Published in final edited form as:
J Comput Chem. 2011 September ; 32(12): 2515–2525. doi:10.1002/jcc.21812.

Three-Residue Loop Closure in Proteins: A New Kinematic Method Reveals a Locus of Connected Loop Conformations

ALI NEKOUZADEH[1,2] and **YORAM RUDY**[1,2]

[1]Cardiac Bioelectricity and Arrhythmia Center, Washington University in St. Louis, St. Louis, Missouri

[2]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri

Abstract

The closure of a three-residue loop was studied using a developed kinematic method. It was shown that there are infinite number of three-residue loops (a locus of conformations), which can connect two segments of a polypeptide. This adds to the current understanding of a finite number of conformations for three-residue loop-closure. In the developed method, some of the equations can be solved analytically to reduce the computation cost. Benefiting from the reduced computation time, we determined all the relative positions of two polypeptide segments that can be connected by a three-residue loop.

Keywords

loop closure; loop conformation; kinematic method; dihedral angles

Introduction

Determining the conformation of amino acid residues in the loop connecting two segments of a protein secondary structure is essential in homology modeling as well as in modeling the dynamics of large scale conformational changes, such as those occurring during ion-channel gating. Prediction of loop structure is performed in two phases: determining all possible combinations of ϕ and ψ dihedral angles that keep the loop connected at both ends to the rest of the polypeptide (referred to as loop–closure problem),[1–9] and selecting the most probable connected conformations.[10–15] Determining the probable conformations is based on scoring the loop conformations according to their energy; it is not the topic of this study, which addresses the loop-closure problem.

Identification of connected loop conformations may be achieved by sampling the loop dihedral angles to determine the combinations of these angles that match the free end of the loop to its target location for loop closure.[10,14,15] Fine sampling of the angles makes the computations extremely time consuming and expensive. Therefore, ϕ and ψ angles are

Correspondence to: A. Nekouzadeh; ali@biomed.wustl.edu.

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

usually sampled coarsely over the more probable regions of the Ramachandran plot[16] to obtain approximate loop closure. In an approximate solution, the loose end of the loop is located close to its target location. Depending on the magnitude of error tolerance, there may be no exact solution in the vicinity of an approximate solution. Also, some exact solutions may be missed, because they are not in the vicinity of an approximate solution. Alternatively, analytical approaches, termed kinematic closure methods, have been used.[1–4] In these approaches, trigonometric[1] or polynomial[2–4] equations for six dihedral angles (usually of three consecutive amino acid residues) are derived and solved numerically for any given combination of the remaining dihedral angles. This reduces the computation time significantly and enables a finer sampling of the remaining angles.

In this article, we first re-evaluate the closure of a three amino acid-residue loop and show that there is a continuous locus of conformations that closes the loop, rather than the commonly assumed finite number of up to 16 conformations. We then present a new kinematic closure technique that can predict this locus of conformations with less computation than existing kinematic closure methods.

Results

Loop-Closure Problem

The loop-closure problem involves locating atoms of the loop residues, while preserving certain properties of the chemical bonds. Bonds lengths, bonds angles, and ω dihedral angle can vary over a small range and cannot account for large deformation of the loop; therefore, they are considered constant in this study. This means that the relative positions of atoms from one C^{α} to the next C^{α} are preserved. We term this preserved structure the backbone structure in this paper. In contrast, dihedral angles can vary over a large range, resulting in large deformations of the loop. The ϕ and ψ backbone dihedral angles are the significant degrees of freedom in the loop-closure problem.[1] Figure 1 shows a three-residue loop that connects two segments of a polypeptide. Residues before and after the loop are termed preloop and target residues, respectively. In a three-residue loop-closure, given the ϕ and ψ angles of the preloop residue and the three residues in the loop, the location of N and C^{α} atoms of the target residue can be predicted. The loop closes properly if and only if these predicted locations coincide with the known locations of N and C^{α} atoms of the target residue. This results in six constraints (equations) for loop closure, as matching the location of each atom requires imposing three constraints (one for each coordinate). However, there is one degree of dependency between the locations of N and C^{α} of the target residue; the distance between these two atoms is constant. Therefore, there are five independent constraints (three coordinates of one atom and only two coordinates of the other atom).

Three-Residue Loop-Closure: A Continuous Locus of Solutions Rather than a Finite Number of Discrete Solutions

Different conformations of the three-residue loop backbone atoms are associated with different dihedral angles: ψ_0, ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, ψ_3, and ϕ_4. Note that the ω dihedral angles are 180° (or in rare cases 0°). Knowing the backbone locations of the preloop and target residues, there is one degree of dependency between these angles and only seven of them

may be considered as independent degrees of freedom. For the loop closure, these dihedral angles should be subjected to five constraints. As the number of degrees of freedom is greater by 2 than the number of constraints, there are infinite number of conformations for a three-residue loop closure (if possible) and the associated dihedral angles form a continuous 2D locus in solution space. In considering ψ_0 as a degree of freedom, it is implicitly assumed that the location of the oxygen atom of the preloop residue is adjustable as well. Arguing that this oxygen atom should maintain its location, ψ_0 cannot be considered variable and there would be only 6 degrees of freedom, resulting in a 1D locus of solutions (There is 1 degree of freedom more than the number of constraints.).

In previously published studies,[1–4] the number of constraints was considered 6 rather than 5, which led to determining a finite number of conformations for three-residue loops. In the method developed by Go and Scheraga,[1] when defining the local coordinates and deriving the governing equations, it was implicitly assumed that the location of the C^a of the third loop residue is known. This location can vary along a circular path depending on the value of ϕ_4. In the method developed later by Wedemeyer and Scheraga[2] and Coutsias et al.,[4] the problem was defined in terms of 3 degrees of freedom and three constraints. In this method, the possible conformations were also limited by assuming a fixed location for C^a of the third loop residue.

Governing Equations

The three residues in the loop are labeled 1, 2, and 3, the residue before the loop (preloop residue) is labeled 0 and the residue after the loop (target residue) is labeled 4. Positions of residues 0 and 4 are fixed, and the goal is to determine the locations of the backbone atoms of residues 1, 2, and 3 so that these three residues connect residues 0 and 4. Knowing the position of residue 0, locations of all the backbone atoms of the loop residues can be determined in terms of ψ_0, ϕ_1, ψ_1, ϕ_2, ψ_2, and ϕ_3. Loop closure can be imposed by calculating locations of N and C^a of the forth residue in terms of theses dihedral angles and ψ_3, and constraining them to the known locations of the fourth residue N and C^a. This approach imposes all the constraints on the peptide bond automatically during the calculation of the N and C^a locations. Instead of matching five coordinates of fourth residue N and C^a, we can match the location of one (three constraints) and the direction of the N—C^a bond (two constraints). In summary, the dihedral angles ψ_0, ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 should be selected such that the prediction for the location of C^a (termed location vector, Δ) and the direction of N—C^a bond (termed direction vector, $\boldsymbol{r}$) of the fourth residue coincide with the known location and direction of the target residue (Fig. 2). To formulate the problem, we define a set of local coordinates for the loop residues. For each residue, two local coordinate systems are considered: R and R′ coordinate systems (Fig. 3). Center of the R coordinate system is located on C^a; its third axis is along the N—C^a bond and its first axis is on the surface that passes through N and C^a of the residue and the C of the previous residue. Center of the R′ coordinate system is located on C; its third axis is along the C^a—C bond and its first axis is in the surface that passes through N, C^a, and C.

