$\gamma_{<n+1,n>}$	Angle between $\hat{N}_n$ and $\hat{N}_{n+1}$ (angle between planes <*n*> and <*n* + 1>).
δ	Angle between planes formed by atoms $\{C_n, C_{\alpha,n}, C_{\alpha,n+1}\}$ and $\{N_{n+1}, C_{\alpha,n+1}, C_{\alpha,n}\}$
ε_0	Perturbation angle about the virtual $C_\alpha - C_\alpha$ bond.
η	$\measuredangle C_n C_{\alpha,n} C_{\alpha,n+1} = 19.014°$.
θ	$\measuredangle N_n C_{\alpha,n} C_n = 109.323°$.
μ	Angle subtended at the center of $\mathcal{C}$ by the projection of $\hat{a}_n$ on the X – Y plane.
ξ	$\measuredangle C_{\alpha,n} C_{\alpha,n+1} N_{n+1} = 16.537°$.
$\sigma_{1,<n>}$	Angle between the planes formed by atoms $\{N_{n+1}, C_{\alpha,n+1}, C_{\alpha,n}\}$ and <*n*>.
$w_{1,<n>}$	Half tangent of $\sigma_{1,<n>}$ that is, $w_{1,<n>} = \tan\frac{\sigma_{1,<n>}}{2}$.
$\tau_{1,<n>}$	Angle between the planes formed by atoms $\{C_n, C_{\alpha,n}, C_{\alpha,n+1}\}$ and <*n*>.
$u_{1,<n>}$	Half tangent of $\tau_{1,<n>}$ that is, $u_{1,<n>} = \tan\frac{\tau_{1,<n>}}{2}$.
$\tau_{2,<n>}$	Angle between the planes formed by atoms $\{C_{n+1}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$ and <*n*>.
$u_{2,<n>}$	Half tangent of $\tau_{2,<n>}$ that is, $u_{2,<n>} = \tan\frac{\tau_{2,<n>}}{2}$.
$\tilde{\tau}_{1,<n>}$	Torsional perturbation of $\tau_{1,<n>}$ about the virtual $C_\alpha - C_\alpha$ bond.
$\tilde{\tau}_{2,<n>}$	Torsional perturbation of $\tau_{2,<n>}$ about the virtual $C_\alpha - C_\alpha$ bond.
ϕ	Backbone torsional angle about the $N - C_\alpha$ bond.
ψ	Backbone torsional angle about $C_\alpha - C$ bond.
Ω_1	$\measuredangle N_n C_{\alpha,n} C_{\alpha,n+1}$.
Ω_2	$\measuredangle C_{\alpha,n} C_{\alpha,n+1} C_{n+1}$.

$$\Delta_1 \equiv \tan\frac{\delta_1}{2} = \frac{\sin\delta_1}{1+\cos\delta_1}, \Delta_1^2 \equiv \tan^2\frac{\delta_1}{2} = \frac{1-\cos\delta_1}{1+\cos\delta_1},$$
$$\mathcal{B}_n(\Delta_1, u_1, u_2, A, B, C, D, E) = 0, \tag{16}$$

where

$$\begin{aligned} A &= -\cos\theta_{n+1} - \cos(\alpha_{n+1} - \xi_n - \eta_{n+1}), \\ B &= -\cos\theta_{n+1} - \cos(\alpha_{n+1} - \xi_n + \eta_{n+1}), \\ C &= 4\sin\xi_n \sin\eta_{n+1}, \\ D &= -\cos\theta_{n+1} - \cos(\alpha_{n+1} + \xi_n - \eta_{n+1}), \\ E &= -\cos\theta_{n+1} - \cos(\alpha_{n+1} + \xi_n + \eta_{n+1}). \end{aligned} \tag{17}$$

2.5 | Polynomial form for helix and β-strand condition

Converting Equation (9) to a polynomial form and using half tangent formulas defined in Equation (14) we have,

$$u_{1,<n>}\, u_{2,<n>} = \pm 1 \forall n. \tag{18}$$

In Equation (18), we distinguish the positive relation as a geometrical condition for the construction of helices and the negative as that for β-strands. Next, we delineate the two cases considered with the different ancillary geometrical assumptions for the modeling of the SSE's.

2.6 | Identical repeated units

In this section, we assume the SSE's are composed of identical repeated *unit-SSE's*, a unit formed by the triad $\{C_{\alpha,n}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$, and as such $\tau_{1,<n+1>} \equiv \tau_{1,<n>}$. To construct SSE's under this assumption it is sufficient to solve the kinematics for a given plane. The dihedrals ϕ and ψ calculated using the solution to the kinematic equation will thus remain same throughout the peptide chain. Our assumption here also allows us to express the relations between γ and τ_1 as plane invariant (see Figure 6). Consequently, we drop the suffix which explicitly identifies this plane. Conditions for the formation of SSE's, Equation (18) gives us,

$$\begin{aligned} &\text{A.}u_1u_2 = 1, && \text{B.}u_1u_2 = -1, \\ &2\pi - \gamma + \tau_1 = \tau_2, && 2\pi - \gamma + \tau_1 = \tau_2, \\ \Rightarrow\ &2\pi - \gamma + \tau_1 = \pi - \tau_1, && 2\pi - \gamma + \tau_1 = \pi + \tau_1, \\ &\Rightarrow \gamma = \pi + 2\tau_1, && \gamma = \pi. \end{aligned} \tag{19}$$

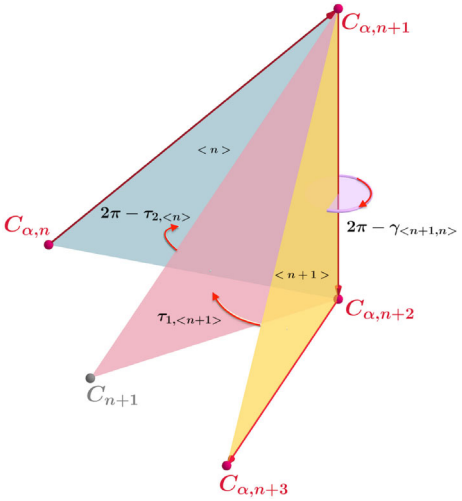

FIGURE 6 The figure here shows the simultaneous angles formed by a tetrahedral plane, triad $\{C_{n+1}, C_{\alpha,n+1}, C_{\alpha,n+2}\}$, with planes <*n*> and <*n* + 1>

2.6.1 | General tetrahedral equation

Following the definition of *unit-SSE's* and our assumption of identical repeated units, all the four angles ξ, θ, η, and α are assumed same across all planes. As above, we drop subscripts which explicitly identify the plane in Equation (16). The substitutions, $u_1 = u$ and $u_2 = v$ lend to the following simplified form,

$$\mathscr{B}(\Delta, u, v, A, B, C, D, E) = 0. \tag{20}$$

Simplification and grouping of terms in Equation (20) results in the following,

$$\begin{aligned} &\Delta^2\left(Av^2 + Du^2v^2 + B - Cuv + Eu^2\right) + \Delta\left(2Auv^2 - 2Duv^2 + 2Bu + Cv - Cu^2v - 2Eu\right) \\ &+ \left(Au^2v^2 + Bu^2 + Cuv + Dv^2 + E\right) = 0, \end{aligned} \tag{21}$$

where the coefficients $A - E$ have the same definition as that of Equation (17), with the subscripts dropped. The above is a simplification of eq. (11) in Coutsias et al., using the relation as defined in eq. (12) for $i = 2$.[58,59] The general case for repeated identical units, $\delta \neq \pi$, although not attempted here, could be implemented using Equation (21) with relations derived in Equation (18) that is, $uv = \pm 1$. The next section demonstrates the kinematic implementation undertaken for the limiting case $\delta = \pi$.

2.6.2 | Tetrahedral solution for helices and beta sheets

In addition to the geometrical symmetry for the SSE's, a strict *trans* configuration is also assumed across the peptide chain that is, $\delta = \pi$ for simplicity, thus $\frac{1}{\Delta} \approx 0$. In this case, tetra1 reduces to

$$Du^2v^2 + Eu^2 - Cuv + Av^2 + B = 0. \tag{22}$$

Substituting for v using Equation (18) we have

$$\begin{aligned} &D + Eu^2 \pm C + A\frac{1}{u^2} + B = 0, \\ \Rightarrow\ &Eu^4 + (D \pm C + B)u^2 + A = 0, \\ \Rightarrow\ &EU^2 + (D \pm C + B)U + A = 0, \text{where } U \equiv u^2 \\ &\Rightarrow\ U = \frac{-(D \pm C + B) \pm \sqrt{(D \pm C + B)^2 - 4EA}}{2E}, \end{aligned} \tag{23}$$

where the solutions are enumerated as follows,

$$\begin{aligned} U_1 &= \frac{-(D - C + B) + \sqrt{(D - C + B)^2 - 4EA}}{2E}, (uv = +1) \\ U_2 &= \frac{-(D - C + B) - \sqrt{(D - C + B)^2 - 4EA}}{2E}, (uv = +1) \\ U_3 &= \frac{-(D + C + B) + \sqrt{(D + C + B)^2 - 4EA}}{2E}, (uv = -1) \\ U_4 &= \frac{-(D + C + B) - \sqrt{(D + C + B)^2 - 4EA}}{2E}, (uv = -1). \end{aligned} \tag{24}$$

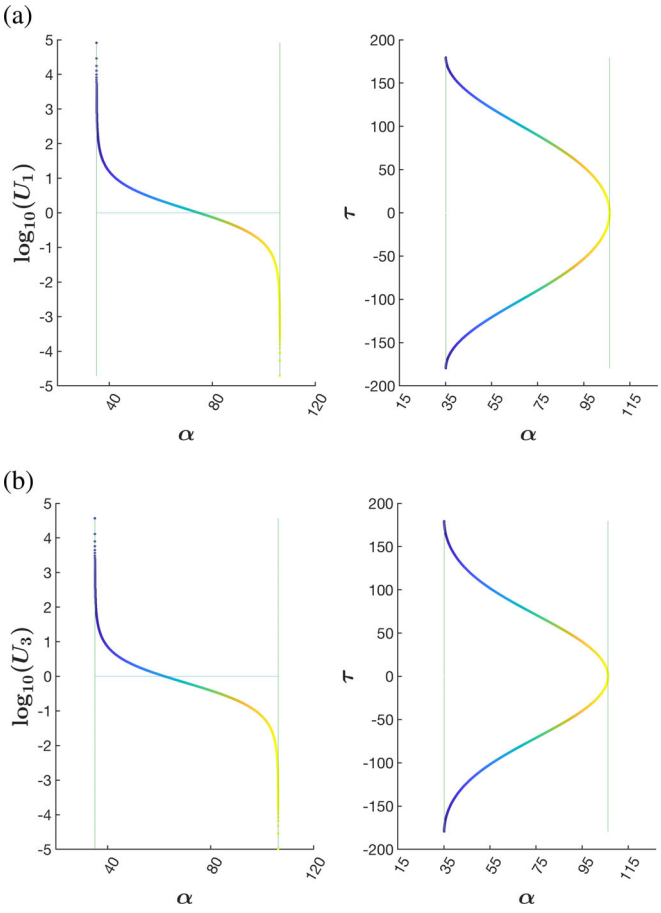

FIGURE 7 Torsional angles for helices and β-strands as a function of the parameter α are shown. The *physical* region of the solutions are marked by the vertical lines and correspond to the geometrically obtained bounds, $\alpha \in \{35.126^\circ, 106.228^\circ\}$. There are two parts to the torsion $\tau(\alpha)$ curves shown here. $\tau^+(\alpha) > 0^\circ$ corresponds to the solution $+\sqrt{U_1}$ or $+\sqrt{U_3}$ whereas $\tau^-(\alpha) < 0^\circ$ corresponds to $-\sqrt{U_1}$ or $-\sqrt{U_3}$. (A) The point of inflection, $\log_{10}(U_1) = 0$ corresponds to $\tau^+ \approx 90^\circ$, $\tau^- \approx -90^\circ$, and parameter $\alpha \approx 74.7^\circ$. Thus, as the parameter α varies from 35.28° to 106.2° the logarithm of the solution, $\log_{10}(U_1)$ passes from maximum positive to minimum negative value while $\tau^+(\alpha)$ varies from $\approx 180^\circ$ to 0° and $\tau^-(\alpha)$ from $\approx -180^\circ$ to 0°. (B) The point of inflection here, $\log_{10}(U_3) = 0$, corresponds to $\tau^+ \approx 90^\circ$, $\tau^- \approx -90^\circ$ and parameter $\alpha \approx 62^\circ$. Variation in the $\tau(\alpha)$ curve is similar to that in part (A)

Substituting for the values of ξ, θ, and η in Equation (23) and using α as a parameter, one may obtain U (see Figure 7(A),(B)); μ is obtained as a function of α and τ_1 using Equation (26), derived from the Whitworth's formulas given by eqs. (5) and (6) (see Figure 4). Here, we have taken $\xi = 16.537^\circ$, $\theta = 109.323^\circ$, and $\eta = 19.014^\circ$ consistent with the trans-planarity of the peptide bond.

