Using Internal and Collective Variables in Monte Carlo Simulations of Nucleic Acid Structures: Chain Breakage/Closure Algorithm and Associated Jacobians

HEINZ SKLENAR,[1] DANIEL WÜSTNER,[1] REMO ROHS[2]
[1]*Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Str. 10, 13092 Berlin, Germany*
[2]*Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel*

Received 5 July 2005; Accepted 9 September 2005
DOI 10.1002/jcc.20345
Published online in Wiley InterScience (www.interscience.wiley.com).

Abstract: This article describes a method for solving the geometric closure problem for simplified models of nucleic acid structures by using the constant bond lengths approximation. The resulting chain breakage/closure equations, formulated in the space of variable torsion and bond angles, are easy to solve, and have only two solutions. The analytical simplicity is in contrast with the high complexity of the closure problem in the torsion angle space with at most 16 solutions, which has been dealt with by several authors and was solved analytically by Wu and Deem (J. Chem. Phys. 1999, 111, 6625). The discussion on the choice of variables and associated Jacobians is focussed on the question of how conformational equilibration is affected in Monte Carlo simulations of molecular systems. In addition to the closure of the phosphate backbone, it is necessary to also solve the closure problem for the five-membered flexible furanose sugar ring. Explicit closure equations and the resulting Jacobians are given both for the complete four-variable model of the furanose ring and simulations in the phase-amplitude space of the five-membered ring, which are based on the approximate two-variable model of furanose introduced by Gabb et al. (J. Comput. Chem. 1995, 16, 667). The suggested closure algorithm can be combined with collective variables defined by translations and rotations of the monomeric nucleotide units. In comparison with simple internal coordinate moves, the resulting concerted moves describe local structural changes that have high acceptance rates and enable fast conformational equilibration. Appropriate molecular models are put forward for prospective Monte Carlo simulations of nucleic acids, but can be easily adapted to other biomolecular systems, such as proteins and lipid structures in biological membranes.

 J Comput Chem 27: 309–315, 2006

Key words: Monte Carlo simulations; nucleic acid structures; Jacobians

Introduction

Simplified molecular models with a reduced number of degrees of freedom can be derived from internal coordinate descriptions that are adapted to the energetic hierarchy of conformational changes. The stiffness of chemical bonds, in comparison with the much higher flexibility related to rotations about single bonds, justify the constant bond length approximation (CBLA) for describing molecular conformations. In contrast with this approximation, the additional use of fixed bond angles must be considered more cautiously. In the case of molecular rings, endocyclic and exocyclic bond angles have to be distinguished because using fixed endocyclic bond angles in rings with less than seven ring atoms would lead to a complete loss of internal flexibility. Moreover, it was observed that already small changes of bond angles could considerably affect the energy landscape. It is therefore not surprising that variation of both dihedral angles and bond angles leads to significantly improved sampling for polypeptides including the reduction of energy barriers between alternate conformers.[1] This is another reason why the use of molecular models with constant bond lengths and fixed bond angles appears to be problematic. Such models generally require to modify the force fields that were developed for simulations in the Cartesian space.[2,3] The work presented in this article is therefore focused on molecular models using only constant bond lengths as constraints.

Correspondence to: H. Sklenar; e-mail: sklenar@mdc-berlin.de

Contract/grant sponsor: the Minerva Foundation (Max Planck Society for the Advancement of Science) through a Minerva Postdoctoral Fellowship (to R.R.)

Contract/grant sponsor: the Max Delbrück Center for Molecular Medicine through MDC Postdoctoral Fellowship (to D.W.)

The advantage of using the CBLA in force field-based molecular simulations is not only due to the reduction of degrees of freedom, but results, in addition and more importantly, from a considerable softening of the energetic landscape. In this way, particularly Monte Carlo (MC) simulations can profit from high acceptance rates of large conformational changes leading to much faster equilibration in comparison to simulations in Cartesian space. For Molecular Dynamics simulations, this advantage is not as clear because of the complex dynamic equations that need to be solved. Implementations of the CBLA in MC simulations, however, face problems of ring and chain closure, and require the calculation of associated Jacobians. The chain closure problem in polymers results from the need of local chain breaking to avoid energetically unfavorable structures, if angular moves would be applied to the full chain. The concerted rotations (CONROT)[4] and the symmetric rebridging[5] algorithms offer solutions to this problem in the torsion angle space (i.e., with constant bond lengths and fixed bond angles) and have been applied to simulations of aliphatic polymers. Similar techniques, such as window moves,[2] an analytical rebridging scheme,[6] and an extension of CONROT that includes bond angles as degrees of freedom,[1] were also used in MC simulations of polypeptide structures. These algorithms have in common that complex chain closure equations have to be solved, with at most 16 solutions in the general case. Hence, computationally demanding numerical procedures are involved. The Newton–Raphson algorithm is the mostly simple variant, but suffers from the fact that all solutions are not readily provided. It should also be noticed that the algorithms suggested in the given references could not be applied to the more complex closure problems in nucleic acid structures where, in addition to the chain closure, the closure of the five-membered furanose ring must be maintained.

