JCTC Journal of Chemical Theory and Computation

Article

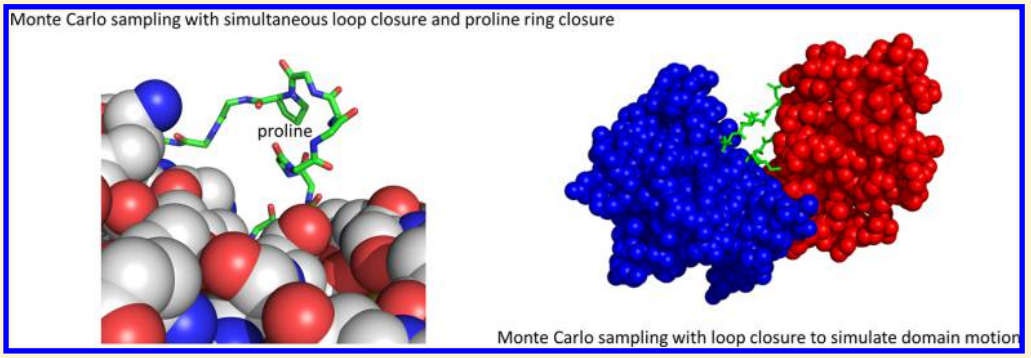
pubs.acs.org/JCTC

Monte Carlo Sampling with Linear Inverse Kinematics for Simulation of Protein Flexible Regions

Steven Hayward*,† and Akio Kitao*,‡

†D'Arcy Thompson Centre for Computational Biology, School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K.

‡Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo 113-0032, Japan



ABSTRACT: A Monte Carlo linear inverse-kinematics method for the simulation of protein chains with fixed ends is introduced. It includes backbone bond-angle bending and simultaneous loop and ring closure to allow full proline ring flexibility. An obstacle to linear null-space methods is the eventual drift of the end group. Maintenance of the end group at its initial position by occasional reset is performed in a way that is consistent with the overall methodology and minimally disruptive to the current conformation. The implementation permitted multiple rigid regions within the chain, enabling the simulation of domain movements where domains are rigid bodies connected by flexible interdomain regions. The method was tested on polyalanine, polyglycine, loop 6 of triosephosphate isomerase, and glutamine binding protein. Simulations of glutamine binding protein, where only 11 of the 226 residues at the interdomain bending regions were flexible, accurately reproduced the experimentally determined domain movement.

INTRODUCTION

Many proteins comprise flexible regions that allow, for the purpose of function, the movement of different parts of the protein. An example are the regions situated between two domains in an enzyme to allow a hinge bending movement for capture of the substrate and release of the product. They may also exist as loops on the protein surface where flexibility can also be engaged for function.[1] Decomposition of a protein into flexible and rigid regions would be advantageous not only for the purpose of efficient computer simulation but also for the understanding of how a protein's structure relates to its function via dynamical behavior. For example, does functional loop movement require conformational change in the rest of the protein, or can the rest of the protein be conceptually regarded as rigid? Similarly, for proteins undergoing hinge bending, is it necessary to consider conformational change within the domains, or can the domains be conceptually viewed as being rigid? Simulation allows us to test these concepts, and some methods have already been developed for this purpose.[2] In the case of protein loops, inverse-kinematics methods can be used to simulate conformational change within the loop without affecting the rest of the structure, i.e., the rest of the protein can remain rigid. Specifically, inverse kinematics allows one to change internal degrees of freedom of the loop, most commonly the backbone torsion angles, in a way that does not change the relative position and orientation of its end groups. This is often referred to as loop closure. Loop closure methods are often used for the purpose of modeling loops in protein comparative modeling[3] when there is an insertion in the target sequence relative to the template sequence. These same methods can be used to model loop motions. For a comprehensive review of these methods within both contexts, see the review by Shehu and Kavraki.[4] There are two main types of inverse-kinematics techniques: nonlinear and linear. The nonlinear method, first developed by Go and Scheraga,[5] finds all possible bridging loop conformations. For a loop of N torsions, $N - 6$ of these, the controls, would be varied, whereas the remaining 6, the compensators, would be determined by solving a set of equations numerically. Researchers in the area of robotics have shown that the problem amounts to solving a polynomial of degree 16,[6] and most applications in the area of protein research now use this approach. A polynomial of degree 16 implies that for a loop of 6 degrees of freedom there are at

Received: March 4, 2015
Published: June 25, 2015

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

most 16 different noninterconvertible conformations of the loop. Linear methods are simpler and require the determination of the null space or inverse of a Jacobian. They are applied iteratively with small incremental changes in internal variables of the loop.

Inverse-kinematics methods have been used within a Monte Carlo scheme for sampling conformations not only for proteins[7] but also within polymers in general.[8] They employ the nonlinear method and all, in one way or another, select a set of variables as the controls (if there is only one control, then it is referred to as the driver), with the remaining 6 as the compensators. These methods therefore vary at least 7 torsion angles within a window, although, depending on the method, the window may span more torsions if some are fixed. While linear inverse-kinematics methods have been used in other applications such as X-ray refinement,[9] sampling from conformations generated by nonlinear methods,[10] and for the purpose of understanding the general effect of the end constraint on loop conformational sampling,[11] there are no reports of a pure linear inverse kinematics being used within a Monte Carlo scheme in the same way that the nonlinear method has been. However, there are distinct advantages in the linear approach. One is that it is relatively simple to implement, and another is that all variables are on the same footing, i.e., there is no arbitrary division of the variables into controls and compensators. An expected disadvantage would be that they do not sample as extensively as the nonlinear methods. However, there is no reason to believe that linear methods would not yield full sampling of all possible conformations under permanently maintained fixed end constraints. Indeed, it has already been shown using Ramachandran plots that, even within a rather idealized model where bond lengths, bond angles, and ω torsions are held rigid and nonbonded interactions are turned off, there is a wide sampling of conformations even for loops of just 5 residues (two of which are brace residues).[11] As shown here, a further advantage of the linear approach is that there is no need to modify the acceptance criteria using Jacobians, as detailed balance is satisfied with a symmetric proposal matrix.

Variation of bond angles is often not implemented in many loop closure applications, although Bruccoleri and Karplus[12] extended Go and Scheraga's initial work to include it, and Coutsias et al. showed that its inclusion is important for bridging short gaps in proteins.[7a] Here, we also include bond angle variation along the backbone. Furthermore, we include the possibility of using a fully flexible proline model. Rotation about the proline ϕ angle is restricted by its ring. High-resolution structures give an average value for the proline ϕ of $-65.4°$ with a standard deviation of $11.2°$.[13] This means that variations in ϕ of over $\pm 20°$ from the average occur in about 5% of prolines. It is therefore a rather crude approximation to fix the proline ϕ angle, although, presumably, in many in loop modeling applications this is usually done when there is no description of how variation in the proline ϕ angle is handled.

THEORETICAL BASIS

Here, we show how to write a linear approximation for the displacement and rotation of a coordinate system at the end of a segment relative to coordinate system at the beginning of the segment due to bond length, bond angle, and torsion angle variations. Our vector analysis approach makes it particularly easy to write the equations for end group displacement in terms of internal variable changes.

End Group Displacement Due to Torsion Angle Variation. In the linear approximation, the rotation of the coordinate system on the end atom[11] due to torsion angle changes along a chain comprising N bonds is given by

$$\boldsymbol{\delta\varphi}_\tau = \sum_{i=1}^{N} \delta\tau_i \mathbf{n}_{\tau_i} \tag{1}$$

The displacement of the end atom is given as

$$\boldsymbol{\delta d}_\tau = \sum_{i=1}^{N} \delta\tau_i \mathbf{n}_{\tau_i} \times \left(\sum_{j=i+1}^{N} l_j \mathbf{n}_{\tau_j} \right) \tag{2}$$

Where $\delta\tau_i$ is the change in the ith torsion, l_j is the length of the jth bond, and $\mathbf{n}_{\tau_j}$ is the unit vector along the jth bond, a torsion axis direction.

The $\mathbf{n}_{\tau j}$ are dependent on the structure of the segment.

End Group Displacement Due to Bond Angle Variation. Similar expressions exist for the bond angle variations. For a set of changes in angle along a chain comprising N bonds, then

$$\boldsymbol{\delta\varphi}_\theta = \sum_{i=1}^{N} \delta\theta_i \mathbf{n}_{\theta_i} \tag{3}$$

where $\delta\theta_i$ is the change in the ith bond angle and $\mathbf{n}_{\theta_i}$ is the unit vector for rotation about a bond angle axis that is perpendicular to the plane defined by three atoms forming the bond angle.

