Efficient conformational space sampling for nucleosides using internal coordinate Monte Carlo simulations and a modified furanose description

H. Gabb, R. Lavery, Chantal Prévost

Efficient Conformational Space Sampling for Nucleosides Using Internal Coordinate Monte Carlo Simulations and a Modified Furanose Description

H. A. GABB and R. LAVERY*

Institut de Biologie Physico-Chimique, 13 Rue Pierre et Marie Curie, 75005 Paris, France

C. PRÉVOST

Centre de Biophysique Moléculaire, 1A Avenue de la Recherche Scientifique, 45071 Orléans, France

Received 22 August 1994; accepted 5 December 1994

ABSTRACT

Internal coordinates can be very helpful in modeling large biomacromolecules because freezing stiffer degrees of freedom, such as bond lengths, strongly reduces the number of variables describing the system. This, however, leads to difficulties in treating flexible rings such as the furanose sugars of nucleic acids or the proline residues of proteins, for which internal coordinates are an overcomplete description. We present here a new, internal coordinate furanose model based on the pseudorotational variables phase and amplitude which avoids having to solve a ring closure problem. The choice of a two- rather than a four-variable description is justified by a detailed analysis of molecular dynamic simulations. The efficiency and accuracy of the method are also demonstrated using extensive Monte Carlo simulations. This method of ring treatment is fast and well adapted to macromolecular simulations. © 1995 by John Wiley & Sons, Inc.

Introduction

Increasing attention has been given to modeling sugar rings in recent years. It has long been known that the key differences between the main nucleic acid families, A, B, and Z, reside in their sugar conformations.[1] Sugar repuckering can efficiently drive the B to A transition,[2,3] and sugars can often discriminate between the conformational substates of both double[3,4] and triple helices.[5] The furanose ring description is therefore particularly important for nucleoside and nucleic acid modeling.

With a constant bond length approximation, an N-atom ring presents $2N - 6$ internal degrees of

*Author to whom all correspondence should be addressed.

CCC 0192-8651 / 95 / 060667-14

freedom ($3N$ degrees of freedom for atomic motion, minus 6 for rotational/translational movement of the whole molecule, and minus another N to account for the bond constraints). The five-atom furanose ring therefore has four degrees of freedom. This can be taken into account in internal coordinate modeling by temporarily breaking one ring bond leading to a five-variable linear molecule (described by three valence angles and two torsions). The correct four-variable model is then generated by adding a distance constraint to reform the broken ring bond.[6] Furanose geometry is, however, commonly described by only two parameters: pseudorotation phase angle (P) and pucker amplitude (ν_m).[7] This is necessarily an approximation, and no general method exists for accurately obtaining ring atom positions from the pseudorotational parameters. Approximate solutions have been proposed by Merritt and Sundaralingam[8] and, for fixed amplitude $\nu_m = 39°$, by Pearlman and Kim.[9] However, the former method only leads to approximate ring closure and the latter fails to account for amplitude variations, which are important for puckers between the canonical $C_{2'}$-endo and $C_{3'}$-endo forms and also for a correct modeling of furanose flexibility.[10] Recently, an exact four-variable solution was presented by Marzec and Day,[11,12] but for the case of large nucleic acids, which may contain hundreds of sugar rings, it would clearly be advantageous to adopt a two-variable model if possible. This choice seems to be justified by the wide use of the classical pseudorotational parameters for describing sugar pucker, but it can also be justified quantitatively by an analysis of sugar ring dynamics, as we will show here. Based on this result, we have developed a rapid two-variable representation for use within internal coordinate modeling algorithms. The representation is tested for nucleosides using Monte Carlo (MC) simulations, which are also shown to be considerably more efficient than earlier molecular dynamics studies.[10] It is expected that this efficiency will be maintained in simulations of large macromolecular systems in which, in contrast to the situation for Cartesian variables,[13] internal coordinate models favor Monte Carlo techniques by allowing large sizes and reduced dimensions of conformational space.

Methods

As mentioned earlier, our basic internal coordinate furanose model (developed within the CINFLEX program[6]) consists of temporarily breaking the sugar ring, leading to a five independent variable system (two torsions, $\tau_{1,2}$, and three valence angles, θ_{1-3}), to which a distance constraint d is subsequently added (Fig. 1, Residue 2). The remaining variables of the ring are dependent. When this technique is used during energy minimization, ring closure is achieved by the distance constraint on the broken ring bond. It is clear that the same approach cannot be used in Monte Carlo simulations, where the ring remain closed at every step. In addition, as discussed earlier, it would be advantageous to use as few variables as possible to describe ring flexibility. Therefore, modifications to the original method are necessary.

Because recent versions of our programs [14] include constraints which enable sugar ring conformation to be controlled by imposing either phase angle or amplitude as defined in Rao et al.,[15] it has been possible to construct a series of adiabatic maps for both 2'-deoxynucleosides and ribonucle-

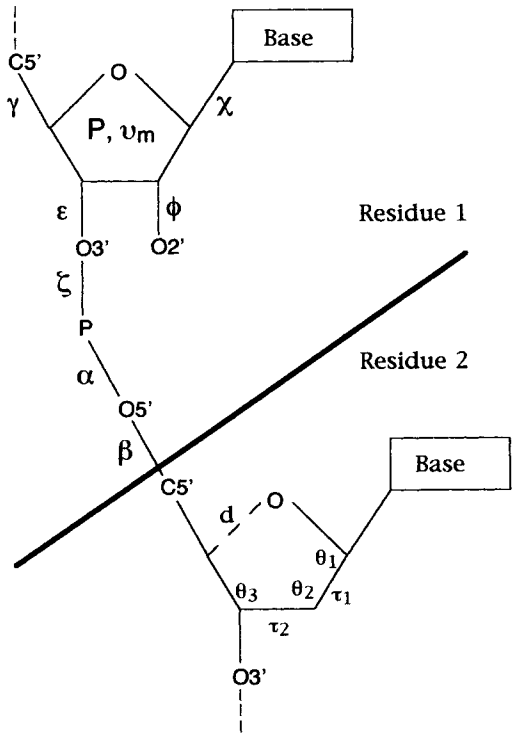

FIGURE 1. Internal coordinate nucleic acid model. Residue 1 shows the complete set of independent variables used to describe nucleosides in this study. Residue 2 shows the internal coordinates and closure constraint bond used by the CINFLEX algorithm to describe the furanose.

osides using the original CINFLEX method.[6] Each point in these maps is energy minimized [with respect to internal relaxation of the sugar, rotation about the *N*-glycosidic bond (χ), the $C_{3'}$—$O_{3'}$ (ϵ) and $C_{4'}$—$C_{5'}$ (γ) bonds, and, in the case of ribonucleosides, the $C_{2'}$—$O_{2'}$ (ϕ) bond] at a fixed value of phase and amplitude (Fig. 1, Residue 1), with phase varying from 0° to 360° in 10° increments and amplitude from 5° to 45° in 2.5° increments. Potential energy and structural information was extracted from the adiabatic maps to produce a series of $E(P, \nu_m)$, $\theta_i(P, \nu_m)$, and $\tau_i(P, \nu_m)$ surfaces.