In energy minimization algorithms for nucleic acid structures[7] and molecular rings,[8] the geometrical solution of the closure problem has been circumvented by using pseudoenergetic penalty terms. It is clear that this approach cannot be used in MC simulations where rings and chain have to be kept closed in each step. For this purpose, but still in the spirit of using penalty terms to avoid solving the closure problem of the furanose ring, an overdetermined five-variable model of the furanose ring has been suggested.[9] This model is based on the pseudorotation description in the phase-amplitude space, and was obtained by fitting equations for five variables (two torsion and three bond angles) to the results of adiabatic mapping in dependence of phase and amplitude, where an energy term is used for ring closure. Apart from avoiding ring closure, conformational changes of the sugar ring are described by only two parameters (phase and amplitude) instead of using four variables that are necessary for a complete description, with some advantage for large-scale simulations of nucleic acid structures. The model, however, is biased by the force field used in adiabatic mapping, and both ring closure and model fitting only approximately reproduce the target values of phase and amplitude. Moreover, using this model in MC simulations also requires a Jacobian factor to be included in the Metropolis acceptance criterion that has been neglected in a first approach of internal coordinate MC simulations to DNA oligomers.[10] In this former work, the efficiency of MC moves was improved by applying the scaled collective variable (SCV) technique developed by Noguti and Go,[11] without using the chain breakage-closure formalism suggested for the treatment of simple polymers.

In the following section, the closure problem is formulated in the torsion and bond angle space, where two molecular fragments have to be geometrically connected under the condition that two chemical bonds of given lengths to a common bridging atom are formed. A simple solution is obtained for the special case of using a torsion angle in one fragment as an independent variable for moving the bridging atom by its remaining third degree of freedom. The resulting chain breakage/closure (CBC) equations for closure and associated Jacobians can be applied to the closure of any type of nonbranched molecular chains or rings. In particular, it will be shown that the closure algorithm can be considered as an appropriate tool for MC simulations of nucleic acid structures. With that perspective, the design of a new scheme for collective variable simulations of these polymers will be discussed. In this approach, translations and rotations of the monomeric units serve as independent variables, which describe local conformational changes and replace the internal variables resulting from chain closure.

Methodology

CBC Algorithm

The geometrical closure problem in the torsion and bond angle space (i.e., by using the CBLA) is shown in Figure 1. Throughout this article the following notations are used: The position of atom i is described by the vector $\mathbf{r}_i$ and the distance vector from atom i to atom j is denoted by $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ with the lengths $r_{ij} = |\mathbf{r}_{ij}|$. Bond angles between bonds (i–j) and (j–k) are defined by $\omega_j = \mathrm{acos}(\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}/(r_{ij}r_{rk}))$. With the bond rotation axes $\mathbf{u}_j = (\mathbf{r}_{ij} \times \mathbf{r}_{jk})/(r_{ij}r_{rk}\sin(\omega_j))$, the dihedral (torsion) angles (i–j–k–l) are defined by $\tau_{jk} = \mathrm{acos}(\mathbf{u}_j \cdot \mathbf{u}_k)$.

The number of atoms in the bridging unit, which are necessary for the closure of two molecular fragments, depends on the independent variables of the molecular model. Varying the coordinates of m atoms in a chain of covalently bound atoms involves variation of $m + 1$ bond lengths, $m + 2$ bond angles, and $m + 3$ torsion angles of the chain, where six degrees of freedom are related to the relative position and orientation of the two fragments. This relationship yields the condition $m + 3 \geq 6$, if both bond lengths and bond angles are constrained, and $2m + 5 \geq 6$, if only bond lengths constraints are used. Therefore, in the former case the atoms of a trimer have to repositioned,[4,5] whereas the new position of only one atom must be found if variable bond angles are allowed. In the latter case, 2 degrees of freedom of this atom (number 4 in Fig. 1) define a closure plane where at most two solutions are located, which fulfill the constraints on the bond distances r_{34} and r_{45}. The third degree of freedom remains an independent variable. In principle, one can choose these variables arbitrary, but that choice has some consequences for the associated Jacobians. In this section, as shown in Figure 1, the torsion angle τ_{23} is chosen as the independent variable that defines the plane of atoms 2, 3, and 4 as the closure plane. For the sake of clarity, it is convenient to use a rectangular axis system with $\mathbf{e}_z$ equal to the normal $\mathbf{u}_3$ of the