2.6.3 | Helical parameters from Whitworth's relation

Here, we derive the relation of the helical parameters, μ, β, and r to the torsion $\tau_{1,<n>} \equiv \tau_1$ for identical repeated units. τ_1 is obtained from Equation (24) for given parametric values of α. From Equation SI.6 we have

$$\sin^2\tau_1 = \kappa^2 \sin^2\mu \cos^2\beta. \tag{25}$$

Using Equation (5) and substituting for κ we have,

$$\tan\frac{\mu}{2} = \pm\sqrt{\frac{1}{\sin^2\tau_1 \cos^2\alpha/2} - 1} \tag{26}$$

It should be noted here that corresponding to the sign of τ_1 there is a two part evaluation for the angle μ using Equation (26). Under the assumption that the parameter $\beta \in \{0, 90^\circ\}$ for right handed helices, Equation (5) may again be used to obtain parameter β. The radius r of the helix is obtained from Equation (1) (see Figure 8(A),(B)).

2.7 | Torsional relations for trans-planarity of the peptide bond

To calculate the torsions about the actual bonds, $N - C_\alpha$ and $C_\alpha - C$, in terms of the torsion about the virtual $C_\alpha - C_\alpha$ bonds we employ the use of the fundamental formulas of spherical geometry as discussed by Bricard.[53] The resulting polynomial form of these relations, with variable definitions $x = \tan\frac{\phi}{2}, y = \tan\frac{\psi}{2}, u = \tan\frac{\tau_1}{2}$, and $v = \tan\frac{\tau_2}{2}$, are biquadratic in nature (see Equations (27) and (28)) and provide an explicit representation in terms of the torsion, u and v, about the virtual $C_\alpha - C_\alpha$ bond. However, the solutions obtained at this stage are not unique. These explicit relations are therefore substituted in the corresponding tetrahedral equations (see Equations (29) and (30)), $\delta_{1,<n>} = \pi$ that is, $\frac{1}{\Delta} \approx 0$, to obtain a unique solution for the backbone torsions (ϕ, ψ). In the ensuing discussion, the *unit-SSE's* are assumed identical. Hence, it is enough to illustrate the relation between the different torsions and tetrahedral angles for a single unit and as such only the first unit is considered (see Figure 9).

Biquadratic polynomial forms obtained through spherical geometry derived relations take on a compact form,

$$F_2 v^2 x^2 + G_2 v^2 + H_2 x^2 + J_2 = 0, \tag{27}$$

$$F_1 u^2 y^2 + G_1 u^2 + H_1 y^2 + J_1 = 0, \tag{28}$$

while the *Tetrahedral Equations*, arrived at in a manner similar to Tetra-angle, for the *unit-SSE* under consideration take on the following form,

$$(\phi, \tau_{1,<n>}) \;:\; B_{\phi,\tau_1} u^2 x^2 + A_{\phi,\tau_1} x^2 - C_{\phi,\tau_1} ux + E_{\phi,\tau_1} u^2 + D_{\phi,\tau_1} = 0, \tag{29}$$

$$(\tau_{2,<n>}, \psi) \;:\; A_{\tau_2,\psi} v^2 y^2 + B_{\tau_2,\psi} v^2 - C_{\tau_2,\psi} vy + D_{\tau_2,\psi} y^2 + E_{\tau_2,\psi} = 0. \tag{30}$$

Substituting the biquadratics into the relevant tetrahedral equation we have,

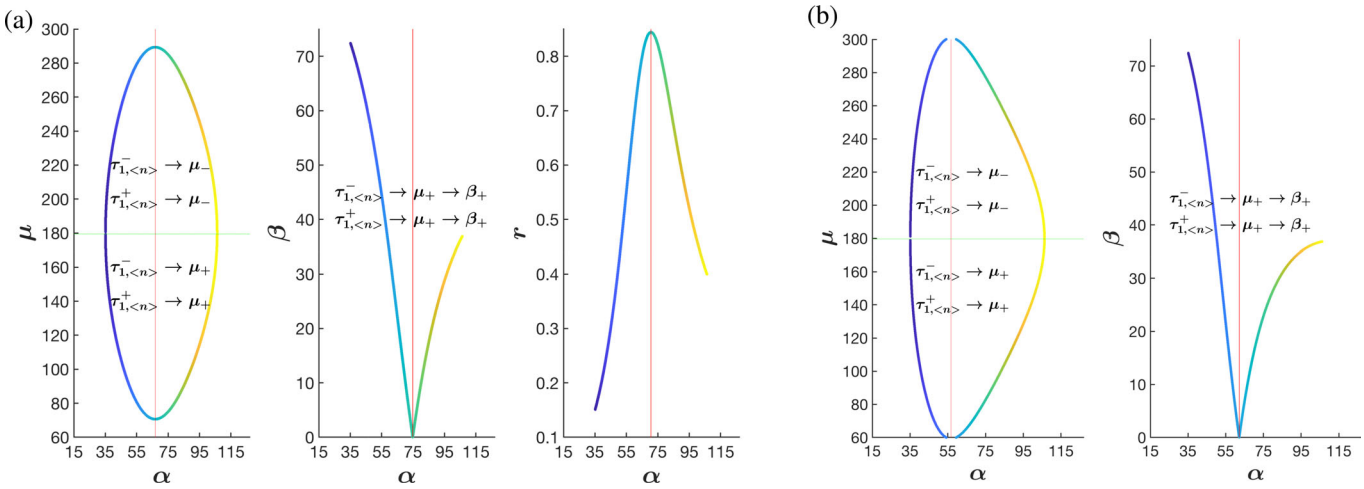

FIGURE 8 (A) Variations in μ, angle projected at the center, β, angular rise per unit residue, and r, radius of the helix as a function of the parameter α. The graph for μ versus parameter α has two parts. The part below the green line is obtained for $\tau_1^+ \to \mu_+$ and $\tau_1^- \to \mu_+$ whereas the part above the green line is obtained for $\tau_1^+ \to \mu_-$ and $\tau_1^- \to \mu_-$. The red line corresponds to $\alpha = 66.82^\circ$ here. Since there are four parts to μ, that is, μ_+, μ_- from τ_1^+ and μ_+, μ_- from τ_1^-, there are four parts associated with β. The β_+ values corresponding to μ_+ obtained from either sign of τ_1 is shown here. The β_- values corresponding to μ_- from either sign of τ_1 are reflections about a horizontal line parallel to the Y-axis and have values equal to $180 - \beta_+$. The red line here corresponds to $\alpha = 74.77^\circ$ and hence $\mu = 74.77^\circ$. The radius r corresponding to the 4 parts of β result in same values. Hence, the graph r versus parameter α is independent of the sign of τ_1. The red line here corresponds to $\alpha = 70.63^\circ$. The graph for r assumes the $C_\alpha - C_\alpha$ distance as unit. (B) In the graph for μ versus parameter α the parts above and below the green line have the same definition as described in Figure 8(A). The red line here corresponds to $\alpha = 56.68^\circ$. The description in Figure 8(A) for β versus parameter α is also relevant for the solution U_3. The red line passing through the minimum value of β corresponds to $\alpha = 62.14^\circ$

$$x = \frac{1}{C_{\phi,\tau_1} u}\left\{E_{\phi,\tau_1}u^2 + D_{\phi,\tau_1} - \left[\frac{G_2 v^2 + J_2}{F_2 v^2 + H_2}\right]\left(B_{\phi,\tau_1}u^2 + A_{\phi,\tau_1}\right)\right\}, \tag{31}$$

$$y = \frac{1}{C_{\tau_2,\psi} v}\left\{B_{\tau_2,\psi}v^2 + E_{\tau_2,\psi} - \left[\frac{G_1 u^2 + J_1}{F_1 u^2 + H_1}\right]\left(A_{\tau_2,\psi}v^2 + D_{\tau_2,\psi}\right)\right\}. \tag{32}$$

Backbone torsional expressions for helices and β-strands are obtained from Equations (31) and (32) using Equation (18). For the coefficient definitions, in terms of the tetrahedral angles, see Section SI.3.

2.8 | Torsional perturbation about the $C_\alpha - C_\alpha$ virtual bond

Torsional perturbation about the virtual $C_\alpha - C_\alpha$ bond by angle $-\varepsilon_0$ with the assumption that $\frac{1}{\Delta} \approx 0$ is shown in Figure 10. The sense of direction of the dihedral angles is same as that used by Coutsias et al.[59] In an equation form we have,

$$\tilde{\tau}_{1,<n>} = \tau_{1,<n>} - \varepsilon_0. \tag{33}$$

Since, the peptide planes are linked at the C_α atoms and their mutual orientation constrained by the θ angle, we will require the tetrahedral Equation (22) written as a quadratic in $\tilde{v}$ where $\tilde{v}$ denotes the perturbation to $\tau_{2,<n>}$ by an unknown dihedral angle (say ε_0'). Hence,

$$\tilde{\tau}_{2,<n>} = \tau_{2,<n>} \pm \varepsilon_0'. \tag{34}$$

A proof for the perturbational relations among the torsions given by Equations (33) and (34) and the conditions under which they exist in relation to the prescribed geometry is presented in Supporting Information (see Section SI.4). The task we undertake here (1) serves to construct conformations which do not have the peptide planes parallel to the helical axis and (2) describes the perturbational range corresponding to the parameter α. Denoting $\tilde{u}$ as the half-tangent of $\tilde{\tau}_{1,<n>}$, we have

$$D\tilde{u}^2\tilde{v}^2 + E\tilde{u}^2 - C\tilde{u}\tilde{v} + A\tilde{v}^2 + B = 0, \tag{35}$$

The quadratic form in $\tilde{v}$ gives us the following two solutions

$$\begin{aligned}\tilde{v}_a &= \frac{C\tilde{u} + \sqrt{(C\tilde{u})^2 - 4\left(D\tilde{u}^2 + A\right)\left(E\tilde{u}^2 + B\right)}}{2\left(D\tilde{u}^2 + A\right)},\\ \tilde{v}_b &= \frac{C\tilde{u} - \sqrt{(C\tilde{u})^2 - 4\left(D\tilde{u}^2 + A\right)\left(E\tilde{u}^2 + B\right)}}{2\left(D\tilde{u}^2 + A\right)}.\end{aligned} \tag{36}$$

All the applied perturbations to $\tau_{1,<n>}$ may not necessarily result in a viable solution of the quadratic Equation (36). Thus, for a given value of the perturbation ε_0 and some fixed parameter α we have two orientations for the torsion $\tilde{\tau}_{2,<n>} \equiv \tilde{\tau}_2$ (see Figure 11(A),(B)). In the case of no perturbation $\varepsilon_0 = 0^\circ$, it can be checked that the calculations

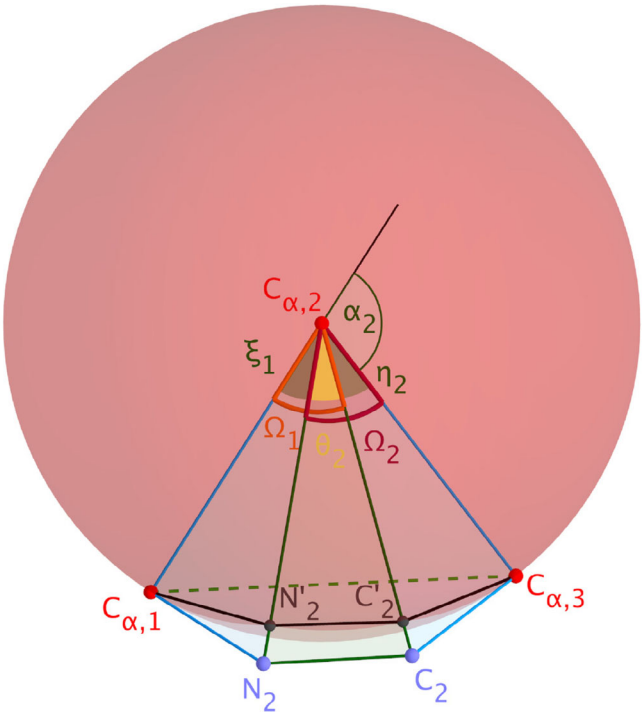

FIGURE 9 Spherical angles Ω_1 and Ω_2 projected at the center $C_{\alpha,2}$ of the sphere $\mathcal{S}$

do revert back to the original conformation of helices or β-strands for a given parameter α that is, the specific conformation available to us on the application of the helix-strand condition $uv = \pm 1$. However, an alternative conformation exists for the same value of $\tau_{1,<n>}$. Hence, the general nature of the solutions we seek requires Equations (31) and (32) to obtain perturbed dihedrals $(\tilde{\phi},\tilde{\psi})$. The coefficient definitions for *A*, *B*, *C*, *D*, and *E* are as expressed in Equation (17).