If we move along the loop residues from the preloop residue (residue 0) toward the target residue (residue 4), then variation of ϕ_i changes the location of C_i and the following residues

but has no effect on R_i (the *i*th R coordinate system); it will change the relative position of R_i' (the *i*th R′ coordinate system) with respect to R_i. Also, variation of ψ_i has no effect on either R_i or R_i'; it will change the relative position of R_{i+1} with respect to R_i'. The relative position of any two of these local coordinate systems can be expressed by a rotational matrix and a displacement vector. The displacement vector is the relative location of the centers of the two coordinate systems. Because the backbone atoms of all the residues are similar, the relative rotation of local coordinates can be formulated in terms of two rotation matrices termed **S** and **F**. The **F** matrix transforms the coordinates of a vector in R′ to R; the **S** matrix transforms coordinates of a vector in R to the preceding R′. As explained above, the **F** matrix is a function of the ϕ angle, and the **S** matrix is a function of the ψ angle.

A spatial location can be transformed from R_i' to R_i through:

$$[\,]_i = \mathbf{F}(\varphi_i)\,([\,]_{i'} - \boldsymbol{d1}) \quad (1)$$

and from R_{i+1} to R_i' through:

$$[\,]_{i'} = \mathbf{S}(\psi_i)\,([\,]_{i+1} - \boldsymbol{d2}) \quad (2)$$

where ***d1*** is the location of C_i^α in R_i' and ***d2*** is the location of C_i in R_{i+1}. ***d1*** and ***d2*** do not depend on the ϕ or ψ angles (Fig. 3). **F**, **S**, ***d1***, and ***d2*** can be determined in terms of bond lengths and bond angles that are shown in Figure 4. Assuming $\omega = 180°$, we derived the rotation matrices as:

$$\mathbf{F}(\varphi) = \begin{bmatrix} \cos(\pi-\gamma_2)\cos\varphi & -\sin\varphi & \sin(\pi-\gamma_2)\cos\varphi \\ \cos(\pi-\gamma_2)\sin\varphi & \cos\varphi & \sin(\pi-\gamma_2)\sin\varphi \\ -\sin(\pi-\gamma_2) & 0 & \cos(\pi-\gamma_2) \end{bmatrix} \quad (3)$$

$$\mathbf{S}(\psi) = \begin{bmatrix} -\cos(\gamma_1-\gamma_3)\cos\psi & -\sin\psi & \sin(\gamma_1-\gamma_3)\cos\psi \\ -\cos(\gamma_1-\gamma_3)\sin\psi & -\cos\psi & \sin(\gamma_1-\gamma_3)\sin\psi \\ \sin(\gamma_1-\gamma_3) & 0 & \cos(\gamma_1-\gamma_3) \end{bmatrix} \quad (4)$$

Similar rotation matrices can be derived for the rare case of $\omega = 0°$. For the ideal SP^2 and SP^3 orbital: $\gamma_1 = \gamma_3 = 120°$, $\gamma_2 = 109.5°$ and $\mathbf{F}(\phi)$ and $\mathbf{S}(\psi)$ can be simplified as:

$$\mathbf{F}(\varphi) = \begin{bmatrix} \cos 70.5\cos\varphi & -\sin\varphi & \sin 70.5\cos\varphi \\ \cos 70.5\sin\varphi & \cos\varphi & \sin 70.5\sin\varphi \\ -\sin 70.5 & 0 & \cos 70.5 \end{bmatrix} \quad (5)$$

$$\mathbf{S}(\psi) = \begin{bmatrix} -\cos\psi & \sin\psi & 0 \\ -\sin\psi & -\cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

d1 and ***d2*** are:

$$\boldsymbol{d_1} = \begin{bmatrix} 0 \\ 0 \\ -l_2 \end{bmatrix} \quad (7)$$

$$\boldsymbol{d_2} = \begin{bmatrix} l_3 \sin\gamma_1 \\ 0 \\ -l_1 + l_3 \cos\gamma_1 \end{bmatrix} \quad (8)$$

In this ideal case, the C^a—C bond is parallel to the N—C^a bond.[9]

To determine the required constraints (equations) for loop closure, we calculate the location of $C^a{}_4$ and the direction of the N_4—$C^a{}_4$ bond in terms of the dihedral angles. We assumed that the location and direction of the target residue is given in the R′ coordinate system of the preloop residue (R'_0) by Δ and ***r***, respectively (Fig. 2). This coordinate system does not change with ψ_0, ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 and is the global coordinate in our problem. The location of $C^a{}_4$ and the direction of N_4—$C^a{}_4$ in the R coordinate system of the target residue (R_4) are: [0 0 0] and [0 0 1], respectively. We can transform this location and direction to the global coordinate system by consecutive transforms between the local coordinate systems:

$$\overrightarrow{\Delta} = \mathbf{S}(\psi_0)\left(\mathbf{F}(\varphi_1)\left(\mathbf{S}(\psi_1)\left(\mathbf{F}(\varphi_2)\left(\mathbf{S}(\psi_2)\left(\mathbf{F}(\varphi_3)\left(\mathbf{S}(\psi_3)\times\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_4 - \boldsymbol{d_2}\right) - \boldsymbol{d_1}\right) - \boldsymbol{d_2}\right) - \boldsymbol{d_1}\right) - \boldsymbol{d_2}\right) - \boldsymbol{d_1}\right) - \boldsymbol{d_2}\right) \quad (9)$$

$$\mathbf{r} = \mathbf{S}(\varphi_0)\mathbf{F}(\varphi_1)\mathbf{S}(\psi_1)\mathbf{F}(\varphi_2)\mathbf{S}(\psi_2)\mathbf{F}(\varphi_3)\mathbf{S}(\psi_3)\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_4 \quad (10)$$

As discussed in before, in a three-residue loop-closure problem there are seven variables (ψ_0, ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3) and only five constraints. The two additional degrees of freedom lead to a 2D locus of solutions. We consider ψ_0 and ϕ_1 as free changing variables and calculate the values of ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 that lead to a successful loop-closure for a given combination of ψ_0 and ϕ_1. Therefore, in eqs. (9) and (10), in addition to **Δ** and ***r***, ψ_0 and ϕ_1 are known. Equations (9) and (10) may be rewritten as:

$$\mathbf{F}^{\mathrm{T}}(\varphi_1)\left(\mathbf{S}^{\mathrm{T}}(\psi_0)\boldsymbol{\Delta} + \boldsymbol{d_2}\right) = \mathbf{S}(\psi_1)\mathbf{F}(\varphi_2)\left(\mathbf{S}(\psi_2)\mathbf{F}(\varphi_3)\boldsymbol{d}(\psi_3) + \boldsymbol{d}(\psi_2)\right) + \boldsymbol{d}(\psi_1) \quad (11)$$

$$\mathbf{F}^{\mathrm{T}}(\varphi_1)\mathbf{S}^{\mathrm{T}}(\psi_0)\mathbf{r} = \mathbf{S}(\psi_1)\mathbf{F}(\varphi_2)\mathbf{S}(\psi_2)\mathbf{F}(\varphi_3)\mathbf{S}(\psi_3)\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_4 \quad (12)$$

where

$$\boldsymbol{d}(\psi)=\begin{bmatrix} a\cos\psi \\ a\sin\psi \\ b \end{bmatrix} \quad (13)$$

$$a=l_3\sin(\pi-\gamma_3)+l_1\sin(\gamma_1-\gamma_3) \quad (14)$$

$$b=l_2-l_3\cos(\gamma_3)+l_1\cos(\gamma_1-\gamma_3) \quad (15)$$

Note that the left sides of both eqs. (11) and (12) are known. For ideal orbital angles, eqs. (11) and (12) can be simplified to:

$$\mathbf{F}^{\mathrm{T}}(\varphi_1)\left(\mathbf{S}^{\mathrm{T}}(\psi_0)\overrightarrow{\Delta}+\boldsymbol{d_2}\right)=\mathbf{F}(\alpha)(\mathbf{F}(\beta)\boldsymbol{d}(\psi_3)+\boldsymbol{d}(\psi_2))+\boldsymbol{d}(\psi_1) \quad (16)$$

$$\mathbf{F}^{\mathrm{T}}(\varphi_1)\mathbf{S}^{\mathrm{T}}(\psi_0)\overrightarrow{r}=\mathbf{F}(\alpha)\mathbf{F}(\beta)\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_{3'} \quad (17)$$

where

$$\alpha=\psi_1+\varphi_2+\pi \quad \text{and} \quad \beta=\psi_2+\varphi_3+\pi \quad (18)$$

Both **F** and **S** are rotational matrices that when multiply a vector, do not change its magnitude. Therefore, the magnitude of the direction vector ***r*** should be normalized to one, as the magnitude of the right side of eq. (17) is one. Equation (17) imposes two constraints (it has only two independent equations) and eq. (16) imposes three constraints, providing together five equations for the five unknowns ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3. An interesting feature of eqs. (16) and (17) is that the system of five equations and five unknown is broken down to two systems: two equations and two unknowns in terms of α and β [eq. (17)], and three equations and three unknowns in terms of ψ_1, ψ_2, and ψ_3 [eq. (16)]. This reduces the required computations significantly.