All energy minimizations were performed using our standard FLEX force field,[6,16] using the following expression:

$$E = \frac{1}{2}\sum k_\theta(\theta - \theta_0)^2 + \frac{1}{2}\sum K_\phi[1 + \cos(n\phi - \delta)] + \sum_{ij}\left(-\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}\right) + \sum_{ij^*}\left[\cos\theta\left(-\frac{A_{ij}^{HB}}{r_{ij}^6} + \frac{B_{ij}^{HB}}{r_{ij}^{12}}\right) + (1 - \cos\theta)\left(-\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}\right)\right] + \sum_{ij}\frac{q_i q_j}{\epsilon(r) r_{ij}}$$

Its terms represent the energy contributions for valence and dihedral angle distortions and pairwise van der Waals[17] and electrostatic interactions.[16] Extra Lennard-Jones terms have been added to the potential function to account for the angular dependence of hydrogen bonding. The torsional term takes into account anomeric effects when calculating torsions that have terminal oxygens.[18] A distance-dependent dielectric damping function is included in the electrostatic term to model solvent effects[19,20]

$$\epsilon(r) = D - \frac{(D-1)}{2}[(rs)^2 + 2rs + 2]e^{-rs}$$

where D is the maximum value of the dielectric at long distances and s is the slope of the sigmoidal portion of the curve. We used $D = 78$ and $s = 0.356$ in our simulations. A lower slope of 0.16, previously used for B-DNA simulations, leads to an abnormally high anti to syn transition barrier in purine nucleosides, due to insufficiently damped short-range repulsion between the N_3 and $O_{1'}$ atoms. Although this simple dielectric function is only a crude model of solvent effects, it has nevertheless given results in good agreement with experiment concerning nucleic acid static[21] and dynamic conformations[22,23] and counterion condensation.[24,25] Efforts are nevertheless underway to replace this simple function with a more exact Poisson-Boltzmann electrostatic term.[26]

Within FLEX simulations, bond lengths are kept at their crystallographically determined equilibrium values and thus no bond stretching term is included in the potential energy function. Because bonds fluctuate very little at biological temperatures, even in the case of the slightly strained sugar rings,[27] this is a reasonable and widely used approximation. Endocyclic valence angles, however, show considerable flexibility[27] and must be included to model ring flexibility. (Note that within CINFLEX, exocyclic valence angles are not fixed but also are not independent variables. They are initially set to their crystallographically determined equilibrium values, but they are then coupled to the corresponding endocyclic angle to maintain a constant position with respect to the external bisector of this angle.) All nonring valence angles are kept fixed.

It is important to verify that the FLEX force field gives results consistent with experiment. Notably, it is well known that 2'-deoxyribose and ribose have two minima with respect to P, corresponding to the $C_{2'}$-endo and $C_{3'}$-endo pucker families,[18] and that these minima are preserved in nucleosides.[27] Specifically, ribose shows a slight preference for $C_{3'}$-endo ($P \sim 18°$), while the more asymmetric 2'-deoxyribose shows a pronounced preference for the $C_{2'}$-endo ($P \sim 162°$) conformation. Also, the $O_{4'}$-endo barrier between these two pucker modes is about 3.8 kcal/mol and 1.8 kcal/mol for ribose and 2'-deoxyribose, respectively, with a considerably higher barrier in the $O_{4'}$-exo region. Pucker amplitude fluctuates about a mean value of 38.6° ± 3° ranging from 30° to 46° in the furanoside crystal structures.[27] These features can be checked by examination of the pseudorotational potential energy surfaces $E(P, \nu_m)$. Figure 2 clearly shows that the correct P and ν_m preferences are reproduced by our force field and the associated internal coordinate model. Note that in the $O_{4'}$-exo region, lower ν_m is favored to avoid steric clash between the sugar $C_{5'}$ atom and the base. The remaining 2'-deoxynucleosides and ribonucleosides give nearly identical surfaces,[28] and are not shown here. This is consistent with previous experimental[27] and modeling results,[10] which show that the type of base or even the presence of base has little influence on sugar flexibility.

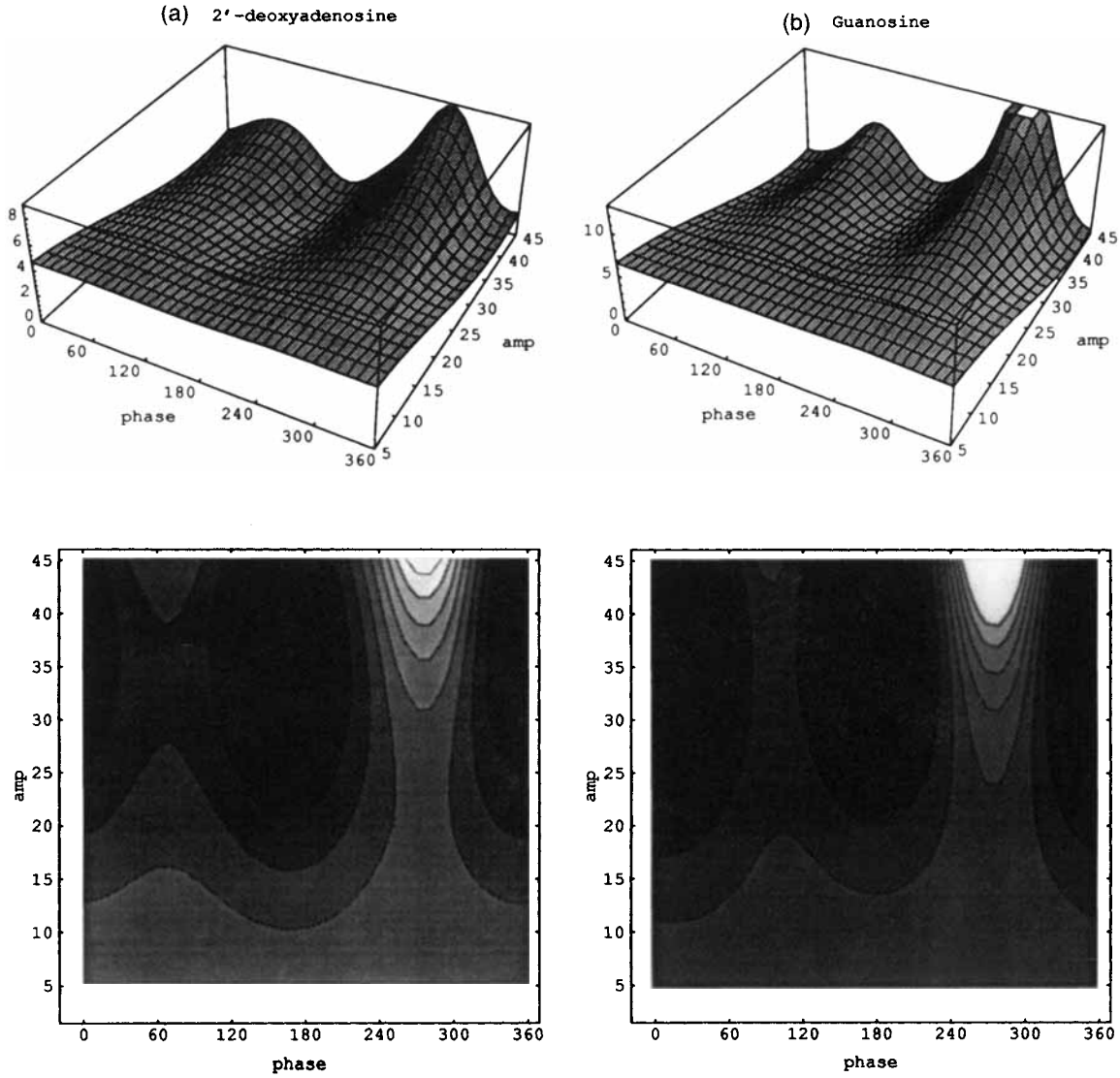

FIGURE 2. Representative 2′-deoxynucleoside and ribonucleoside potential energy surfaces produced by adiabatic mapping with respect to pseudorotation phase and pucker amplitude. (a) 2′-Deoxyadenosine. (b) Guanosine. For the sake of comparison, absolute energies are scaled to zero.

$E(P, \nu_m)$ surfaces created using Cartesian coordinate models and the AMBER force field[29] also agree well with the surfaces shown.[28] Variations of the five independent internal variables (θ_{1-3} and $\tau_{1,2}$) as a function of P at constant ν_m are shown in Figure 3. The valence angle variations (Fig. 3a) concord with results derived by a regression analysis of furanoside crystal structures (see Fig 5a, de Leeuw et al.[27]). The torsion angle variations (Fig. 3b), while similar to the curves expected from the pseudorotation equations, are not perfect cosine functions, as expected for a ring with unequal bond lengths.