For the displacement of the Nth atom

$$\boldsymbol{\delta d}_\theta = \sum_{i=1}^{N} \delta\theta_i \mathbf{n}_{\theta_i} \times \left(\sum_{j=i+1}^{N} l_j \mathbf{n}_{\tau_j} \right) \tag{4}$$

End Group Displacement Due to Bond Length Variation. Bond length variation would not affect the orientation of the end group; it would affect only its location. Its displacement due to bond length variation is

$$\boldsymbol{\delta d}_l = \sum_{i=l}^{N} \delta l_i \mathbf{n}_{\tau_i} \tag{5}$$

Matrix Equations for End Group Displacement. In order to use these equations, we need to be able to transform between coordinate systems. A general transformation between two coordinate systems 1 and 2 is given by $\mathbf{X}_1 = \mathbf{A}\mathbf{X}_2 + \mathbf{B}$, where $\mathbf{A}$ is a 3×3 rotation matrix, $\mathbf{B}$ is a 3×1 displacement vector, and $\mathbf{X}_1$ and $\mathbf{X}_2$ are 3×1 position vectors of an object from their respective coordinate system. There is a one-to-one correspondence between internal variables and coordinate system transformations that is conceptually useful. If the initial coordinate system on a three-atom system is defined according to Eyring,[14] then variable associated coordinate systems can be defined as shown in Figure 1. In Figure 1A, the variable concerned is the bond length, variable l. Application of the general transformation formula gives $\mathbf{X}_1 = \mathbf{X}_2 + \mathbf{B}$, where $\mathbf{B} = (l\ 0\ 0)^t$ and t denotes the transpose. In Figure 1B, the variable concerned is the torsion, and $\mathbf{X}_2 = \mathbf{A}_\tau \mathbf{X}_3 + \mathbf{B}$, where $\mathbf{B} = \mathbf{0}$ and $\mathbf{A}_\tau$ is given by

$$\mathbf{A}_\tau = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\tau) & -\sin(\tau) \\ 0 & \sin(\tau) & \cos(\tau) \end{pmatrix} \tag{6}$$

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

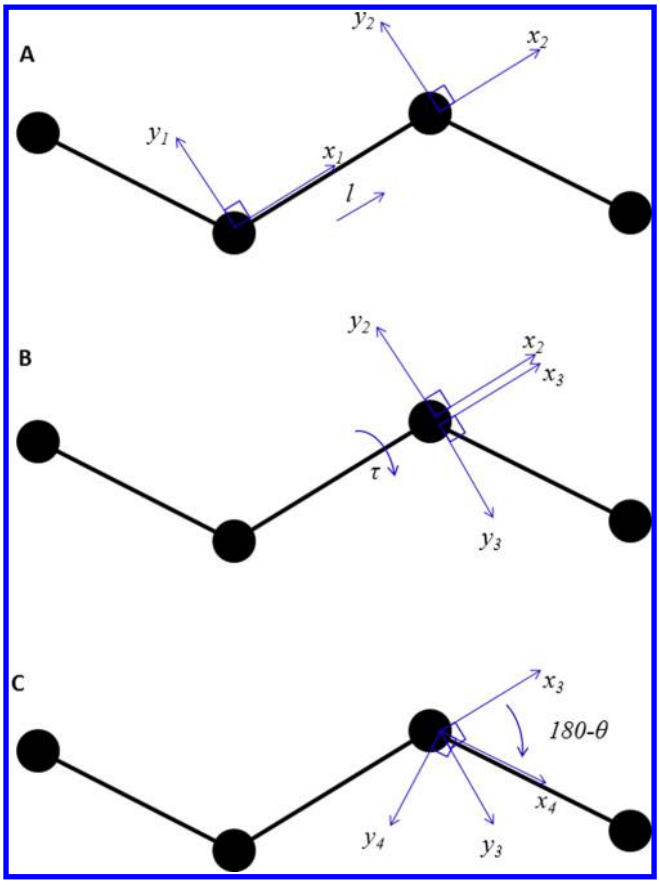


Figure 1. Illustration of how each variable has a coordinate system associated with it. (A) Starting from coordinate system 1, coordinate system 2 is associated with the bond length. (B) Coordinate system 3 is associated with the torsion angle. (C) Coordinate system 4 is associated with the bond angle.

where τ is the torsion angle. This involves a rotation about the x axis. Likewise for the bond angle θ in Figure 1C, the transformation is given by $\mathbf{X}_3 = \mathbf{A}_\theta \mathbf{X}_4 + \mathbf{B}$, where $\mathbf{B} = \mathbf{0}$ and $\mathbf{A}_\theta$ is given by

$$\mathbf{A}_\theta = \begin{pmatrix} -\cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & -\cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

This involves a rotation about the z axis by an angle of $180 - \theta$. Recursive application of the transformation equation allows one to convert between internal coordinates and Cartesian coordinates. These transformations allow us to write $\mathbf{n}_{\tau_i}$ and $\mathbf{n}_{\theta_i}$ in terms of the coordinate system at the beginning of the loop. Further details can be found elsewhere.[11] For the main-chain part of the loop, eqs 1−5 can be written in matrix form as

$$\begin{pmatrix} \delta\boldsymbol{\varphi} \\ \delta\mathbf{d} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_R & \mathbf{\Lambda}_R & \mathbf{0} \\ \mathbf{Y}_D & \mathbf{\Lambda}_D & \mathbf{L}_D \end{pmatrix} \begin{pmatrix} \delta\boldsymbol{\tau} \\ \delta\boldsymbol{\theta} \\ \delta\mathbf{l} \end{pmatrix} = (\mathbf{Y} \quad \mathbf{\Lambda} \quad \mathbf{L}) \begin{pmatrix} \delta\boldsymbol{\tau} \\ \delta\boldsymbol{\theta} \\ \delta\mathbf{l} \end{pmatrix} \tag{8}$$

where $\delta\boldsymbol{\varphi} = \delta\boldsymbol{\varphi}_\tau + \delta\boldsymbol{\varphi}_\theta$, $\delta\mathbf{d} = \delta\mathbf{d}_\tau + \delta\mathbf{d}_\theta + \delta\mathbf{d}_l$; $\mathbf{Y} = (\mathbf{Y}_R\ \mathbf{Y}_D)^t$, $\mathbf{\Lambda} = (\mathbf{\Lambda}_R\ \mathbf{\Lambda}_D)^t$, and $\mathbf{L} = (\mathbf{0}\ \mathbf{L}_D)^t$; $\mathbf{Y}_R$ and $\mathbf{\Lambda}_R$ are $3 \times N$ matrices corresponding to rotation due to torsion angle variation and bond angle variation, respectively; $\mathbf{Y}_D$, $\mathbf{\Lambda}_D$, and $\mathbf{L}_D$ are $3 \times N$ matrices corresponding to displacement due to torsion angle, bond angle, and bond length variation, respectively; and $\delta\boldsymbol{\tau}$, $\delta\boldsymbol{\theta}$, and $\delta\mathbf{l}$ are $N \times 1$ column matrices the ith element being $\delta\tau_i$, $\delta\theta_i$, and δl_i, respectively. $\mathbf{Y}$, $\mathbf{\Lambda}$, and $\mathbf{L}$ are dependent on the structure of the N-bond segment being functions of $\boldsymbol{\tau}$, $\boldsymbol{\theta}$, and $\mathbf{l}$.

End Group Displacement with Constrained Bond Lengths. A common assumption is that the stiff degrees of freedom such as bond lengths can be constrained. Considering torsions and bond angles alone, eq 8 becomes

$$\delta\mathbf{s} = \begin{pmatrix} \delta\boldsymbol{\varphi} \\ \delta\mathbf{d} \end{pmatrix} = (\mathbf{Y}(\boldsymbol{\tau}, \boldsymbol{\theta}) \quad \mathbf{\Lambda}(\boldsymbol{\tau}, \boldsymbol{\theta})) \begin{pmatrix} \delta\boldsymbol{\tau} \\ \delta\boldsymbol{\theta} \end{pmatrix} = \mathbf{\Omega}(\boldsymbol{\nu})\delta\boldsymbol{\nu} \tag{9}$$

where $\delta\boldsymbol{\nu}$ stands for the combined set of torsion and bond angle variable displacements, $\delta\mathbf{s}$ is the combined rotational and translational displacement vector, and $\mathbf{\Omega}(\boldsymbol{\nu})$ is the Jacobian, a $6 \times 2N$ matrix. In our application, the C^α's at the beginning and end of the loop are fixed. For a loop of N_{res} residues, $N = 3(N_{res} - 1)$.

Side Chain Internal Variables. For side chains, no end constraints are applied. In order to convert from internal to Cartesian coordinates, one needs to be able to track back through the structure to the system coordinates. For side chains, this may involve one or more branches and therefore coordinate transformations at branches need to be considered. For example, at the main-chain to side-chain branch at a C^α atom, there are three coordinate transformations from the coordinate system with the x axis along the C^α−C bond to the one with the x axis along the C^α−C^β bond: one for the N−C^α−C angle, one for the branch torsion (usually called an improper torsion) N−C^α−C^β−C (our definition: for atoms i−j−k−l, with atom j at the branch, it is the angle between j−k and j−l projected onto the plane perpendicular to the i−j bond, i.e., in a Newman projection), and finally one for the N−C^α−C^β angle. In order to deal with this extra complexity, each amino acid residue has its own topology file comprising a list of atom numbers from which the preceding coordinate system of any coordinate system in the residue can be derived. Combining this topological information allows us to convert from internal coordinates to Cartesian coordinates for any given loop. In order to transform from a molecule-based coordinate system to a system-based coordinate system, a virtual atom was used, as described elsewhere.[15]

In our implementation, the side-chain branch torsions and all side-chain bond angles are held fixed for all residues, apart from proline (see below). Groups of atoms at the terminal ends of side chains within which single bonds are absent or have no effect are treated as being rigid, e.g., the alanine methyl group and the histidine imidazole group. They remain fixed to the local coordinate system of the first atom in the group, i.e., in the case of alanine, the three hydrogens are in the coordinate system on the C^β atom. All hydrogen atoms are treated as being rigidly attached, for example, the amide hydrogen is fixed to the N−C^α−C coordinate system. Groups of rigidly attached atoms and the atom to which they are attached are stored in each residue's topology file.