Defining $\boldsymbol{r}^*$ as:

$$\mathbf{r}^*=\begin{bmatrix} r_1^* \\ r_2^* \\ r_3^* \end{bmatrix}=\mathbf{F}^{\mathrm{T}}(\varphi_1)\mathbf{S}^{\mathrm{T}}(\psi_0)\overrightarrow{r} \quad (19)$$

we may rewrite eq. (17) as:

$$r_1^*=\cos 70.5\sin 70.5\cos\alpha\cos\beta-\sin 70.5\sin\alpha\sin\beta+\cos 70.5\sin 70.5\cos\alpha \quad (20)$$

$$r_2^*=\cos 70.5\sin 70.5\sin\alpha\cos\beta+\sin 70.5\cos\alpha\sin\beta+\cos 70.5\sin 70.5\sin\alpha \quad (21)$$

$$r_3^*=-\sin^2 70.5\cos\beta+\cos^2 70.5 \quad (22)$$

Equation (22) can be solved analytically for β and substituting for β in eq. (20) or (21), α can be determined analytically as well.

Similarly defining $\mathbf{\Delta}^*$ as:

$$\mathbf{\Delta}^*=\begin{bmatrix}\Delta_1^*\\ \Delta_2^*\\ \Delta_3^*\end{bmatrix}=\mathbf{F}^{\mathrm{T}}(\alpha)\mathbf{F}^{\mathrm{T}}(\varphi_1)(\mathbf{S}^{\mathrm{T}}(\psi_0)\Delta+\boldsymbol{d2}) \quad (23)$$

we may rewrite eq. (16) as:

$$\cos\psi_3=\frac{-\Delta_3^*+b(1+2\cos 70.5)}{a\sin 70.5}+\cos(\psi_1-\alpha) \quad (24)$$

$$\cos\psi_2=\frac{\Delta_1^*+b\sin 70.5(1-\cos\beta)}{\alpha}-\cos 70.5\cos(\psi_1-\alpha)-\cos 70.5\cos\beta\cos\psi_3+\sin\beta\sin\psi_3 \quad (25)$$

$$\sin(\psi_1-\alpha)+\sin\psi_2+\cos 70.5\sin\beta\cos\psi_3+\cos\beta\sin\psi_3+\frac{-\Delta_2^*+b\sin 70.5\sin\beta}{\alpha}=0 \quad (26)$$

If ψ_1 is known, the two variables, ψ_3 and ψ_2, can be calculated analytically in terms of ψ_1 via eqs. (24) and (25), respectively. Therefore, the numerical solution of these equations reduces to search for a ψ_1 that satisfies eq. (26).

Analytical and Numerical Solutions of the Equations

Equations (20)–(22) can be solved analytically for α and β. β can be calculated from eq. (22) as:

$$\beta=\pm\cos^{-1}\left(\frac{\cos^2 70.5-r_3^*}{\sin^2 70.5}\right) \quad (27)$$

Knowing that the term inside the parentheses in eq. (27) can only vary between −1 and +1, the acceptable range of r_3^* for a successful loop-closure can be determined. Knowing β, we can calculate α using eq. (21) or (20). Note that these equations are not identical. Equations (20) and (21) can be rewritten as eqs. (28) and (29), respectively:

$$\cos(\alpha-\lambda)=\frac{r_1^*}{A} \quad (28)$$

$$\sin(\alpha-\lambda)=\frac{r_2^*}{A} \quad (29)$$

where

$$A=\sqrt{\sin^2 70.5\sin^2\beta+(\sin 70.5\cos 70.5\cos\beta+\sin 70.5\cos 70.5)^2} \quad (30)$$

and

$$\begin{array}{c}\sin\lambda=\frac{-\sin 70.5\sin\beta}{A}\&\\ \cos\lambda=\frac{\sin 70.5\cos 70.5\cos\beta+\sin 70.5\cos 70.5}{A}\end{array} \quad (31)$$

Given that the magnitude of $\boldsymbol{r}^*$ is one, eqs. (28) and (29) are consistent. Although each of these equations has up to two solutions for α, only one solution satisfies both equations. Therefore, for each β calculated from eq. (27), α is calculated as:

$$\alpha=\lambda+k\cos^{-1}\frac{r_1^*}{A} \quad k=\begin{cases}-1 & r_2^*<0\\ +1 & r_2^*\geq 0\end{cases} \quad (32)$$

There are up to two pairs of solutions for α and β. For each pair, eqs. (24)–(26) can be solved numerically to determine the three unknowns: ψ_1, ψ_2, and ψ_3. These equations are linear in terms of six variables, the sin and cos of the three ψ angles. Values of the 3 sin terms can be determined in terms of the 3 cos terms, separately. For each cos term, there are two possible values for the sin term:

$$\sin x=\pm\sqrt{1-\cos^2 x} \quad (33)$$

Therefore, eqs. (24)–(26) represents $2^3 = 8$ different sets of three equations and three unknowns and have up to eight solutions for each pair of α and β. Consequently, loop closure has up to 16 solutions for the ideal SP^2 orbital. Note that 16 is an upper bound for the number of solutions. Detailed analysis of the number of solutions in different regions of the variables domain (similar to what was performed by Milgram et al.[9]) is outside the scope of this article; it will be conducted as a separate study to further characterize our new approach.

We could not find analytical solutions for eqs. (24)–(26), however, similar to equations developed by Go and Scheraga,[1] these equations can be simplified analytically into one equation in term of one angle. Here, we used ($\psi_1 - \alpha$) as the independent variable. A possible range of this angle was determined from eq. (24) and the cos($\psi_1 - \alpha$) and sin($\psi_1 - \alpha$) were computed for all the ($\psi_1 - \alpha$) angles within this range. cos ψ_3 and, consequently, sin ψ_3 were computed using eq. (24). Then, cos ψ_2 and sin ψ_2 were computed using eqs. (25) and (26), respectively. The function $f(\psi_1 - \alpha) = \cos^2 \psi_2 + \sin^2 \psi_2 - 1$ was then computed. The angles at which this function crosses zero are the solutions values of ($\psi_1 - \alpha$). We used

very fine sampling of the ψ_1 angle with 1 degree increments, which should capture solutions within ±0.5 degrees of the sampled values. This precision is sufficient for any practical application of loop closure, knowing that the stochastic fluctuations in dihedral angles are much larger than ±0.5 degrees.

