Analytical rebridging Monte Carlo: Application to *cis/trans* isomerization in proline-containing, cyclic peptides

Minghong G. Wu and Michael W. Deem
Department of Chemical Engineering, University of California, Los Angeles, California 90095-1592

We present a new method, the analytical rebridging scheme, for Monte Carlo simulation of proline-containing, cyclic peptides. The *cis/trans* isomerization is accommodated by allowing for two states of the amide bond. We apply our method to five peptides that have previously been characterized by nuclear magnetic resonance methods. Our simulations achieve effective equilibration and agree well with experimental data in all cases. We discuss the importance of effective equilibration as well as the role of bond flexibility and solvent effects in the prediction of equilibrium properties.   © *1999 American Institute of Physics.* [S0021-9606(99)52037-0]

I. INTRODUCTION

One modern approach to therapeutics is to identify and to block the molecular interactions responsible for disease. Within this context, combinatorial peptide screening experiments play an important role in the discovery of inhibitors. Libraries of peptides are used to discover the crucial molecular structure, or pharmacophore, that can block effectively the aberrant binding event. Cyclic peptides are preferred for this purpose since they display higher binding affinities due to their reduced conformational entropy in solution.[1,2] The cyclic scaffolds and templates derived from combinatorial cyclic peptide studies have been used to assemble various spatially-defined functional groups, and highly active analogs have been synthesized in this way.[2–4] A classic example is the blocking of the molecular event responsible for blood platelet aggregation by peptides of the form CRGDxxxC, CxxxRGDC, and CxxxKGDC.[5] Organic analogs of these RGD peptides have been synthesized, and several companies are pursuing therapeutic applications in clinical trials.

The ability of a peptide molecule to bind selectively to a receptor depends on its structural and conformational properties, which are in turn determined by its constituent amino acids. As the only natural secondary amino acid, proline plays a particular role in peptide and protein structural biology as a turn-promoting unit.[6] Although most amide bonds in native peptides and proteins are in the *trans* state, a significant fraction (6%) of the X-Proline amide bonds are in the *cis* state.[6] In some proteins[7] and cyclic peptides,[8] prolyl *cis/trans* isomerization has been detected in solution. This isomerization event is frequently a rate-limiting step in protein folding. The free energy barrier for the *cis/trans* isomerization of typical X-Pro amide bonds is about 19 kcal/mol.[9] While isomerization of nonprolyl amide bonds is rare, the energy barrier to isomerization is similar.[10]

Conventional molecular simulations cannot follow the natural *cis/trans* isomerization dynamics, since the typical equilibration time is in the range of 10–100 s.[11] Monte Carlo simulations have been successful only in the case where isomerization is experimentally known to occur. In such cases, an isomerization reaction coordinate can be defined, and special techniques such as umbrella sampling can be applied. Conventional molecular dynamics for cyclic peptides, even in the absence of *cis/trans* isomerization, turns out to be nonergodic as well, failing to sample the multiple solution conformations during the accessible simulation time. The difficulties arise from the cyclic constraints that, along with the intrinsic high energy barriers, isolate the accessible conformations to several separated regions in phase space. Monte Carlo methods, however, do not require the system to follow the natural trajectory. Therefore, larger and unphysical moves can be performed to overcome these energy barriers. A numerical peptide rebridging method, inspired by the alkane rebridging method[12] and the configurational bias Monte Carlo (CBMC) method,[13,14] has been successfully applied to the simulation of nonproline-containing, cyclic peptides.[15] Parallel tempering was shown to be a key factor in the efficiency of equilibration. Proline poses some geometrical complexity in the rebridging approach, however, and has not yet been treated.

Here, we present a new Monte Carlo method, the analytical rebridging scheme, that is suitable for equilibration of proline-containing, cyclic peptides. Our analytical method was inspired by the solution for a related inverse kinematics problem in robotic control.[16] The method can accommodate any rigid unit geometry. The rebridging method is not restricted to peptides and can be readily applied to other molecules. The *cis/trans* isomerization is naturally incorporated in the method by allowing for two states of the amide bond. As an added benefit, we find the analytical approach to be at least ten times more computationally efficient than the previous numerical method,[15] even in the simplest, nonproline-containing cases. The analytical rebridging method and other components of our simulation methodology are described in Sec. II. We apply our method to five cyclic peptides that have previously been characterized by nuclear magnetic resonance (NMR) methods in Sec. III. We show that our method can effectively equilibrate these molecules, yielding conformations consistent with the NMR analyses. We discuss the results in Sec. IV and conclude in Sec. V.
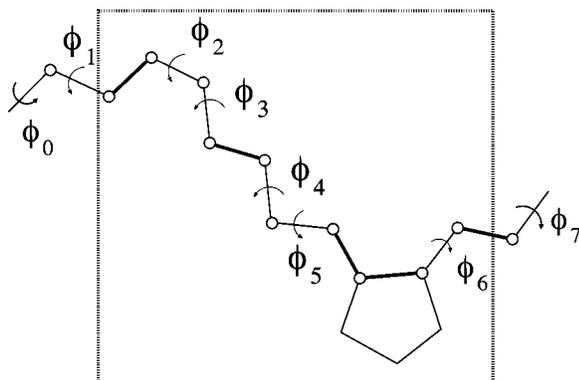
FIG. 1. Backbone segment selected to be rebridged. Only the backbone atoms are shown. A change of the driver angles ϕ_0 and ϕ_7 breaks the connectivity. The dotted area represents the region in which the positions of the backbone atoms must be restored. The thick solid lines represent pi bonds or rigid molecular fragments within which no rotation is possible.

II. SIMULATION METHODS

A. Analytical peptide rebridging

Our equilibration scheme involves three types of moves: rebridging moves, semi-look-ahead (SLA) moves for side chains, and swapping moves. The rebridging move is described below and in Sec. II B. We describe the SLA move in Sec. II C and the swapping move in Sec. II D.

In our system, bond lengths and bond angles are kept at their equilibrium value. We focus on sampling the biologically relevant, torsional degrees of freedom. With this simplification, a molecule is comprised of a set of rigid units.[17] The rebridging scheme can easily be generalized to accommodate flexible bond angles and bond lengths.