Results and Discussion

JUSTIFICATION FOR AN INDEPENDENT VARIABLE REPRESENTATION OF FURANOSE BASED ON PHASE AND AMPLITUDE IN A DYNAMICAL POINT OF VIEW

Recently, Marzec and Day[11,12] proposed a new description of a furanose ring that exactly defines the geometry of a five-membered ring with fixed bond lengths. In common with the CINFLEX algorithm, it also uses four independent variables to

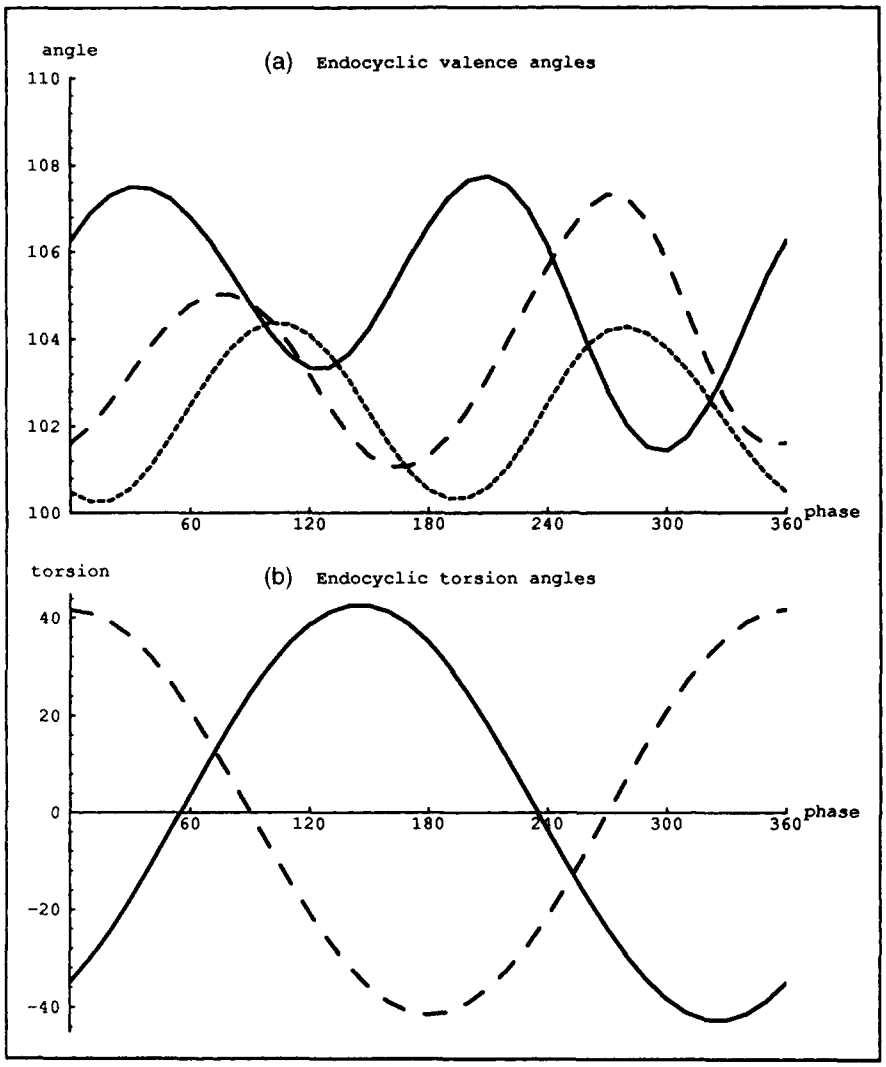

FIGURE 3. Endocyclic valence and torsion angles from CINFLEX at $\nu_m = 39°$. (a) Valence angles, θ_1, $O_{1'} - C_{1'} - C_{2'}$ (solid line), θ_2, $C_{1'} - C_{2'} - C_{3'}$ (dashed line), and θ_3, $C_{2'} - C_{3'} - C_{4'}$ (dotted line). (b) Torsion angles, τ_1, $O_{1'} - C_{1'} - C_{2'} - C_{3'}$ (solid line), and τ_2, $C_{1'} - C_{2'} - C_{3'} - C_{4'}$ (dashed line).

represent the sugar's internal degrees of freedom. However, it will be shown, based on molecular dynamics simulations, that these variables can be divided into two soft and two hard variables according to the formalism of Go and Scheraga.[30] Hard variables correspond to those undergoing rapid oscillations (i.e., with frequencies greater than 400 cm^{-1}), while soft variables undergo slower oscillations. In such cases, it is possible to treat hard variables as parameters which are functions of the soft variables. This in turn justifies the use of a two-variable furanose description within the classical rigid model as defined by Go and Scheraga[31] and already used for the linear parts of molecular within the FLEX force field.

The Marzec and Day furanose description requires four variables, P_{CP}, q, Γ, and S. Their Model is based on the definition of a planar pseudoring or mean ring of Cremer and Pople.[32] P_{CP} and q represent, respectively, the phase and maximum deformation of the ring atoms, j, in the z-direction perpendicular to the plane of the pseudoring:

$$Z_j = \sqrt{2/5} \cdot q \cdot \cos\left(\phi_z + \frac{4\pi j}{5}\right)$$

$$P_{MD} = \phi_z + \frac{\pi}{2}$$

P_{CP} and q were shown to be related to P and ν_m, respectively. P_{CP} and P are identical to within 6°,

and q is related to ν_m by the following equation:

$$\nu_m = C \cdot q$$

Values for C vary from 102.9 to 98.0°/Å depending on the data fitted.[27,33] Γ and S represent the direction and magnitude of the in-plane deformation of the pseudoring.

We have calculated the variations of P_{CP}, q, Γ, and S for each conformation obtained during adiabatic mapping with respect to P and ν_m (Fig. 2) using the program Breakring of Marzec and Day. The results show that P_{CP} is independent of ν_m and oscillates around the line $P_{CP} = P$ with a frequency of 180° and an amplitude of 3°. This is compatible with earlier calculations of Harvey and Prabhakaran.[33] In contrast, we find a slightly different functional form for q which shows a linear dependence with respect to ν_m (with a slope equal to 0.89×10^{-2} Å/°) and a periodic dependence with respect to P with periods of π and 2π.

$$q = (0.89 \times 10^{-2})\nu_m \times \left[1 + 0.061\cos\left(P - \frac{\pi}{2}\right) + 0.073\cos 2P\right]$$

The Γ and S values for the structures from the adiabatic maps are compatible with those obtained from a survey of the X-ray crystallographic coordinates of 665 furanose rings (see Fig. 5, Marzec and Day[11]).

We now turn to an analysis of the frequency components of furanose motion. A 1-ns molecular dynamics simulation of 2'-deoxyguanosine was taken from Gabb and Harvey[10] for Fourier analysis. This simulation was carried out on the free nucleoside at 300 K using the AMBER force field[29] and a 1-fs integration step. The simulation was performed with a constant dielectric of 4. No nonbond cutoff was used, and bond lengths were unconstrained. The Cartesian coordinates of the five ring atoms were output every 20 steps (0.02 ps). We extracted the Marzec and Day variables (P_{CP}, q, Γ, S) from these data using the program Breakring to obtain the time series for each parameter. The endocyclic valence and torsion angles were also extracted from the coordinates along with the values of P and ν_m (calculated according to Rao et al.[15]).

We calculated normalized time correlation functions[34] for fluctuations in all variables extracted from the simulation data. The Fourier transforms of the time correlation functions were then calculated to provide the frequency spectra of each variable. This analysis was made with the Fourier program of Mathematica (Wolfram Research Inc.) using a total of 32,768 data points. Because coordinates were output every 0.02 ps during the simulation, the Fourier transform only gives information for vibrations up to frequencies of 834 cm^{-1}. This is, however, sufficient for distinguishing between hard and soft variables.[30] The resulting spectra for P, ν_m, Γ, S, and representative valence and torsion angles are shown in Figure 4. The spectra of P_{CP} and q are not shown because they are nearly identical to those of P and ν_m. Note that the valence angle, $C_{4'}$—$O_{4'}$—$C_{1'}$, and the torsion angle, $C_{3'}$—$C_{4'}$—$O_{4'}$—$C_{1'}$, are defined here as θ_5 and τ_4, respectively.

The first point to note from these results is the high-frequency oscillations observed for Γ and S (Fig. 4), which exceed 400 cm^{-1} in both cases. S has a single peak at 542 cm^{-1}, while Γ has three main peaks at 414, 485, and 542 cm^{-1}. On the other hand, the important oscillations for P and ν_m occur at frequencies below 200 cm^{-1}, except for a peak at 484 cm^{-1} in the case of ν_m. In fact, 75% of the total signal for ν_m, after subtraction of noise, is centered around a peak at 116 cm^{-1}, and most of the signal for P is concentrated at values lower than 100 cm^{-1} (the two main peaks occur at 14 cm^{-1} and 37 cm^{-1}).