Nonideal Bond Angles

Typically for protein structures, the angles γ_1 and γ_3 are slightly different than the ideal 120° considered in this formulations. Therefore, the term $\gamma_1 - \gamma_3$ may be nonzero and consequently eqs. (16) and (17) are not accurate. We calculated the joint probability distribution of these angles in the 3D structure of the Kv7.1 channel, which was derived previously[17] based on homology to Kv1.2 with a known crystal structure in the open state.[18] Considering only the five loop regions between consecutive transmembrane segments (in all four subunits), the average values of γ_1 and γ_3 are 120.7° and 117.2°, respectively, and their standard deviations are 2.8° and 1.3°, respectively. Figure 5 shows the joint probability distribution of γ_1 and γ_3 in the loop regions of Kv7.1 protein. The white line marks the backbone structures with parallel C^{α}—C and N—C^{α} bonds. Eight percent of the backbone structures have parallel bonds. In particular, the following residues had parallel bonds (within ±0.5°) in all four subunits: Thr153 (~118.3°) in S1–S2 loop, Phe193 (~117.5°) in S2–S3 loop, Ser217 (~116.9°) in S3–S4 loop, and Gly314 (~116.9°) in the loop between P-loop helix and S6. Also, the following residues had parallel bonds in at least two subunits: Val141, Thr144, and Asp 317. The backbone with γ_1 = 120° ± 0.5° and γ_3 = 117° ± 0.5° is the most probable structure, with a probability of about 8% as well. In the following, we present a method to determine the loop closure ϕ and ψ dihedral angles for a backbone structure with nonparallel bonds, using the dihedral angles computed for the backbone structure with parallel bonds as a starting point.

Equations (11) and (12) may be written in the general form of:

$$\begin{aligned}\Delta_1^* &= f_1(\psi_1, \varphi_2, \psi_2, \varphi_3, \psi_3)\\ \Delta_2^* &= f_2(\psi_1, \varphi_2, \psi_2, \varphi_3, \psi_3)\\ \Delta_3^* &= f_3(\psi_1, \varphi_2, \psi_2, \varphi_3, \psi_3)\\ r_1^* &= f_4(\psi_1, \varphi_2, \psi_2, \varphi_3, \psi_3)\\ r_2^* &= f_5(\psi_1, \varphi_2, \psi_2, \varphi_3, \psi_3)\end{aligned} \quad (34)$$

Assume $\Psi_i = [\psi_{1i}\ \phi_{2i}\ \psi_{2i}\ \phi_{3i}\ \psi_{3i}]^T$ is the (exact) solution of eqs. (16) and (17) (the ideal case where $\gamma_1 = \gamma_3$) for a given direction and location of a target residue given by $L=[\Delta_1^*\ \Delta_2^*\ \Delta_3^*\ r_1^*\ r_2^*]^T$. This would be an approximate solution of eqs. (11) and (12) where $\gamma_1 \neq \gamma_3$ (nonideal case). Assume $L_i=[\Delta_{1i}^*\ \Delta_{2i}^*\ \Delta_{3i}^*\ r_{1i}^*\ r_{2i}^*]^T$ is the location and direction of the target residue for Ψ_i dihedral angles (ideal solution which serves as an approximate solution for the nonideal case) when the loop consists of three residues with nonideal backbone structures. L_i is deflected from the desired location and direction by $\delta L = L - L_i$.

We can linearize equation set (34) in the vicinity of the approximate solution using the Jacobean matrix, J:

$$\delta L=\delta \begin{bmatrix} \Delta_1^* \\ \Delta_2^* \\ \Delta_3^* \\ r_1^* \\ r_2^* \end{bmatrix} =[J_{ij}] \left(\begin{bmatrix} \psi_1 \\ \varphi_2 \\ \psi_2 \\ \varphi_3 \\ \psi_3 \end{bmatrix} - \begin{bmatrix} \psi_{1i} \\ \varphi_{2i} \\ \psi_{2i} \\ \varphi_{3i} \\ \psi_{3i} \end{bmatrix} \right) \quad (35)$$

where,

$$J_{i1}=\frac{\partial f_i}{\partial \psi_1}, \quad J_{i2}=\frac{\partial f_i}{\partial \varphi_2}, \quad J_{i3}=\frac{\partial f_i}{\partial \psi_2}, \quad J_{i4}=\frac{\partial f_i}{\partial \varphi_3}, \quad J_{i5}=\frac{\partial f_i}{\partial \psi_3}, \quad (36)$$

Given that $\gamma_1 - \gamma_3$ is small, the solution of eqs. (11) and (12) can be derived through eq. (35) as:

$$\Psi= \begin{bmatrix} \psi_1 \\ \varphi_2 \\ \psi_2 \\ \varphi_3 \\ \psi_3 \end{bmatrix} = \begin{bmatrix} \psi_{1i} \\ \varphi_{2i} \\ \psi_{2i} \\ \varphi_{3i} \\ \psi_{3i} \end{bmatrix} +[J_{ij}]^{-1} \left(\begin{bmatrix} \Delta_1^* \\ \Delta_2^* \\ \Delta_3^* \\ r_1^* \\ r_2^* \end{bmatrix} - \begin{bmatrix} \Delta_{1i}^* \\ \Delta_{2i}^* \\ \Delta_{3i}^* \\ r_{1i}^* \\ r_{2i}^* \end{bmatrix} \right) \quad (37)$$

If the Jacobean matrix is close to singular and its inverse consists of relatively large elements, we may need to use an intermediate backbone structure. The solutions of the intermediate structure will be computed using solutions of the parallel structure as initial estimates. These intermediate solutions can then be used as initial estimates to compute the solutions of the desired structure.

Example of Application

As an example, we determine all possible conformations of a three-residue loop for connecting two segments of a polypeptide when the location of the target C^a is Δ = [8, 2, 3]$_{0'}$ angstrom and the direction of the N—C^a bond is $\boldsymbol{r}$ = [0.8, 0 – 0.6]$_{0'}$ in R′$_0$. Figure 6 shows the 2D locus of all possible loop dihedral angles determined using the new method. ψ_0 and ϕ_1 are the free changing variables and ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 are calculated via eqs. (16) and (17) for any given ψ_0 and ϕ_1 (if possible). The number of possible loop conformations for each combination of the ψ_0 and ϕ_1 angles are shown in Figure 6A. The resultant loop conformations along the $\psi_0 = -1$ path form a 1D locus (Fig. 7) and are visualized in two movies (Supporting Information Movies M1 and M2).

Table 1 provides the numerical values (in degrees) of the eight solutions computed for eqs. (16) and (17) (ideal backbone structure) when $\psi_0 = -\pi/4$ and $\phi_1 = -\pi/2$. For a nonideal backbone structure with γ_1 = 120.7° and γ_3 = 117.2°, using the ideal case dihedral angles of Table 1 results in 3% and 2% average error in matching the direction and location of the target residue. The eight accurate solutions for this nonideal backbone structure were computed using eq. (37) and are presented in Table 2.

Over What Distance Can a Three-Residue Loop Connect Two Polypeptide Segments?

Using this new method, we were able to solve the loop-closure problem for different locations and directions of the target residue and determine over what maximum distance a three-residue loop can connect two segments of a polypeptide. For this purpose, we define the location of the target C^{α} in a spherical coordinate system in terms of r, θ_r, and γ_r (Fig. 8). Both γ_r and ψ_0 are rotations about the same axis; any change in γ_r leads to an equal change in ψ_0 values of the solutions. Therefore, without loss of generality, we can set γ_r to zero. The direction of N—C^{α} can be defined in terms of two angles θ_d and γ_d (Fig. 8). γ_d is the rotation of the direction vector about the $3'_0$ axis with respect to the location vector, and θ_d is the angle between the direction vector and the $3'_0$ axis. Figure 9 shows possible locations of the target C^{α} for three-residue loops. For any possible location of target C^{α}, only a fraction of target N—C^{α} directions lead to a successful loop closure. These fractions are shown by different colors in the figure; if C^{α} is located about 7–10Å from the carbonyl carbon of the preloop residue at an angle between 0 and $\pi/2$, loop closure is possible for about 50% of the target directions. The possible directions are different for different locations as shown in Figure 10. The location associated with each panel of Figure 10 is marked by the same letter in Figure 9.