3 | RESULTS

3.1 | Helices and β-strands

The inverse kinematic construction for the helices and β-strands presented here is based on two main geometrical assumptions (1) the orientation of the peptide plane is such that it is parallel to the helical axis (Z) and the carbonyl group points in its positive direction (2) the helical structure consists of identical repeated units. The single parameter α is varied from 0 ° to 180° in a continuous manner and only those solutions which satisfy the kinematic Equation (22), give real positive solutions, are collected. It was found that for the α-helix condition $uv = +1$, only U_1 produced real positive solutions while for the β-strand condition $uv = -1$, only U_3 produced real positive solutions (see Figure 7(A),(B)). The range of the parameter $\alpha \in \{0°, 180°\}$ was discretized into 1001 values which results in a step-size of 0.18°. Of the 1001 different α values 395 of these result in real positive solutions. These solutions correspond to the parameter

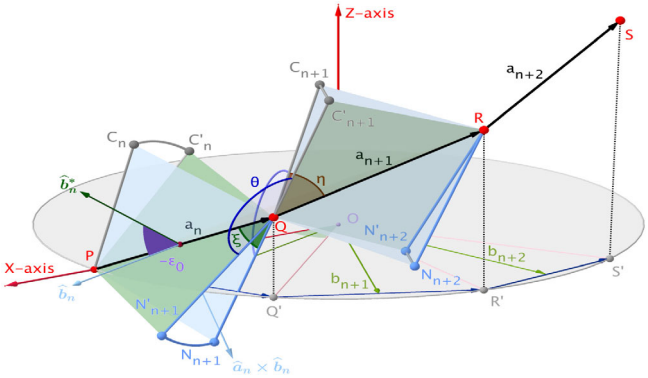

FIGURE 10 Geometrical skeletal of a helix with perturbation about $C_\alpha - C_\alpha$ bond. P, Q, R and S represent the position of the C_α atoms

$\alpha \in \{35.28°, 106.2°\}$ and they are contained withing the defined physical region (see the caption for Figure 7). The range $\alpha \in \{86.76°, 90.9°\}$ and $\alpha \in \{95.58°, 100.62°\}$ corresponds to α/π–helices and 3_{10}–helices, respectively. A relaxation of H···O by 0.4 Å and ∡HNO by 10° relaxes the parameter α bounds to $\alpha \in \{82.62°, 92.16°\}$ and $\alpha \in \{93.42°, 103.32°\}$ for α/π–helices and 3_{10}–helices, respectively. The structures with the relaxation of the H-bond criteria are shown in Figures 12(A),(C), 13(A),(C), and 14(A),(B). The H-bond recognition is based on the hybridization[2] (Figure 12(A)) and geometry of small molecule crystal structures implemented in chimera.[60]

3.2 | Construction protocol

The theoretical framework discussed here details the construction of the SSE's and is independent of the number of the residues considered. This is also true for any generalization of the current methodology. The protocol for identical repeated units is implemented on a relatively large canonical system, 32 residues to figuratively exhibit the geometrical changes in the gradual transition between helical types (see Table 1 and Figure 14 for a transition from π-helix to α-helix). Here, we detail the simple construction procedure adopted.

1. To begin with, a *n*-residues canonical backbone structure where each residue contains the atoms N, C_α, and C is used.
   a. Bond length definitions: $N - C_\alpha \approx 1.453$ Å, $C_\alpha - C \approx 1.53$ Å, and $C - N \approx 1.325$ Å.
   b. Bond Angle definitions: Internal atoms ∡N ≈ 120°, ∡ $C_\alpha \approx 109.36°$, ∡ C ≈ 117.5°
   c. Torsion definition: $\omega \approx 180°$; ϕ, ψ variable
   The corresponding Cartesian coordinates (*x*, *y*, *z*) of each atom in the system gives us a co-ordinate matrix $C_{3\times 3n}$.
2. Co-ordinate matrix $C_{3\times 3n}$ is used to determine the 3*n* dihedral vector, $D_{3n\times 1}$.
3. N_α, the number of discrete α values required is set. In our case, we set this value to 1001 for a reasonable discretization of the range 0 °

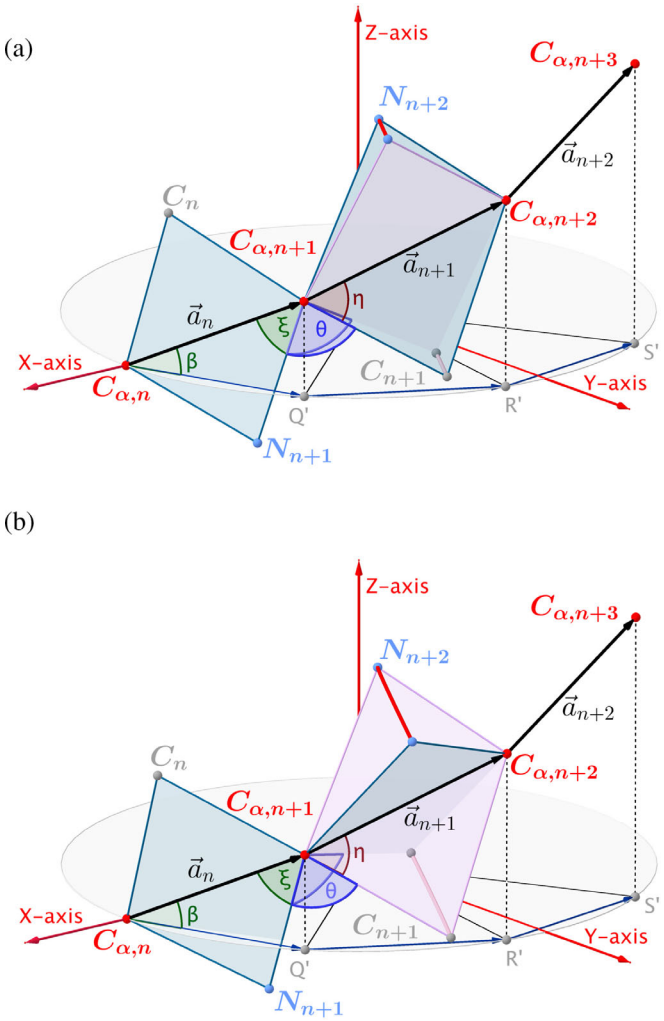

FIGURE 11 The figures shown here are a schematic representation of the peptide plane perturbation, in-between $C_{\alpha,n}$ and $C_{\alpha,n+1}$, under the $\measuredangle\theta$ constraint. Each orientation of this plane, vertical (no perturbation), or off-vertical (ε_0 perturbation), result in two corresponding orientations (see Equation (36)). (A) Shows the two solutions corresponding to the vertical alignment, parallel to the helical axis and hence zero perturbation, of the peptide plane. The blue–blue plane arrangement results in a perfect helical structure while the blue–purple plane arrangement will result in a conformation which forces the $C_{\alpha,n+3}$ to change its position (say $C'_{\alpha,n+3}$ not shown in the figure). (B) Similar to Figure 11(A), the peptide plane is perturbed by a small angle (off-vertical). The corresponding solutions are shown as the blue and purple plane orientations assumed under the $\measuredangle\theta$ constraint

– 180°. This also determines the maximum number of conformations that will be generated. Not all values of the parameter α result in a solution. If S denotes the number of solutions then $S \leq N_\alpha$.

4. Corresponding to each discrete value of α we get a set of five coefficients A, B, C, D, and E. Indexing we have $((\theta, \alpha_i, \xi, \eta) : A_i, B_i, C_i, D_i, E_i)\ \forall\ i \in \{1, \ldots, N_\alpha\}$. The entire procedure for solution finding is implemented for only those indices for which $|B_i| \geq 0$.
5. The solutions U_1 and U_2 are evaluated as given in Equation (24) and those which are real retained. Among the real solutions we locate those which are positive. Indexing again, we have $((\theta, \alpha_j, \xi, \eta) : A_j, B_j, C_j, D_j, E_j)\ \forall\ j \in \{1, \ldots, S\}$.
6. The positive, u_j^+ and negative, u_j^- torsions in polynomial form are evaluated from U_{1j} and then for $U_{2j} \forall\, j \in \{1, \ldots, S\}$. This gives us two sets of positive u_j^+ torsions and two sets of negative, u_j^- torsions. The attributes *positive* and *negative* denote only the part of the solutions U from which they are calculated. They bear no reference to the actual sign of the dihedrals (ϕ, ψ). For instance, $u_j^+ = +\sqrt{U_{1j}}$. In our calculations, U_2 did not result in any real positive solution so instead of four sets of u values we had only two corresponding to a u_j^+ and u_j^- both associated with the solution U_1.
7. From these half-tangents, we convert to dihedrals (ϕ, ψ) as follows

$$u_j^+ \rightarrow \left(x_j^+, y_j^+\right) \underset{2\tan^{-1}\left(y_j^+\right)}{\overset{2\tan^{-1}\left(x_j^+\right)}{\rightarrow}} \left(\phi_j^+, \psi_j^+\right) \forall j \in \{1, \ldots, S\}$$

and

$$u_j^- \rightarrow \left(x_j^-, y_j^-\right) \underset{2\tan^{-1}\left(y_j^-\right)}{\overset{2\tan^{-1}\left(x_j^-\right)}{\rightarrow}} \left(\phi_j^-, \psi_j^-\right) \forall j \in \{1, \ldots, S\}$$

using Equations (31) and (32) with helical condition $uv = 1$. The different combinations of the dihedral values will result in four different ensembles, each of size S. For our case, we use $\left(\phi_j^+, \psi_j^+\right) \forall j \in \{1, \ldots, S\}$ in the construction of right-handed helical structures.

8. The ϕ, ψ dihedrals from residue 1 to n are reassigned to the vector $D_{3n\times 1}$ and the corresponding Cartesian coordinates generated. If $D^j_{3n\times 1}$ denotes the vector of dihedrals (ϕ_j, ψ_j, ω) repeated n-times then its corresponding Cartesian coordinates are represented as $C^j_{3\times 3n} \forall j \in \{1, \ldots, S\}$. Since our objective was not to disturb the transplanarity of the amide bond we do not make any changes to the ω-torsion.
9. The construction for β-strands follows the same procedure as outlined above with the difference that $uv = -1$ is used for solving the quartic equation Equation (24). Similar to steps 6 and 7, u_j^+ and u_j^- corresponding to $U_{3j} \forall\, j \in \{1, \ldots, S\}$ are evaluated. As mentioned earlier, U_4 does not provide us with any real positive solution. The dihedrals $\left(\phi_j^+, \psi_j^+\right)$ for the β-strand solutions are calculated and the procedure defined in step 8 is followed.