If we now compare these spectra with those of valence and torsion angles, it is apparent that, similar to Γ and S, the valence angles mainly exhibit vibrational frequencies higher than 400 cm^{-1}, whereas the torsion angles have most frequencies below 200 cm^{-1}, except for a peak at 484 cm^{-1}, similar to P and ν_m. Strikingly, the spectra of θ_2 and, to a lesser extent, θ_4 (not shown) are analogous to that of Γ, except for low-frequency components (14 and 116 cm^{-1}) that do not appear in the Γ spectrum. These frequencies can, however, be found in the spectra of P (14 cm^{-1}) and ν_m (116 cm^{-1}). In the same way, the signals of θ_5 and S are similar and are analogous to those of θ_1 (not shown), which presents, in addition, small peaks at 14, 238, and 484 cm^{-1} that are not found in the spectra of P and ν_m. θ_5 is the only valence angle which oscillates at unique frequency, 542 cm^{-1}, also found in the spectra of every other valence angle except θ_3. Interestingly, this is also the unique vibrational frequency of S. This can be related to the observation of Marzec and Day[11] that important deformations in S occur primarily for $\Gamma = 90°$—namely, along a line from the center of the pseudoring towards the oxygen atom.

Turning to the torsion angles, τ_5 exhibits the

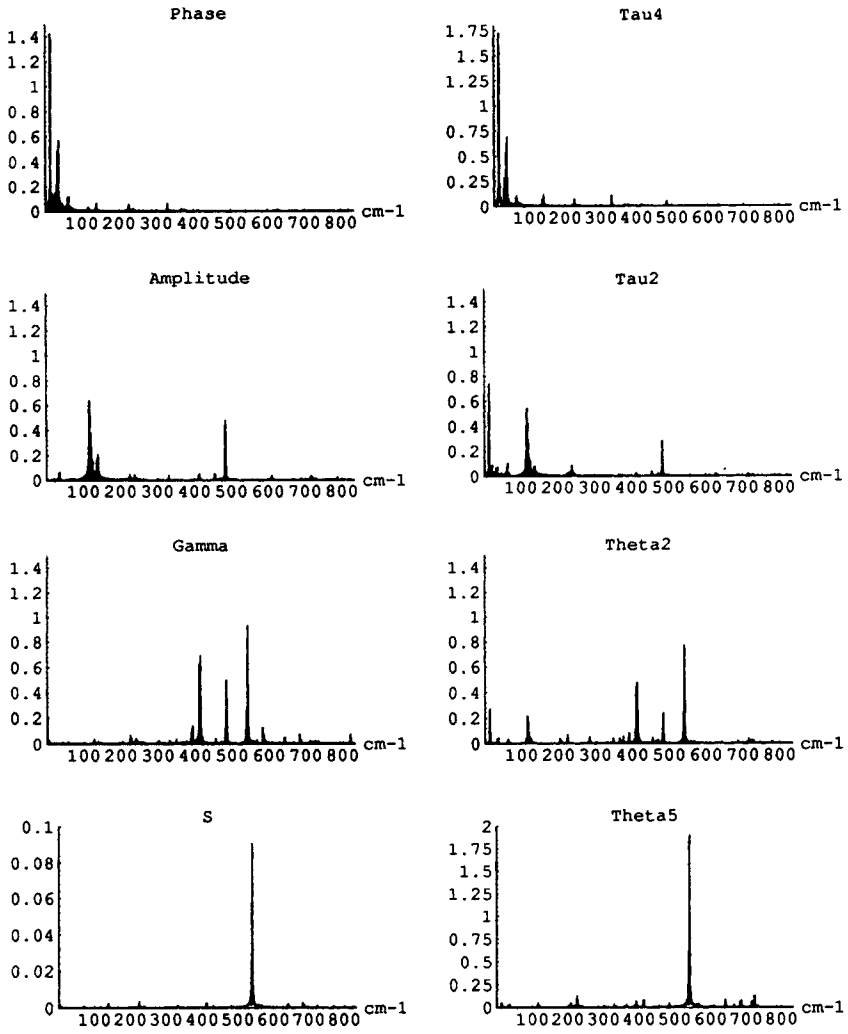

FIGURE 4. Frequency spectra from the Fourier analysis of a 1-ns molecular dynamics simulation of 2′-deoxyguanosine. Left (from top down), pseudorotation phase and amplitude as defined by Altona and Sundaralingam (1972), and the variables Γ and S of Marzec and Day (1993). Right (from top down), two representative dihedral angles, τ_4 ($C_{3'}-C_{4'}-O_{1'}-C_{1'}$) and τ_2 ($C_{1'}-C_{2'}-C_{3'}-C_{4'}$), and two representative valence angles, θ_2 ($C_{1'}-C_{2'}-C_{3'}$) and θ_5 ($C_{5'}-O_{1'}-C_{1'}$).

same vibrational frequencies as P (14 and 37 cm^{-1}), and τ_2, with the exception of a peak at 14 cm^{-1}, exhibits signals similar to those of ν_m (116 and 484 cm^{-1}). The other dihedral angles give spectra which are hybrids of τ_2 and τ_5. The peak at 484 cm^{-1}, the only high-frequency peak for ν_m, appears only in the spectra of τ_2 and τ_3.

This spectral analysis of the furanose ring (Fig. 4) allows us to conclude that the parameters, P_{CP} and q of Marzec and Day, and the pseudorotational parameters, P and ν_m of Rao et al.,[15] are comparable in a dynamic sense to the torsion angle variations. The Γ and S parameters are clearly linked to endocyclic valence angle variations. There are, however, low-frequency signals for the valence angles that do not appear in Γ and S but do appear for P and ν_m (Fig. 4). Therefore, all of the low-frequency motions for the sugar are accounted for by P and ν_m, and the internal movements of the ring can indeed be divided into hard (Γ and S) and soft (P and ν_m) variables following Go and Scheraga.[30] It is thus justified to consider Γ and S as parameters of P and ν_m and to use a two-variable model for studying furanose flexibility.

EMPIRICAL EQUATIONS FOR THE FURANOSE INTERNAL DEGREES OF FREEDOM

Satisfied that a two-variable model will behave in a manner consistent with experiment, we set out to derive a set of equations describing the valence angle and torsional degrees of freedom in terms of P and ν_m. We recall here that the $\theta_i(P, \nu_m)$ and $\tau_i(P, \nu_m)$ surfaces generated with CINFLEX are all similar regardless of the type of base, although the 2′-deoxynucleoside and ribonucleoside $E(P, \nu_m)$ surfaces differ. It is thus reasonable to derive a general set of equations describing each independent variable as a function of P and ν_m. We examined the shape of the $\theta_i(P, \nu_m)$ and $\tau_i(P, \nu_m)$ surfaces (Fig. 5) to determine the functional form of the generalized equations. Notice that the $\theta_i(P, \nu_m)$ surfaces (Figs. 5a, 5b, 5c) are quadratic in the ν_m dimension and periodic in the P dimension. The same is true for the $\tau_i(P, \nu_m)$ surfaces (Figs. 5d, 5e). The Mathematica package (Wolfram Research Inc.) was used to fit the averaged data for each independent variable to the following combined quadratic/periodic functional forms:

$$\text{Valence angles: } (1 + \nu_m + \nu_m^2)(1 + \cos P + \sin P + \cos 2P + \sin 2P)$$