Comparison of the Computation Time with the Polynomial Method

The computational efficiency of this method was compared with the widely accepted polynomial method developed by Coutsias et al.,[4] based on the initial formulation of Wedemeyer and Scheraga.[2] In the polynomial method,[4] the relative positions of the two segments of a polypeptide are characterized by three angles: α_1 (or alternatively α_2), ξ_3, and η_3, and the loop conformation can be determined by solving the constraints for three variable angles: τ_1, τ_2, and τ_3. These parameter sets are different than the kinematic approach developed in this paper. To have a fair comparison, we generated about 3500 different loop-closure problems for each method by choosing different combinations of the α_1, ξ_3, and η_3 in the polynomial method, and different combinations of r, θ_r, θ_d, and γ_d in the kinematic method presented here. Using polynomial method, we solved the generalized eigenvalue formulation of the problem using the efficient built-in function of the MATLAB software (based on QZ factorization). The C.P.U. time was measured (10 times) for each method and divided by the number of loop-closure problems solved, to find the average computing time per loop-closure problem. Table 3 compares the mean and standard deviation of the computing time per loop-closure problem for the kinematic method of this article and the polynomial method. The kinematic method of this article, on average, is about 15 times faster than the polynomial method.

Multiplicity of solutions is taken into account in calculating the average computing time per loop closure problem, which considers all possible solutions (up to 16). In the supplementary MATLAB code used for these computations, all eight equation sets were solved for each loop-closure problem and the average computing time was determined for the entire loop-closure problem with multiple solutions.

Discussion

Determining all possible loops that can connect two segments of a protein secondary structure is important for analysis of protein folding and of large conformational changes of the folded protein. The energy levels among all possible connecting loops can vary by several kT (k is the Boltzmann constant and T is the absolute temperature), and consequently, the probability of the associated protein folded structure can vary by orders of magnitude. Therefore, selecting a limited number of loop conformations by trial and error or underestimating the number of possible loop conformations (up to 16 discrete conformations rather than a 2D continuous locus of conformations) may lead to significant inaccuracy in estimating the likelihood of the folded structure.

The kinematic closure technique derived here for three-residue loops is applicable to two residue loops as well. In two residue loops, the number of constraints and degrees of freedom is 5, and therefore, instead of a 2D locus, there is a finite number of conformations. For one residue loops, the number of degrees of freedom is less than the number of constraints (three degrees of freedom and five constraints); in this case, a solution exists if the five constraint equations can be reduced to three (or less) independent equations. If the number of residues in a loop is greater than three, they may be grouped into several three residue segments, and possibly an additional single residue or double residue segment, and the equations for each group may be solved independently.

The kinematic closure technique of this article (similar to many others) limits the degrees of freedom to the ϕ and ψ dihedral angles and assumes that the bond lengths and bond angles are constant and the ω dihedral angle is 180°. However, in actual protein structures the bond angles (e.g., γ_1, γ_2, and γ_3) and ω dihedral angle may vary by about ±5°, and the bond lengths may vary by about ±0.01 Å. The position of the target residue may be altered slightly by changing the angles within ±5°, but would not change significantly by changing the lengths within ±0.01 Å. Including the four additional bond angles as degrees of freedom will increase computation greatly. For example, if we consider only five possibilities for each of these angles, we need to find the ϕ and ψ dihedral angles for $5^4 = 625$ different backbone structures. Given the limited range of these angles, the resultant dihedral angles and consequently the loop conformations will be close to each other (for each solution of the loop closure). To determine the loop conformations that can connect two segments of a polypeptide (if any), we choose one of the probable backbone structures and determine the closed loop conformations for that backbone structure. Each resultant loop conformation is representative of a family of closely positioned conformations with slightly perturbed backbone structures. For the Kv7.1 ion-channel protein structure, we found that backbone structures with parallel bonds ($\gamma_1 = \gamma_3$) are a good choice, as it accounts for about 8% of all backbone structures and requires less computation than the nonparallel bond structures. Note that the probability of the most probable backbone structure is also about 8% (Fig. 5), so the parallel bonds structures (ideal) are not far from the physiological condition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract/grant sponsor: National Institutes of Health-National Heart, Lung and Blood Institute; contract/grant number: R01-HL 049054-18; contract/grant number: R01-HL 033343-26

Contract/grant sponsor: Children's Discovery Institute Fellowship; contract/grant number: CH-F-2008-121

Many thanks to colleagues in the Rudy Lab and to Dr. Niloufar Ghoreishi for help, advice, and discussions. Y.R. is the Fred Saigh Distinguished Professor at Washington University in St. Louis. Figures 1 and 8 and Supporting Information movies were produced using the UCSF Chimera package (Chimera package was supported by NIH grant P41 RR-01081).

References

1. Go N, Scheraga HA. Macromolecules. 1970; 3:178.
2. Wedemeyer WJ, Scheraga HA. J Comput Chem. 1999; 20:819.
3. Mandell DJ, Coutsias EA, Kortemme T. Nat Methods. 2009; 6:551. [PubMed: 19644455]
4. Coutsias EA, Seok C, Jacobson MP, Dill KA. J Comput Chem. 2004; 25:510. [PubMed: 14735570]
5. Dodd LR, Boone TD, Theodorou DN. Mol Phys. 1993; 78:961.
6. Manocha D, Canny JF. IEEE Trans Rob Autom. 1994; 10:648.
7. Wu MHG, Deem MW. J Chem Phys. 1999; 111:6625.
8. Dinner AR. J Comput Chem. 2000; 21:1132.
9. Milgram RJ, Liu GF, Latombe JC. J Comput Chem. 2008; 29:50. [PubMed: 17542001]
10. Felts AK, Gallicchio E, Chekmarev D, Paris KA, Friesner RA, Levy RM. J Chem Theory Comput. 2008; 4:855. [PubMed: 18787648]
11. Ginalski K. Curr Opin Struct Biol. 2006; 16:172. [PubMed: 16510277]
12. Hornak V, Simmerling C. Proteins. 2003; 51:577. [PubMed: 12784217]
13. Hu XZ, Wang HC, Ke HM, Kuhlman B. Proc Natl Acad Sci USA. 2007; 104:17668. [PubMed: 17971437]
14. Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, Friesner RA. Proteins. 2004; 55:351. [PubMed: 15048827]
15. Sellers BD, Zhu K, Zhao S, Friesner RA, Jacobson MP. Proteins. 2008; 72:959. [PubMed: 18300241]
16. Ramachandran GN, Ramakrishnan C, Sasisekharan V. J Mol Biol. 1963; 7:95. [PubMed: 13990617]
17. Silva JR, Pan H, Wu D, Nekouzadeh A, Decker KF, Cui J, Baker NA, Sept D, Rudy Y. Proc Natl Acad Sci USA. 2009; 106:11102. [PubMed: 19549851]
18. Long SB, Campbell EB, Mackinnon R. Science. 2005; 309:897. [PubMed: 16002581]