Simultaneous Proline Ring and Loop Closure. Ring closure can be written in the same way as loop closure. In our implementation, the initial coordinate system is tied to the main-chain N of the proline with the x axis along the N−C^α bond, y axis in the C−N−C^α plane, and z axis perpendicular to this plane. The final coordinate system is determined by carrying out transformations around the ring until one reaches this initial coordinate system. These transformations are associated not only with the usual bond length, bond angle,

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

and proper torsion variables but also the branch torsions, N−C^{α}−C^{β}−C and C^{δ}−N−C−C^{α}. The equation for the movement of the end group has the same form as eq 9, as we also allow bond angle variation (this is necessary for ring flexibility), but it also includes these branch torsions. A key problem to solve is the simultaneous solution of the loop closure and proline ring closure equations, as they are not independent through a coupling of rotation about the N−C^{α} bond. Within our development, the solution to this is simple and can be easily extended to cases where there is more than one proline in the loop. Equation 10 shows the structure of the Jacobian for a single proline in the loop

$$\begin{pmatrix} \delta \mathbf{s}_l \\ \delta \mathbf{s}_p \end{pmatrix} = \begin{pmatrix} \mathbf{\Omega'}_l & \mathbf{\Omega}_{l,\phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{p,\phi} & \mathbf{\Omega'}_p \end{pmatrix} \begin{pmatrix} \delta \boldsymbol{\nu'}_l \\ \delta\phi \\ \delta \boldsymbol{\nu'}_p \end{pmatrix} \tag{10}$$

where $\delta\boldsymbol{\nu'}_l$ and $\mathbf{\Omega'}_l$ are the displacement vector of loop variables and Jacobian, respectively, excluding terms for proline ϕ angle variation. The terms $\delta\boldsymbol{\nu'}_p$ and $\mathbf{\Omega'}_p$ are the displacement vector of proline ring variables and the Jacobian, respectively, which exclude terms for proline ϕ angle variation. The terms $\mathbf{\Omega}_{l,\phi}$ and $\mathbf{\Omega}_{p,\phi}$ are the Jacobians for the loop and proline ring, respectively, which contain terms for variation in the proline ϕ angle only. The terms $\delta\mathbf{s}_l$ and $\delta\mathbf{s}_p$ give the rotational and translational displacements of the end groups for the loop and proline ring, respectively. There are 12 variables involved in proline ring closure: the 5 proper torsions around the ring including the ϕ angle, the 5 bond angle variables around the ring, and the two branch torsions that couple the ϕ angle rotation to the ring. It should be clear how to extend this method to deal with multiple prolines. Our approach integrates loop closure and proline ring closure equations, which is expected to be more efficient than approaches where loop closure and proline ring closure equations are solved separately. If we write the equivalent of eq 10 for a loop with N_{pro} prolines in shorthand as $\delta\mathbf{s} = \mathbf{J}(\boldsymbol{\nu})\delta\boldsymbol{\nu}$, then simultaneous loop and multiple proline ring closure will be given by solving the following

$$\mathbf{J}(\boldsymbol{\nu})\delta\boldsymbol{\nu}_0 = \mathbf{0} \tag{11}$$

where $\delta\boldsymbol{\nu}_0 = [\delta\boldsymbol{\nu}^1{}_0\ \delta\boldsymbol{\nu}^2{}_0\ ...]$, with $\delta\boldsymbol{\nu}^i{}_0$ being $N_{var} \times 1$ column vectors that span the null space of the Jacobian, $\mathbf{J}(\boldsymbol{\nu})$. For a loop of N_{res} residues braced at C^{α} of residue 1 and C^{α} of residue N_{res}, N_{var} is given by $N_{var} = 6(N_{res} - 1) + 11N_{pro}$. The maximum rank of $\mathbf{J}(\boldsymbol{\nu})$ is $6(N_{pro} + 1)$, so the minimum dimension of the null space, N_{null}, is $6(N_{res} - 2) + 5N_{pro}$. In most cases, $N_{null} = 6(N_{res} - 2) + 5N_{pro}$.

In order to compare a proline with a flexible ring to one with a rigid ring, a rigid-ring proline model was also created where the ϕ angle is fixed and side-chain atoms are rigidly attached to the coordinate system defined by C−N−C^{α}.

Rigid Segments within the Loop. Consider a loop within which a there is a rigid segment. Each variable has its own column in the Jacobian, **J**, and because variables within the rigid segment do not change, their corresponding columns in **J** can be deleted. This can make for efficient simulation. In order to determine the movement of the rigid segment, coordinates of the rigid segment atoms are determined in the local coordinate system of an atom adjacent to the rigid segment. The trajectory of the rigid segment can then be determined from the trajectory of the local coordinate system. Our implementation allows multiple rigid segments within a loop.

Correction for End Drift. Sampling of protein loop conformations involves solving eq 11 at each Monte Carlo step (see below). Due to the linear approximation, after a number of steps there is an eventual drift of the end group, which needs to be corrected. At each step, the mean-square displacement (MSD_{end}) between atoms on an end group, be it on the loop or a proline, in their current and starting positions is calculated. If the MSD_{end} of *any* end group exceeds a threshold value (MSD_{end_thresh}), then *all* end groups are brought back from their current to their original positions. In order to perform the correction, the rotational and translational displacements of the end groups from their original positions to their current positions are calculated for the loop and all prolines and combined in the displacement vector, $\delta\mathbf{s}_{drift}$. The correction can then be found as

$$\delta\boldsymbol{\nu}_{drift} = -\mathbf{J}(\boldsymbol{\nu})^{-1}\delta\mathbf{s}_{drift} \tag{12}$$

where $\delta\boldsymbol{\nu}_{drift}$ denotes a particular solution. However, any vector $\delta\boldsymbol{\nu}_{drift} + \sum c_i\ \delta\boldsymbol{\nu}^i{}_0$ is also a solution to eq 12, where c_i is a constant. The approach taken is the one that minimizes the effect on the current conformation by selecting the solution that is orthogonal to the current null space. This is achieved when $c_i = -\delta\boldsymbol{\nu}_{drift}\cdot\delta\boldsymbol{\nu}^i{}_0$ for all i. Adding this correction to $\boldsymbol{\nu}$ will reset all end groups to their original positions.

Monte Carlo on Constraint Surface. Equation 11 is solved at each step of Monte Carlo sampling. Moves are made in the space of all backbone torsions (ϕ,ψ,ω), all backbone and proline bond angles, those side-chain torsions not assigned to be rigid, and branch torsions of prolines. A random direction in the null space for backbone and proline moves is taken by generating a normalized N_{null} component column vector, $\boldsymbol{\lambda}$, with components, before normalization, generated from the normal distribution with mean 0 and standard deviation 1.0. $\delta\boldsymbol{\nu}_{rand} = \delta\boldsymbol{\nu}_0\boldsymbol{\lambda}$ randomly combines the null space vectors, giving on each step a random direction in the null space. The step size, $|\delta\boldsymbol{\nu}_{rand}|$, for main-chain and proline sampling is randomly selected (on a uniform distribution) within the range $[0,\delta\nu_{max}]$. For side-chain sampling at each step, a single side-chain torsion is selected randomly for a move with a step size randomly (on a uniform distribution) within the range $[-\delta s_{max},\delta s_{max}]$. $\delta\nu_{max}$ and δs_{max} are parameters to be specified. A basic random walk algorithm would follow:

(1) Generate random direction $\delta\boldsymbol{\nu}_{rand}$, where $\mathbf{J}(\boldsymbol{\nu}_t)\delta\boldsymbol{\nu}_{rand} = \mathbf{0}$
(2) $\boldsymbol{\nu}_{t+1} = \boldsymbol{\nu}_t + \delta\boldsymbol{\nu}_{rand}$ and analogously for side-chain torsions
(3) $\mathbf{J}(\boldsymbol{\nu}_t) \rightarrow \mathbf{J}(\boldsymbol{\nu}_{t+1})$
(4) $t \rightarrow t + 1$
(5) Return to step 1

Figure 2 shows a flow diagram for the detailed Monte Carlo algorithm with end drift correction.