From the generated ensemble it is possible now to calculate the different H-bond parameters as shown in Figure 15. The concentration of scatter data points as seen in Figure 15 is obtained from experimentally derived x-ray structures. We detail below the procedure adopted for the experimental results.

3.3 | Experimental data plot

1. The structures are obtained from a culled library maintained by the Dunbrack Lab,[3] `cullpdb_pc20_res1.6_R0.25_d200109_`

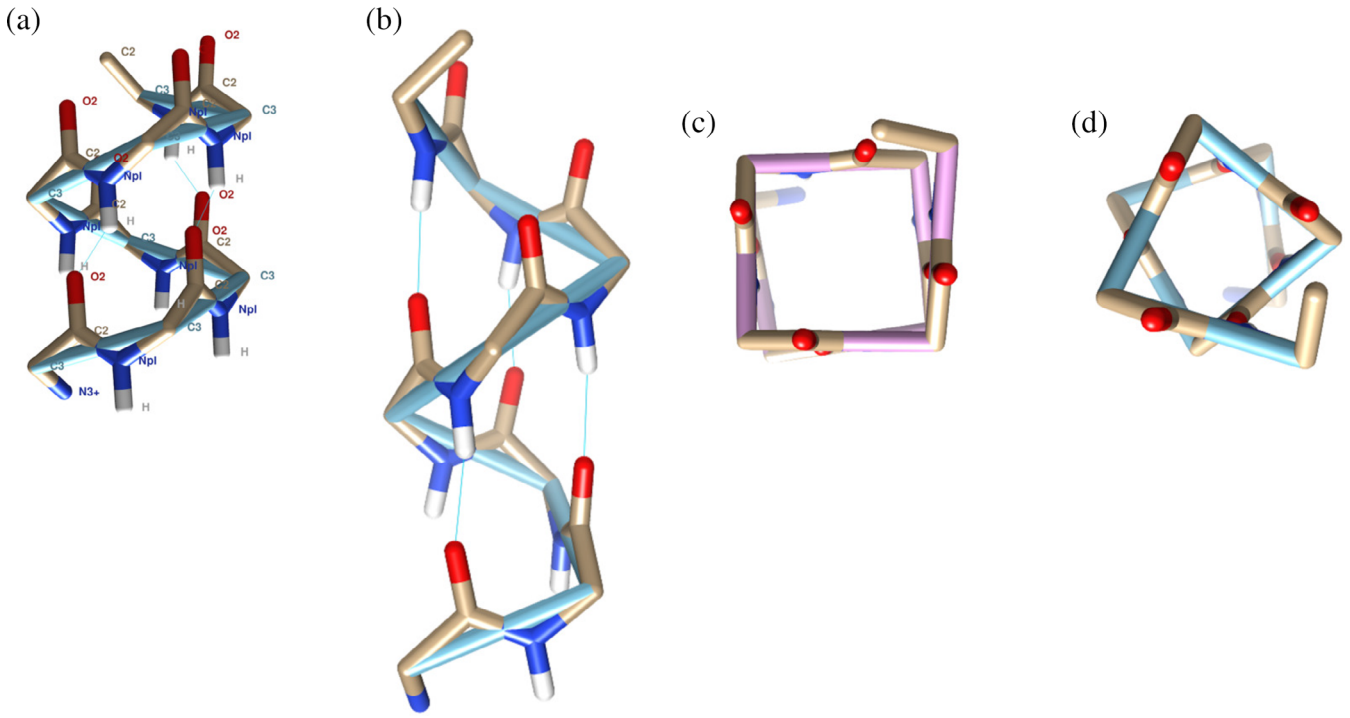

FIGURE 12 The atom types are same across all constructed conformations that is, conformations obtained for different values of the parameter α. All conformations are constructed under the assumption $\delta = \pi$. Results are shown for two values of the parameter, $\alpha = 83.7^\circ$ and 90°. H-bonds are shown as wires for the 8 residue system (see Table 1 for frequently observed helix length). The torsional definition used for the starting structure is $\omega \approx 180^\circ$; ϕ, ψ variable. (A) An 8 residue system showing the different atom types. The virtual bonds between C_α atoms are shown in cyan. H-bonds are formed between $O_n \cdot\cdot H_{n+5}$, $n \in \{1,2,3\}$. (B) Same construction procedure as that of Figure 12(A), parameter $\alpha = 90^\circ$. The H-bonds are between $O_n, \cdots, H_{n+4}$, $n \in \{1,2,3,4\}$. The Figure 12(C),(D) show the top view of the conformation shown in Figure 12 (A),(B), respectively

`chains3743` from which we chose the first 800 entries of resolution <1.6 Å and R-factor <0.25. These were checked for helices of length greater than or equal to 8 residues which resulted in $P_\alpha = 698$ different peptide chains. The P_α peptide chains contain $N_\alpha = 4031$ helices of varying lengths. 98% of these helices, as identified in the pdb format, were α-helical while the remaining 2% were of π-helical nature.

2. For each H_i, $i \in \{1, \ldots, N_\alpha\}$ of length Nr_i residues, $Nr_i - 2$ parameter α values were calculates and saved incrementally in a vector V_α such that length of the vector

$$L_\alpha \equiv l(V_\alpha) = \sum_{i=1}^{N_\alpha} (Nr_i - 2).$$

In our case $L_\alpha = 49160$. Since, to each parameter α value there is an associated (ϕ, ψ) dihedral the vector length of the dihedrals formed in a similar manner is also L_α. This provides us with the concentration of the scatter points in Figure 15.

3.4 | Construction protocol for perturbation about the $C_\alpha - C_\alpha$ virtual bond

For a fixed parameter value α_j, $j \in \{1, \ldots, S\}$ a perturbation to the corresponding dihedral, $\tau_{1,<n>}$ about the virtual $C_{\alpha,n} - C_{\alpha,n+1}$ bond, in a tripeptide unit as shown in Figure 5, is applied to allow different orientations, $\tilde{\tau}_{1,<n>}$. The kinematic Equation (37) shows that for each orientation there exists exactly two unique conformations. However, the sign of the solutions determine the right or left handedness of the identical unit assembled conformation. This adds two additional sets of conformations. For example for U_1,

$$u_j^+ \overset{2\tan^{-1}(u_j^+)}{\rightarrow} \tau_{1,<n>}^+ \overset{\varepsilon_{0,k}}{\rightarrow} \tilde{\tau}_{1,<n>}^+ \overset{\tan\left(\frac{\tilde{\tau}_{1,<n>}}{2}\right)}{\rightarrow} \tilde{u}_j^+ \rightarrow \left(\tilde{v}_{a,jk}^+; \tilde{v}_{b,jk}^+\right), k \in \{1, \ldots, N_\varepsilon\}$$

and

$$u_j^- \overset{2\tan^{-1}(u_j^-)}{\rightarrow} \tau_{1,<n>}^- \overset{-\varepsilon_{0,k}}{\rightarrow} \tilde{\tau}_{1,<n>}^- \overset{\tan\left(\frac{\tilde{\tau}_{1,<n>}}{2}\right)}{\rightarrow} \tilde{u}_j^- \rightarrow \left(\tilde{v}_{a,jk}^-; \tilde{v}_{b,jk}^-\right), k \in \{1, \ldots, N_\varepsilon\}, \quad (37)$$

where $\varepsilon_0 \in \{-180^\circ, 180^\circ\}$ and $N_\varepsilon = 101$ is the total number of perturbations considered (see Figure SI.1). Now, for a given value of (j, k) there are three possibilities for the existence of a solution of $\tilde{v}$,

a. $\left(\tilde{v}_{a,jk}^+, \tilde{v}_{b,jk}^+\right)$ and $\left(\tilde{v}_{a,jk}^-, \tilde{v}_{b,jk}^-\right), j \in \{1, \ldots, S\}, k \in \{1, \ldots, N_\varepsilon\}$ exist.
b. $\left(\tilde{v}_{a,jk}^+, \tilde{v}_{b,jk}^+\right)$ exists but $\left(\tilde{v}_{a,jk}^-, \tilde{v}_{b,jk}^-\right), j \in \{1, \ldots, S\}, k \in \{1, \ldots, N_\varepsilon\}$ does not.
c. $\left(\tilde{v}_{a,jk}^+, \tilde{v}_{b,jk}^+\right)$ does not exist but $\left(\tilde{v}_{a,jk}^-, \tilde{v}_{b,jk}^-\right), j \in \{1, \ldots, S\}, k \in \{1, \ldots, N_\varepsilon\}$ does.

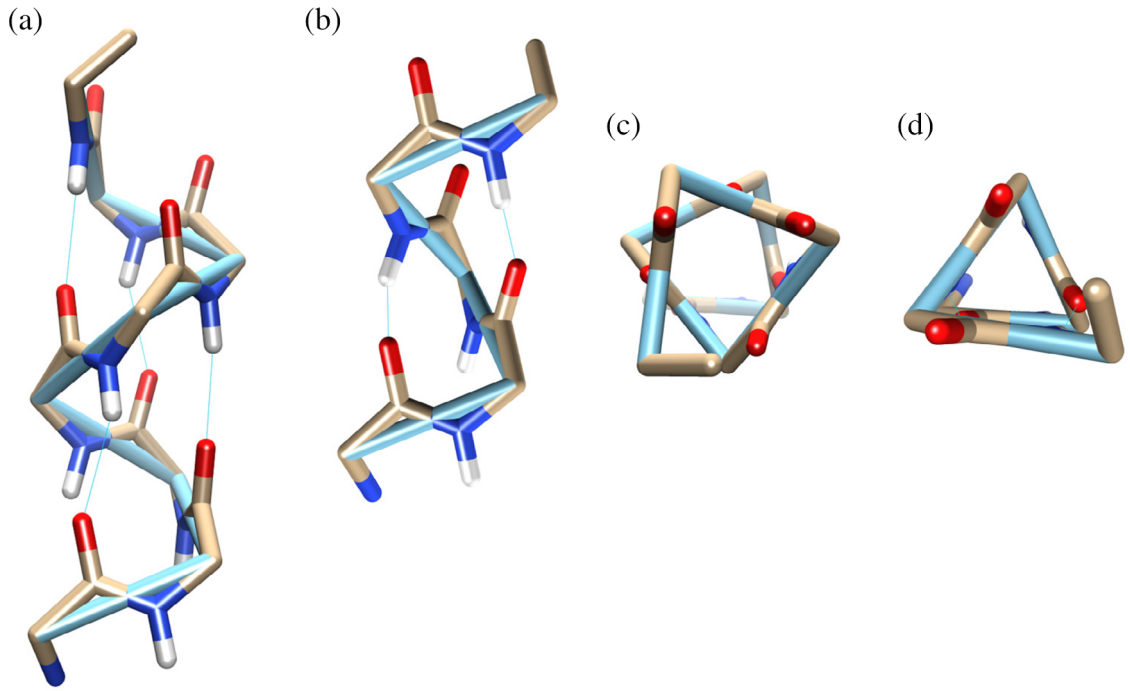

FIGURE 13 All conformations are constructed under the assumption $\delta = \pi$. Here, the results are shown for two values of the Figures 12 and 13, $\alpha = 91.26^\circ$ and $\alpha = 96.3^\circ$. The virtual bonds between the C_α atoms are shown in cyan. H-bonds are shown as wires for both the 8 residue system, α-helix, and the 5 residue system, 3_{10}-helix (see Table 1 for frequently observed helix length). The torsional definition of the starting structure is same as in Figure 12. (A) The constructed conformation shows H-bonds formed between $O_n \cdots H_{n+4}$, $n \in \{1,2,3,4\}$. (B) Similar to Figure 13(A), H-bonds are formed between $O_n \cdots H_{n+3}$, $n \in \{1, 2\}$ for parameter $\alpha = 96.3^\circ$. The Figure 13(C),(D) show the top view of the conformations shown in Figure 13(A),(B), respectively