A peptide rebridging move causes a local conformational change within the molecule, leaving the rest of the molecule fixed. Consider the segment of a peptide backbone shown in Fig. 1. The angles ϕ_0 and ϕ_7 are rotated, causing the rigid units between 0 and 6 to change. The range of rotation is within $\pm\Delta\phi_{max}$. These two rotations break the connectivity of the molecule. We then find all the solutions that reinsert the backbone units in a valid way between rigid units 1 and 6. The rebridging move is based upon enforcing six geometrical constraints, and this is why we choose to modify six rigid units. Modification of more than six units in a single move is possible by the rotation of more than two angles. Such a move is likely, however, to lead to an infeasible geometry most of the time and so to result in ineffective equilibration. Therefore, we choose to rotate ϕ_0 and ϕ_7 only.

Our peptide rebridging scheme features an analytical solution of the geometrical problem arising from the reconnection of the backbone units, a problem previously solved in a numerical way.[15,17] The side chains are rigidly rotated for each of the solutions. For rebridging moves, the solutions for both the new and the old configurations are needed so as to satisfy detailed balance. The analytical solution involves the reduction of twenty linear equations to an 8×8 determinant equation of one torsional angle. The determinant equation is equivalent to a polynomial of degree sixteen. Therefore, the maximum number of new geometrical solutions is strictly limited to 16, a bound that is obeyed in previous simulations

using the numerical rebridging method.[18] The determinant equation is reformulated as an eigenvalue problem[16] and solved using the QR algorithm.[19] The details of the analytical rebridging method are described in Appendix A.

Following the ''with Jacobian'' (WJ) biasing,[15] one of the solutions is picked with a probability proportional to $J^{(n)}\exp(-\beta U^{(n)})$, the product of the Jacobian and the Boltzmann factor of the solution. The Jacobian J accounts for the correction to the nonuniform distribution generated by rebridging moves, and can be expressed in several ways. Here, we present one form, involving the determinant of a 5×5 matrix $\mathbf{B}$:

$$J=\frac{\hat{\mathbf{u}}_6\cdot\hat{\mathbf{e}}_3}{\det|B|}$$

$$\mathbf{B}_{ij}=[\hat{\mathbf{u}}_j\times(\mathbf{r}_6-\mathbf{r}_j)]_i,\quad\text{if}\ \ i\leqslant3,$$
$$[\hat{\mathbf{u}}_j\times\hat{\mathbf{u}}_6]_{i-3},\quad\text{if}\ \ i=4,5. \tag{1}$$

Here, $\hat{\mathbf{u}}_i$ is the unit vector about which the torsional angle ϕ_i is measured, $\mathbf{r}_i$ is the position of the atom in unit $i-1$ that is bonded by a sigma bond to unit i, and $\hat{\mathbf{e}}_3$ is a unit vector along the laboratory z axis. The subscript outside the brackets refers to the component of the bracketed vector in the laboratory frame. The Jacobian can also be written in terms of a determinant of a 4×4 matrix, although the definition of the components of the 4×4 matrix is more involved.[15] The attempted move is accepted with the probability

$$\text{acc}(o\rightarrow n)=\min\left(1,\frac{W^{(n)}}{W^{(o)}}\right), \tag{2}$$

where $W^{(n)}$ and $W^{(o)}$ are the normalization (Rosenbluth) factors for the new and old solutions, respectively.[15,20]

B. The cis/trans isomerization

For each amide bond that we wish to allow to isomerize, we assign two discrete states to the corresponding rigid unit. As shown in Fig. 2, the amide unit takes the torsional values of $\omega=0°$ or $\omega=180°$ in the cis and trans conformations, respectively. The partition function includes a summation over both states. Because of the sum over states, solutions corresponding to all the possible cis/trans states between unit 0 and unit 6 are included in the calculation of the Rosenbluth factor. The same approach can be applied to both nonprolyl and prolyl amide bonds.

C. Side groups

The chemical functionality of peptides lies mostly in the side chains. Efficient equilibration for the side chains, therefore, is very important. The SLA move, based on CBMC methods, has been shown to equilibrate effectively long and bulky chains.[15] We use the SLA method to equilibrate side chains and end groups in this work. A SLA move proceeds by regrowing a randomly selected side chain, unit by unit, beginning from the bond that connects the backbone to the side chain. The reverse move is performed so as to satisfy detailed balance. The Jacobian for each solution is unity. The attempted move is accepted with the probability
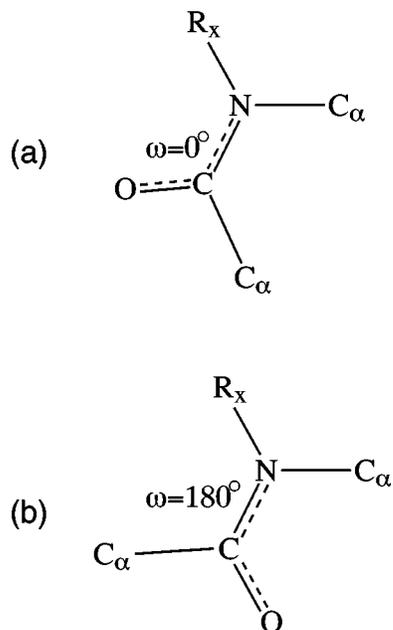
FIG. 2. (a) *cis* and (b) *trans* conformation of the amide bond. Only backbone atoms are depicted. The R_x atom is hydrogen for a normal, non-prolyl amide bond or carbon for a prolyl or methylated amide bond.