For the general Metropolis−Hastings procedure,[16] the probability p of acceptance of the trial conformation is given by

$$p = \min\left[1,\ \alpha(\boldsymbol{\nu}_{t+1}, \boldsymbol{\nu}_t)\frac{\exp(-\beta E(\boldsymbol{\nu}_{t+1}))}{\exp(-\beta E(\boldsymbol{\nu}_t))}\right] \tag{13}$$

where $\beta = 1/k_B T$ and k_B and T are the Boltzmann constant and temperature, respectively. $\alpha(\boldsymbol{\nu}_{t+1},\boldsymbol{\nu}_t)$ is $Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t)/Q(\boldsymbol{\nu}_t|\boldsymbol{\nu}_{t+1})$, where $Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t)$ is the proposal matrix and $(Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t)$ is the probability of making the trial transition from conformation $\boldsymbol{\nu}_t$ to $\boldsymbol{\nu}_{t+1}$). For the general Metropolis−Hastings procedure,

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

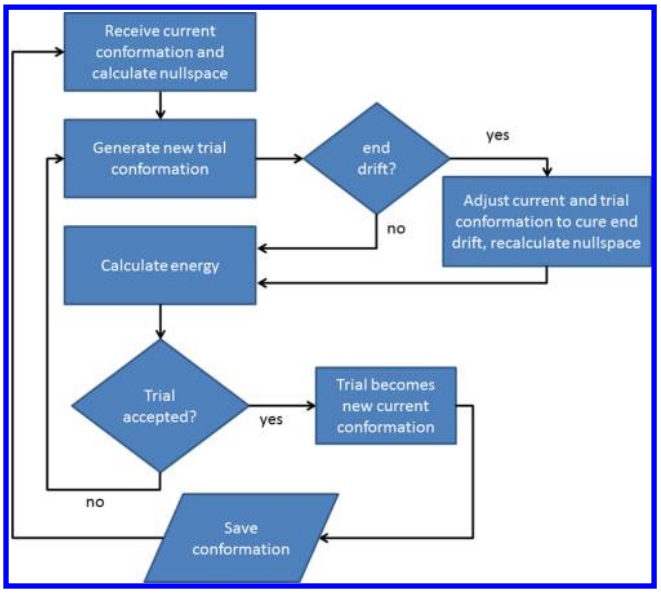


Figure 2. Flow diagram of the algorithm.

$\alpha(\boldsymbol{\nu}_{t+1},\boldsymbol{\nu}_t)$ must be known. For the Metropolis procedure,[17] the proposal matrix is symmetric ($Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t) = Q(\boldsymbol{\nu}_t|\boldsymbol{\nu}_{t+1})$) and therefore $\alpha(\boldsymbol{\nu}_{t+1},\boldsymbol{\nu}_t) = 1$.

Evaluation of Conformational Energy. To evaluate the conformational energy, the AMBER ff99SB-ILDN force field[18] was employed. The nonbonded energy was calculated without a cutoff except for glutamine binding protein (GBP), where a shifted-force cutoff was made at 10 Å.

Implementation. The software has been implemented in two parts. Conformer generation as described above was implemented in a MATLAB program, LoopGen. The null space and inverse of the Jacobian were determined by singular value decomposition (SVD). The evaluation of the energy was implemented in FORTRAN. The FORTRAN program, ParaDym,[19] is the driving program, calling LoopGen. Practically, this was achieved by compiling LoopGen in C using the MATLAB *mcc* and *mbuild* functions from MATLAB *Compiler*, which created library files for linking to ParaDym.

RESULTS

Our targets systems were polyalanine and polyglycine models, loop 6 from triosephosphate isomerase (TIM), and GBP. Before applying the software to these target systems, the symmetry of the proposal matrix, the frequency of end drift, and the acceptance ratio under the Monte Carlo scheme were considered.

Validity of Metropolis Procedure. For the purpose of determining the form of the proposal matrix, a short five residue loop was used. To use the Metropolis procedure, we need to show $Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t) = Q(\boldsymbol{\nu}_t|\boldsymbol{\nu}_{t+1})$. Although for side-chain moves alone it is clear that the proposal matrix would be symmetric, the highly curved nature of the conformational space that the end constraints create for the backbone degrees of freedom[11] means it is possible that the proposal matrix would not be symmetric, i.e., $Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t) \neq Q(\boldsymbol{\nu}_t|\boldsymbol{\nu}_{t+1})$. In order to test this, backbone moves were made with a fixed step size, namely, $\delta\nu = 5°$. Let $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_t$ and $\boldsymbol{\nu}_2 = \boldsymbol{\nu}_{t+1}$ represent two neighboring conformations after t steps of random walk and $\boldsymbol{\nu}_{1-2}$, the conformation of a random trial attempt to reach conformation 2 from conformation 1. A transition from conformation 1 to conformation 2 was considered to have occurred when $|\boldsymbol{\nu}_{1-2} - \boldsymbol{\nu}_2| \leq 5°$. If so, then the conformation was set to conformation 2; otherwise, it was reset to conformation 1. The number of transitions, n_{1-2}, from conformation 1 to 2 was counted against the number of trial attempts, n_1. The same procedure was followed for attempts to reach conformation 1 from conformation 2 to give n_2 and n_{2-1}. The quantities $p_{12} = n_{1-2}/n_1$ and $p_{21} = n_{2-1}/n_2$ provide estimates of $Q(\boldsymbol{\nu}_2|\boldsymbol{\nu}_1)$ and $Q(\boldsymbol{\nu}_1|\boldsymbol{\nu}_2)$, respectively. The quantities p_{12} and p_{21} were evaluated after t = 100, 1000, 2000, 3000, and 4000 steps of random walk. At each of these five conformations, 1 000 000 steps of trial attempts were made and the trajectory of p_{12} and p_{21} was plotted as shown in Figure 3. All show

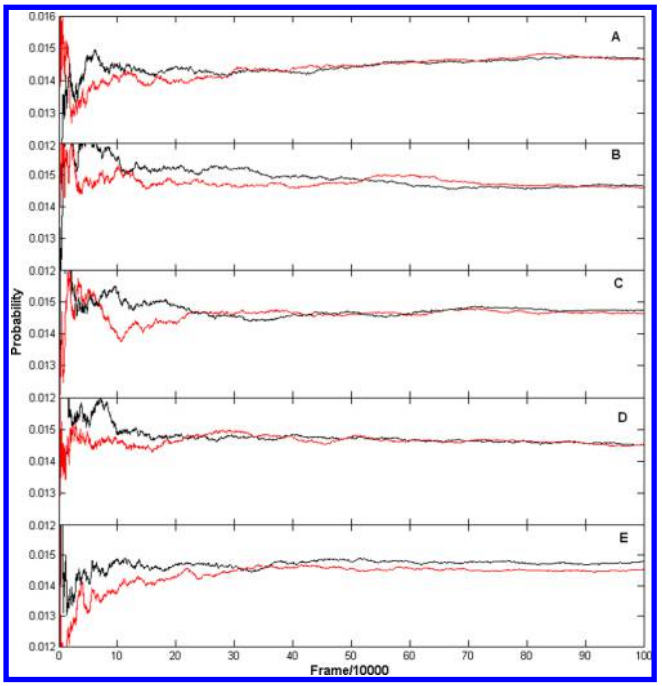


Figure 3. Trajectories of p_{12} (black lines) and p_{21} (red lines), the ratio of transitions to attempts from conformation 1 to conformation 2 and from to conformation 2 to conformation 1, respectively, after (A) 100 steps, (B) 1000 steps, (C) 2000 steps, (D) 3000 steps, and (E) 4000 steps of random walk.

convergence, and for all apart for t = 4000, $100 \times \min(p_{12},p_{21})/\max(p_{12},p_{21})$ was greater than 99%. For t = 4000, the value was 98.3%, but extending the number of trial attempts to 2 000 000 showed further convergence with a final value of 99.5%. These results indicate a convergence of p_{12} and p_{21} to the same value, implying $Q(\boldsymbol{\nu}_{t+1}|\boldsymbol{\nu}_t) = Q(\boldsymbol{\nu}_t|\boldsymbol{\nu}_{t+1})$. On the basis of this result, we use the Metropolis procedure for Monte Carlo sampling.

Drift of End Group and Reset Frequency. A further consideration is the degree of end drift and frequency with which the correction needs to be made. If MSD_{end_thresh} is too high or it is violated considerably before reset, as would happen if $\delta\nu_{max}$ were set too high, then it is possible that conformations occur that are incompatible with permanently maintained end constraints, i.e., they are on different self-motion manifolds.[20] We used TIM loop 6 from the closed conformation[21] (PDB: 1N55) for testing the software. Loop 6 comprises residues 167−179 with sequence EPVWAIGTGKVAT. The loop was fixed at E167 and T179 and it therefore contains a proline. With a MSD_{end_thresh} value of 0.01 Å^2 (MSD_{end} is calculated on a group of four virtual atoms on the end coordinate system

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

located at (0,0,0), (1,0,0), (0,1,0), and (0,0,1)), the proportion of end drift violations (for the loop or the proline) versus $\delta\nu_{max}$ is shown in Figure 4 for 1 000 000 steps of Metropolis Monte

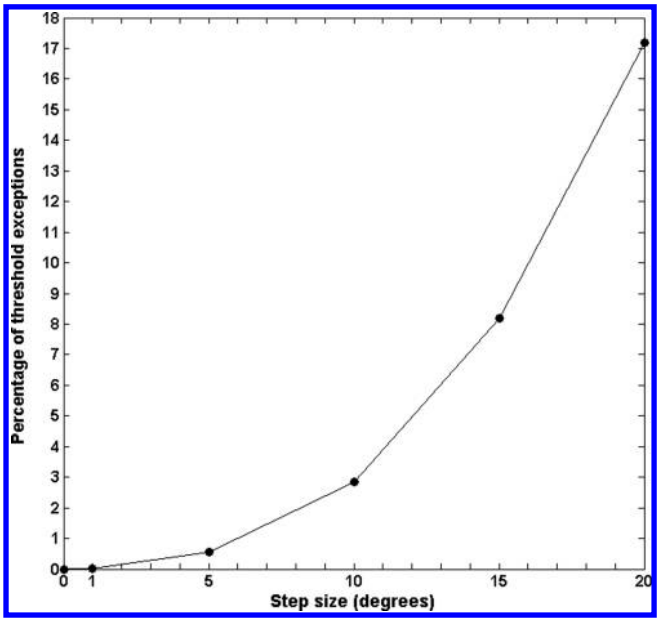


Figure 4. TIM loop percentage of steps when MSD_{end_thresh} is exceeded for either the loop or the proline (Pro168) plotted versus maximum step size, $\delta\nu_{max}$, for 1 000 000 Metropolis Monte Carlo steps at 300 K. When the threshold is exceeded, the end groups are reset to their original positions.