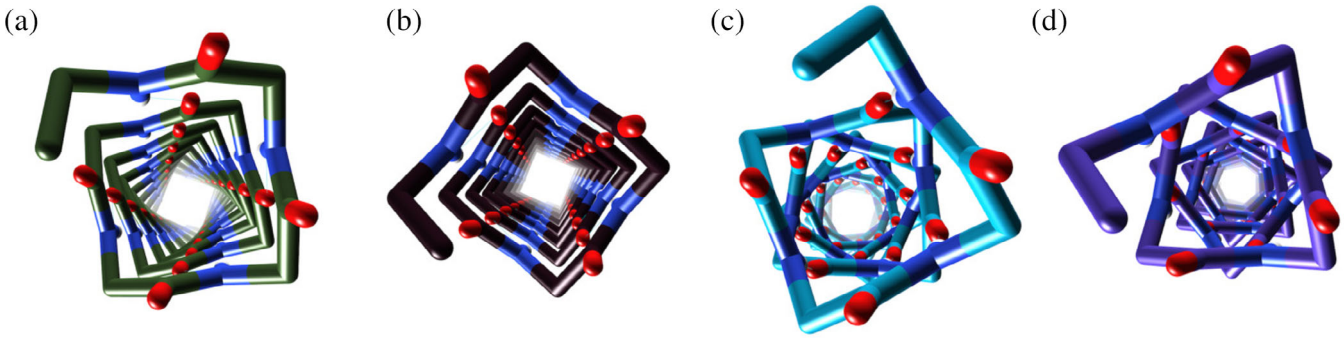

FIGURE 14 All conformations are constructed under the assumption that $\delta = \pi$. The construction procedure is same as that for Figures 12 and 13. Torsional definition used for the starting structure is $\phi, \psi, \omega \approx 180^\circ$. The results are shown for four different parameter values, $\alpha = \{83.7^\circ, 84.6^\circ, 88.56^\circ, 90.9^\circ\}$. The figures attempt to show a gradual transformation from a π-helix to an α-helix across the values the parameter α. A similar transition from an α-helix to a 3_{10}-helix also exists for a different range of the parameter α values (not shown here). However, we did not find one range of sequential parameter α values for a transition from a π-helix to a 3_{10}-helix. H-bonds are shown as wires for the 32 residue system

Due to the symmetry of the perturbed torsions $\tilde{\tau}^{+}_{1,<n>}$ and $\tilde{\tau}^{-}_{1,<n>}$ considered in our calculations the latter two cases do not exist. For both solutions, expressed in Equation (37), not all values of $k \in \{1, \ldots, N_\varepsilon\}$ result in solutions. Hence, there exists for each $j \in \{1, \ldots, S\}$ a variable range of allowed perturbations, $k \in \left\{1, \ldots, \underbrace{k_0, \ldots, N_{\varepsilon'}}, \ldots, N_\varepsilon\right\}$ that is, the values of k_0 and $N_{\varepsilon'}$ change for each $j \in \{1, \ldots, S\}$. In addition, the range $k \in \{k_0, \ldots, N_{\varepsilon'}\}$ may be a single or a bifurcated range. In our case, for the four sets of conformations available, corresponding to the allowed range of perturbations, two sets of conformations are of an opposite handedness. For example,

1. the sets $CF_a : \left(\tilde{u}_j^{+}, \tilde{v}_{a,jk}^{+}\right)$ and $CF_c : \left(\tilde{u}_j^{-}, \tilde{v}_{a,jk}^{-}\right)$, $j \in \{1, \ldots, S\}, k \in \{1, \ldots, N_\varepsilon\}$
2. and sets $CF_b : \left(\tilde{u}_j^{+}, \tilde{v}_{b,jk}^{+}\right)$ and $CF_d : \left(\tilde{u}_j^{-}, \tilde{v}_{b,jk}^{-}\right)$, $j \in \{1, \ldots, S\}$, $k \in \{1, \ldots, N_\varepsilon\}$,

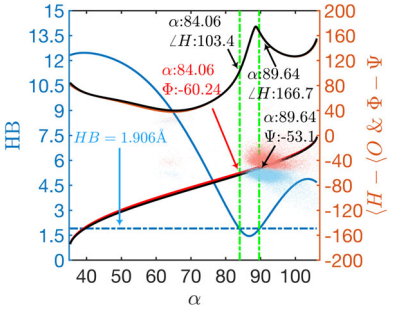

FIGURE 15 H-bond geometries calculated for 395 conformations generated through the geometrical construction for $n = 8$ residues. The $\measuredangle O = \measuredangle C_nO_nH_{n+4}$, $\measuredangle H = \measuredangle O_nH_{n+4}N_{n+4}$ and $HB = O_n \cdots H_{n+4}$. The peak point (black) in-between the green lines corresponds to $\measuredangle O_nH_{n+4}N_{n+4} = 174.572°$ tending toward linearity. The $\measuredangle O$ curve is shown in orange and follows closely the black curve of $\measuredangle H$. The colored, blue and red, scattered concentration of points correspond to experimentally derived (ϕ, ψ) dihedrals, respectively. These are plotted against their corresponding parameter α values. The total number of parameter α values plotted here are $L_\alpha = 49160$. Refer to Section 3.3 for details

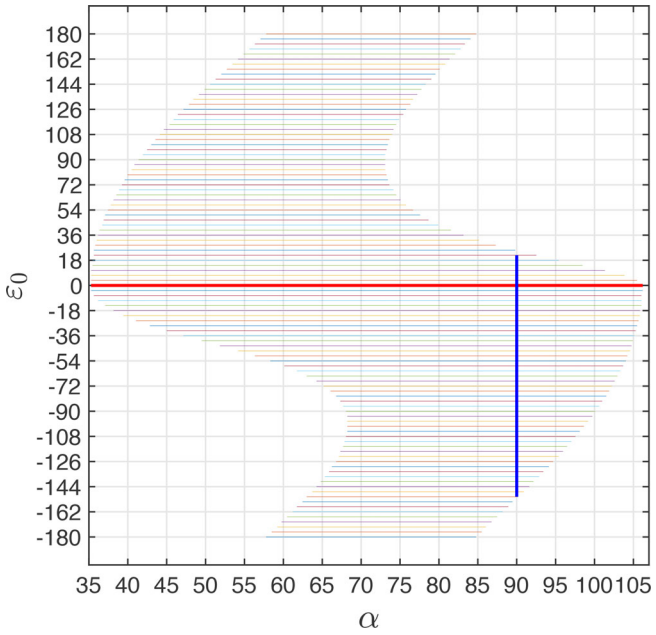

FIGURE 16 Conformations $\alpha_j, j \in \{1, ..., S\}$ which result in a solution for $\tilde{v}_a$ and $\tilde{v}_b$ with respect to the different perturbations $\varepsilon_{0,k}$, $k \in \{1, ..., N_\varepsilon\}$. The red line here corresponds to $\varepsilon_{0,51} = 0°$. The blue line corresponds the range of allowed perturbations $\varepsilon_0 \in \{-151.2°, 21.6°\}$, $k \in \{9, ..., 57\}$ for $\alpha_{305} = 90°$

result in conformations of opposite handedness. Similarly, 2 sets of alternative conformations for the same handedness result from

1. the sets $CF_a : \left(\tilde{u}_j^+, \tilde{v}_{a,jk}^+\right)$ and $CF_b : \left(\tilde{u}_j^+, \tilde{v}_{b,jk}^+\right)$, $j \in \{1,...,S\}$, $k \in \{1,...,N_\varepsilon\}$,
2. and sets $CF_c : \left(\tilde{u}_j^-, \tilde{v}_{a,jk}^-\right)$ and $CF_d : \left(\tilde{u}_j^-, \tilde{v}_{b,jk}^-\right)$, $j \in \{1,...,S\}$, $k \in \{1,...,N_\varepsilon\}$.

In a manner similar to that adopted for the construction of helices and β-strands the perturbed *units-SSE's* are repeated to obtain perturbed conformations. The computational procedure undertaken is described as follows:

1. A range for the perturbation ε_0 is first selected and discretized into N_ε points.
2. Vectors $T_{S\times1}^+$ and $T_{S\times1}^-$ containing corresponding values of $\tau_{1,<n>}$ is obtained from u_j^+ and $u_j^-, j \in \{1,...,S\}$, respectively.
3. The vectors are now given some perturbation $\varepsilon_{0,k}$, $k \in \{1, ..., N_\varepsilon\}$ to obtain new vectors $\tilde{T}_{S\times1}^+$ and $\tilde{T}_{S\times1}^-$. Matrices $\tilde{U}_{S\times N_\varepsilon}^+$ and $\tilde{U}_{S\times N_\varepsilon}^-$ are formed such that the columns of these matrices contain the half-tangents of vectors $\tilde{T}_{S\times1}^+$ and $\tilde{T}_{S\times1}^-$ formed for the different values of $\varepsilon_{0,k}$, $k \in \{1, ..., N_\varepsilon\}$ as expressed in Equation (37).
4. $\tilde{U}_{S\times N_\varepsilon}^+$ is used to form two matrices $\tilde{V}_{a,S\times N_\varepsilon}^+$ and $\tilde{V}_{b,S\times N_\varepsilon}^+$ while $\tilde{U}_{S\times N_\varepsilon}^-$ is used to form the two matrices $\tilde{V}_{a,S\times N_\varepsilon}^-$ and $\tilde{V}_{b,S\times N_\varepsilon}^-$ using Equation (36). Since, not all solutions are possible the indices which result in solutions are collected in an index matrix $I_{S\times N_\varepsilon}$ and populated column-wise. In general, there are two such index matrices, an $I_{S\times N_\varepsilon}^+$ which corresponds to solutions obtained from $\tilde{U}_{S\times N_\varepsilon}^+$ and an $I_{S\times N_\varepsilon}^-$ which corresponds to solutions obtained from $\tilde{U}_{S\times N_\varepsilon}^-$. In our case $I_{S\times N_\varepsilon}^+ = I_{S\times N_\varepsilon}^- \equiv I_{S\times N_\varepsilon}$. The column values of the index matrix correspond to the indices of the horizontal lines shown in Figure 16.
5. The column-wise combination of the matrices

$$\left(\tilde{U}_{S\times N_\varepsilon}^+, \tilde{V}_{a,S\times N_\varepsilon}^+\right), \left(\tilde{U}_{S\times N_\varepsilon}^+, \tilde{V}_{b,S\times N_\varepsilon}^+\right), \left(\tilde{U}_{S\times N_\varepsilon}^-, \tilde{V}_{a,S\times N_\varepsilon}^-\right) \text{ and } \left(\tilde{U}_{S\times N_\varepsilon}^-, \tilde{V}_{b,S\times N_\varepsilon}^-\right)$$

result in four matrices for ϕ and four matrices for the ψ dihedrals. For instance,

a. $\left(\tilde{U}_{S\times N_\varepsilon}^+, \tilde{V}_{a,S\times N_\varepsilon}^+\right) \to \tilde{\Phi}_{a,S\times N_\varepsilon}^+$ and $\left(\tilde{U}_{S\times N_\varepsilon}^+, \tilde{V}_{a,S\times N_\varepsilon}^+\right) \to \tilde{\Psi}_{a,S\times N_\varepsilon}^+$,
b. $\left(\tilde{U}_{S\times N_\varepsilon}^+, \tilde{V}_{b,S\times N_\varepsilon}^+\right) \to \tilde{\Phi}_{b,S\times N_\varepsilon}^+$ and $\left(\tilde{U}_{S\times N_\varepsilon}^+, \tilde{V}_{b,S\times N_\varepsilon}^+\right) \to \tilde{\Psi}_{b,S\times N_\varepsilon}^+$,
c. $\left(\tilde{U}_{S\times N_\varepsilon}^-, \tilde{V}_{a,S\times N_\varepsilon}^-\right) \to \tilde{\Phi}_{a,S\times N_\varepsilon}^-$ and $\left(\tilde{U}_{S\times N_\varepsilon}^-, \tilde{V}_{a,S\times N_\varepsilon}^-\right) \to \tilde{\Psi}_{a,S\times N_\varepsilon}^-$,
d. $\left(\tilde{U}_{S\times N_\varepsilon}^-, \tilde{V}_{b,S\times N_\varepsilon}^-\right) \to \tilde{\Phi}_{b,S\times N_\varepsilon}^-$ and $\left(\tilde{U}_{S\times N_\varepsilon}^-, \tilde{V}_{b,S\times N_\varepsilon}^-\right) \to \tilde{\Psi}_{b,S\times N_\varepsilon}^-$.