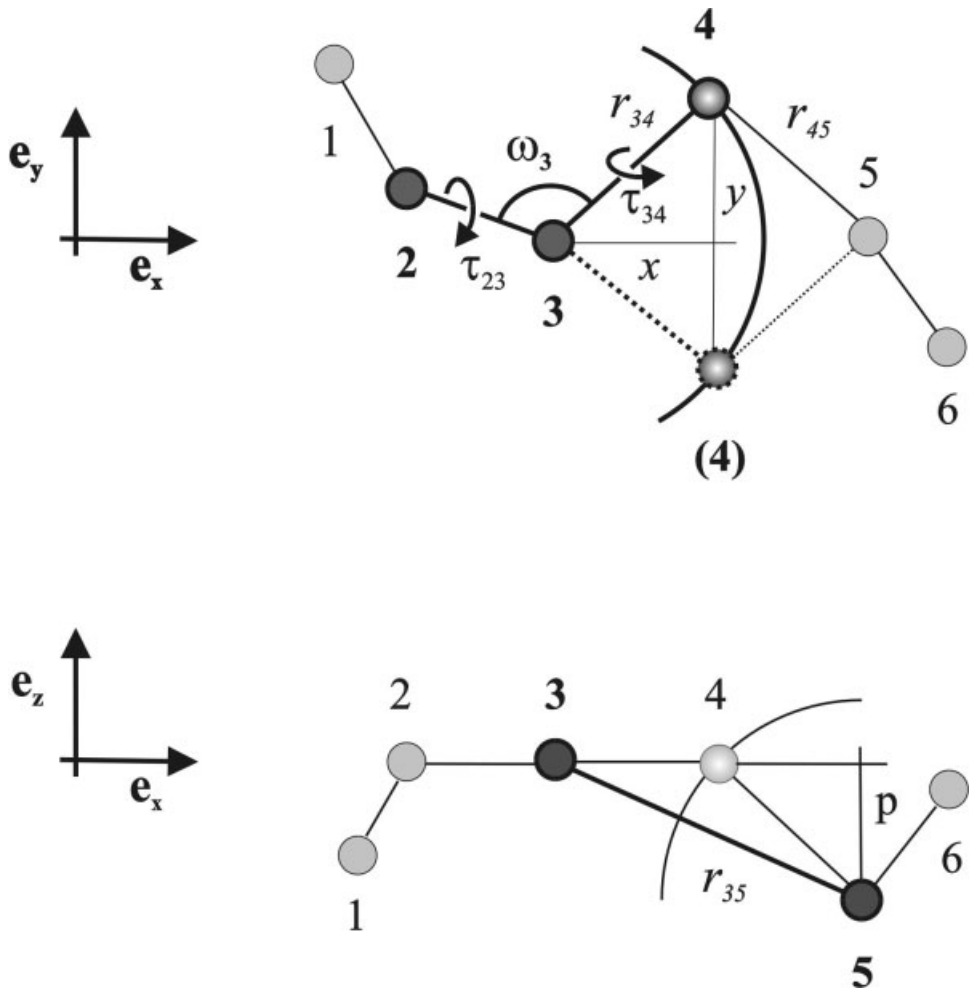

Figure 1. The closure problem is shown in two perpendicular views, with dark atoms lying in the drawing plane. Atoms 1, 2, 3, 5, and 6 are fixed in space, and atom 4 is repositioned to fulfill the given bond distances r_{34} and r_{45}. The torsion angle τ_{23} is an independent variable, and the bond angle ω_3 is used for the closure. One of the two solutions for atom 4, shown in the upper panel, is plotted with dotted lines.

closure plane, and $\mathbf{e}_x$ being oriented in the direction of the projected distance vector $(1 - \mathbf{e}_z \circ \boldsymbol{e}_z) \cdot \mathbf{r}_{35}$:

$$\mathbf{e}_z = \mathbf{u}_3, \quad \mathbf{e}_x = \frac{(\mathbf{r}_{35} - p\mathbf{e}_z)}{|\mathbf{r}_{35} - p\mathbf{e}_z|}, \quad \mathbf{e}_y = \mathbf{e}_z \times \mathbf{e}_x, \tag{1}$$

where $p = \mathbf{e}_z \cdot \mathbf{r}_{35}$. Using the notations of Figure 1, simple geometric calculus yields the two solutions for the position of atom 4 by adding the respective difference vectors $\mathbf{r}_{34}$ to $\mathbf{r}_3$:

$$\mathbf{r}_4 = \mathbf{r}_3 + x\mathbf{e}_x \pm y\mathbf{e}_y$$

$$x = \frac{\mathbf{r}_{34}^2 + \mathbf{r}_{35}^2 - \mathbf{r}_{45}^2}{2\sqrt{\mathbf{r}_{35}^2 - p^2}}, \quad y = \sqrt{\mathbf{r}_{34}^2 - x^2}. \tag{2}$$

The closure problem is solved if $r_{34}^2 \geq x^2$. Otherwise, no viable solution exists. It should be noted that under special geometric conditions, if $p^2 = r_{35}^2$, an infinite number of solutions is obtained. This case is numerically rare, and can be rejected in MC simulations. For an existing structure, with bond lengths r_{34} and r_{45}, the closure condition is always fulfilled, and the position of atom 4 is exactly reproduced by one solution given by eq. (2).