FIG. 3. Conformation of c(Pro-Phe-D-Trp-Lys-Thr-Phe). The colors green, red, and blue denote carbon, oxygen, and nitrogen atoms, respectively. The hydrogen atoms are omitted.

$$\mathrm{acc}(o \rightarrow n) = \min\left(1, \frac{W^{(n)}}{W^{(o)}}\right), \tag{3}$$

where $W^{(n)}$ and $W^{(o)}$ are the normalization (Rosenbluth) factors for the new and old geometries, respectively.[20]

D. Parallel tempering

Parallel tempering was first proposed for the study of glassy systems with large free energy barriers.[21] It has since been successfully applied to a variety of systems.[22–26] This method achieves rigorously correct canonical sampling, and it significantly reduces the equilibration time in a simulation. Instead of a single system, we consider in parallel tempering a larger ensemble with n systems, each equilibrated at a distinct temperature T_i, $i = 1, \ldots, n$. The system with the lowest temperature is the one of our interest; the higher temperature systems are added to aid in the equilibration of the system of interest. In addition to the normal Monte Carlo moves performed in each system, swapping moves are proposed that exchange the configurations between two systems i and $j = i + 1$, $1 \leq i < n$. A swapping move is accepted with the probability

$$\mathrm{acc}[(i,j) \rightarrow (j,i)] = \min[1, \exp(-\Delta\beta\Delta U)], \tag{4}$$

where $\Delta\beta$ and ΔU are the difference of the reciprocal temperatures and energies, respectively. The higher temperature systems are included solely to help the lowest temperature system to escape from local energy minima via the swapping moves. To achieve efficient sampling, the highest temperature should be such that no significant free energy barriers are observed. So that the swapping moves are accepted with a reasonable probability, the energy histograms of systems adjacent in the temperature ladder should overlap.
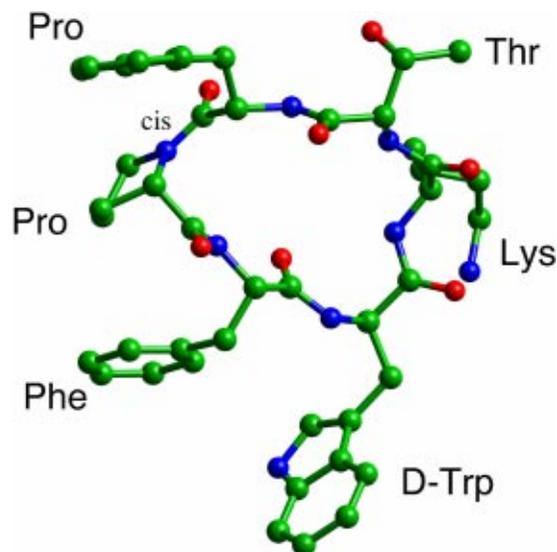
III. COMPARISON WITH EXPERIMENTAL RESULTS

We perform simulations on five distinct cyclic peptides that have been previously characterized by NMR methods. We focus on the backbone structure of proline-containing, cyclic peptides that were observed experimentally to undergo *cis/trans* isomerization. Molecular interactions are described by the AMBER force field with explicit atoms.[27] Aqueous solvent effects are estimated by simple dielectric theory.[28] Five or six systems were used in the parallel tempering for each simulation, with the highest temperatures ranging from 10^5 to 10^7 K. The lowest temperature system in each case is 298 K. For the first three peptides, *cis/trans* isomerization is allowed in prolyl amide bonds only. For the last two peptides, isomerization of all amide bonds is allowed. The simulations take 5–8 CPU hours for the first three peptides and 15–20 CPU hours for the last two peptides. All the simulations were performed on an Intel Pentium II 450 MHz Linux workstation. Rapid equilibration is achieved for all peptides. The results for each of the peptides are presented below:

(1) c(Pro-Phe-D-Trp-Lys-Thr-Phe). This analog of somatostatin displays high activity in inhibiting the release of a growth hormone.[4,29] Analysis of the NMR spectrum in D_2O solution indicated a unique backbone conformation.[29] The prolyl amide bond at Phe-Pro adopted a *cis* conformation. In our simulation, we find an essentially unique conformation, possessing the same amide bond *cis/trans* sequence. A representative conformation for this peptide is shown in Fig. 3.

(2) c(Phe-Phe-Aib-Leu-Pro). This pentapeptide contains the Pro-Phe-Phe sequence that has been proposed to be responsible for the cytoprotective ability of antamanide and cyclolinopeptides.[30,31] NMR analysis indicated that the peptide is conformationally non-homogeneous at room temperature. Two predominant *cis/trans* isomers for the Leu-Pro amide bond were identified in acetonitrile at 240 K.[31] Our simulation led to two interconverting conformers in the simulation, as shown in Fig. 4. The *cis* and *trans* conformers
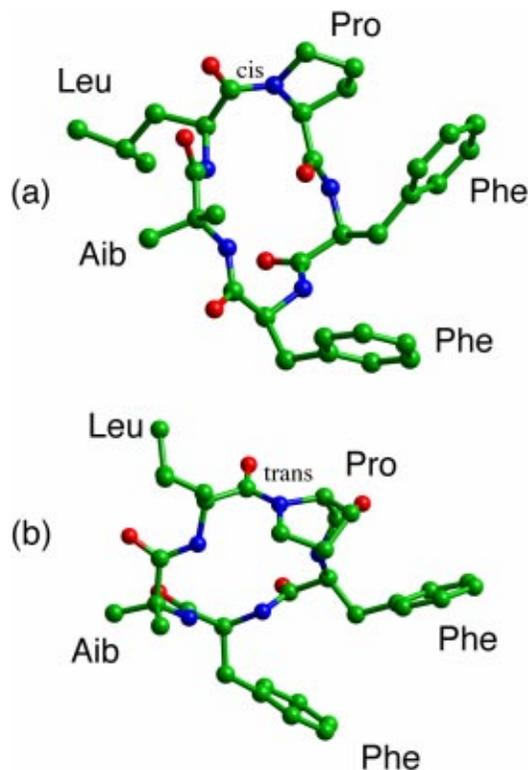
FIG. 4. (a) *cis* and (b) *trans* conformations of c(Phe-Phe-Aib-Leu-Pro).

FIG. 6. Conformation of c(Pro-Ala-Pro-Ala-Ala).

occur with probability 58% and 42%, respectively.