$$\text{Torsion angles: } (1 + \nu_m + \nu_m^2)(1 + \cos P + \sin P)$$

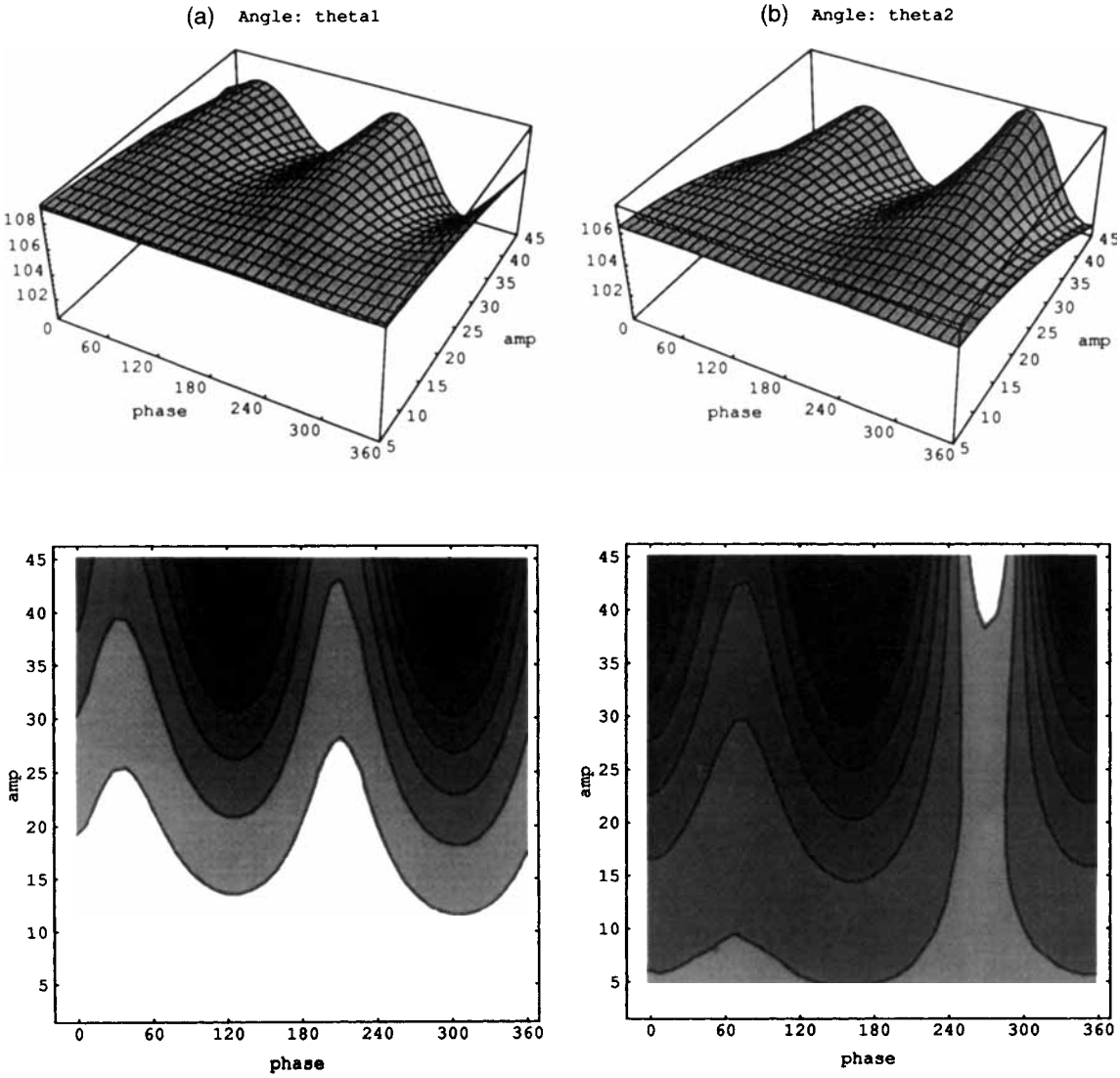

FIGURE 5. Representative surfaces for the endocyclic valence and torsion angles produced by adiabatic mapping with respect to pseudorotation phase and pucker amplitude. (a) θ_1, $O_{1'}-C_{1'}-C_{2'}$. (b) θ_2, $C_{1'}-C_{2'}-C_{3'}$. (c) θ_3, $C_{2'}-C_{3'}-C_{4'}$. (d) τ_1, $O_{1'}-C_{1'}-C_{2'}-C_{3'}$. (e) τ_2, $C_{1'}-C_{2'}-C_{3'}-C_{4'}$.

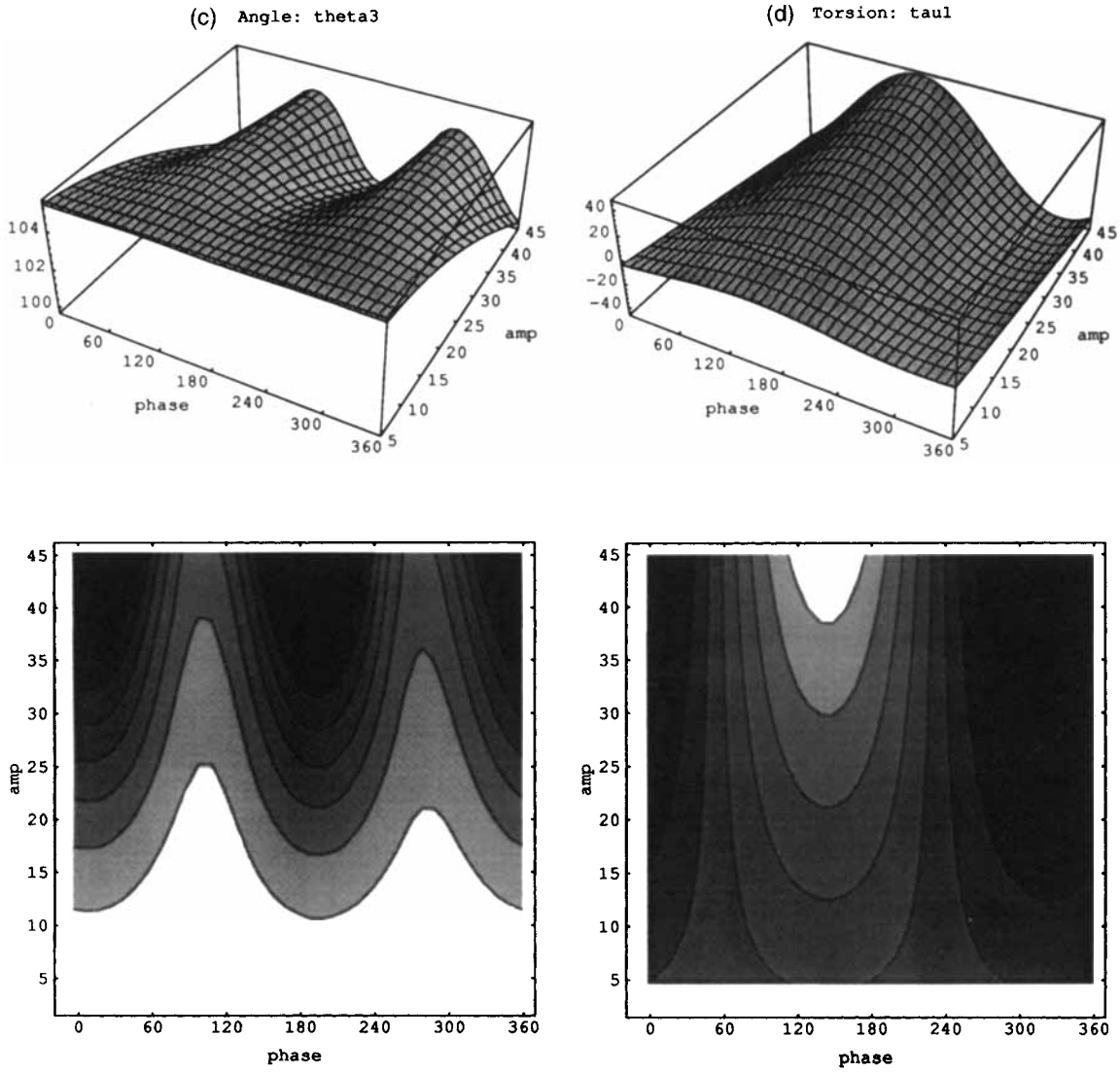

FIGURE 5. (Continued)

Because the periodicity of the $\tau_i(P, \nu_m)$ surfaces in the P dimension is only 2π are opposed to π and 2π for the $\theta_i(P, \nu_m)$ surfaces, a slightly less complicated functional form suffices. These functions, when expanded, give the following cross-terms, each with its own coefficient, a_n:

Valence angles:

$$a_1 + a_2\nu_m + a_3\nu_m^2 + a_4 \cos P + a_5\nu_m \cos P + a_6\nu_m^2 \cos P + a_7 \sin P + a_8\nu_m \sin P + a_9\nu_m^2 \sin P + a_{10} \cos 2P + a_{11}\nu_m \cos 2P + a_{12}\nu_m^2 \cos 2 + a_{13} \sin 2P + a_{14}\nu_m \sin 2P + a_{15}\nu_m^2 \sin 2P$$

Torsion angles:

$$a_1 + a_2\nu_m + a_3\nu_m^2 + a_4 \cos P + a_5\nu_m \cos P + a_6\nu_m^2 \cos P + a_7 \sin P + a_8\nu_m \sin P + a_9\nu_m^2 \sin P$$

The coefficients obtained by a least-squares fit to the surface data are listed in Table 1. Tests including third- and fourth-order cross-terms were also made, but such terms were found to give very little increase in accuracy. Also, coefficients smaller than 0.001 have virtually no impact on the final sugar conformation. They are included for completeness.