It should be noticed that any closure of two molecular fragments involves six dependent internal variables due to three translations and three rotations of a rigid body in space. In the closure model described above, the dependent variables are the torsion angles τ_{34}, τ_{45}, and τ_{56}, and the bond angles ω_3, ω_4, and ω_5, while six dependent torsion angles are obtained by closure in the torsion space. This means that, instead of using some internal variables as driving variables in MC sampling,[2,4,5] one can also directly use six rigid body variables as independent variables, namely three translations and three rotations of given fragments. In the case of biopolymers, a natural choice of such fragments is to use the monomeric units. In addition to their position and orientation in

space, flexible elements of the fragments can be described by internal variables. The respective molecular model for MC simulations of nucleic acid structures will be discussed below.

Associated Jacobians and Metropolis Acceptance Criterion

Using generalized coordinates for describing the configuration space of a molecular system requires a transformation of the volume element into the configuration integral of the partition function. A subset of the new coordinates is usually considered as constraints. In general form, the transformation of the volume element from the canonical Cartesian space into a new set of transformed coordinates reads

$$d\Gamma = \prod_{i=1,N} dx_i dy_i dz_i = J \prod_{i=1,M} dq_i \prod_{j=1,3N-M} dc_j, \tag{3}$$

where $\{q_i\}$ is the subset of independent generalized coordinates and $\{c_j\}$ are the constrained variables. The Jacobian J is the functional determinant of the transformation and is most conveniently calculated as the inverse

$$J^{-1} = \left| \frac{\partial(q_1,\cdots,q_M,c_1,\cdots,c_{3N-M})}{\partial(\mathbf{r}_1,\cdots,\mathbf{r}_N)} \right|. \tag{4}$$

In the case of uniform MC sampling in the $\{q_i\}$ space, as it is common practice, the Jacobian has to be included in the Metropolis acceptance criterion. In the following, respective Jacobians will be calculated by starting from the general expression given by eq. (4). Using the notations of Figure 1, the inverse of the Jacobian J_c, associated with chain closure reads

$$J_c^{-1} = \left| \frac{\partial(\tau_{23},r_{34},r_{45})}{\partial(\mathbf{r}_4)} \right| = \left| \frac{\partial\tau_{23}}{\partial\mathbf{r}_4} \cdot \left(\frac{\partial r_{34}}{\partial\mathbf{r}_4} \times \frac{\partial r_{45}}{\partial\mathbf{r}_4} \right) \right|. \tag{5}$$

The Jacobian depends on the selected solution l (l = 1,2), because the position of atom 4 is different for the two solutions. Inserting the formulas for the derivatives of the torsion angle τ_{jk} and of the bond length r_{ij},

$$\frac{\partial\tau_{jk}}{\partial\mathbf{r}_l} = \frac{\mathbf{r}_{jk} \times \mathbf{r}_{kl}}{r_{jk} r_{kl}^2 \sin^2\omega_k}, \quad \text{and} \quad \frac{\partial r_{ij}}{\partial\mathbf{r}_j} = \mathbf{r}_{ij}/\mathbf{r}_{ij},$$

into eq. (5) yields

$$J_c^{-1} = \frac{\cos\tau_{34} \sin\omega_4}{r_{34} \sin\omega_3}. \tag{6}$$

Note that both $\sin\omega_3$ and $\sin\omega_4$ are different for the two solutions, whereas $\cos\tau_{34}$ changes only its sign but J_c becomes singular for $\tau_{34} = = \pm\pi/2$, as mentioned above.

The acceptance rate $P(m \rightarrow n)$ of the MC Metropolis algorithm[12] is given by

$$P(m \rightarrow n) = \min\left(1, \frac{J_n(l)\exp(-E_n/RT)}{J_m(l)\exp(-E_m/RT)}\right). \tag{7}$$

Here, the probability model $w(k,l)$, chosen for the selection of one of the two solutions, is included in the factors $J_k(l) = J_c/w(k,l)$, and E_k is the energy calculated for the structure with the respective solution in state k. This renders the move microscopically reversible with any chosen $w(k,l)$. An appropriate choice of $w(k,l)$ is based on weighting the solutions with the Boltzmann factor of the angular energy $E_{\text{ang}}(k,l)$[5] that yields

$$w(k,l) = \frac{\exp(-E_{\text{ang}}(k,l)/RT)}{\exp(-E_{\text{ang}}(k,1)/RT) + \exp(-E_{\text{ang}}(k,2)/RT)}. \tag{8}$$