(3) c(Gly-Glu(OBzl)-Pro-Phe-Leu-Pro). This cyclic hexapeptide was synthesized for use as a possible chiral site for enantiomeric separation.[32] NMR studies in dimethyl sulfoxide (DMSO) reported two isomers, one having two *cis* prolyl bonds, and the other having all-*trans* bonds. We find only the 2-*cis* conformer in simulation, as shown in Fig. 5. All the torsional angles fluctuate around mean values, except that the amide group between Pro and Gly flips between two opposite orientations. The all-*trans* conformer was found at higher temperatures.

(4) c(Pro-Ala-Pro-Ala-Ala). This cyclic peptide has been designed to serve as a rigid structural template. An unique conformation with two *cis* prolyl amide bonds in DMSO solution was found, according to NMR analysis.[33] The backbone consists of two intertwined type-VIb β turns, centered about the two prolyl amide bonds respectively. We

find the same unique conformation. The torsional angles are close to the values derived from restrained molecular dynamics.[33] Figure 6 depicts the geometry of this peptide.

(5) Tentoxin, c(MeAla-Leu-MePhe[(Z)$\triangle$]-Gly). This tetrapeptide selectively induces chlorosis in the seedlings of plants. Although tentoxin lacks proline, its two methylated amide bonds were found to adopt the *cis* conformation in a nearly saturated aqueous solution.[34] The other two non-methylated amide bonds adopt the *trans* conformation. The observation of this *cis-trans-cis-trans* sequence of the backbone, along with other experimental data, led to a proposed boat-like conformation, with the two *cis* bonds located on the same side of the mean plane. In our simulation, we find the same amide bond sequence and boat-like conformation that was found in the experimental structure. The conformation is shown in Fig. 7. All the carbonyl groups lie in the same side of the mean plane, which implies that we have found the the third major (conformer C, 8% abundance) of the four conformers found at 268 K in Ref. 34. The four conformers differ only in the orientation of the two nonmethylated amide groups. The other three conformers were found at higher temperatures in the simulation.

IV. DISCUSSION

By analyzing the energy trajectories and conformational data, we find that all the peptides are effectively equilibrated with relatively few Monte Carlo steps. For example, we find
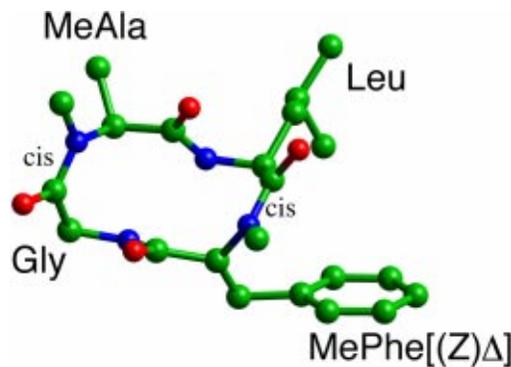

FIG. 5. Conformation of c(Gly-Glu(OBzl)-Pro-Phe-Leu-Pro).

FIG. 7. Conformation of c(MeAla-Leu-MePhe[(Z)$\triangle$]-Gly).

that the *cis/trans* equilibrium for peptide 2 was attained within the first 10% of the simulation time. Parallel tempering is crucial for this interconversion, since the peptide essentially does not undergo *cis/trans* isomerization in a single, room-temperature, canonical simulation.

For peptide 1, the NMR-based conformational study indicated a type-II'β turn in the Phe-D-Trp-Lys-Thr region. This turn is characterized by the hydrogen bond between the C=O of Phe and the N–H of Thr. In our simulation, the C=O of Phe and the N–H of Thr are close to each other, but are not in precise alignment. Such disorder is expected at finite temperature.

Bond angle inflexibility may influence the predicted equilibrium properties of these highly strained molecules. We have investigated the dependence on bond angles by changing the angle between C_α–C and N–C_α in the amide bond on peptide 2 from 6° to 0°. The predicted *cis/trans* equilibrium shifts from 42% *trans* to 14% *trans*, which is a nontrivial, although energetically small, effect. We suspect, therefore, that the absence of the all-*trans* conformer of peptide 3 in our simulation may be due partly to inflexibility of the bond angles. We used a rigid proline ring, with a $\phi_{\text{Pro}} = -75°$. This constraint suppresses the small fluctuations that occur in the proline ring. In exceptionally constrained, cyclic peptides, ϕ_{Pro} can take on other values. For example, $\phi_{\text{Pro}} \simeq -50°$ has been observed for some unusual prolines in the *trans* state.[33] Fluctuations in the proline ring may, therefore, be important in some highly strained systems.

Although experimental conformational analysis of peptide 3 in water was not performed, NMR analysis in CHCl$_3$ indicated a minor conformer in addition to the two major ones found in DMSO solution.[32] It is clear, then, that solvent effects play a role in the equilibrium structure of peptide 3. Note that the dielectric constants are $\epsilon_{\text{CHCl}_3} = 4.8$,[35] $\epsilon_{\text{DMSO}} = 45$,[36] and $\epsilon_{\text{H}_2\text{O}} = 78$.[35] The low-dielectric CHCl$_3$ solution favored the all-*trans* conformer. Water, with a high dielectric constant, may favor the 2-*cis* conformer, which is what we observe in simulation. Indeed, upon reducing the dielectric constant in our implicit solvent model, we find a small amount of the all-*trans* state. However, the detailed structure of the solvent molecules around the peptide is likely to be important, and a more accurate description of solvent is likely necessary to account fully for the solvent effects.

Tentoxin (peptide 5) in water at 268 K was experimentally found to aggregate in a way that suggested micellar organization. The critical micelle concentration (CMC) was estimated to be roughly 35 μM.[34] Conformations observed at concentrations above the CMC may differ from that of the dilute-limit, monomeric form. In fact, of all four conformers found in Ref. 34, only conformer C yielded chemical shifts for the δ and γ protons of Leucine close to that expected for a monomeric form. The other two major conformers, A and B, displayed strong shielding effects. The shielding constants for the minor conformer D were not reported. These shielding effects were explained by an aggregated structure of these two conformers. Since the concentration of conformer C (250 μM) is still well above the critical micelle concentration, it is not entirely clear why the chemical shifts for conformer C were relatively unaffected by potential

aggregation.[34] Nonetheless, the experimental data suggests that conformer C may be either in monomeric form or in an environment similar to aqueous solution. This gives one possible explanation for the absence of conformers A and B in our simulation, since these conformers are certainly not in monomeric form.