Carlo at 300 K. The low value for MSD_{end_thresh} ensures virtually no drift of the end groups, and for $\delta\nu_{max} \leq 10°$, reset occurs infrequently. Note that for these small step sizes we did not encounter singularities in the determination of the inverse of the Jacobian.[4]

Acceptance Ratio. Figure 5 shows the acceptance ratio for different values of $\delta\nu_{max}$ and δs_{max} determined from simulations on the TIM loop (also fixed at E167 and T179) at 300 K for 1 000 000 steps of Metropolis Monte Carlo. These acceptance ratios are commensurate with those found for other Monte Carlo methods applied to protein loop simulations, e.g., Nilmeier et al.,[7b] who also applied their loop closure method to the TIM loop, quote 14%, and Carlacci and Englander,[22] who used a simulated annealing based approach, quote 25%.

On the basis of these results, we have selected $\delta\nu_{max} = 5°$ as optimal, and all simulations reported from hereon have this value.

Polyalanine and Polyglycine. We performed 5 000 000 steps with $\delta\nu_{max} = 5°$ and $\delta s_{max} = 10°$ of Metropolis Monte Carlo at 300 K on a 14-mer alanine model and a 14-mer glycine model with residues 5 and 10 fixed. Four simulations were performed starting each model from both the α-helix and β-sheet conformations. Figure 6 shows Ramachandran plots for residue 7 in these simulations. As one can see, the distributions are consistent with what one would expect for trajectories starting from the α-helix conformation remaining in that conformation due to hydrogen bonding, and those starting from the extended β-sheet conformation exploring more widely, including, for the glycine model, regions with positive ϕ angles. Similar trajectories were observed for residues 6, 8, and 9 (residues 5 and 10 had their ϕ and ψ angles, respectively, fixed). We also checked that bond angle fluctuations were in

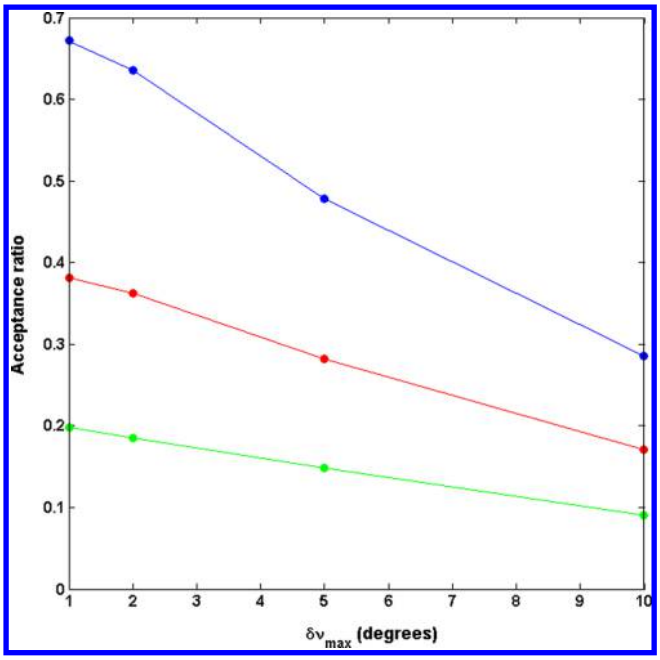


Figure 5. Acceptance ratio plotted versus $\delta\nu_{max}$ for TIM loop. Each point is evaluated after 1 000 000 Metropolis Monte Carlo steps at 300 K. Blue line is for $\delta s_{max} = 2°$, red, $\delta s_{max} = 5°$, and green, $\delta s_{max} = 10°$.

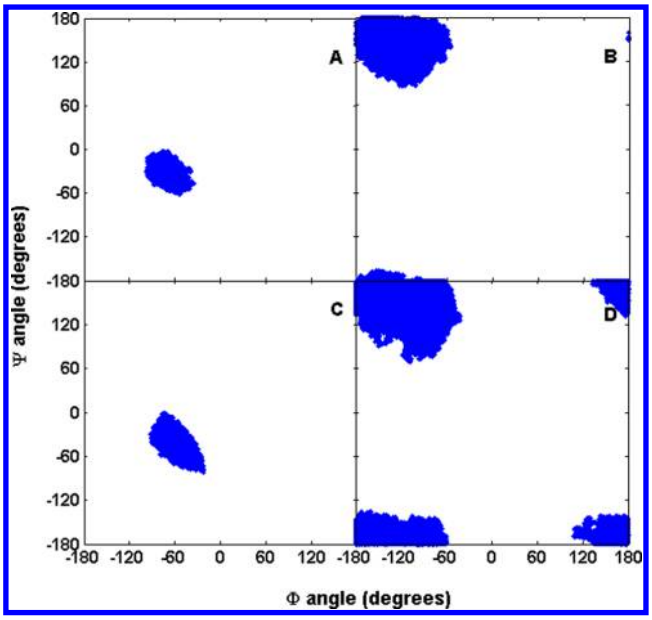


Figure 6. Ramachandran plots of trajectories of residue 7 in a 14-mer alanine model and a 14-mer glycine model with residues 5 and 10 fixed. (A) Polyalanine starting from the α-helix conformation. (B) Polyalanine starting from the β-sheet conformation. (C) Polyglycine starting from the α-helix conformation. (D) Polyglycine starting from the β-sheet conformation.

reasonable ranges. For example, for the 14-mer glycine model starting from the extended conformation, we found the following means and standard deviations for the main-chain bond angles at residue 7: C−N−C^{α} 124.0° (2.8°), N−C^{α}−C 111.1° (2.9°), and C^{α}−C−N 116.3° (2.4°). The corresponding values at residues 6, 8, and 9 were very similar to these, with the standard deviations all within 0.2° of these figures. These values correspond reasonably well with those of Engh and Huber:[23]

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

C−N−C^{α} 120.6° (1.7°), N−C^{α}−C 112.5° (2.9°), and C^{α}−C−N 116.4° (2.1°).

Loop Movement in TIM. TIM catalyzes the isomerization of its substrate, D-glyceraldehyde 3-phosphate, to dihydroxyacetone. Function involves the movement of loop 6 to cover the phosphodianion group on the substrate. In the closed structure, loop 6 covers the transition state analogue, 2-phosphoglycolate. As above, the segment selected for simulation comprised residues 167−179 with the sequence EPVWAIGTGKVAT. Glu167 acts as the catalytic base in the reaction. In our simulations, the loop is fixed at Glu167 and Thr179 situated adjacent to the N-terminal and C-terminal hinges of the loop, respectively.[1] In going from the open loop conformation, as found in subunit 1 of the PDB structure 5TIM,[24] to the closed conformation, loop 6 undergoes a 37° rotation, as determined by the DynDom program (minimum domain size set to 10 residues).[25] The hinge axis passes close to both the bending regions. Loop 6 is not the only loop that undergoes conformational change: neighboring loop 7 is also seen to change conformation.[24] The ring of Pro168 has an unusually flat conformation ($\chi_1 = -2°$) in the closed structure, which contrasts with the puckered conformations dominated by up ($\chi_1 = -25°$) and down ($\chi_1 = 25°$) forms.[26] The open structure, however, has the puckered form, with $\chi_1 = 41°$. It has been stated that the relaxation of the strained flat conformation of the Pro168 ring in the closed loop conformation drives loop opening.[21] If this were true, then it would mean that if we were to simulate the loop in isolation of the rest of the protein relaxation of the strain in the ring of Pro168 would drive the conformation of the loop toward the open structure. However, if the loop were to open, then it may be due to relaxation of strain elsewhere within the loop. If so, then simulation with our rigid-ring proline model would also show the loop move toward the open conformation. In order to investigate this, we carried out two sets of Monte Carlo simulations, both starting with the closed loop conformation. In one set of simulations (three independent runs using different seeds for the random number generator), we used our flexible ring proline model, and in the other (three independent runs), the rigid-ring proline model. In all simulations, all other parts of the protein and any ligands were removed. In order to preserve the flat conformation of the proline ring for the flexible case prior to the Monte Carlo simulations, minimizations were performed with positional restraints on the heavy atoms. The simulations were carried out at 300 K for 1 000 000 steps with $\delta\nu_{max} = 5°$ and $\delta s_{max} = 2°$. Before proceeding to analyze the loop movement, we analyzed the ϕ and χ_1 angle distributions of Pro168 in order to check the method for simultaneous proline ring and loop closure. Figure 7 shows their distributions. The ϕ angle has a mean value of −69.8° with a standard deviation of 9.9° and is similar to experimentally derived distributions.[13,26b] The χ_1 shows some indication of the expected bimodal distribution with its overall range close to the experimentally derived distribution.[26b] These results lead us to conclude that our sampling method for simultaneous ring and loop closure works correctly.