One would think intuitively that ribose and 2′-deoxyribose require two different sets of coeffi-

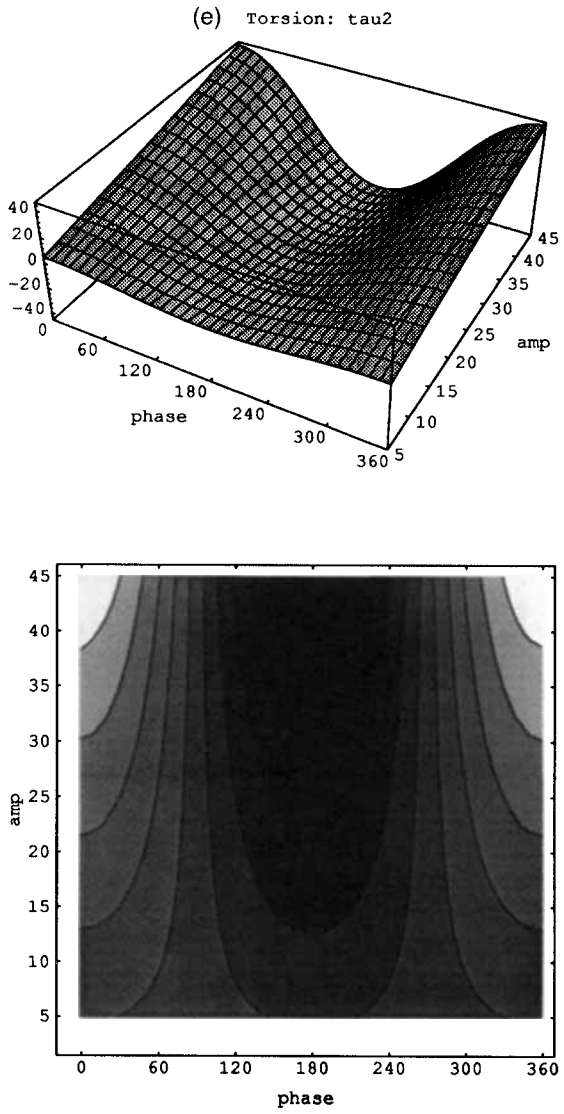

FIGURE 5. (Continued)

cients given their structural differences. However, when the generalized coefficients are compared to the original data for each nucleoside and 2′-deoxynucleoside, the average differences between the calculated angle values and the actual values from the adiabatic maps is always less than 0.8° (in most cases less than 0.5°), with the largest errors occuring in the high ν_m/O_4′-exo region. If this region, which is rarely visited in reality, is excluded from the error calculations, the average errors fall to less than 0.5°. For reasons of space, we are unable to include all of the valence and torsion angle data. These data are available from the authors.

Using these equations, we first verified that the closure bond, d (Fig. 1, Residue 2), does not deviate significantly from its equilibrium value. In all cases, the fitted endocyclic valence and torsion angle functions satisfy the ring closure constraint to within 1% of the equilibrium distance (roughly 0.01 Å). To test the ability of the equations to generate a ring with the desired values of P and ν_m, structures were built without minimization throughout the range of P (0°–360°) and ν_m (0°–50°). Back-calculating the pseudorotational parameters from the Cartesian coordinates using the Rao et al.[15] equations showed that the empirical equations perform well. Nearly all structures were 0.5° of their P (mean deviation = $0.16° \pm 0.16°$) and ν_m (mean deviation = $0.15° \pm 0.15°$) target values, with larger errors (1.15° in P and 1.85° in ν_m) only occuring in the O_4′-exo high-amplitude region (beyond $\nu_m = 45°$), which is energetically unfavorable. We can thus conclude that our simple two-variable model is perfectly compatible with the complete internal coordinate model.

INTERNAL COORDINATE MONTE CARLO SIMULATIONS

We performed extensive Monte Carlo simulations on both free 2′-deoxynucleosides and ribonucleosides to test both the efficiency and accuracy of our model. Following the standard Monte Carlo method,[35,36] a representative Boltzmann ensemble of conformations is generated by a series of steps, which are accepted or rejected with the following probabilities:

$$\text{Probability} = \begin{cases} e^{-\Delta E/RT} & \text{if } \Delta E > 0 \\ 1 & \text{if } \Delta E \leq 0 \end{cases}$$

where ΔE represents the potential energy difference between two states ($E_{i+1} - E_i$). R is the gas constant and T the absolute temperature. In all the simulations reported here, T equals 300 K. We maintain a 50% acceptance rate for each independent variable, and individual variables are monitored throughout the simulation. Our simulations satisfy both the detailed balance and microreversibility conditions: There is an equal probability of choosing a trial conformation $i + 1$ from conformation i as there is of choosing a trial conformation i from conformation $i + 1$; and for any step $i \rightarrow i + 1$, the reverse step $i + 1 \rightarrow i$ will exactly reproduce i. To improve sampling, multiple simulations were performed, each starting from a randomly chosen, minimized conformation. Ten simu-

TABLE I.
Coefficients for the Generalized Endocyclic Valence and Torsion Angle Functions.

	θ_1	θ_2	θ_3	τ_1	τ_2
a_1	1.9028	1.8621	1.8418	0.0001	0.0000
a_2	−0.0065	−0.0157	−0.0185	0.0003	−0.0007
a_3	−0.1277	−0.0604	−0.0770	0.0029	0.0026
a_4	0.0001	0.0000	−0.0005	0.0004	0.0005
a_5	−0.0077	0.0062	0.0046	−0.8310	0.9765
a_6	−0.0030	0.0091	−0.0089	0.0095	0.0037
a_7	0.0010	−0.0011	−0.0002	−0.0007	0.0001
a_8	0.0036	−0.0042	0.0088	0.5800	0.0059
a_9	0.0295	−0.0367	−0.0067	−0.0137	−0.0010
a_{10}	0.0013	0.0004	0.0007		
a_{11}	−0.0110	−0.0036	−0.0061		
a_{12}	0.0547	−0.0726	−0.0471		
a_{13}	−0.0010	−0.0007	0.0011		
a_{14}	0.0106	0.0067	−0.0087		
a_{15}	0.0676	0.0042	−0.0174		

lations of 2×10^6 steps (approximately 15 minutes computation on a Hewlett–Packard 735 workstation) were run for the eight commonly occurring 2′-deoxynucleosides and ribonucleosides. The results of each set of simulations were averaged together to give the free energy or, more correctly, the potential of mean force[37] for each molecule (Figs. 6 and 7). For space considerations, only