6. Collecting matrices together as in step 5-pert the column-wise combination of the dihedral matrices give us the following four sets of conformations.

a. $CF_a : \left(\tilde{\Phi}_{a,S\times N_\varepsilon}^+, \tilde{\Psi}_{a,S\times N_\varepsilon}^+\right)$,

b. $CF_b : \left(\tilde{\Phi}_{b,S\times N_\varepsilon}^+, \tilde{\Psi}_{b,S\times N_\varepsilon}^+\right)$,

c. $CF_c : \left(\tilde{\Phi}_{a,S\times N_\varepsilon}^-, \tilde{\Psi}_{a,S\times N_\varepsilon}^-\right)$,

d. $CF_d : \left(\tilde{\Phi}_{b,S\times N_\varepsilon}^-, \tilde{\Psi}_{b,S\times N_\varepsilon}^-\right)$.

The conformation sets CF_a, CF_b, CF_c, and CF_d conclude the consideration of perturbation $\varepsilon_0 \in \{-180°, 180°\}$ for all conformations

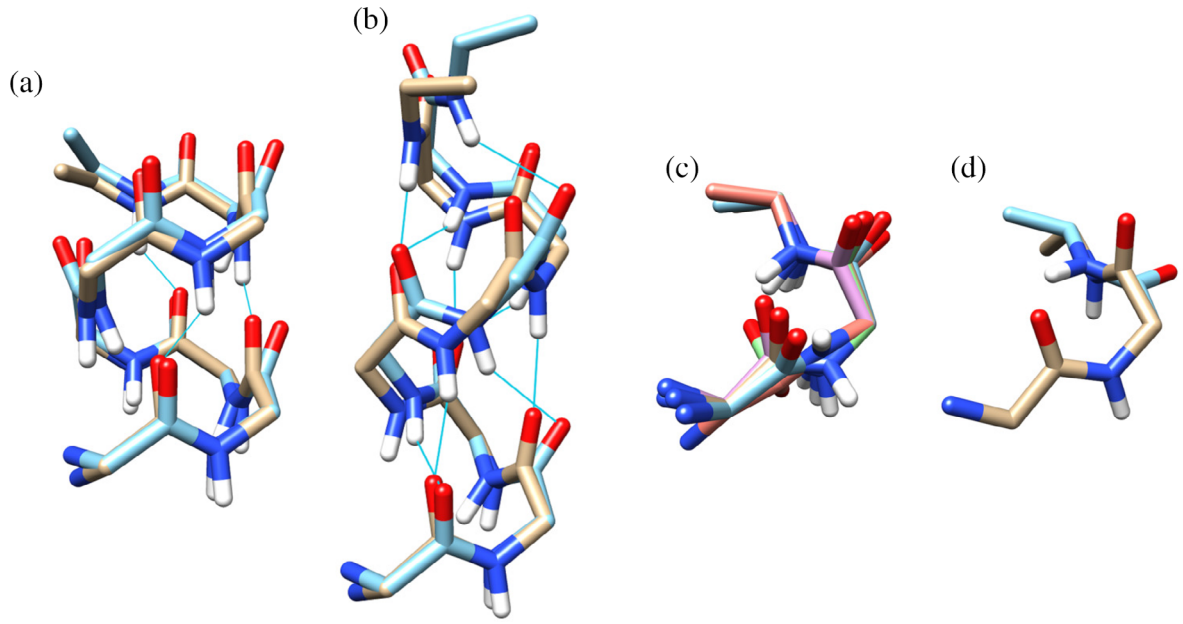

FIGURE 17 (A) Two 8 residue systems are shown where the original structure $\alpha = 83.7°$, is shown in brown and the perturbed conformation $\alpha = 83.7°$, $\varepsilon_{0,57} = 21.6°$ is shown in cyan. H-bonds were not found in the perturbed conformation. (B) Similar to Figure 17(A), 8 residue systems are shown where the original structure $\alpha = 90°$, is shown in brown and the perturbed conformation $\alpha = 90°$, $\varepsilon_{0,57} = 21.6°$ is shown in cyan the H-bonds in the original structure are formed between $O_n \cdots H_{n+4}$, $n \in \{1,2,3,4\}$ whereas in the perturbed structure the H-bonds are formed between atoms $O_n \cdots H_{n+3}$, $n \in \{1,2,3,4\}$. All structure are constructed under the assumption $\delta = \pi$. Torsional definition of the starting structure used for construction is $\omega \approx 180°$, ϕ, ψ variable.Figure 17: (C) five 3 residue systems $\alpha = 90°$ are shown for different perturbations $\varepsilon_0 = \{-151.2°, -90°, -21.6°, 0°, 21.6°\}$, $k = \{9,26,45,51,57\}$. (D) An alternative conformation for a 3 residue system with $\alpha = 90°$ is shown for $\varepsilon_{0,51} = 0°$. The conformation in cyan is obtained from the conformational set CF_a whereas the structure in brown is obtained from the conformational set CF_b and is overlapped with the original starting structure. Thus, although three structure are shown here only two are visible. Torsional definition for the starting structure used here is $\phi, \psi, \omega \approx 180°$

identified in the construction of helices and β-strands. It should be noted here that although the matrices formed are of dimension $S \times N_\varepsilon$ it is only the feasible perturbation range, single or bifurcated, for which the dihedral matrices $\tilde{\Phi}$ and $\tilde{\Psi}$ are evaluated.

7. Among the sets of conformations available as a result of the perturbations, conformations corresponding to $\varepsilon_{0,51} = 0°$ and α_j, $j \in \{1, \ldots, S\}$ contained in the set CF_b revert back to the original conformation for helices as shown in Figure 17(d).
8. A three dimensional plot of each of the dihedral sets described in step 6 are plotted and shown in Figure SI.2 in Supporting Information (see Section SI.5). The similarity between the Ψ dihedral matrix plots $\left(\tilde{\Psi}^{+}_{a,S\times N_\varepsilon}, \tilde{\Psi}^{-}_{a,S\times N_\varepsilon}\right)$ and $\left(\tilde{\Psi}^{+}_{b,S\times N_\varepsilon}, \tilde{\Psi}^{-}_{b,S\times N_\varepsilon}\right)$ is noticeable among these graphs. Thus, it is reasoned that the opposite handedness of the conformations sets CF_a and CF_c is the sole consequence of the opposite sign of the dihedral matrix set $\tilde{\Phi}^{+}_{a,S\times N_\varepsilon}$ and $\tilde{\Phi}^{-}_{a,S\times N_\varepsilon}$. A similar reasoning is applied to the opposite handedness of the conformational set CF_b and CF_d with sign switch now between the dihedral sets $\tilde{\Phi}^{+}_{b,S\times N_\varepsilon}$ and $\tilde{\Phi}^{-}_{b,S\times N_\varepsilon}$.

All the constructed structures in Figure 17(A)–(D) have their first 3 C_α-atoms superimposed. The starting structure used here is different from that used in Figures 12 and 13 and shows the invariance of the kinematic principle to the Cartesian coordinates of the starting structure and its dependence only on consistent values of ξ, θ, and η.

3.5 | β-beta-turns

In addition to the ability of the kinematic principle to generate *SSE*'s ab-initio it can also be used to sample exhaustively the geometrically feasible β-turn backbone conformations (see Figure 18). This requires the inverse kinematic solution of a loop closure problem, achieved here through an implementation of the BRIKARD program,[49] and is used to generate an ensemble of $N_{\text{sample}} = 10{,}000$ conformations. The initial conformation used, to generate N_{sample}, is a canonical 4 residue GLY structure obtained using *construct*, a subprogram of the BRIKARD suite. The β-turn H-bond between residues (n, $n+3$), of the form $C_n - O_n \cdots H_{n+3} - N_{n+3}$, is modeled together with the intervening backbone as a closed loop. Using inverse kinematics and referring to Figure 19, we may set the bond angles $\measuredangle C_1O_1H_4$ and $\measuredangle O_1H_4N_4$ and the distance $O_1 \cdots H_4$ arbitrarily. These three loop closure conditions, part of the H-bond parameter set, are kept constant for a particular ensemble but they may be prescribed to any desired value(s). Including the torsion about $O_1 \cdots H_4$, there are seven rotatable bonds available. For each triplet of parameter values, scanning the H-bond torsion results in a 1-parameter family of closed loop solutions for all the backbone torsions in the β-turn. We refer to the generated curves as *Bricard Curves*.[4] These Bricard curves, giving the distribution of the dihedrals over N_{sample}, are superimposed on the experimental data derived for that particular dihedral from a set $N_{\text{Dunbrack}} = 1043$

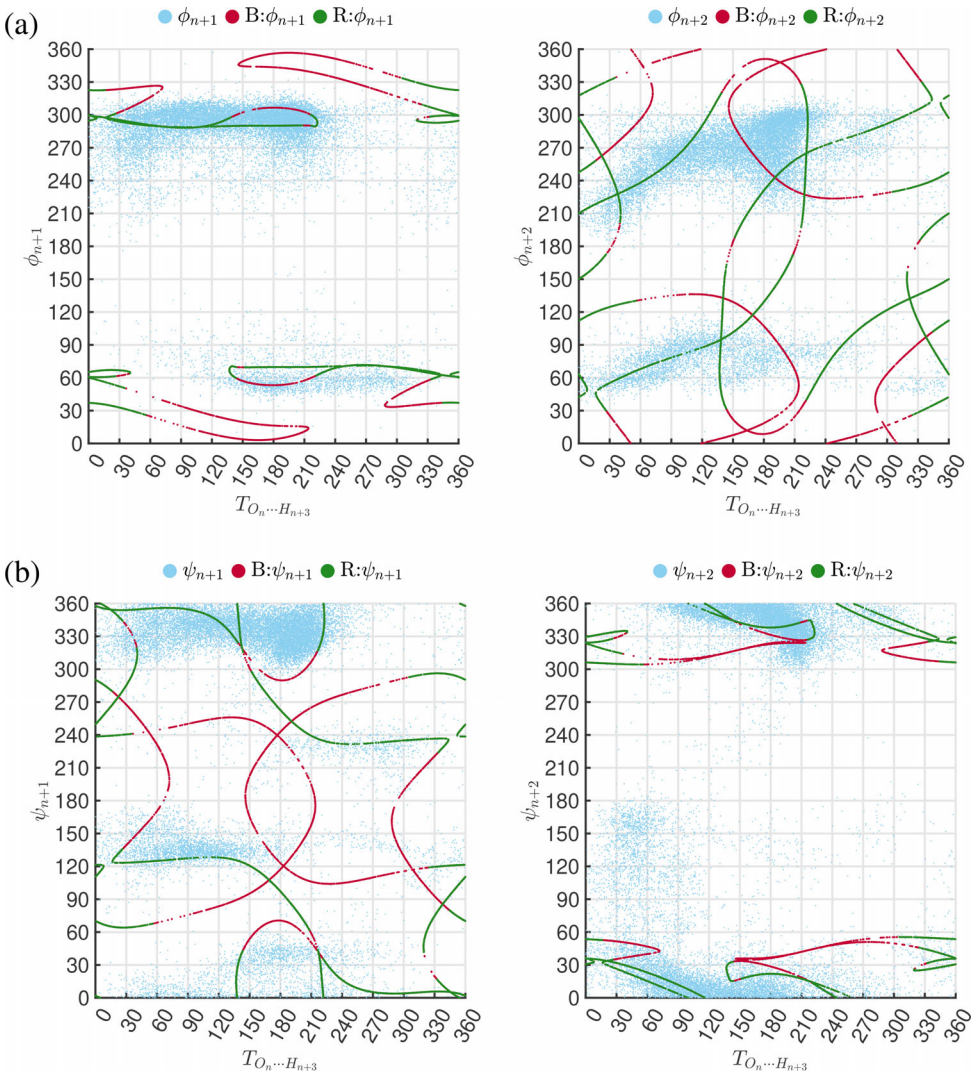

FIGURE 18 β-turn *Bricard Curves* generated for variable C_α positions and variable parameter α but fixed H-bond geometry. $\alpha \in \{81.82^\circ, 109.78^\circ\}$ at $C_{\alpha,2}$ and $\alpha \in \{78.97^\circ, 108.80^\circ\}$ at $C_{\alpha,3}$. BRIKARD generated conformations, *B*: Without Ramachandran criteria and *R*: With Ramachandran criteria, are shown as red and green curves, respectively in the figure. 10,000 points each are used to generate these curves. Experimentally derived values are shown as dots and are 21,091 in number. The H-bond geometry is the closure criteria predefined for sampling and is given in step 5-brik-curve of the validation process. *Bricard Curves* with respect to the torsions about $O_n \ldots C_n : 4 - 3$ and $N_{n+3} \ldots H_{n+3} : 13 - 17$ are provided in the Supporting Information (see also Figure SI.3)

different protein structures as shown in Figure 18. This database of PDB structures, including different models of the same structure, was examined earlier by the Dunbrack Lab for new types of protein turns.[13] There is a good agreement between the kinematically obtained backbone torsions and the experimental data. However, it should be noted that this agreement depends critically on the criteria used for the detection of β-turns in protein structures. The general procedure followed for this study is outlined in the next section.