V. CONCLUSION

We presented a new method, the analytical rebridging scheme, for the simulation of chemically diverse, chain-like molecules. The method naturally accommodates *cis/trans* isomerizations. Our rebridging scheme, combined with parallel tempering and biased Monte Carlo, is very successful at equilibration of proline-containing, cyclic peptides. Our method is not limited to cyclic peptides and can simulate any chain-like molecule. We compared our simulations with experimental data on five cyclic peptides and found the predicted conformations to be reasonably accurate. We were able to sample multiple, relevant conformations separated by high energy barriers, a feat not possible with conventional molecular dynamics or Monte Carlo. The numerical quality of our predictions, while not limited by sampling issues, may be limited by our choice of a simple forcefield. Nonetheless, our method can be easily extended to accommodate flexible bond angles and bond lengths. In addition, solvent effects may be represented more accurately by better implicit solvent models.[37,38]

The methods described here are powerful enough and general enough to influence the preferred approach to simulating biological systems. For example, our method should be a valuable tool for the fitting of new potential parameters for biological systems. New simulation methods for long alkanes have made possible the optimization of force fields that significantly reduce the discrepancies between simulation and experiment.[39,40] The same approach should lead to improved forcefields for biological systems. In addition, we expect that our peptide rebridging scheme, combined with parallel tempering, should replace high temperature molecular dynamics and is readily suitable for use in NMR-based conformational analyses of biomolecules.

ACKNOWLEDGMENT

APPENDIX A: ANALYTICAL METHOD FOR SOLUTION OF THE REBRIDGING PROBLEM

Our analytical method was inspired by the inverse kinematics problem of six-revolute manipulators, which is important for automatic control of robotic arms. This problem has been proved to have at most sixteen solutions in the general case.[41] Lee and Liang reduced the problem to a polynomial of one variable of degree sixteen.[42,43] The polynomial is derived by equating the determinant of an 8×8 matrix to zero, each element of the matrix being a quadratic polynomial. A different closed form was obtained later by Raghavan and Roth.[44] An excellent review of this subject is given by Manocha and Canny.[16] In the following, we apply the
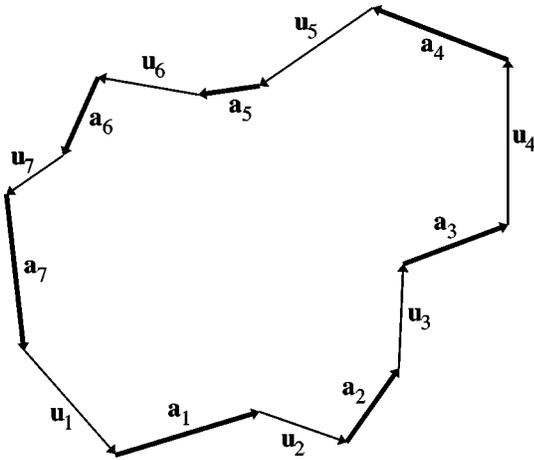
FIG. 8. Geometry of the closed, seven-revolute mechanism, consisting of seven joints and seven links. The joints are represents by $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_7$. The links are represented by $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_7$. Each link is perpendicular in three dimensions to the two adjacent joints. The unit axes of the joints and links are defined as $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{a}}_i$, respectively.

symbolic formulation of Lee and Liang to reduce our peptide rebridging problem to an eigenvalue problem.

The six-revolute inverse kinematics problem can be formulated as an equivalent closed, seven-revolute mechanism,[43] as shown in Fig. 8. The closed loop consists of seven joints with offsets $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_7$ and of seven links with vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_7$. For a given backbone fragment to be rebridged, the corresponding closed loop is uniquely determined by calculating the link-joint intersections. We first draw the seven joint lines that are parallel to the incoming sigma bonds of the seven rigid units. The links are defined as the shortest vectors that connect consecutive joint lines. Therefore, each link is perpendicular to the two adjacent joint vectors. We denote the unit axes of the joint $\mathbf{u}_i$ and the link $\mathbf{a}_i$ by $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{a}}_i$, respectively. The joint rotation angles $\phi_1, ..., \phi_7$ are the torsional angles measured around the joints; they are equal to the corresponding biological torsional angles plus constant offsets. The joint rotation angle ϕ_7 and other parameters are determined by the given backbone geometry, and we need to calculate only the six unknown joint rotation angles.

The idea is to find an overconstrained set of twenty equations that are linear in $x_6 \equiv \tan \phi_6/2$ and the sines of cosines of ϕ_1, ϕ_2, ϕ_4, and ϕ_5. These equations are obtained by equating scalar and vector products of the loop axes in both directions of the loop. For instance, we equate $\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6 (\phi_4, \phi_5) = \hat{\mathbf{u}}_6 \cdot \hat{\mathbf{u}}_3 (\phi_1, \phi_2)$ in the first equation (see below). These equations were first derived using a recursive notation.[43] We define the chain vector $\mathbf{R}^{\alpha,\beta}$ as the vector summation of the consecutive joints and links from unit α to unit β. The summation always goes in the direction of units 1,2,...,7. The indices α and β can take one of two forms: i denotes either starting from or ending at $\mathbf{u}_i$, and i' denotes either starting from or ending at $\mathbf{a}_i$. The index β can be less than α, and this indicates wrapping around the closed loop. For example,

$$\mathbf{R}^{6',2} = \mathbf{a}_6 + \mathbf{u}_7 + \mathbf{a}_7 + \mathbf{u}_1 + \mathbf{a}_1 + \mathbf{u}_2,$$
$$\mathbf{R}^{i',i} = \mathbf{0}. \tag{A1}$$