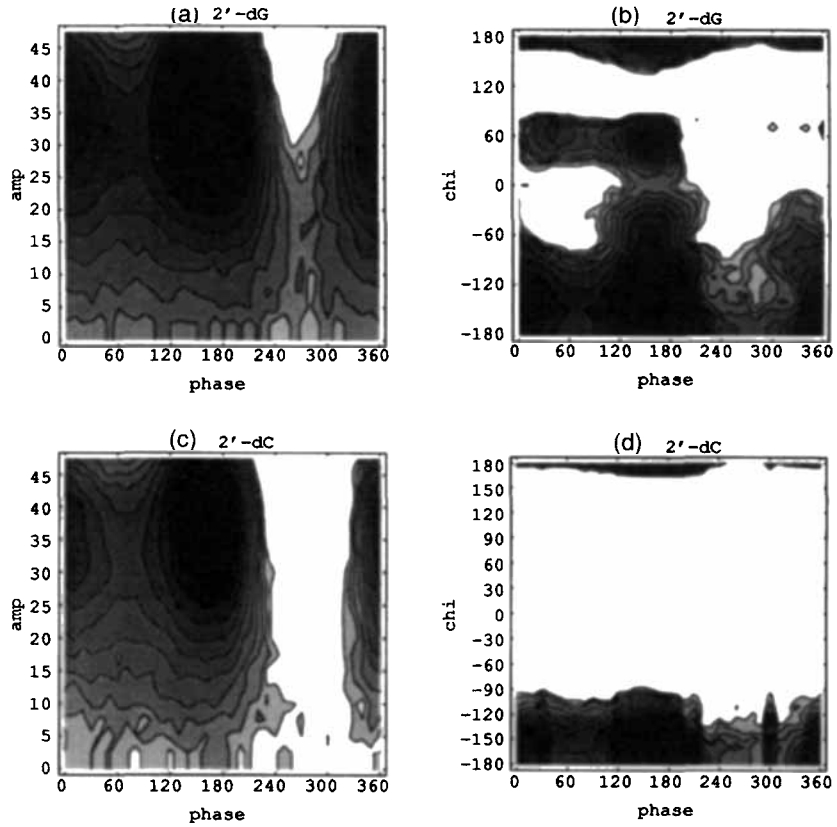

FIGURE 6. Nucleoside free energy surfaces calculated from Metropolis Monte Carlo simulations at 300 K. (a) 2′-dG, $E(P, \nu_m)$. (b) 2′-dG, $E(P, \chi)$. (c) 2′-dC, $E(P, \nu_m)$. (d) 2′-dC, $E(P, \chi)$. (e) rG, $E(P, \nu_m)$. (f) rG, $E(P, \chi)$. (g) rU $E(P, \nu_m)$. (h) rU, $E(P, \chi)$. The surfaces are shown from 0 to 6 kcal / mol with 0.5 kcal / mol per contour.

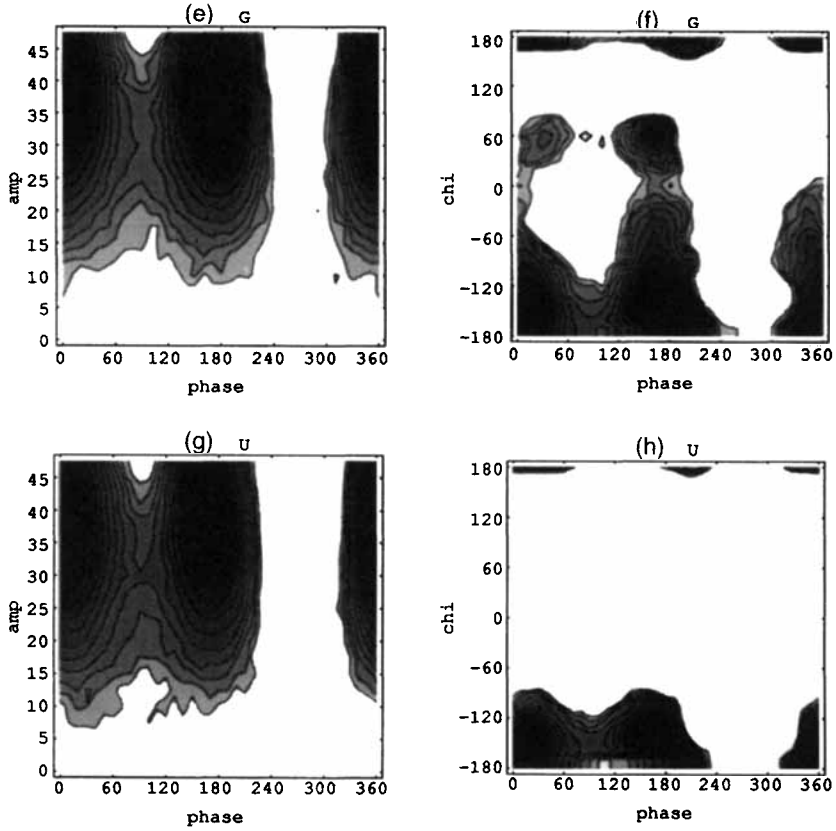

FIGURE 6. (Continued)

selected examples will be discussed here. The remaining data are available from the authors.

Comparing the free energy and the potential energy surfaces as a function of the pseudorotational parameters shows that the new furanose description behaves consistently with experimental results and does not distort the shape of the surfaces. A comparison of Figures 6a and 6c with Figure 2a and of Figures 6e and 6g with Figure 2b shows that the minima are in the same positions and that their ordering is conserved. It is also seen that the 2′-deoxynucleosides can sample lower ν_m ranges than the ribonucleotides (Fig. 7) due to the absence of repulsions between eclipsed vicinal oxygens at the 2′ and 3′ positions. The present results also support previous experimental[27] and theoretical[10] evidence showing that the type of base has very little influence on sugar flexibility. The $\Delta G(P, \chi)$ surfaces (Figures 6b, 6d, 6f, and 6h) demonstrate that the force field correctly models sugar and base interactions. First, pyrimidines rarely transit from anti ($\chi = -180°$ to $-120°$) to syn ($\chi = 0°$ to 60°) geometries[27,38] due to a steric clash between the pyrimidine O_2 and the sugar $C_{5'}$ atoms. Second, both potential and free energy calculations show that of the two possible pathways for the anti to syn glycosidic rotation (see Saenger[39]), the transition occurs preferentially through high anti (χ near 0°) intermediates, which exhibit significantly lower energy barriers than high syn[10,39,40] (χ near 120°). The $\Delta G(P, \chi)$ surfaces also show that when χ is syn, there is a distinct preference for the $C_{2'}$-endo sugar pucker. A survey of furanoside crystal structures shows the same tendency.[27]

Although we performed extensive simulations for the sake of completeness, it was found that a single simulation as short as 4×10^5 steps was in fact sufficient to reproduce the significant features of Figure 6. We attribute this increased efficiency to the larger steps which can be made in our internal coordinate model. Average step sizes were $\langle \Delta P \rangle = 15.4° \pm 4.9°$, $\langle \Delta \nu_m \rangle = 5.1° \pm 0.8°$, $\langle \Delta \chi \rangle = 14.2° \pm 5.5°$, $\langle \Delta \gamma \rangle = 12.3° \pm 2.2°$, $\langle \Delta \epsilon \rangle = 32.2° \pm 3.9°$, and for ribonucleosides $\langle \Delta \phi \rangle = 25.8° \pm 2.3°$. The results presented here are, encouragingly, nearly identical to earlier molecular dynamics simulations using the AMBER force field.[10] The latter simulations, however, required hours of computation, while our protocol needs only minutes.

CONCLUSIONS

A system of semiempirical equations has been derived which enable furanose endocyclic valence and torsion angles to be calculated from the pseudorotational parameters P and ν_m. These equations have been derived from energy minimized nucleoside conformations using the full internal coordinate model of the CINFLEX algorithm. By a detailed analysis of molecular dynamic simulations, it is demonstrated that a two-variable furanose model is perfectly justifiable. In addition, comparisons with experimental data[27] and previ-