Four-Variable Model of the Furanose Ring

Using the CBLA, the conformation of a five-membered ring is described by four independent variables. In the molecular model shown in Figure 2, the ring conformation is completely defined by the torsion angles τ_{12} and τ_{23}, the bond angles ω_1 and ω_2, and the five bond lengths constraints. These 9 degrees of freedom correspond to 15 Cartesians of the ring atoms minus 6 degrees of freedom for overall translations and rotations. Exocyclic bond angles can be treated in a simplified manner.[9] Bond angles involving two exocyclic atoms are held fixed, and the others are coupled with the corresponding endocyclic angle to maintain a symmetric position with respect to the bisector of this angle. Conformational moves by the four independent variables temporarily break the bond between atoms 4 and 5, and therefore, have to be combined with a ring-closure algorithm that restores the target bond length d_{45} and defines the remaining six dependent angular variables. The numbers of equivalent atoms are chosen identically in Figures 1 and 2, so that eq. (2) can be used for ring closure without changing notations. The only difference is that atoms 1 and 6 in Figure 1 are now the same atom and correspond to atom 1 in Figure 2.

To remove overall translations and rotations, without loss of generality, atom 1 is put to the origin of a Cartesian coordinate system, and the moves of atoms 2 and 5 are restrained to a fixed, say (x,y), plane. In addition, to remove rotations in this plane, the symmetry constraint $c_{\text{sym}} = x_2y_5 + x_5y_2 = 0$ is used. This convention is most convenient and has already been used in previous work.[7–10] With that choice of constraints, the complete Jacobian of the four-variable model reads

$$J^{-1}(\omega_1,\tau_{12},\omega_2,\tau_{23}) = \left| \frac{\partial(\omega_1,\tau_{12},\omega_2,\tau_{23},r_{12},r_{23},r_{34},r_{45},r_{51},c_{\text{sym}})}{\partial(x_5,y_5,x_2,y_2,\mathbf{r}_3,\mathbf{r}_4)} \right|. \tag{9}$$

This 10×10 determinant can be written as a product of a 4×4 determinant D_{52} and two 3×3 determinants, D_3 and D_4:

$$J^{-1} = |D_{52} \cdot D_3 \cdot D_4|. \tag{10}$$

With the formulas for the derivatives of the torsion angle and the bond length (see above), and the derivatives of the bond angle ω_k,

$$\frac{\partial\omega_k}{\partial\mathbf{r}_l} = (\mathbf{r}_{kl} \times \mathbf{u}_k)/r_{kl}^2,$$

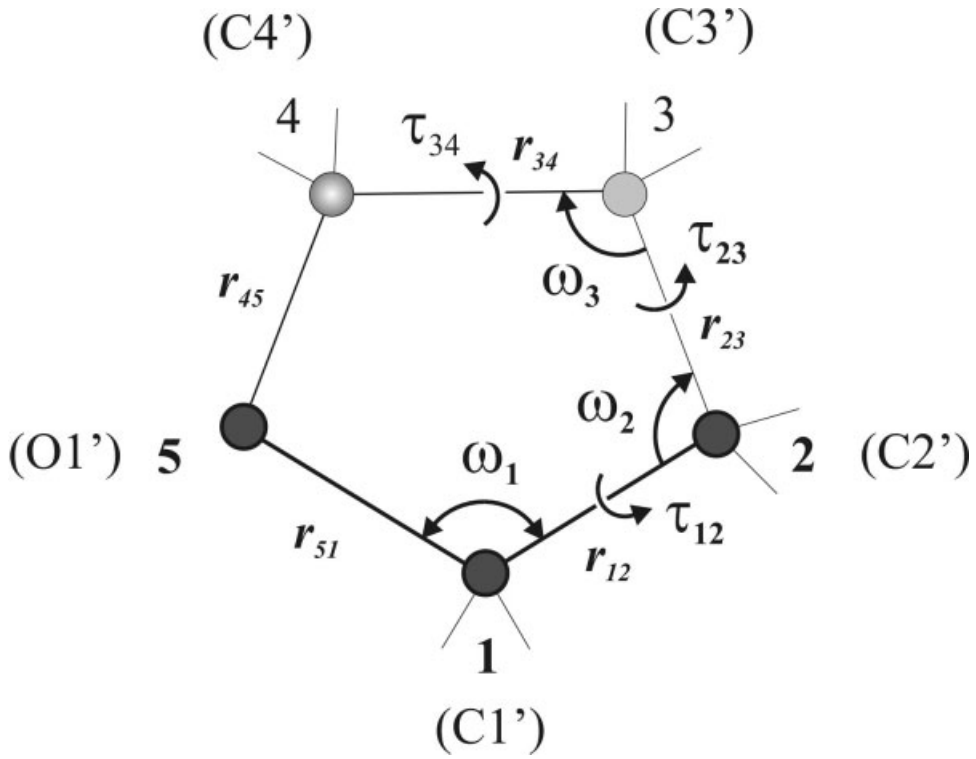

Figure 2. View of the five-membered furanose ring. With constant bond lengths, the structure of this ring is completely defined by the angles ω_1, τ_{12} ω_2, and τ_{23}. The angles ω_3 and τ_{34} are dependent variables. The atoms and bonds lying in the (x, y) plane are shown in dark.