3.6 | Validation of the generated Bricard curves

1. A list of N_{Dunbrack} PDB-ID's is used to obtain the structures in a PDB format from the RCSB-Protein Data Bank.[5] All model types of these structures, that were examined for determining new protein turns, were collected into separate PDB's resulting in an ensemble $N_{exp} = 1075$ PDB structures.
2. Dunbrack's code for turn detection is applied to each structure s_i, $i \in \{1, \ldots, N_{exp}\}$ in the ensemble and the corresponding output of the turn detection, OP_i, $i \in \{1, \ldots, N_{exp}\}$ stored in separate files. These output files, OP_i, $i \in \{1, \ldots, N_{exp}\}$ contain details such as the starting and ending residues for different turn nomenclatures. This is the only information we make use of in our study here. Sample output files generated by `DunbrackTurn_detect.sh` can be found in the `Sample Output` folder within the downloadable `BetaTurn18` folder provided by the Dunbrack Lab at https://github.com/sh-maxim/BetaTurn18.
3. First, hydrogens are added to each s_i, $i \in \{1, \ldots, N_{exp}\}$ and second, the corresponding output files OP_i, $i \in \{1, \ldots, N_{exp}\}$ are scanned for all types of turns of length four residues. For each s_i, $i \in \{1, \ldots, N_{exp}\}$, our objective is to collected the Cartesian coordinates of full backbone atoms that is, N, C_α, C, and O of residues n through $n + 3$. The hydrogen attached to the N_{n+3} atom is also included in the coordinate list. A schematic representation of all atoms considered in our calculations is shown in Figure 19. As each s_i, $i \in \{1, \ldots, N_{exp}\}$ may contain more than one β-turn, n and thereby $n + 3$ take on different values within each s_i, $i \in \{1, \ldots, N_{exp}\}$. Consequently, we identified, $N_\beta = 21{,}091$ β-turns in N_{exp} structures. We are interested in the hydrogen atom attached to N_{n+3} atom for (1) calculation

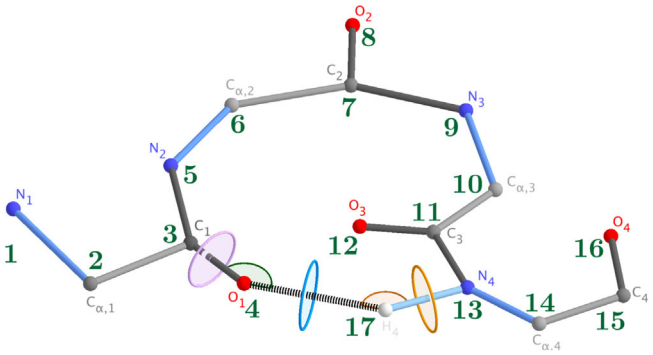

FIGURE 19 A schematic representation of the atoms considered in the analysis of β-turn type, for both experimentally derived and BRIKARD generated structures, are numbered 1 through 17. The ring model used in BRIKARD is defined as R : {3,5,6,7,9,10,11,13,17,4}. The parameters for the formation of H-bonds, considered in the study of backbone torsional variations through *Bricard Curves*, are three torsions, two angles and one distance. The torsions are shown as circles, the angles $\measuredangle C_1O_1H_4$, and $\measuredangle N_4H_4O_1$ are marked and the distance $O_1 \cdots H_4$ is shown as a broken line. Torsion definition, of the three circles shown, with respect to the closed ring model are (a) 13 – 17 : 11 – 13 – 17 – 4, (b) 17 – 4 : 13 – 17 – 4 – 3, and (c) 4 – 3 : 17 – 4 – 3 – 5

of the torsion about the covalent bond $N_{n+3} - H_{n+3}$ and (2) angle at H_{n+3} that is, $\measuredangle N_{n+3}H_{n+3}O_n$. Since, these calculations are not possible with proline in the $n + 3$ position scan results of such kind are discarded.

4. For each of the β-turns B_j, $j \in \{1, \ldots, N_\beta\}$ all 6 H-bond parameters can easily be computed from the Cartesian coordinates of the 17 atoms listed. Additionally, the backbone dihedrals $(\phi_x, \psi_x)_{x = \{2,3\}}$ can also be easily computed. In Figure 18 a scattered plot of torsion about 5 – 6 and 9 – 10 versus torsion about 17 – 4 is shown. The vector length of $(\phi_x, \psi_x)_{x = \{2,3\}}$ and the torsion about 17 – 4 is N_β.
5. The closure criteria defined for modeling of a H-bond in $(n, n + 3)$ type turns is set to the following values: $O_n \cdots H_{n+3}$ = 2.095 Å, $\measuredangle C_nO_nH_{n+3}$ = 116.991° and $\measuredangle N_{n+3}H_{n+3}O_n$ = 156° (see Figure SI.4). With this criteria set, BRIKARD is now used to randomly sample the torsion about 4 – 3. All generated conformations s_b, $b \in \{1, \ldots, N_{\text{sample}}\}$ satisfy the closure criteria and provide us with Cartesian coordinates for the calculation of the H-bond parameters as well as the backbone torsions mentioned in step 4. Alternatively, it is also possible through this design to sample the torsion about the H-bond 17 – 4 instead of the 4 – 3 torsion. From the ring design perspective there is no superiority in the choice of the torsion used for sampling.
6. **This step is optional**: For the sake of computational detail, the conformations s_b, $b \in \{1, \ldots, N_{\text{sample}}\}$ generated through BRIKARD contains the side-chain information. Since, we are only interested in the backbone atoms there is a backbone atoms extraction procedure applied to each s_b, $b \in \{1, \ldots, N_{\text{sample}}\}$ similar to that defined in step 3 to obtain the specific Cartesian coordinates of the 17 atoms shown in Figure 19. The process of obtaining the H-bond parameters and torsions is same as that outlined in step 4.
7. Next, the same closure criteria defined in step 5 is used in conjunction with the Ramachandran's criteria for allowable backbone torsions to generate an ensemble, N_R = 10,000. The conformations generated undergo the same process as those outlined in steps 5 and 6 for the calculation of H-bond parameters and relevant torsions.

The $C_{\alpha,1} - C_{\alpha,4}$ distance for all s_b, $b \in \{1, \ldots, N_{\text{sample}}\}$ and all Ramachandran feasible conformations lie within the range of 4.68–5.95 Å, well within the cut-off range of 7 Å used for experimentally derived structures. The results shown in this section demonstrate the ability to explore backbone torsional dependence for a given set of acceptable H-bond parameters. Repeated application with different closure criteria can be used to investigate the limits of the H-bond geometries.

4 | CONCLUSIONS

In this work, ideas from Kinematics were used to construct a geometrical characterization of three common SSE motifs, helices, β-strands and β-turns. While for a simple helical polygonal line it suffices to consider simple analytic geometry formulas, transitioning to a periodic chain with multiatom, identical units naturally leads to kinematics. Whitworth's relations given by Equations (5) and (6) were utilized to segregate the interrelation of the helical parameters H_p to the analysis of helical structures. Its derivation utilizes a set of intrinsic parameters μ, β, γ, and α to relate the rise per residue, h, and radius r, of the helix as functions of these parameters (see Figure 20). A well known tool from the theory of spherical mechanisms in Kinematics,[52] the *Tetrahedral Equation* (Equation (21)) was then used to relate the orientations of peptide planes about successive virtual $C_\alpha - C_\alpha$ axes. Finally, the molecular geometry software BRIKARD was used to solve the loop closure problem for β-turns.

Some of our methods resemble closely the mathematical treatment of the loop closure problem, developed by Coutsias et al.[58] However, the formulation of the periodic backbone problem for the former two SSE motifs is new: an analytical solution is derived for ideal helical polypeptide backbones, composed of planar peptide units having prescribed inclinations to the helical axis. The Equations (18), (23), (31), and (32) applied together express the backbone torsions (ϕ, ψ) in terms of the distance between next-nearest neighboring C_α atoms, that is, distance between $C_{\alpha,n} - C_{\alpha,n+2}$ atoms. This is parametrized by the angle α subtended at their shared neighboring C_α atom, that is, $C_{\alpha,n+1}$. Favorable peptide plane orientations align $C_n = O_n$ and $N_{n+4} - H_{n+4}$ groups and are consistent with experimentally observed H-bond favorable conformations, which are arrived at without necessitating an a priori distance dependent constraint imposed for an H-bond formation. In this sense, our

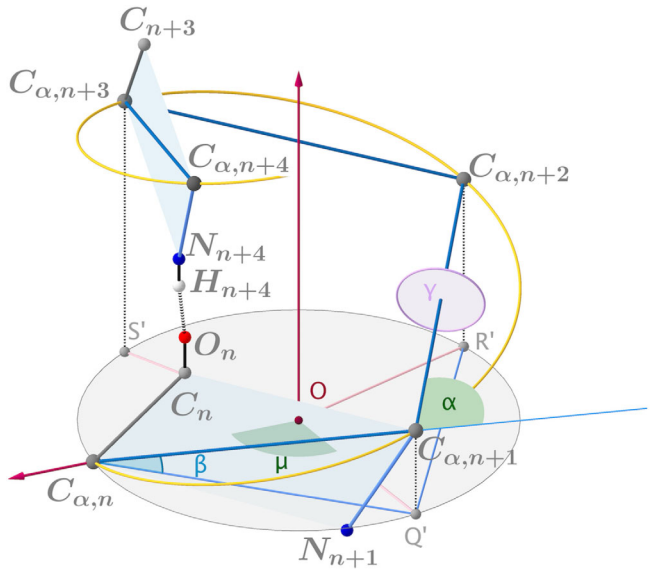

FIGURE 20 A schematic representation of the intramolecular helical H-bond, $O_n\cdots H_{n+4}$

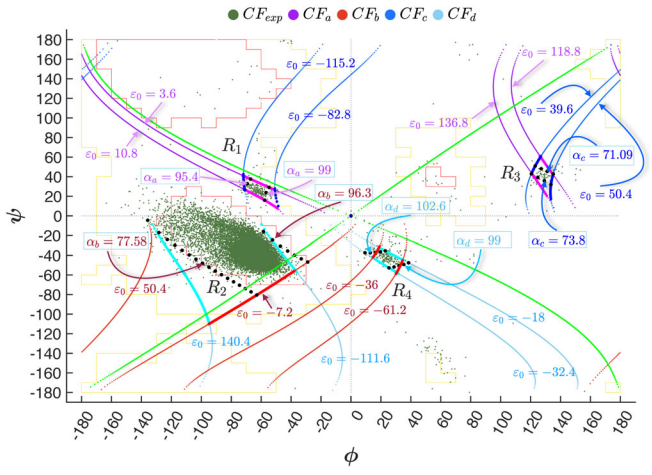

FIGURE 21 A backbone dihedral (ϕ, ψ) map is shown here. The green curves correspond to $\varepsilon_0 = 0°$ for the different sets CF_a, CF_b, CF_c and CF_d. The colored curves represent some of the perturbations in the different sets and are marked by the values of ε_0. The bold colored curves form bounds for the four regions labeled R_1, R_2, R_3, and R_4. It should be noticed that the same regions may be bound by single colored bold curves by traversing across the perturbation curves. A parameter α bound with subscripts denoting the conformational sets it belongs to is also given, α_a refers to bounds for the CF_a conformational set. The two colored curve bounds used here highlight the fact that the same region can be accessed by different conformational sets. For example, R_1 is completely accessible by either CF_b or CF_d conformational sets. As mentioned earlier, these sets are of opposite handedness and their individual accessibility is a consequence of the symmetry of the conformations. The red contour lines correspond to the frequently observed regions of helices (right and left handed) and β-sheets while the yellow regions mark the allowed regions for these motifs

approach is also one of construct-and-test, similar to the classical search for helical conformations. However, ours is a 1-parameter search which yields H-bond suitable solutions for a given range of allowable α values. Geometrically, the allowable region is given by $\alpha \in \{180^\circ - (\xi + \theta + \eta),\ 180^\circ - (\theta - \xi - \eta)\}$, that is, $\alpha \in \{35.126^\circ, 106.228^\circ\}$ (see Table 3 for the definition and values of ξ, θ, η used in this work), for a zero volume constrained motion tetrahedral similar to that shown in Figure 5. The numerical range, obtained post kinematic consideration, of α values consistent with H-bond formation, $\alpha \in \{35.28^\circ, 106.2^\circ\}$, is contained within this geometrically feasible range. It should be noted here that the values of ξ, θ, and η affect both the geometrical and numerical range of allowable α parameter values.