In order to measure how far the loop moves from the closed to the open conformation, we use a projection method. First, the open experimental structure is superposed on the closed structure on regions 164−167 and 179−182 flanking the mobile part. There are 11 C^{α} atoms in the mobile part of the loop, and we construct a vector $\mathbf{\Delta d}_{exp} = (\mathbf{\Delta d}^{t}_{exp,1} \ldots \mathbf{\Delta d}^{t}_{exp,i} \ldots \mathbf{\Delta d}^{t}_{exp,11})^{t}$, where $\mathbf{\Delta d}_{exp,i}$ is a 3 × 1 column vector giving the displacement (x,y,z components) of the ith C^{α} atom from the closed experimental to open experimental structure. An equivalent quantity exists for the simulated loop $\mathbf{\Delta d}(k)_{sim} = (\mathbf{\Delta d}(k)^{t}_{sim,1} \ldots \mathbf{\Delta d}(k)^{t}_{sim,i} \ldots \mathbf{\Delta d}(k)^{t}_{sim,11})^{t}$, where $\mathbf{\Delta d}(k)_{sim,i}$ is a 3 × 1 column vector giving the displacement of the ith C^{α} atom from the closed experimental to simulated structure at the kth step. The projection value, $p(k)$ is given by

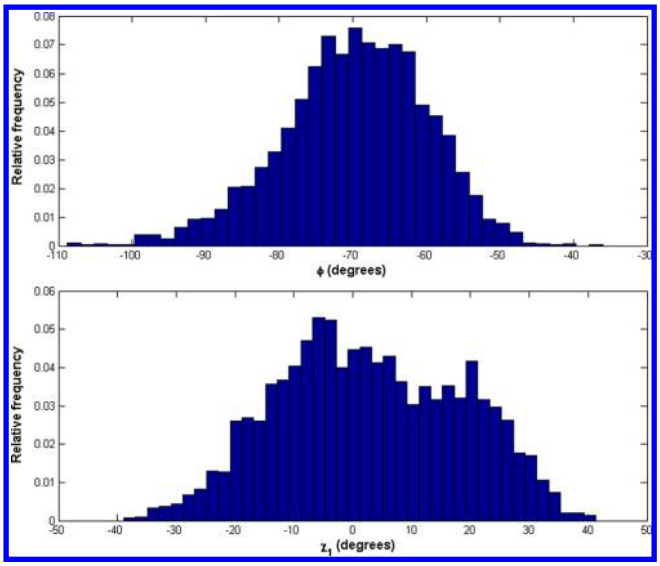


Figure 7. Distributions of Pro168 angles derived by combining results from three independent 1 000 000 step runs at 300 K of the TIM loop in the closed conformation using flexible proline model. Top: ϕ angle. Bottom: χ_1 angle.

$$p(k) = \frac{\mathbf{\Delta d}_{sim}(k)\cdot\mathbf{\Delta d}_{exp}}{|\mathbf{\Delta d}_{exp}|} \quad (14)$$

$p(k)$ is equal to 0 when the loop is closed and equal to 1 when it is open. Figure 8 shows the trajectory of $p(k)$ for all six simulations, together with the trajectory of the χ_1 angle of Pro168. Obviously, for the rigid-ring proline model, this angle is constant and remains close to 0, as expected for the planar conformation of the ring. The value of $p(k)$ fluctuates around 0, indicating that the loop remains largely in the closed conformation. Although for the flexible ring proline model the χ_1 angle of Pro168 fluctuates over a large range of angles, $p(k)$ fluctuates around 0, meaning that the loop does not relax toward the open structure and that loop opening is not related to relaxation of strain in the Pro168 ring. This result shows that the proposed mechanism for loop opening in TIM is incorrect and that it must be more complex, involving other parts of the protein and/or the exiting product (see Discussion).

These simulations also give us the opportunity to study the effect of proline ring flexibility on loop conformations. The root-mean-square fluctuation (RMSF) of the loop for the flexible-ring proline model is larger in all three simulations (3.4, 3.1, 3.5 Å) than that in any of the three simulations with the rigid-ring proline model (2.8, 2.8, 3.0 Å). Although it has been shown that the loop does not relax toward the open conformation, we have also analyzed whether the fluctuations are more in the direction of the open conformation for the flexible-ring proline model than for the rigid-ring proline model. In order to evaluate this, we calculate

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

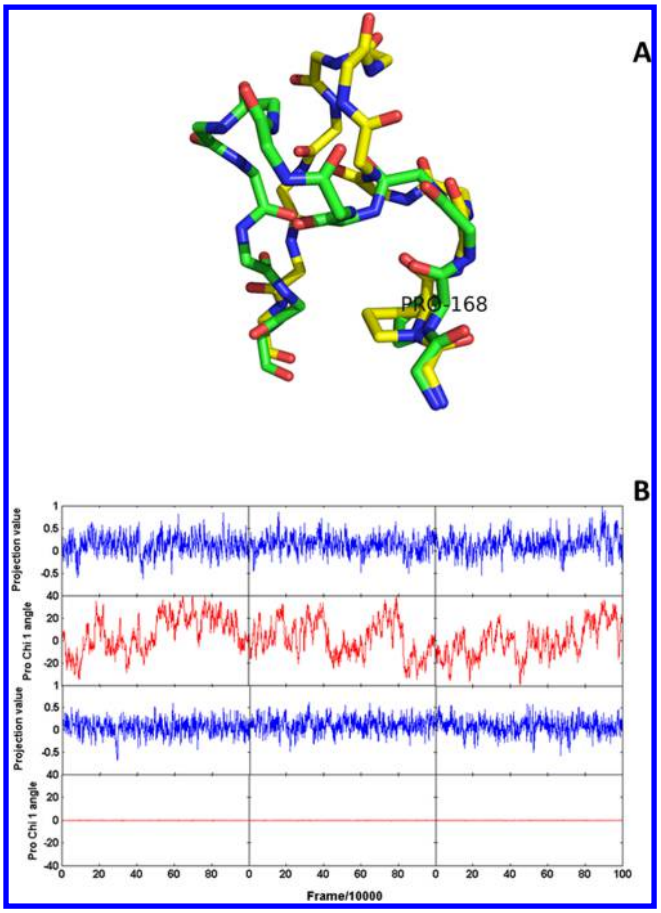


Figure 8. (A) TIM loop 6 residues 167−179 showing the location of Pro168 and the conformational difference between closed (green) and open (yellow) when structures are fitted on all residues not in loop 6 (figure produced using PyMOL). (B) Projection and χ_1 angle trajectories for TIM loop 6. Each column shows the result of an independent simulation. Top two rows are projection values $p(k)$ (blue) and χ_1 angle (red) for the flexible-ring proline model. Bottom two rows are projection values $p(k)$ (blue) and χ_1 angle (red) for the rigid-ring proline model. A projection value of 0.0 indicates a closed conformation, and a value of 1.0, an open conformation of the loop.

$$z = \frac{\mu_{\text{flex}} - \mu_{\text{rig}}}{\sqrt{\frac{\sigma_{\text{flex}}^2}{N} + \frac{\sigma_{\text{rig}}^2}{N}}} \tag{15}$$

where N (=1000) is the number of points sampled, μ_{flex} and σ_{flex} are the mean and standard deviation of the projection value for the flexible-ring proline, respectively, and μ_{rig} and σ_{rig} are the mean and standard deviation of the projection value for the rigid-ring proline, respectively. Table 1 shows z values for all nine pairs of runs. In a test where the null hypothesis is that there is no difference between the two means, these values are all highly significant (e.g., probability(z) $\geq$ 1.97) = 0.024). This shows clearly that limited rotation about the proline ϕ axis is important.

Table 1. z Values for Differences in Projection Values between Flexible- And Rigid-Ring Proline Model Runs

		rigid-ring proline model		
		run 1	run 2	run 3
flexible-ring proline model	run 1	8.8	2.0	5.4
	run 2	10.4	3.2	6.8
	run 3	10.1	3.4	6.8

Domain Movement in GBP. It might at first sight appear inappropriate to use this software for simulation of domain movement. The present implementation allows us to simulate a single loop within which one or more rigid regions may exist. In the case of GBP, one domain is the rigid region from which the loop extends and the other is a rigid region within the loop.