one obtains

$$D_{52} = \frac{\partial(\omega_1, r_{12}, r_{51}, c_{\text{sym}})}{\partial(x_2, y_2, x_5, y_5)} = \begin{vmatrix} y_2 r_{51}^2 & -x_2 r_{51}^2 & -y_5 r_{12}^2 & x_5 r_{12}^2 \\ x_2 & y_2 & 0 & 0 \\ 0 & 0 & x_5 & y_5 \\ y_5 & x_5 & y_2 & x_2 \end{vmatrix} / (r_{12}^3 r_{51}^3) = 2,$$

$$D_3 = \frac{\partial(\tau_{12}, \omega_2, r_{23})}{\partial(\mathbf{r}_3)} = \frac{1}{r_{23}^2 \sin \omega_2}, \quad \text{and } D_4 = \frac{\partial(\tau_{23}, r_{34}, r_{45})}{\partial(\mathbf{r}_4)} = \frac{\cos\tau_{34} \sin \omega_4}{r_{34} \sin \omega_3}.$$

The determinant D_4 results from ring closure [see eq. (6)], that is, the torsion angle τ_{34} and the bond angle ω_3 are dependent variables and have to be calculated after closure. The determinant D_{52} and also the Jacobians, related to the repositioning of exocyclic atoms with constant exocyclic bond angles (see above), are constant. Because the acceptance rate (7) depends on the ratio of Jacobians in state n and m, all constant terms can be removed. Hence, the relative Jacobians for MC moves of the bond angle ω_1 and of both torsion angles are

$$J = \sin \omega_3 / (|\cos \tau_{34}| \sin \omega_4), \qquad (10')$$

and the relative Jacobian for moving the bond angle ω_2 reads

$$J = \sin \omega_2 \sin \omega_3 / (|\cos \tau_{34}| \sin \omega_4). \qquad (10'')$$

Two-Variable Model of the Furanose Ring

The approximate two-variable model of the furanose ring[9] is based on five equations that define the angles τ_{12}, τ_{23}, ω_1, ω_2, and ω_3 as functions of phase P and amplitude A:

$$\tau_{ij} = g_i(P,A) \quad i = 1,2 \quad j = i + 1 \qquad (11)$$

$$\omega_k = f_k(P,A) \quad k = 1,2,3.$$

This model reduces the number of independent variables from four to two. In the original version, an overdetermined description by five dependent variables is used, which allows for removing the bond length constraint r_{45} to avoid the ring-closure problem. One can, however, easily recover the description in the space of four dependent variables, if the two-variable model is combined with the ring closure algorithm described above. Such reformulation does not require demanding computations and has the advantage that the bond length constraint r_{45} is exactly fulfilled (see below). Using the two-variable model in MC simulations, with uniform sampling in the (P, A) space, requires a Jacobian that is more complex than in the four-variable model and has been neglected in

earlier applications.[9,10] In the following derivations, the laws for partial derivatives of functions of two variables are used. In particular, one should observe that these independent variables could be written as functions of different pairs of variables because P and A are overdetermined by eq. (11). Therefore, the partial derivative with respect to a given variable depends on the choice of the second variable that is held constant [see eq. (17) below].

In its general form, the Jacobian of the two-variable model reads

$$J^{-1}(P,A) = \left|\frac{\partial(r_{51},r_{12},c_{\mathrm{sym}},r_{23},r_{34},P,A,c_1,c_2,c_3)}{\partial(x_5,y_5,x_2,y_2,\mathbf{r}_3,\mathbf{r}_4)}\right|. \qquad (12)$$

Without loss of generality, eq. (11) for τ_{12} and τ_{23} are used to implicitly define $P(\tau_{12},\tau_{23})$ and $A(\tau_{12},\tau_{23})$, and the equations for ω_i serve as constraints defined by

$$c_i = f_i(P,A) - \omega_i = 0 \quad i = 1,2,3. \qquad (13)$$

By transforming the Cartesians at first into the five-variable angle space and then from this angle space into the (P, A) space, the Jacobian (10) can be written as the product

$$J^{-1}(P,A) = \left|\frac{\partial(r_{51},r_{12},c_{\mathrm{sym}},r_{23},r_{34},\tau_{12},\tau_{23},\omega_1,\omega_2,\omega_3)}{\partial(x_5,y_5,x_2,y_2,\mathbf{r}_3,\mathbf{r}_4)}\right| \cdot \left|\frac{\partial(P,A,c_1,c_2,c_3)}{\partial(\tau_{12},\tau_{23},\omega_1,\omega_2,\omega_3)}\right|$$

$$= |D_{\mathrm{ang}} \cdot D_{\mathrm{PA}}|. \qquad (14)$$

The first factor corresponds to eq. (9), with the only difference that no ring closure is involved:

$$D_{\mathrm{ang}} = \frac{1}{r_{23}r_{34}\sin\omega_2\sin\omega_3}. \qquad (15)$$