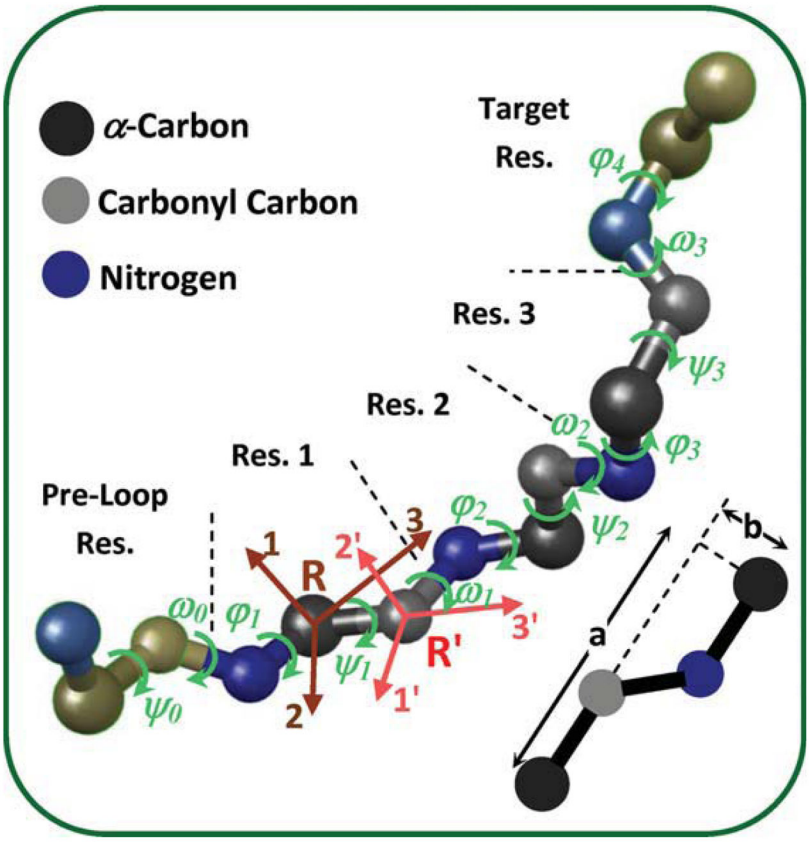

Figure 1.

Backbone atoms of loop residues (Res. 1–3), preloop residue (Res. 0), and target residue (Res. 4) in a three-residue loop-closure problem. The ϕ and ψ dihedral angles of the residues are shown with green arrows. Two local coordinate systems are defined for each residue: the 1, 2, and 3 coordinates are located on α-carbon (R coordinate system), and the 1′, 2′, and 3′ coordinates on carbonyl carbon (R′ coordinate system).

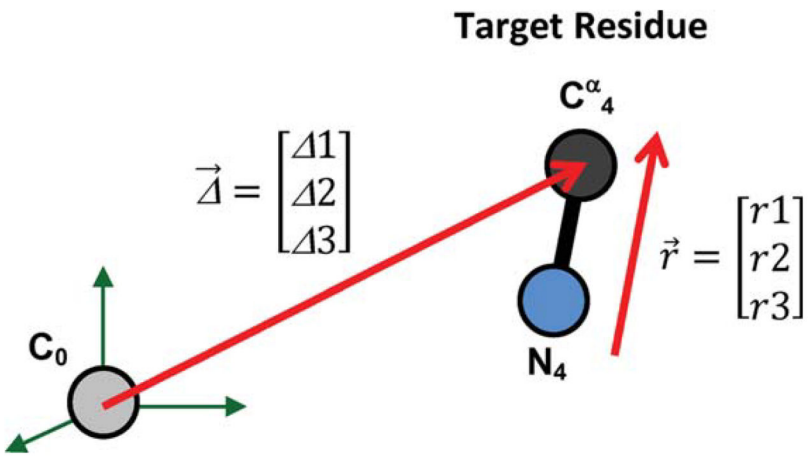

Figure 2.

Position of N_4 and C^{α}_4 (N and C^{α} of the target residue) can be quantified in terms of the location of C^{α}_4, presented by the **Δ** vector, and direction of the N_4—C^{α}_4 bond, presented by the ***r*** vector. Note that **Δ** has three independent components, but ***r*** has only two independent components as it is a direction vector and its magnitude is normalized to one. C_0 is the carbonyl carbon of the preloop residue.

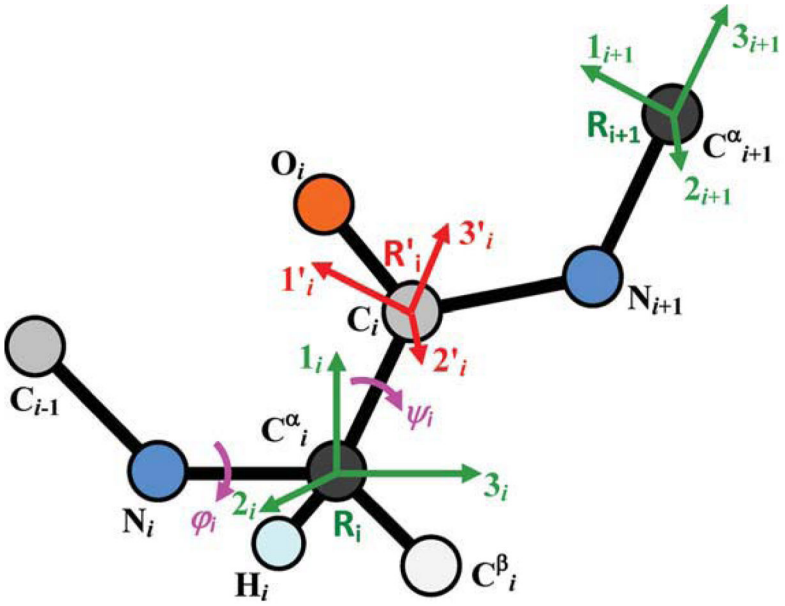

Figure 3.

Local coordinate systems R and R′ of the *i*th residue (R_i and R_i'). 3_i is along the $N_i - C_i^{\alpha}$ bond, 1_i is in the surface that passes through C_{i-1}, N_i, and C_i^{α}. Similarly, $3'_i$ is along $C_i^{\alpha} - C_i$ bond and $1'_i$ is in the surface that passes through N_i, C_i^{α}, and C_i. Variation of ϕ_i changes the position of R_i' with respect to R_i. Variation of ψ_i changes the position of R_{i+1} with respect to R_i'.

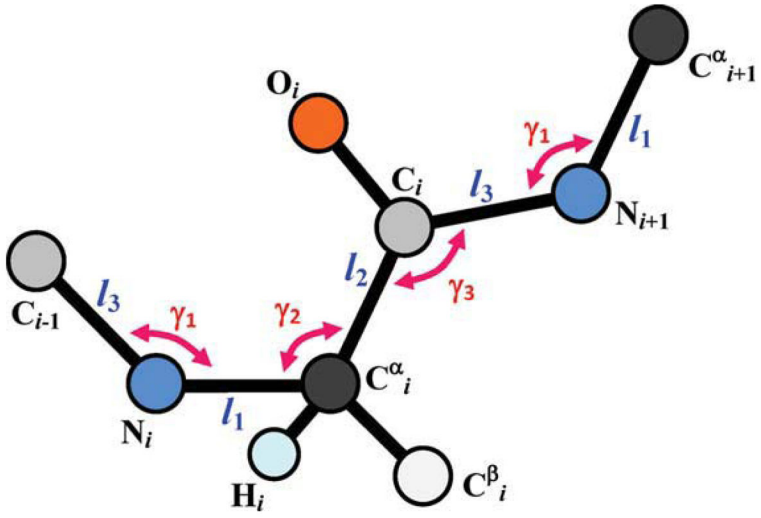

Figure 4.

Geometrical parameters of the polypeptide backbone that are required for determining the rotation matrices F and S, and the translation vectors ***d1*** and ***d2***.