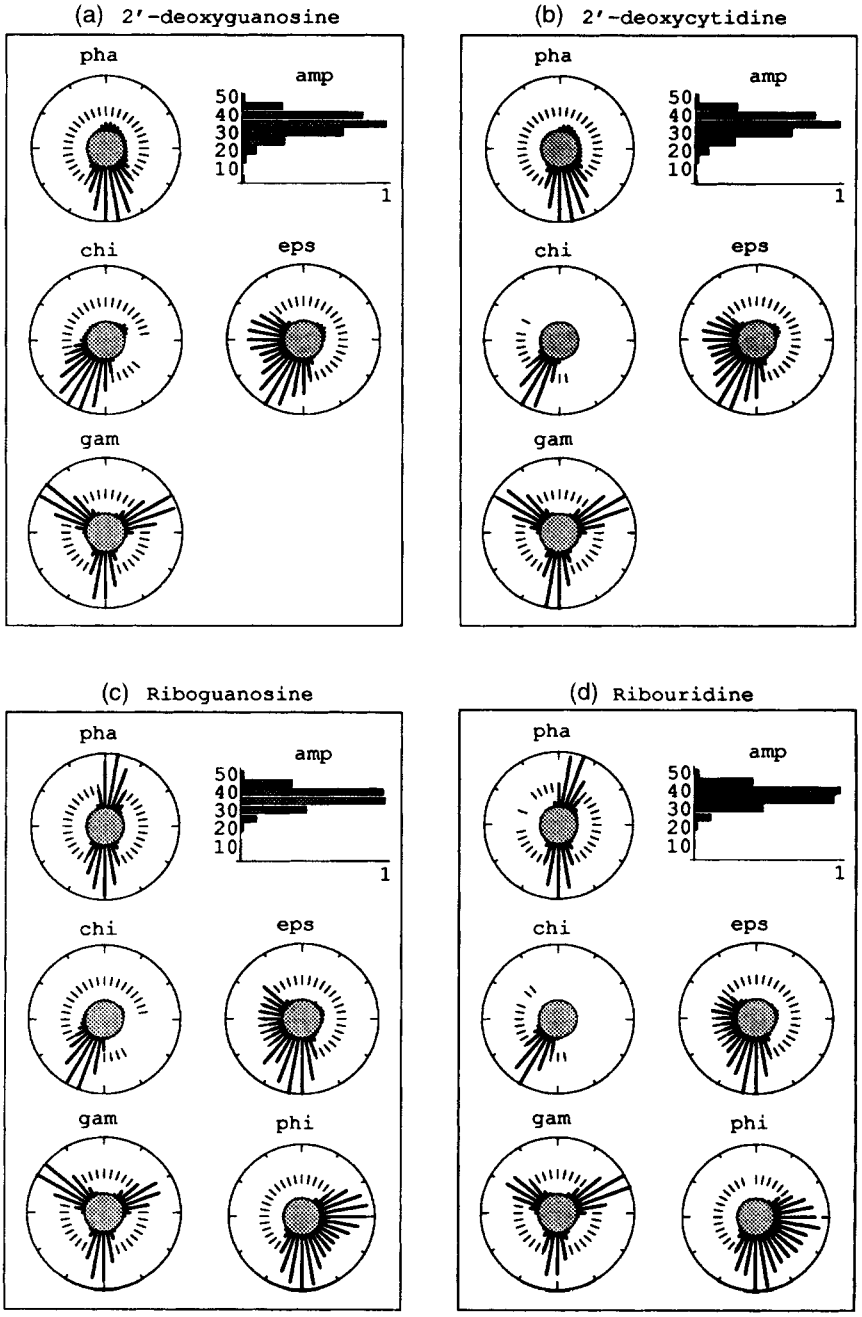

FIGURE 7. Pseudorotation phase, amplitude, and torsional data for representative nucleosides constructed from the Boltzmann distributions of ten 2.0×10^6 step Monte Carlo simulations, averaged and normalized to 1. In the radial plots, the thin lines not connected to the inner circle indicate areas of very low sampling. (a) 2'-dG. (b) 2'-dC. (c) rG. (d) rU.

ous molecular modeling[10] demonstrate that the new furanose description accurately models nucleoside behavior.

This method for treating furanose flexibility is also computationally efficient. It avoids costly ring closure problems and, by reducing the number of independent variables, will lead to considerable savings in simulations of large nucleic acid systems. Tests on both deoxy- and ribonucleosides using internal coordinate Monte Carlo simulations indeed demonstrate rapid conformational sampling with large step sizes in the ring puckering variables. Results close to those of all-atom molecular dynamics are obtained, but with much less computational expense.

Acknowledgments

We would like to thank W. J. Zakrewski (University of Durham, U.K.) for his assistance in determining the functional form for our empirical equations. We would also like to thank C. J. Marzec and L. A. Day (The Public Health Research Institute, New York, USA) for generously providing the program Breakring. This research was generously supported by a grant from the Association for International Cancer research (St. Andrews University, U.K.). H. A. G. thanks the French Ministère des Affaires Estrangères for the award of a postdoctoral fellowship.

References

1. R. E. Dickerson, *Meth. Enzymol.*, **211A**, 67 (1992).
2. C.-S. Tung. *J. Biomol. Struct. Dyn.*, **9**, 1185 (1992).
3. R. Lavery and B. Hartmann, *Biophys. Chem.*, **50**, 33 (1994).
4. M. Poncin, B. Hartmann, and R. Lavery, *J. Mol. Biol.*, **226**, 775 (1992).
5. M. Ouali, R. Letellier, F. Adnet, J. Liquier, J.-S. Sun, R. Lavery, and E. Taillandier, *Biochemistry*, **32**, 2098 (1993).
6. R. Lavery, I. Parker, and J. Kendrick, *J. Biomol. Struct. Dyn.*, **4**, 443 (1986).
7. C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, **94**, 8205 (1972).
8. E. A. Merritt and M. Sundaralingam, *J. Biomol. Struct. Dyn.*, **3**, 559 (1985).
9. D. A. Pearlman and S.-H. Kim, *J. Biomol. Struct. Dyn.*, **3**, 85 (1985).
10. H. A. Gabb and S. C. Harvey, *J. Am. Chem. Soc.*, **115**, 4218 (1993).
11. C. J. Marzec and L. A. Day, *J. Biomol. Struct. Dyn.*, **10**, 1091 (1993).
12. C. J. Marzec and L. A. Day, *J. Biomol. Struct. Dyn.*, **10**, 1125 (1993).
13. S. H. Northrup and J. A. McCammon, *Biopolymers*, **19**, 1001 (1980).
14. S. R. Sanghani and R. Lavery, *Nucleic Acids Res.*, **22**, 1444 (1994).
15. S. T. Rao, E. Westhof, and M. Sundaralingham, *Acta Crystallogr. Sect. A*, **37**, 421 (1981).
16. R. Lavery, H. Sklenar, K. Zakrzewska, and A. Pullman, *J. Biomol. Struct. Dyn.*, **3**, 989 (1986).
17. V. B. Zhurkin, V. I. Poltiev, and V. L. Florent'ev, *Molekulyarnaya Biologiya*, **14**, 116 (1980).
18. W. Olson, *J. Am. Chem. Soc.* **104**, 278 (1982).
19. B. E. Hingerty, R. H. Ritchie, T. L. Ferrell, and J. E. Turner, *Biopolymers*, **24**, 427 (1985).
20. R. Lavery, In *DNA Bending and Curvature*, W. K. Olson, R. H. Sarma, M. H. Sarma, and M. Sundaralingam, Eds., Adenine Press, New York, 1988, p. 191.
21. R. Lavery, *Advances in Computational Biology*, **1**, 69 (1994).
22. V. Fritsch and E. Westhof, *J. Am. Chem. Soc.*, **113**, 8271 (1991).
23. V. Fritsch and E. Westhof, *J. Comp. Chem.*, **12**, 147 (1991).
24. B. Jayaram, S. Swaminathan, D. L. Beveridge, K. Sharp, and B. Honig, *Macromolecules*, **23**, 3156 (1990).
25. M. O. Fenley, G. S. Manning, and W. K. Olson, *Biopolymers*, **30**, 1191 (1990).
26. K. Sharp, *Curr. Opin. Struct. Biol.*, **4**, 234 (1994).
27. H. P. M. de Leeuw, C. A. G. Haasnoot, and C. Altona, *Isr. J. Chem.*, **20**, 108 (1980).
28. H. A. Gabb, Ph.D. thesis, "Modeling Conformational Transitions in 2'-Deoxynucleosides and Nucleic Acids," University of Alabama at Birmingham, Schools of Medicine and Dentistry, Birmingham, AL, 1993.
29. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, **7**, 230 (1986).
30. N. Go and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969).
31. N. Go and H. A. Scheraga, *Macromolecules*, **9**, 535 (1976).
32. D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, **97**, 1358 (1975).
33. S. C. Harvey and M. Prabhakaran, *J. Am. Chem. Soc.*, **108**, 6128 (1986).
34. S. Swaminathan, T. Ichiye, W. van Gunsteren, and M. Karplus, *Biochemistry*, **21**, 5230 (1982).
35. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
36. M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods Volume I: Basics*, John Wiley & Sons, New York, 1986.
37. D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Chem.*, **18**, 431 (1989).
38. A. E. V. Haschemeyer and A. Rich, *J. Mol. Biol.*, **27**, 369 (1967).
39. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
40. D. A. Pearlman and S.-H. Kim, *J. Biomol. Struct. Dyn.*, **3**, 99 (1985).