The second determinant D_{PA} can be factorized and is expressed by the product of a 2×2 determinant D_1 and a 3×3 determinant D_2:

$$D_{\mathrm{PA}} = |D_1 \cdot D_2| \qquad (16)$$

where

$$D_1 = \begin{vmatrix} \frac{\partial P(\tau_1,\tau_2)}{\partial\tau_1} & \frac{\partial P(\tau_1,\tau_2)}{\partial\tau_2} \\ \frac{\partial A(\tau_1,\tau_2)}{\partial\tau_1} & \frac{\partial A(\tau_1,\tau_2)}{\partial\tau_2} \end{vmatrix} = 1/\left(\frac{\partial g_1}{\partial P}\cdot\frac{\partial g_2}{\partial A} - \frac{\partial g_1}{\partial A}\cdot\frac{\partial g_2}{\partial P}\right),$$

and

$$D_2 = \det\left(\frac{\partial c_i}{\partial\omega_k}\right) \quad i,k = 1,2,3.$$

As one can expect, the inverse of D_1 is the Jacobian determinant of $g_1(P,A)$ and $g_2(P,A)$. This determinant vanishes if g_1 and g_2 are dependent functions. In the pseudorotation description of the five-membered ring, such dependency is related to $A = 0$. In this case, the Jacobian becomes equal to zero, because the flat ring structure is independent of P.

The elements of the second determinant are calculated by

$$\frac{\partial c_i}{\partial\omega_k} = \frac{\partial f_i}{\partial\omega_k} - \delta_{ik} = \frac{\partial f_i}{\partial P}\frac{\partial P(\omega_k,A)}{\partial\omega_k} + \frac{\partial f_i}{\partial A}\cdot\frac{\partial A(\omega_k,P)}{\partial\omega_k} - \delta_{ik}$$

$$= \frac{\partial f_i}{\partial P}/\frac{\partial f_k}{\partial P} + \frac{\partial f_i}{\partial A}/\frac{\partial f_k}{\partial A} - \delta_{ik} \qquad (17)$$

where the Kronecker symbol δ_{ik} is equal to one for $i = k$, and equal to zero elsewhere. The partial derivatives of c_i with respect to τ_{12} and τ_{23} do not affect the Jacobian determinant and need therefore not to be evaluated.

As mentioned above, a more rigorous implementation of the two-variable model results from replacing one of the constraints, say the equation for ω_3, by the bond lengths constraint d_{45}. In this case, the ring-closure algorithm determines the bond angle ω_3. To retain the concept of the two-variable model, the solution with the bond angle ω_3 has to be selected, which is closest to the constraint c_3 defined by eq. (13). With this modification, the Jacobian is only slightly changed. The first factor in eq. (14) becomes exactly the Jacobian that was obtained for the four-variable model,

$$D_{\mathrm{ang}} = \frac{\cos\tau_{34}\sin\omega_4}{r_{23}r_{34}\sin\omega_2\sin\omega_3}, \qquad (18)$$

and the second factor D_{PA} changes to the product of two 2×2 determinants.

Implementation of the CBC Algorithm into an MC Simulation Program for Nucleic Acid Structures

The currently implemented molecular model for nucleic acids is shown in Figure 3. It includes 14 independent chain and ring variables per nucleotide: six rigid body variables of the base, four sugar ring variables, three torsion angles (χ, ε, γ), and the bond angle at O3′. The P-O5′ and O1′–C4′ bonds are chosen for CBC, where the positions of atoms O5′ and C4′, respectively, are determined by the closure equations. Alternatively, the four sugar ring variables can be replaced by the sugar phase and amplitude resulting in 12 independent variables per nucleotide.[13] Associated Jacobians and a suitable probability model for the selection of one of the two closure solutions is included in the Metropolis acceptance criterion for MC moves.