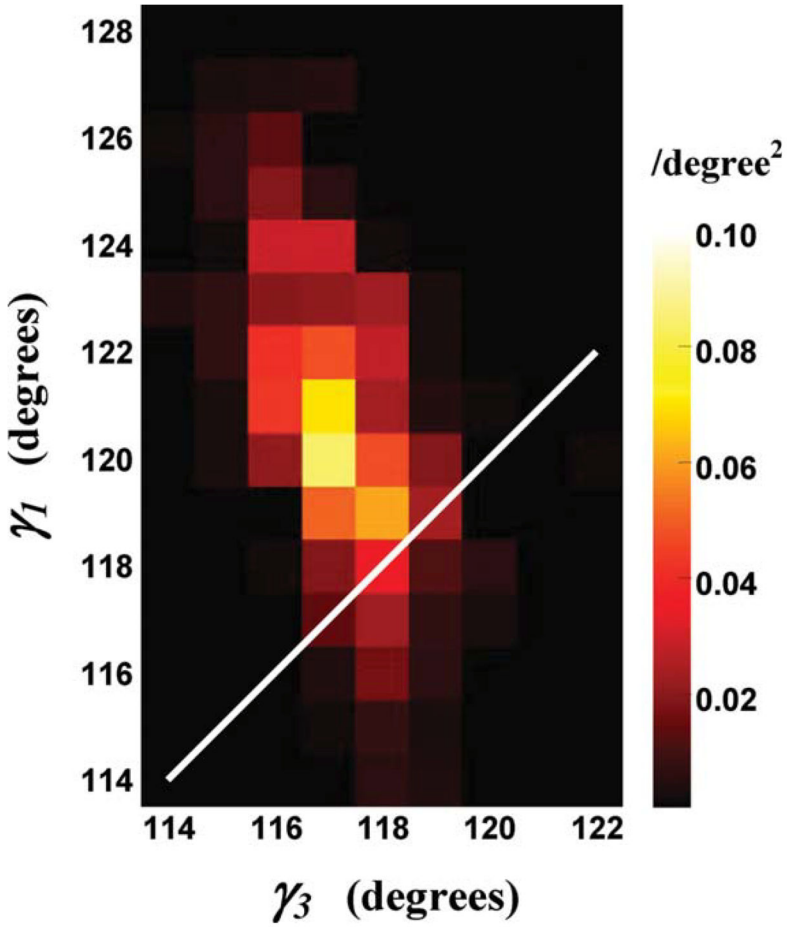

Figure 5.

Joint probability distribution of the of two consecutive γ_1 and γ_3 bond angles in the loop regions of the Kv7.1 ion-channel protein. The white line marks the backbone structures with $\gamma_1 = \gamma_3$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

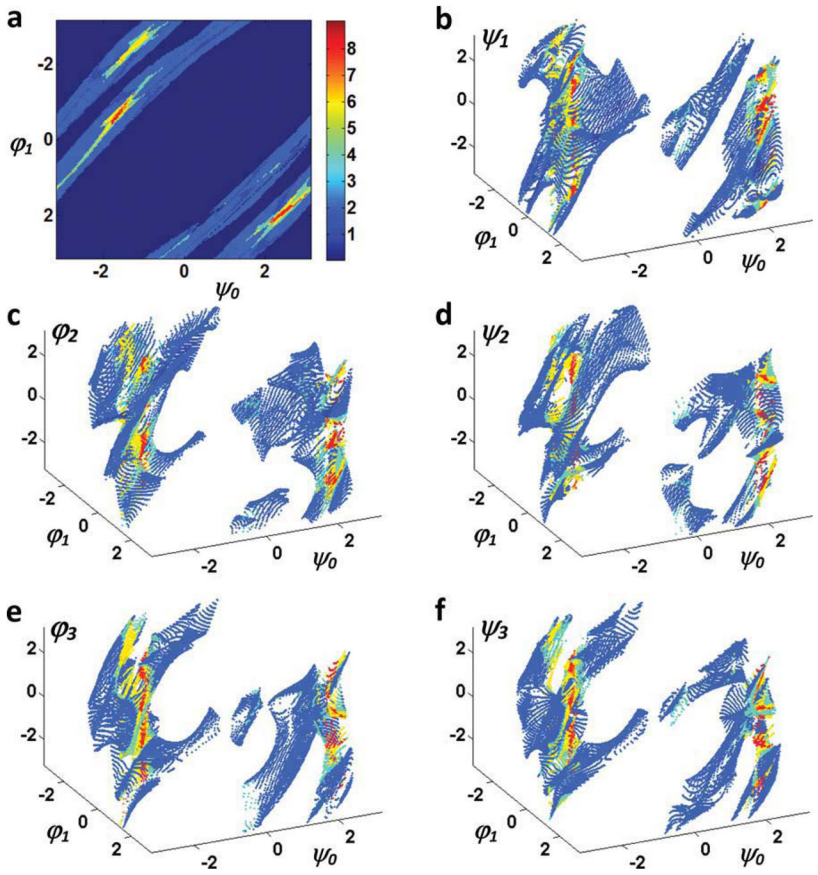

Figure 6.

2D locus of loop dihedral angles for the location vector [8,2,3] and direction vector [0.8,0.0,−0.6]. (a) The number of solutions at each combination of ψ_0 and ϕ_1 and (b–f) are ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 angles, respectively, for different choices of ψ_0 and ϕ_1. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

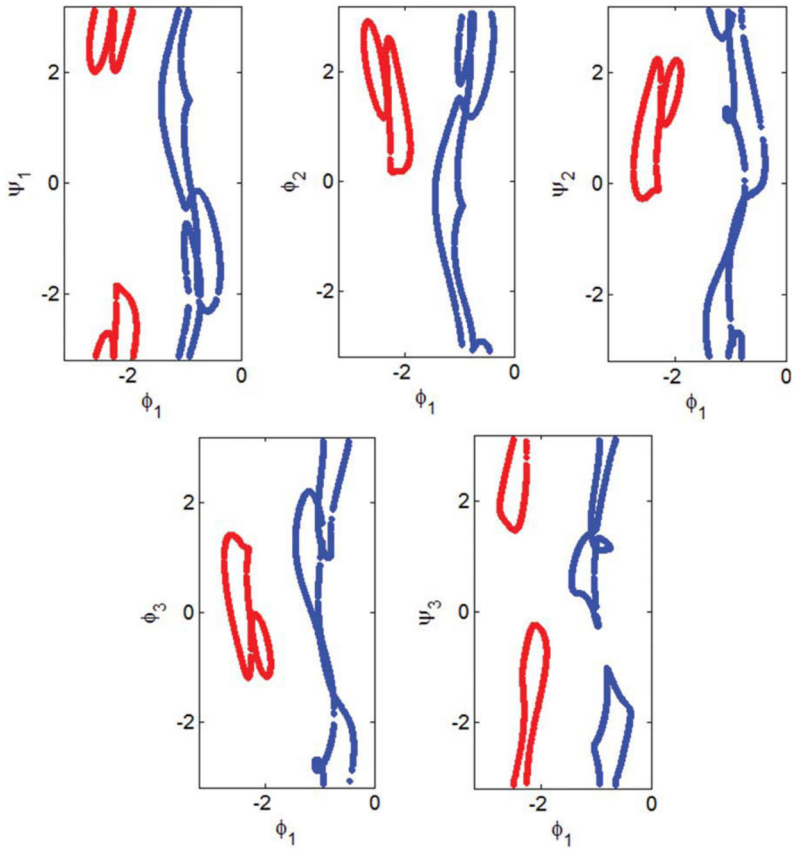

Figure 7.

1D locus of the dihedral angles ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3 in terms of the free changing variable ϕ_1 when $\psi_0 = -1$ in Figure 6. The locus consists of two closed curves colored blue and red. Note that as π and $-\pi$ are the same angles the top and bottom of each panel is connected. Supporting Information Movies M1 and M2 are associated with the red and blue curves, respectively.