The twenty equations are listed below. The left-hand side of each equation is a linear function of $\cos \phi_1$, $\sin \phi_1$, $\cos \phi_2$, $\sin \phi_2$, and x_6. The right-hand side of each equation is a linear function of $\cos \phi_4$, $\sin \phi_4$, $\cos \phi_5$, $\sin \phi_5$, and x_6:

$$\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6 = \hat{\mathbf{u}}_6 \cdot \hat{\mathbf{u}}_3,$$

$$\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6 x_6 = \hat{\mathbf{u}}_6 \cdot \hat{\mathbf{u}}_3 x_6,$$

$$\mathbf{R}^{3',5'} \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{u}}_3 = \mathbf{R}^{6',2'} \cdot \hat{\mathbf{u}}_3 \times \hat{\mathbf{u}}_6,$$

$$\mathbf{R}^{3',5'} \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{u}}_3 x_6 = \mathbf{R}^{6',2'} \cdot \hat{\mathbf{u}}_3 \times \hat{\mathbf{u}}_6 x_6,$$

$$\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_3 = -\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_3,$$

$$\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_3 x_6 = -\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_3 x_6,$$

$$\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_6 = -\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_6,$$

$$\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_6 x_6 = -\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_6 x_6,$$

$$\mathbf{R}^{3,5'} \cdot \mathbf{R}^{3,5'} = \mathbf{R}^{6,2'} \cdot \mathbf{R}^{6,2'},$$

$$\mathbf{R}^{3,5'} \cdot \mathbf{R}^{3,5'} x_6 = \mathbf{R}^{6,2'} \cdot \mathbf{R}^{6,2'} x_6,$$

$$\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{a}}_5 x_6 - \hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_5 = -\hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3 x_6 - \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3,$$

$$\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_5 x_6 + \hat{\mathbf{u}}_3 \cdot \hat{\mathbf{a}}_5 = -\hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3 x_6 + \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3,$$

$$-\mathbf{R}^{3,5'} \cdot \hat{\mathbf{a}}_5 x_6 + \mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_5$$
$$\quad = -\mathbf{R}^{6,2'} \cdot \hat{\mathbf{a}}_6 x_6 - \mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_6,$$

$$\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_5 x_6 + \mathbf{R}^{3,5'} \cdot \hat{\mathbf{a}}_5 = \mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_6 x_6 - \mathbf{R}^{6,2'} \cdot \hat{\mathbf{a}}_6,$$

$$\mathbf{R}^{3',5} \cdot \hat{\mathbf{a}}_5 \times \hat{\mathbf{u}}_3 x_6 - \mathbf{R}^{3',5'} \cdot (\hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_5) \times \hat{\mathbf{u}}_3$$
$$\quad = -(\mathbf{R}^{7,2'} \cdot \hat{\mathbf{u}}_3 \times \hat{\mathbf{a}}_6 - \mathbf{u}_6 \times \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3) x_6$$
$$\quad -[\mathbf{R}^{6',2'} \cdot \hat{\mathbf{u}}_3 \times (\hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_6) + |\mathbf{u}_6| \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3],$$

$$\mathbf{R}^{3',5'} \cdot (\hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_5) \times \hat{\mathbf{u}}_3 x_6 + \mathbf{R}^{3',5} \cdot \hat{\mathbf{a}}_5 \times \hat{\mathbf{u}}_3$$
$$\quad = -(\mathbf{R}^{6',2'} \cdot \hat{\mathbf{u}}_3 \times (\hat{\mathbf{u}}_6 \times \hat{\mathbf{a}}_6) + |\mathbf{u}_6| \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3) x_6$$
$$\quad + (\mathbf{R}^{7,2'} \cdot \hat{\mathbf{u}}_3 \times \hat{\mathbf{a}}_6 - \mathbf{u}_6 \times \hat{\mathbf{a}}_6 \cdot \hat{\mathbf{u}}_3),$$

$$1/2(\mathbf{R}^{3,5'} \cdot \mathbf{R}^{3,5'})(\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6) - (\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_3)(\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_6)$$
$$\quad = 1/2(\mathbf{R}^{6,2'} \cdot \mathbf{R}^{6,2'})(\hat{\mathbf{u}}_6 \cdot \hat{\mathbf{u}}_3) - (\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_6)(\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_3),$$

$$[1/2(\mathbf{R}^{3,5'} \cdot \mathbf{R}^{3,5'})(\hat{\mathbf{u}}_3 \cdot \hat{\mathbf{u}}_6) - (\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_3)(\mathbf{R}^{3,5'} \cdot \hat{\mathbf{u}}_6)]x_6$$
$$\quad = [1/2(\mathbf{R}^{6,2'} \cdot \mathbf{R}^{6,2'})(\hat{\mathbf{u}}_6 \cdot \hat{\mathbf{u}}_3) - (\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_6)(\mathbf{R}^{6,2'} \cdot \hat{\mathbf{u}}_3)]x_6,$$

$$[1/2(\mathbf{R}^{3,5'}\cdot\mathbf{R}^{3,5'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{a}}_5)-(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{a}}_5)(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{u}}_3)]x_6$$

$$-[1/2(\mathbf{R}^{3,5'}\cdot\mathbf{R}^{3,5'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_5)$$

$$-(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{u}}_3)(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_5)]$$

$$=-[1/2(\mathbf{R}^{6,2'}\cdot\mathbf{R}^{6,2'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{a}}_6)$$

$$-(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{a}}_6)(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{u}}_3)]x_6$$

$$-[1/2(\mathbf{R}^{6,2'}\cdot\mathbf{R}^{6,2'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_6)$$

$$-(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{u}}_3)(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_6)],$$

$$[1/2(\mathbf{R}^{3,5'}\cdot\mathbf{R}^{3,5'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_5)$$

$$-(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{u}}_3)(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_5)]x_6$$

$$+[1/2(\mathbf{R}^{3,5'}\cdot\mathbf{R}^{3,5'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{a}}_5)$$

$$-(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{a}}_5)(\mathbf{R}^{3,5'}\cdot\hat{\mathbf{u}}_3)]$$

$$=-[1/2(\mathbf{R}^{6,2'}\cdot\mathbf{R}^{6,2'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_6)$$

$$-(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{u}}_3)(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{u}}_6\times\hat{\mathbf{a}}_6)]x_6$$