DynDom analysis of the open X-ray structure (PDB code: 1GGG, chain A) and the closed X-ray structure bound to glutamine (PDB code: 1WDN, chain A) reveals a two-domain protein whereby domain B rotates 56° relative to domain A between the open and closed conformations. The domain/bending region order along the chain is domain A 1−86; bending 87−89; domain B 90−180; bending 181−184; domain A 185−226. The domain movement is controlled by a double-hinged β-sheet, a motif found in other hinge bending proteins such as lactoferrin.[27] One can regard the bending regions together with domain B as a single loop extending from domain A, with domain B as a rigid segment within the loop flanked by flexible bending regions. For the purpose of simulation, the loop is defined as 85−186, with residues 85 and 186 fixed and region 91−179 being a rigid segment. This means that domain B will move relative to domain A only through flexibility at regions 86−90 and 180−185 (there are no prolines in these regions). If we are able to see the correct domain movement in our simulations, then it will show that these regions alone control the movement. All simulations were started from the open domain conformation (1GGG, chain A). After minimization, 1 000 000 Monte Carlo steps with $\delta\nu_{\text{max}} = 5°$ and $\delta s_{\text{max}} = 1°$ were carried out. In order to measure the degree of domain closure, the projection value as defined in eq 14 was again employed but with $p(k)$ defined by 102 C^{α} atoms in the region 85−186 inclusive and with $p = 0$ indicating the open domain conformation and $p = 1$ indicating the closed domain conformation. Simulations carried out at 300 K showed no closure. Given that in these simulations the domains are held rigid and are effectively held at 0 K, it would seem reasonable to try the simulations at a higher temperature. The simulations were rerun at 1000 K, but the domains remained in the open domain conformation. A closer examination of the contacts between the domains was made. DynDom analysis of the X-ray structures shows that the rotational transition between the domains[28] occurs between 88 and 89 in the first bending region and between 182 and 183 in the second bending region. For the purpose of the following analysis, domain A is defined as 1−88 and 183−226 and domain B, as 89−182. In going from the open to closed conformation, the following contacts (based on a 4 Å cutoff between any pair of heavy atoms) are broken (domain A residue−domain B residue): Lys87−Asp178, Tyr217−Leu162, Lys219−Lys166, Trp220−Tyr163, Trp220−Lys166, and Phe22−-Leu162. Lys8−-Asp178 is a salt bridge, and Lys219−Lys166 is a hydrogen bond between Lys166's side chain and the main chain CO of Lys219. It is conceivable that if the environment of Lys87 and Lys166 were to change so as to reduce the pK_a sufficiently then they could become deprotonated, breaking the Lys87−Asp178 and Lys219−Lys166 bonds. The breaking of these bonds might allow the domains to close. An example of deprotonated lysines has been found in engineered staphylococcal nuclease where the lysines were introduced into its hydrophobic core.[29] In order to test this idea, a Lys87Ala, Lys166Ala double mutant was created. Although simulation of this mutant at 300 K did not produce

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

closure, a simulation at 1000 K did. A further two wild-type (WT) simulations and a further two mutant simulations were carried out at 1000 K to bring the number of independent simulations to three for both. Figure 9 shows the projection

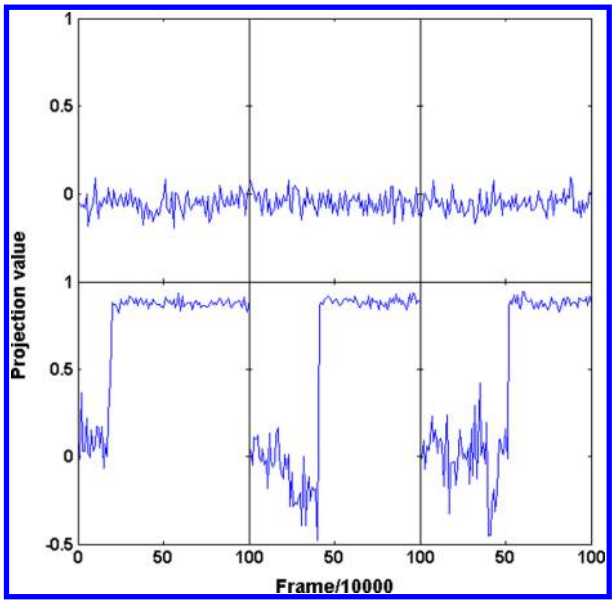


Figure 9. Projection trajectories for GBP, where a projection value of 0.0 indicates an open domain conformation and a value of 1.0, a closed domain conformation. Top row: three independent wild-type simulations at 1000 K. Bottom row: three independent Lys87Ala, Lys166Ala double mutant simulations at 1000 K.

value trajectories, $p(k)$, for these simulations. All of the WT simulations remain locked in the open-domain conformation, whereas for all the mutant simulations, after period of large fluctuation around the open-domain conformation, the domains close rapidly and remain closed. How similar is the simulated domain movement in comparison to the experimentally observed domain movement? DynDom was applied to the first and last structures of the mutant simulations. The rotation angles for the three independent simulations were 69°, 64°, and 66° and are somewhat larger than the 56° for the experimental transition, which may be explained by the absence of glutamine in the interdomain binding site in the simulations. Figure 10 shows a comparison of the DynDom hinge axis for the experimental movement with the hinge axes of the three closing simulations. As can be seen, all three simulation axes are similar in location and orientation to the experimental axis. These results could indicate that an environmental switch operates in the process of glutamine binding in GBP.

DISCUSSION

A Monte Carlo linear inverse-kinematics method for simulation of flexible regions in proteins under loop closure conditions has been described and implemented. The implementation allows for bond angle variation along the backbone and fully flexible prolines. Our results show good correspondence to experimentally derived distributions of ϕ, ψ angles in polyalanine and polyglycine models and proline ϕ and χ_1 angles. A particular feature of the implementation is that the loop can have multiple rigid regions flanked by flexible regions. These rigid regions add no cost to the simulation other than their contribution to the evaluation of the total energy.

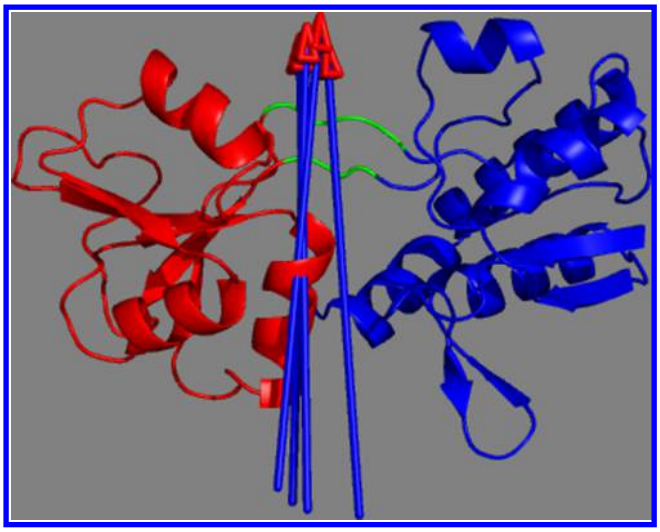


Figure 10. DynDom comparisons between the experimental transition and the mutant simulations for GBP. The arrow to the right depicts the hinge axis from the experimental movement, with the other arrows being those from the simulation. The DynDom coloring is from the analysis of the experimental movement, with the domains colored blue and red and the interdomain bending regions colored green.

Linear inverse-kinematics methods are not as popular as nonlinear methods. However, for sampling of loop conformations under fixed end constraints, it is likely that nonlinear methods will find solutions that are not consistent with permanently maintained end constraints, i.e., they are on different self-motion manifolds.[20] If so, then linear methods would be preferable to nonlinear methods for sampling conformations under permanently maintained end constraints. However, for polymer simulations that perform loop closure on randomly selected segments within the polymer chain (i.e., the constraint on the end of the segments is not permanently maintained), nonlinear methods, including those that use configurational bias rebridging methods,[30] may be more efficient, as at each step they may be able to access conformations not attainable using the linear method. A further advantage of the linear approach is that there is no need to use Jacobians to modify the acceptance probability of a trial move in the Monte Carlo procedure. As we have shown, it is relatively straightforward to apply simultaneous loop closure and proline ring closure within the linear method. It is not clear how this would be achieved so easily within the nonlinear approach.

A disadvantage of the linear null-space approach is that finite step iteration eventually leads to end drift. This means that the end group must be reset occasionally to its original position. The reset method is one that is firmly based within the methodology for determining movements that do not disturb the end group and resets it with minimal effect on the current conformation. The reset is simultaneously applied to the loop and all prolines and does not add appreciable extra computational cost.

A further disincentive for the development of linear methods may be the apparent high computational load in evaluating the null-space vectors through SVD. The rigid segments add no appreciable cost, so the total cost relates directly to the total number of residues in the flexible regions N_{res}. The Jacobian without prolines is of order $6 \times (6N_{res} - 6)$. Because N_{res} would be small in applications appropriate for this kind of method-

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

ology, the computational load would be low. For loops with prolines, the Jacobian becomes larger but sparse, so fast and efficient methods for performing SVD on sparse matrices can be used.[31] Furthermore, with the advent of GPU computing, methods have been developed for fast SVD on GPUs.[32]

For proteins, molecular dynamics (MD) has become the established computational tool for simulating protein dynamics. It is important, therefore, to explain under what circumstances a new simulation method might offer an advantage over MD. It is clear that Monte Carlo offers advantages in being easier to implement than MD, and in some systems, it has been found that Monte Carlo gives an advantage over MD in sampling speed.[33] Although our implementation precludes us from making this kind of direct comparison, it offers the advantage of being able to rigidify regions easily. This is not possible in standard atom-based integration MD applied to proteins, as constraints are applied to distances between pairs of atoms, causing the simulations to become prohibitively slow when a large number constraints are required to completely rigidify a region.

Rigidifying groups allows one to test concepts of protein motion based on models of proteins comprising rigid regions connected by flexible regions. Although proteins or parts of proteins are never completely rigid, simulation methods do allow us to construct rigid−flexible models to test these concepts. In contrast to the all-atom MD simulations reported on GBP,[34] our simulation is of two perfectly rigid domains connected by two short flexible regions. These flexible regions comprised only 11 out of the 226 residues in the protein. We showed, albeit using a high-temperature simulation of a mutant protein, that the experimental domain movement could be reproduced accurately. This leads us to the conclusion that the domain movement is largely controlled through these short flexible regions and that the rest of the protein can be conceptually regarded as comprising two rigid domains. This finding conforms to the so-called door closing model of protein closure, where two separated interdomain bending regions create a stable hinge axis for control of the domain motion.[27] This is not to suggest that the rigid regions and residues within them do not have an important role to play. However, it does suggest that the rigid−flexible model is useful for understanding mechanism in GBP.