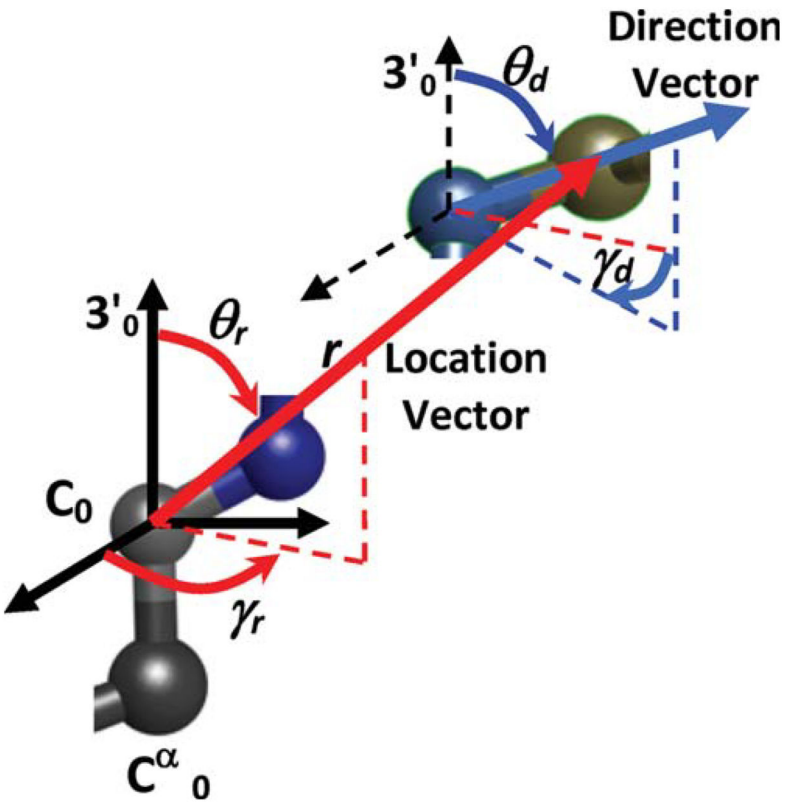

Figure 8.
Direction vector and location vector of the target residue and their characteristic parameters.

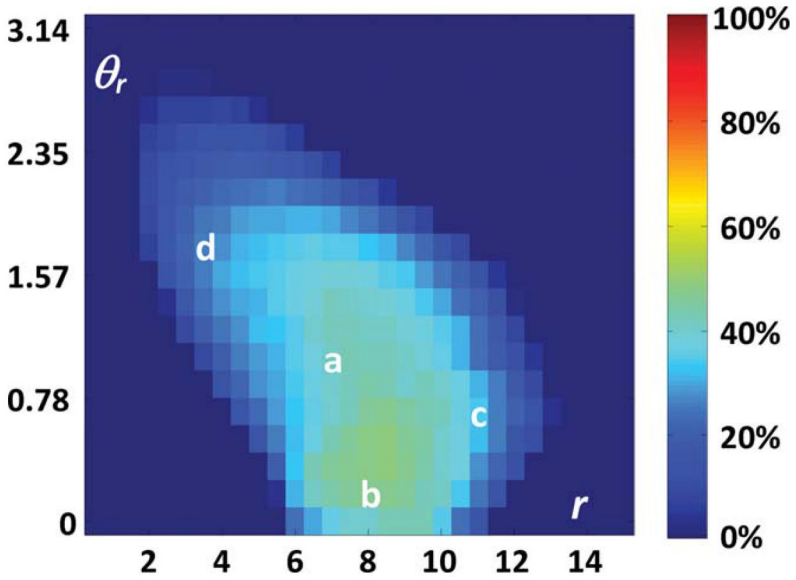

Figure 9.

Possible locations of the target C^a. The percentage of possible directions at each location is shown by its color. a–d mark indicate four arbitrary locations of C^a.

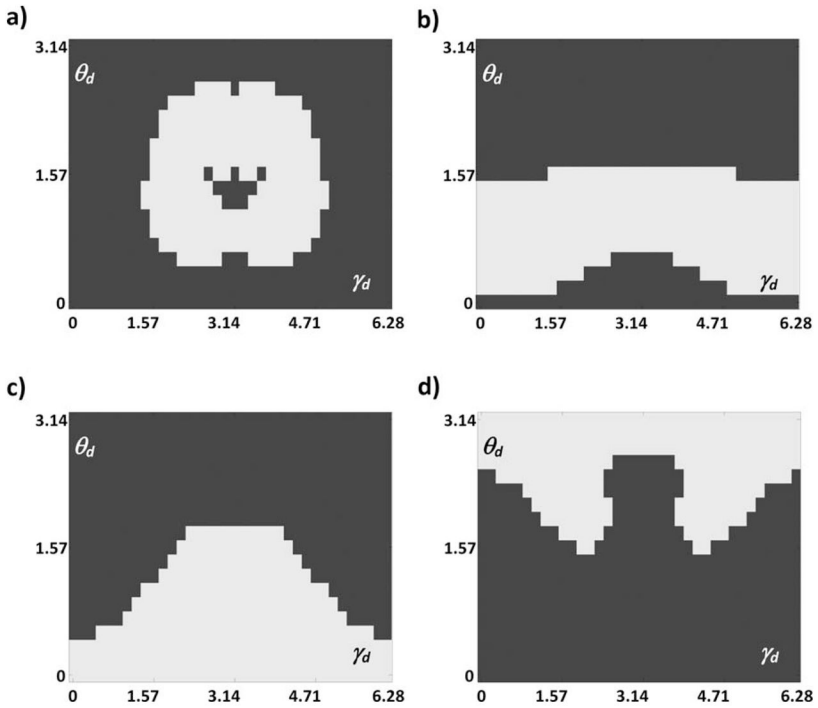

Figure 10.

Possible directions (white area) of the target N—C^{α} at four arbitrary locations a–d (shown in respective panels) of the target C^{α}. Locations correspond to those marked in Figure 9.

Table 1

Example of Application in the Ideal ($\gamma_1 = \gamma_3$) Case.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
ψ_1	122.2	−156.7	101.7	−176.2	25.6	−97.3	14.1	−68.9
ϕ_2	−54.3	−135.4	−33.8	−115.9	42.3	165.2	53.8	136.8
ψ_2	142.0	173.4	51.4	52.1	−128.2	126.0	−37.0	−71.0
ϕ_3	142.1	110.7	−127.2	−127.9	52.3	50.1	−38.8	−4.8
ψ_3	−124.9	−43.7	145.1	63.5	−136.8	−12.2	125.4	42.3

The eight sets of dihedral angles for the loop-closure when $\psi_0 = -\pi/4$, $\phi_1 = -\pi/2$, and $\gamma_1 = \gamma_3 = 120°$. The target residue is located at $\Delta = [8,2,3]_{0'}$ along direction $\boldsymbol{r} = [0.8,0, -0.6]_{0'}$.

Table 2

Example of Application in the Nonideal ($\gamma_1 \neq \gamma_3$) Case.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
ψ_1	128.9	−156.0	113.9	−173.6	25.4	−83.5	22.3	−69.7
ϕ_2	−61.4	−140.3	−45.1	−122.6	47.6	150.7	50.7	137.9
ψ_2	122.9	154.0	60.4	72.9	−103.7	−116.3	−54.4	−90.5
ϕ_3	160.2	130.8	−140.5	−152.5	29.1	42.2	−22.7	15.4
ψ_3	−129.7	−29.7	155.2	54.1	−150.4	5.8	145.2	33.3

The eight sets of dihedral angles for the loop-closure when $\psi 0 = -\pi/4$, $\phi 1 = -\pi/2$, $\gamma 1 = 120.7°$ and $\gamma 3 = 117.2°$. The target residue is located at $\Delta = [8,2,3]0'$ along direction $\boldsymbol{r} = [0.8,0, -0.6]0'$.

Table 3

Comparison of Computation Time per Loop Between the Kinematic Method of This Article and the Polynomial Method.

C.P.U. time per loop	Mean	Standard deviation
New kinematic (this article)	4.7×10^{-5} s	4.4×10^{-6} s
Polynomial (ref. 4)	7.2×10^{-4} s	6.7×10^{-6} s