$$+[1/2(\mathbf{R}^{6,2'}\cdot\mathbf{R}^{6,2'})(\hat{\mathbf{u}}_3\cdot\hat{\mathbf{a}}_6)-(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{a}}_6)(\mathbf{R}^{6,2'}\cdot\hat{\mathbf{u}}_3)].$$

$$\text{(A2)}$$

This set of equations can be put into matrix form as

$$\mathbf{AB}=\mathbf{P}x_6+\mathbf{Q}, \tag{A3}$$

where $\mathbf{A}$ is a 20×16 constant matrix, and $\mathbf{B}$ is a 16×1 matrix with the following variables as its elements:

$$\mathbf{B}^\perp=[c_4c_5x_6,s_4c_5x_6,c_5x_6,c_4s_5x_6,s_4s_5x_6,s_5x_6,c_4x_6,$$

$$s_4x_6,c_4c_5,s_4c_5,c_5,c_4s_5,s_4s_5,s_5,c_4,s_4]. \tag{A4}$$

Here $c_i=\cos\phi_i$ and $s_i=\sin\phi_i$. The 16×1 matrices $\mathbf{P}$ and $\mathbf{Q}$ have elements that are linear in the sines and cosines of ϕ_1 and ϕ_2. Sixteen equations are chosen from the twenty equations in Eq. (A2) and used to express $\mathbf{B}$ as functions of ϕ_1, ϕ_2, and x_6. This is accomplished by defining $\mathbf{A}_s$ as the corresponding 16×16 submatrix of $\mathbf{A}$ and multiplying both sides of these sixteen equations by the inverse of A_s to obtain

$$\mathbf{B}=\mathbf{A}_s^{-1}\mathbf{P}_sx_6+\mathbf{A}_s^{-1}\mathbf{Q}_s. \tag{A5}$$

Here $\mathbf{P}_s$ and $\mathbf{Q}_s$ are the corresponding submatrices of $\mathbf{P}$ and $\mathbf{Q}$. We find that special geometries, such as vanishing lengths of links, may render some choices for $\mathbf{A}_s$ singular. The linear dependence is identified by singular value decomposition[19] and avoided by choosing 16 linearly independent equations. The expression for $\mathbf{B}$ is inserted into the other four unused equations to yield four equations that are linear functions of x_6 and the sines and cosines of ϕ_1 and ϕ_2. Replacing the sines and cosines of ϕ_1 and ϕ_2 with $x_1\equiv\tan\phi_1/2$ and $x_2\equiv\tan\phi_2/2$, these four equations can be expressed as

$$(d_ix_2^2+e_ix_2+f_i)x_6+(g_ix_2^2+h_ix_2+p_i)=0, \quad i=1,\ldots,4, \tag{A6}$$

where $d_i,e_i,\ldots,p_i$ are quadratic in x_1. Multiplying Eq. (A6) by x_2 gives four additional equations, which together with Eq. (A6) can be used to set up a linear set of equations:

$$\begin{bmatrix} 0 & 0 & d_1 & e_1 & f_1 & g_1 & h_1 & p_1 \\ 0 & 0 & d_2 & e_2 & f_2 & g_2 & h_2 & p_2 \\ 0 & 0 & d_3 & e_3 & f_3 & g_3 & h_3 & p_3 \\ 0 & 0 & d_4 & e_4 & f_4 & g_4 & h_4 & p_4 \\ d_1 & g_1 & e_1 & f_1 & 0 & h_1 & p_1 & 0 \\ d_2 & g_2 & e_2 & f_2 & 0 & h_2 & p_2 & 0 \\ d_3 & g_3 & e_3 & f_3 & 0 & h_3 & p_3 & 0 \\ d_4 & g_4 & e_4 & f_4 & 0 & h_4 & p_4 & 0 \end{bmatrix} \begin{bmatrix} x_2^3x_6 \\ x_2^3 \\ x_2^2x_6 \\ x_2x_6 \\ x_6 \\ x_2^2 \\ x_2 \\ 1 \end{bmatrix}=\mathbf{0}. \tag{A7}$$

Here $\mathbf{0}$ is an 8×1 null vector. Solutions are found by equating to zero the corresponding determinant of the 8×8 matrix, which can be expanded to a polynomial of degree sixteen in x_1. The determinant equation is best solved by reformulating it as an eigenvalue problem. This is achieved by rewriting Eq. (A7) as

$$|\mathbf{A}_0+\mathbf{A}_1x_1+\mathbf{A}_2x_1^2|=0, \tag{A8}$$

where $\mathbf{A}_0$, $\mathbf{A}_1$, and $\mathbf{A}_2$ are 8×8 numerical matrices. Note that in the crankshaft case, $\mathbf{A}_0$, $\mathbf{A}_1$, and $\mathbf{A}_2$ are identically zero, and so any solution for x_1 is possible.[18] The roots of the determinant equation are the eigenvalues of the matrix[16]

$$\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{A}_2^{-1}\mathbf{A}_0 & -\mathbf{A}_2^{-1}\mathbf{A}_1 \end{bmatrix}, \tag{A9}$$

where $\mathbf{0}$ and $\mathbf{I}$ are 8×8 null and identity matrices, respectively. The matrix is first reduced to an upper Hessenberg matrix, and then the eigenvalues are found by the QR algorithm[19] to obtain x_1. Occasionally the matrix $\mathbf{A}_2$ may be almost singular. This occurs when one of the roots is $\phi_1\approx\pm\pi$. This singularity can be avoided by the transformation

$$x_1=\frac{t_1\bar{x}_1+t_2}{t_3\bar{x}_1+t_4}, \tag{A10}$$

where t_1, t_2, t_3, and t_4 are random numbers uniformly distributed in $[-1,1]$. Solutions for x_1 are substituted into Eq. (A7) to calculate x_2 and x_6. These values are substituted back to Eq. (A5) to calculate ϕ_4 and ϕ_5. The angle ϕ_3 is not needed for our purpose; it can be determined by calculating the positions of $\hat{\mathbf{a}}_2$ and $\hat{\mathbf{a}}_3$ from the ottorsional angles.