The parameter α coupled with the off vertical planar peptide inclination ε_0 enter in the kinematic solution, Equation (36), and the general Equations (32) and (31). Our formulas result in families of ideal helices, right handed helices, left handed helices, and their opposite handed conformations. These occupy different regions of the $\phi - \psi$ space and are a consequence of the different branches of the kinematic solution (see step 5 in Section 3.4). The $\phi - \psi$ map, shown in Figure 21, is a 2-D projection of 4-D surface formed by α, ε_0, ϕ, and ψ where α and ε_0 serve as a coordinate system. This coordinate system is superposed on the $\phi - \psi$ map and helps to delineate bounds relevant to experimental helical conformations. Mathematically, it will be of relevance to show that the Ramachandran feasible region for SSE motifs or the $\phi - \psi$ map as a whole is a conformal map of the coordinate system introduced here. The torsional pairs associated with ideal helices, within the $\alpha - \varepsilon_0$ bounds, of these types fall neatly in a restricted sector of the helical region and lends to the biological relevance of the constructed conformations. The general nature of the applied perturbation ε_0 is evident from Figure 16 which shows that the allowable range of the parameter α decreases with an increase in perturbation and is viewed as the shortening of the horizontal lines. Moreover, we provide a step-by-step numerical protocol for deriving all such ideal and nonideal helices.

It is possible to view an H-bond arrangement between (n, $n + k$), k = {3,4,5} residues as a closed kinematic chain, Figure 20 shows such an arrangement for $k = 4$. Thus, the problem of finding H-bond capable ideal helices could also be approached as a problem in inverse kinematics with appropriate constraints, distance $O_n\cdots H_{n+4}$, $\measuredangle H_{n+4}O_nC_n$, and $\measuredangle N_{n+4}H_{n+4}O_n$, to include perturbations to periodicity. This is precisely the path undertaken for β-turn (see Figure 19). By comparing with Dunbrack's database of structures,[13] we show that using parameter values in the central voxel of three-parameter space - two angles and a distance (see Figure SI.4)—we get Bricard curves for torsional correlations that span well the regions of physically occurring structures (Figure 18, 2-D Figure SI.5, 3-D Figure SI.6). Thus the kinematic analysis successfully recapitulates the physical space of β-turns (for a

distribution of the experimentally derived parameter α values and dihedrals (ϕ, ψ) see Figure SI.7). The spread of the physical points is of course to be expected due to the variability consistent with variations in the closure parameters.

ACKNOWLEDGMENTS

This work was supported in part by NIH grant GM107104. The authors acknowledges support from the Laufer Center for Physical and Quantitative Biology at Stony Brook University. The authors also thank Kevin Hauser for sharing important references and Ken A. Dill for helpful suggestions on an initial draft of the work.

ENDNOTES

[1] The variable KM here refers to parameter names used by Kortemme et al. [47]

[2] http://www.cgl.ucsf.edu/chimera/1.4.1/docs/UsersGuide/idatm.html

[3] http://dunbrack.fccc.edu/Guoli/pisces_download.php

[4] *Bricard Curves* since they are generated using the flexibility analysis of tetrahedral equations pioneered by R. Bricard. [3]

[5] https://www.rcsb.org

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Mosavverul Hassan https://orcid.org/0000-0002-9540-0490
Evangelos A. Coutsias https://orcid.org/0000-0003-2910-9125

REFERENCES

[1] K. Cahill, *Phys. Rev. E* **2005**, *72*, 062901.
[2] R. B. Cooley, D. J. Arp, P. A. Karplus, *J. Mol. Biol.* **2010**, *404*, 232.
[3] R. B. Corey, L. C. Pauling, *Proc. Royal Soc. London. Series B-Biol. Sci.* **1953**, *141*, 10.
[4] G. Ramachandran, C. Venkatachalam, S. Krimm, *Biophys. J.* **1966**, *6*, 849.
[5] L. Pauling, R. B. Corey, H. R. Branson, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205.
[6] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 235.
[7] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 241.
[8] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 729.
[9] C. Ramakrishnan, *Proc. Indiana Acad. Sci.* **1964**, *59*, 327.
[10] C. Ramakrishnan, G. Ramachandran, *Biophys. J.* **1965**, *5*, 909.
[11] M. B. Hossain, D. Van der Helm, *J. Am. Chem. Soc.* **1978**, *100*, 5191.
[12] C. Venkatachalam, *Biopolymers* **1968**, *6*, 1425.
[13] M. Shapovalov, S. Vucetic, R. L. Dunbrack Jr., *PLoS Comput. Biol.* **2019**, *15*, e1006844.
[14] P. A. Kollman, *Acc. Chem. Res.* **1977**, *10*, 365.
[15] L. Pauling, *The nature of the chemical bond*, Cornell, New York **1939**.
[16] T. Oroguchi, M. Nakasako, *Sci. Rep.* **2017**, *7*, 15859.
[17] G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer Science & Business Media, Berlin, Heidelberg **2012**.
[18] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, *Pure Appl. Chem.* **2011**, *83*, 1619.
[19] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, *Pure Appl. Chem.* **2011**, *83*, 1637.
[20] D. Santos-Martins, S. Forli, *J. Chem. Theory Comput.* **2020**, *16*, 2846.
[21] J. Emsley, *Chem. Soc. Rev.* **1980**, *9*, 91.
[22] G. R. Desiraju, *Acc. Chem. Res.* **2002**, *35*, 565.
[23] G. R. Desiraju, T. Steiner, *Chem. Commun.* **1998**, 891.
[24] L. Paoloni, *J. Chem. Phys.* **1959**, *30*, 1045.
[25] S. Scheiner, E. A. Hillenbrand, *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 2741.
[26] S. Scheiner, P. Redfern, E. A. Hillenbrand, *Int. J. Quantum Chem.* **1986**, *29*, 817.
[27] H. Tsubomura, *Bull. Chem. Soc. Jpn.* **1954**, *27*, 445.
[28] C. Coulson, U. Danielsson, *Ark. Fysik* **1954**, *8*, 239.
[29] K. Morokuma, *Acc. Chem. Res.* **1977**, *10*, 294.
[30] L. Kroon-Batenburg, J. Kanters, *Acta Crystallog. Sect. B: Struct. Sci.* **1983**, *39*, 749.
[31] Foloppe, N.; Chen, J., I *Future Med. Chem.* **2019**, *11*, 97–118.
[32] C. Coulson, *Hydrogen bonding*, Elsevier, Imprint, Pergamon **1959**, p. 339.
[33] P. A. Kollman, *J. Am. Chem. Soc.* **1972**, *94*, 1837.
[34] H. Umeyama, K. Morokuma, *J. Am. Chem. Soc.* **1977**, *99*, 1316.
[35] A. V. Morozov, T. Kortemme, *Adv. Protein Chem.* **2005**, *72*, 1.
[36] Z. Lu, N. Zhou, Q. Wu, Y. Zhang, *J. Chem. Theory Comput.* **2011**, *7*, 4038.
[37] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, *J. Chem. Theory Comput.* **2016**, *12*, 281.
[38] P. Cieplak, J. Caldwell, P. Kollman, *J. Comput. Chem.* **2001**, *22*, 1048.
[39] T. Bereau, C. Kramer, M. Meuwly, *J. Chem. Theory Comput.* **2013**, *9*, 5450.
[40] P. G. Karamertzanis, C. C. Pantelides, *Mol. Simul.* **2004**, *30*, 413.
[41] K. Vanommeslaeghe, O. Guvench, A. D. MacKerell, *Curr. Pharm. Des.* **2014**, *20*, 3281.
[42] J.-H. Lii, N. L. Allinger, *J. Phys. Org. Chem.* **1994**, *7*, 591.
[43] F. Fabiola, R. Bertram, A. Korostelev, M. S. Chapman, *Protein Sci.* **2002**, *11*, 1415.
[44] M. J. Sippl, *Curr. Opin. Struct. Biol.* **1995**, *5*, 229.
[45] A. Grishaev, A. Bax, *J. Am. Chem. Soc.* **2004**, *126*, 7281.
[46] B. A. Grzybowski, A. V. Ishchenko, R. S. DeWitte, G. M. Whitesides, E. I. Shakhnovich, *J. Phys. Chem. B* **2000**, *104*, 7293.
[47] T. Kortemme, A. V. Morozov, D. Baker, *J. Mol. Biol.* **2003**, *326*, 1239.
[48] A. V. Morozov, T. Kortemme, K. Tsemekhman, D. Baker, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6946.
[49] E. A. Coutsias, K. W. Lexa, M. J. Wester, S. N. Pollock, M. P. Jacobson, *J. Chem. Theory Comput.* **2016**, *12*, 4674.
[50] A. L. Morris, M. W. MacArthur, E. G. Hutchinson, J. M. Thornton, *Proteins: Struct., Func., Bioinfor.* **1992**, *12*, 345.
[51] Hohenwarter, M. *M.Sc. thesis*, Paris Lodron University, (Salzburg, Austria), **2002**; (In German).
[52] K. H. Hunt, *Kinematic Geometry of Mechanisms*, Vol. 7, Oxford University Press, USA **1978**.
[53] R. Bricard, *J.Math.Pures Appl.* **1897**, *3*, 113.
[54] W. Whitworth, *The Oxford, Cambridge and Dublin Messenger of Mathematics* **1875**, *4*, 88.
[55] T. Shimanouchi, S.-i. Mizushima, *J. Chem. Phys.* **1955**, *23*, 707.
[56] T. Miyazawa, *J. Polym. Sci.* **1961**, *55*, 215.
[57] U. Langel, B. F. Cravatt, A. Graslund, N. Von Heijne, M. Zorko, T. Land, S. Niessen, *Introduction to peptides and proteins*, CRC Press, Boca Raton **2009**, p. 36.
[58] E. A. Coutsias, C. Seok, M. P. Jacobson, K. A. Dill, *J. Comput. Chem.* **2004**, *25*, 510.

[59] E. A. Coutsias, C. Seok, M. J. Wester, K. A. Dill, *Int. J. Quantum Chem.* **2006**, *106*, 176.
[60] J. E. Mills, P. M. Dean, *J. Comput. Aided Mol. Des.* **1996**, *10*, 607.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Hassan M, Coutsias EA. Protein secondary structure motifs: A kinematic construction. *J Comput Chem*. 2021;42:271–292. https://doi.org/10.1002/jcc.26448