For the TIM loop (loop 6), we showed that the proposed model of opening being driven by energy stored in a high-energy flat conformation of the proline ring cannot be correct. These simulations showed that the proline ring sampled extensively puckered conformations without an associated movement of the loop toward the open conformation. Given that these simulations were performed on an isolated loop, it is very likely that loop opening in TIM involves other regions of the protein. In particular, it has been noted that neighboring loop 7 also undergoes conformational change,[24] and this loop may be more directly involved in the movement of loop 6 than expected. While opening did not occur, comparison of the flexible- and rigid-ring proline simulations showed clearly that proline ring flexibility is crucial in allowing the loop to move toward the open conformation.

The current implementation of this method is limited to one loop, but it could, in principle, be extended to more than one loop and indeed to a more general purpose simulation method for proteins comprising flexible and rigid regions. The basic method for simultaneous loop−ring closure could also be developed for simultaneous ring−ring closure for double-ring molecules or double rings created by two disulfide bridges or extended to molecules with even more complex ring structures.

AUTHOR INFORMATION

Corresponding Authors

*(S.H.) Tel: +44-1603-593542; Fax: +44-1603-593345; E-mail: sjh@cmp.uea.ac.uk.

*(A.K.) Tel: +81-3-5841-2297; Fax: +81-3-5841-2297; E-mail: kitao@iam.u-tokyo.ac.jp.

Funding

S.H. is grateful to the The Japan Society for the Promotion of Science (JSPS) for a Bridge Fellowship awarded in Financial Year 2014. This work was supported by the Computational Material Science Initiative (CMSI), Grants-in-Aid for Science Research in Innovative Areas (25104002), from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and Grants-in-Aid for Science Research B (15H04357) from JSPS to A.K.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Malabanan, M. M.; Amyes, T. L.; Richard, J. P. A role for flexible loops in enzyme catalysis. *Curr. Opin. Struct. Biol.* **2010**, *20*, 702−710.

(2) (a) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. Protein flexibility predictions using graph theory. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 150−165. (b) Wells, S. A.; Menor, S.; Hespenheide, B.; Thorpe, M. F. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* **2005**, *2*, S127−S136.

(3) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.

(4) Shehu, A.; Kavraki, L. E. Modeling Structures and Motions of Loops in Protein Molecules. *Entropy* **2012**, *14*, 252−290.

(5) Go, N.; Scheraga, H. A. Ring Closure and local conformational deformations of chain molecules. *Macromolecules* **1970**, *3*, 178−187.

(6) Lee, H. Y.; Liang, C. G. A new vector theory for the analysis of spatial mechanisms. *Mechanism and Machine Theory* **1988**, *23*, 209−217.

(7) (a) Coutsias, E. A.; Seok, C.; Jacobson, M. P.; Dill, K. A. A kinematic view of loop closure. *J. Comput. Chem.* **2004**, *25* (4), 510−528. (b) Nilmeier, J.; Hua, L.; Coutsias, E. A.; Jacobson, M. P. Assessing Protein Loop Flexibility by Hierarchical Monte Carlo Sampling. *J. Chem. Theory Comput.* **2011**, *7*, 1564−1574. (c) Dinner, A. R. Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *J. Comput. Chem.* **2000**, *21*, 1132−1144. (d) Hoffmann, D.; Knapp, E. W. Polypeptide folding with off-lattice Monte Carlo dynamics: The method. *Eur. Biophys. J.* **1996**, *24*, 387−403.

(8) (a) Dodd, L. R.; Boone, T. D.; Theodorou, D. N. A Concerted rotation algorithm for atomistic monte-carlo simulation of polymer melts and glasses. *Mol. Phys.* **1993**, *78*, 961−996. (b) Wu, M. G.; Deem, M. W. Efficient Monte Carlo methods for cyclic peptides. *Mol. Phys.* **1999**, *97*, 559−580.

(9) van den Bedem, H.; Lotan, I.; Latombe, J. C.; Deacon, A. M. Real-space protein-model completion: an inverse-kinematics approach. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 2−13.

(10) (a) Yao, P.; Dhanik, A.; Marz, N.; Propper, R.; Kou, C.; Liu, G. F.; van den Bedem, H.; Latombe, J. C.; Halperin-Landsberg, I.; Altman, R. B. Efficient Algorithms to Explore Conformation Spaces of Flexible Protein Loops. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2008**, *5*, 534−545. (b) Shehu, A.; Clementi, C.; Kavraki, L. E. Modeling protein conformational ensembles: From missing loops to equilibrium fluctuations. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 164−179.

(11) Hayward, S.; Kitao, A. The Effect of End Constraints on Protein Loop Kinematics. *Biophys. J.* **2010**, *98*, 1976−1985.

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905

(12) Bruccoleri, R. E.; Karplus, M. Chain closure with bond angle variations. *Macromolecules* **1985**, *18*, 2767−2773.
(13) Morris, A. L.; Macarthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein-structure coordinates. *Proteins: Struct., Funct., Genet.* **1992**, *12*, 345−364.
(14) Eyring, H. The Resultant Electric Moment of Complex Molecules. *Phys. Rev.* **1932**, *39*, 746−748.
(15) Hayward, S.; Milner-White, E. J. Simulation of the beta- to alpha-sheet transition results in a twisted sheet for antiparallel and an alpha-nanotube for parallel strands: Implications for amyloid formation. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 3193−3207.
(16) Hastings, W. K. Monte-Carlo sampling methods using markov chains and their applications. *Biometrika* **1970**, *57*, 97−109.
(17) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(18) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.
(19) Kitao, A.; Wagner, C. A space-time structure determination of human CD2 reveals the CD58-binding mode. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 2064−2068.
(20) Burdick, J. W. On the inverse kinematics of redundant manipulators: characterization of the self-motion manifolds. In *Advanced Robotics, 1989*; Springer: Berlin, 1989; pp 25−34.
(21) Kursula, I.; Wierenga, R. K. Crystal structure of triosephosphate isomerase complexed with 2-phosphoglycolate at 0.83-Å resolution. *J. Biol. Chem.* **2003**, *278*, 9544−9551.
(22) Carlacci, L.; Englander, S. W. The loop problem in proteins - a monte-carlo simulated annealing approach. *Biopolymers* **1993**, *33*, 1271−1286.
(23) Engh, R. A.; Huber, R. Accurate bond and angle parameters for x-ray protein-structure refinement. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1991**, *47*, 392−400.
(24) Wierenga, R. K.; Noble, M. E. M.; Vriend, G.; Nauche, S.; Hol, W. G. J. Refined 1.83-Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 m-ammonium sulfate - a comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. *J. Mol. Biol.* **1991**, *220*, 995−1015.
(25) Hayward, S.; Berendsen, H. J. C. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins: Struct., Funct., Genet.* **1998**, *30*, 144−154.
(26) (a) Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. Energy parameters in polypeptides 0.10. improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **1992**, *96*, 6472−6484. (b) Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Sci.* **2001**, *10*, 2627−2632.
(27) Hayward, S. Structural principles governing domain motions in proteins. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 425−435.
(28) Hayward, S.; Lee, R. A. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graphics Modell.* **2002**, *21*, 181−183.
(29) (a) Takayama, Y.; Castaneda, C. A.; Chimenti, M.; Garcia-Moreno, B.; Iwahara, J. Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *J. Am. Chem. Soc.* **2008**, *130*, 6714−6715. (b) Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Garcia-Moreno, B. E. Large shifts in pK(a) values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5260−5265.
(30) (a) Chen, Z.; Escobedo, F. A. A configurational-bias approach for the simulation of inner sections of linear and cyclic molecules. *J. Chem. Phys.* **2000**, *113*, 11382−11392. (b) Uhlherr, A. Monte Carlo conformational sampling of the internal degrees of freedom of chain molecules. *Macromolecules* **2000**, *33*, 1351−1360. (c) Wick, C. D.; Siepmann, J. I. Self-adapting fixed-end-point configurational-bias Monte Carlo method for the regrowth of interior segments of chain molecules with strong intramolecular interactions. *Macromolecules* **2000**, *33*, 7207−7218.
(31) Berry, M. W. Large-scale sparse singular value computations. *Int. J. Supercomput. Appl.* **1992**, *6*, 13−49.
(32) Tomov, S.; Nath, R.; Dongarra, J. Accelerating the reduction to upper Hessenberg, tridiagonal, and bidiagonal forms through hybrid GPU-based computing. *Parallel Computing* **2010**, *36*, 645−654.
(33) Jorgensen, W. L.; Tirado-Rives, J. Monte Carlo vs molecular dynamics for conformational sampling. *J. Phys. Chem.* **1996**, *100*, 14508−14513.
(34) (a) Loeffler, H. H.; Kitao, A. Collective Dynamics of Periplasmic Glutamine Binding Protein upon Domain Closure. *Biophys. J.* **2009**, *97*, 2541−2549. (b) Pang, A.; Arinaminpathy, Y.; Sansom, M. S. P.; Biggin, P. C. Interdomain dynamics and ligand binding: molecular dynamics simulations of glutamine binding protein. *FEBS Lett.* **2003**, *550*, 168−174.

DOI: 10.1021/acs.jctc.5b00215
J. Chem. Theory Comput. 2015, 11, 3895−3905