[1] S. Jackson, W. Degrado, A. Dwivedi, A. Parthasarathy, A. Higley, J. Krywko, A. Rockwell, J. Markwalder, G. Wells, R. Wexler, S. Mousa, and R. Harlow, J. Am. Chem. Soc. **116**, 3220 (1994).

[2] C. Bisang, L. Jiang, E. Freund, F. Emery, C. Bauch, H. Matile, G. Pluschke, and J. A. Robinson, J. Am. Chem. Soc. **120**, 7439 (1998).

[3] R. Haubner, R. Gratias, B. Diefenbach, S. L. Goodman, A. Jonczyk, and H. Kessler, J. Am. Chem. Soc. **118**, 7461 (1996).

[4] Z. Huang, A. Pröbstl, J. R. Spencer, T. Yamazaki, and M. Goodman, Int. J. Pept. Protein Res. **42**, 352 (1993).

[5] K. T. O'Neil, R. H. Hoess, S. A. Jackson, N. Ramachandran, S. A. Mousa, and W. F. DeGrado, Proteins: Struct., Funct., Genet. **14**, 509 (1992).

[6] L. Stryer, *Biochemistry*, 3rd ed. (W. H. Freeman, New York, 1988).

[7] K. A. Higgins, D. J. Craik, J. G. Hall, and P. R. Andrews, Drug Design Deliv. **3**, 159 (1988).

[8] H. Weißhoff, T. Wieprecht, P. Henklein, C. Frömmel, C. Antz, and C. Mügge, Fed. Euro. Bioc. Soc. Lett. **387**, 201 (1996).

[9] R. L. Stein, Adv. Quantum Chem. **11**, 1 (1993).

[10] G. Scherer, M. L. Kramer, M. Schutkowski, and G. Fischer, J. Am. Chem. Soc. **120**, 5568 (1998).

[11] F. X. Schmid, L. M. Mayr, M. Mueche, and E. R. Schoenbrunner, Adv. Quantum Chem. **11**, 25 (1993).

[12] L. R. Dodd, T. D. Boone, and D. N. Theodorou, Mol. Phys. **78**, 961 (1993).

[13] D. Frenkel and B. Smit, Mol. Phys. **75**, 983 (1992).

[14] J. J. de Pablo, M. Laso, and U. W. Suter, J. Chem. Phys. **96**, 6157 (1992).

[15] M. G. Wu and M. W. Deem, Mol. Phys. **97**, 559 (1999).

[16] D. Manocha and J. F. Canny, IEEE Trans. Rob. Autom. **10**, 648 (1994).

[17] M. W. Deem and J. S. Bader, Mol. Phys. **87**, 1245 (1996).

[18] The analytical rebridging equations are singular for the geometry which allows for an infinite number of ''crankshaft-type'' solutions. This geometry occurs with probability zero in the general case.

[19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge, New York, 1992).

[20] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, San Diego, 1996).

[21] C. J. Geyer, *Computing Science and Statistics* (American Statistical Association, New York, 1991), pp. 156–163.

[22] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).

[23] E. Marinari, G. Parisi, and J. Ruiz-Lorenzo, in *Spin Glasses and Random Fields*, Vol. 12 of *Directions in Condensed Matter Physics*, edited by A. Young (World Scientific, Singapore, 1998).

[24] M. C. Tesi, E. J. Janse van Rensburg, E. Orlandini, and S. G. Whittington, J. Stat. Phys. **82**, 155 (1996).

[25] G. Boyd, Nucl. Phys. B (Proc. Suppl.) **60A**, 341 (1998).

[26] M. Falcioni and M. W. Deem, J. Chem. Phys. **110**, 1754 (1999).

[27] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, J. Comput. Chem. **7**, 230 (1986).

[28] P. E. Smith, B. M. Pettitt, and M. Karplus, J. Phys. Chem. **97**, 6907 (1993).

[29] D. F. Veber, in *Peptides, Synthesis, Structure and Function, Proceedings of the Seventh American Peptide Symposium*, edited by D. H. Rich and V. J. Gross (Pierce Chemical Co., Rockford, IL, 1981), p. 685.

[30] H. Kessler, M. Klein, A. Müller, K. Wagner, J. W. Bats, K. Ziegler, and M. Frimmer, Angew. Chem. Int. Ed. Engl. **25**, 997 (1986).

[31] G. Zanotti, M. Saviano, G. Saviano, T. Tancredi, F. Rossi, C. Pedone, and E. Benedetti, J. Peptide Research **51**, 460 (1998).

[32] I. McEwen, Biopolymers **33**, 693 (1993).

[33] G. Müller, M. Gurrath, M. Kurz, and M. Kessler, Proteins: Struct., Funct., Genet. **15**, 235 (1993).

[34] E. Pinet, J. M. Neumann, I. Dahse, G. Girault, and F. André, Biopolymers **36**, 135 (1995).

[35] *CRC Handbook of Chemistry and Physics*, 68th ed., edited by R. C. Weast, M. J. Astle, and W. H. Beyer (CRC, Boca Raton, FL, 1987), p. E-50.

[36] *The Merck Index*, 10th ed., edited by M. Windholz, S. Budavari, R. F. Blumetti, and E. S. Otterbein (Merck & Co. Inc., New Jersey, 1983), p. 475.

[37] H. S. Ashbaugh and M. E. Paulaitis, J. Phys. Chem. B **102**, 5029 (1998).

[38] M. E. Paulaitis, Curr. Opin. Colloid Interface Sci. **2**, 315 (1997).

[39] B. Smit, S. Karaborni, and J. Siepmann, J. Chem. Phys. **102**, 2126 (1995).

[40] S. K. Nath, F. A. Escobedo, and J. J. de Pablo, J. Chem. Phys. **108**, 9905 (1998).

[41] E. J. F. Primrose, Mech. Mach. Theory **21**, 509 (1986).

[42] H.-Y. Lee and C.-G. Liang, Mech. Mach. Theory **23**, 209 (1988).

[43] H.-Y. Lee and C.-G. Liang, Mech. Mach. Theory **23**, 219 (1988).

[44] M. Raghavan and B. Roth, Trans. ASME, J. Mech. Design **115**, 502 (1993).