Performance and results of this new MC approach have been demonstrated by applications to several DNA decamers with palindromic sequences and to a ligand–DNA complex.[13] Results for base-sequence effects on intrinsic DNA bending and flexibility of papillomavirus E2 protein–DNA binding sites were published recently.[14] In these applications the Cornell et al. AMBER force field has been used for energy calculations,[15] and solvent electrostatic effects were taken into account by a continuum model of the aqueous solvent[16] with explicit counterions for neutralizing the phosphate charges. Fast equilibration of counterions was found to

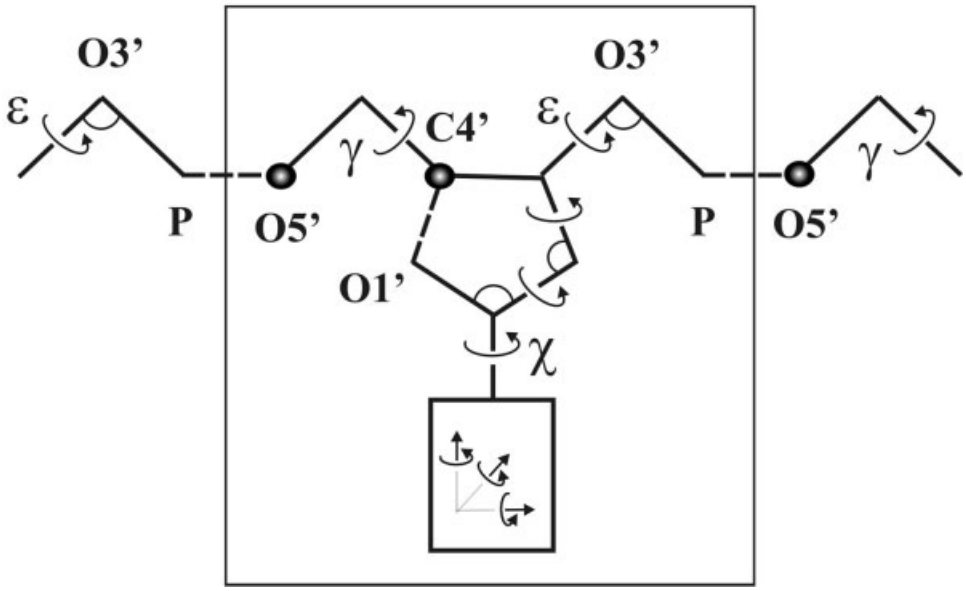

Figure 3. Molecular model of nucleic acid structures with 14 independent variables per nucleotide. Closure atoms are marked by spheres and closure bonds by dashed lines.

be important for observing frequent conformational transitions in the DNA oligomers. The MC simulations equilibrate within less than 10^6 MC cycles, and the results uniquely reflect the behavior of the molecular model under the force field used for energy calculations. Accordingly, the suggested MC simulation technique easily allows for exploring the effect of modified force fields and, in particular, of the approximations used for describing solute/solvent interactions on the results in comparison with available experimental data.

Conclusions

By using the CBLA and solving the CBC problem in the bond/torsion angle space, collective variables have been defined for nucleic acid structures that maintain structural moves entirely local and allow for large conformational changes in MC simulations. It is essentially this choice of independent variables that enables conformational equilibration on an acceptable CPU time scale, which is considered to be a necessary condition for deriving meaningful structural data from the trajectories of molecular simulations. It should be noted that the suggested CBC algorithm is quite general, so that it can be used in MC simulations of all types of chain molecules including molecular rings. For example, an adapted form of this algorithm has been successfully implemented to simulate phospholipids membranes.[17]

References

1. Ulmschneider, J. P.; Jorgensen, W. L. J Chem Phys 2003, 118, 4261.
2. Hoffmann, D.; Knapp, E. W. Eur Biophys J 1996, 24, 387.
3. Katritch, V.; Totrov, M.; Abagyan, R. J Comput Chem 2003, 24, 254.
4. Dodd, L. R.; Boone, T. D.; Theodorou, D. N. Mol Phys 1993, 78, 961.
5. Pant, P. V. K.; Theodorou, D. N. Macromolecules 1995, 28, 7224.
6. Wu, M. H. G.; Deem, M. W. J Chem Phys 1999, 111, 6625.
7. Lavery, R.; Zakrzewska, K.; Sklenar, H. Comput Phys Commun 1995, 91, 135.
8. Lavery, R.; Parker, I.; Kendrick, J. J Biomol Struct Dyn 1986, 4, 443.
9. Gabb, H. A.; Lavery, R.; Prevost, C. J Comput Chem 1995, 16, 667.
10. Gabb, H. A.; Prevost, C.; Bertucat, G.; Robert, C. H.; Lavery, R. J Comput Chem 1997, 18, 2001.
11. Noguti, T.; Go, N. Biopolymers 1985, 24, 527.
12. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. J Chem Phys 1953, 21, 1087.
13. Rohs, R. PhD Thesis; Freie Universität Berlin: http://www.diss.fu-berlin.de/2003/3/, 2002.
14. Rohs, R.; Sklenar, H.; Shakked, Z. Structure (Camb) 2005, 13, 1499.
15. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1996, 117, 5179.
16. Hingerty, B. E.; Ritchie, R. H.; Ferrell, T. L.; Turner, J. E. Biopolymers 1985, 24, 427.
17. Wüstner, D.; Sklenar, H. 2005, in preparation.