Principal torsion angles of collective motions in biomolecules: A study of a single base opening in DNA duplexes

Y. Z. Chen, V. Mohan, and R. H. Griffey
ISIS Pharmaceuticals, 2292 Faraday Avenue, Carlsbad, California 92008
(Received 29 December 1997)

Large amplitude collective motions in biomolecules primarily involve the displacement of a number of torsion angles of rotatable bonds. This work tests the hypothesis that one or a few of these angles are the principal variables that act as the reaction coordinate for a particular motion. Our study on a low energy single base opening in DNA duplexes identified one such coordinate. This torsion angle together with the accompanying geometric requirement on other angles have been used to determine the pathway of the opening process. The energetics along the pathway has also been analyzed based on widely used molecular mechanics force field parameters. Our results support the proposed hypothesis and further tests on other collective motions in different systems are warranted. [S1063-651X(98)11607-0]

PACS number(s): 87.10.+e, 87.15.By

INTRODUCTION

Collective motions in biomolecules play important roles in biological processes [1]. The resulting conformational changes are facilitated primarily by a low energy, concerted displacement of a number of torsion angles of rotatable bonds in the molecular chain (a torsion angle is defined as the angle between the *A-B-C* plane and *B-C-D* plane of a four atom chain *A-B-C-D*). Knowledge of microscopic details of the motions may aid in elucidating the underlying mechanism of these biological events. Theoretical investigations of such processes, however, are hindered by the complexity and the large number of degrees of freedom involved. Novel methods need to be developed to overcome this difficulty.

The collective nature of these motions implies a disproportionate role of the associated torsion angles. Localized modes of certain torsions promote particular group motion while the restoring forces drive the displacement of other torsions. We hypothesize that only one or a few of the torsion angles are principal variables that play a major role in the collective motion. The identification of these principal torsion angles makes it easier to determine the motion pathway, as the variation of the rest of the local torsion angles can be deduced by the geometric constraints along the molecular chain.

In the present work we test this hypothesis by studying a low energy single base opening process in DNA duplexes. Our focus is on the opening of a single base with minimum structural distortion in the neighboring nucleotides. We examine to what extent a single base can be opened without displacement of its neighbors. We also demonstrate how the identification of a principal torsion angle along with helix geometric constraint can be used to determine the variation of all the rotatable torsion angles responsible for the opening. Because of the localized nature, the pathway for the above discussed opening is likely to be a low energy route that might initiate other premelting and melting events in DNA duplexes.

The kinetics and thermodynamics of base pair opening has been extensively investigated by a variety of experimental techniques at both premelting [2–4] and melting [5] regions. Theoretical methods based on simple mathematical models [6] and atomic level lattice dynamics approach [7,8] have been used to study hydrogen bond (H bond) breaking dynamics and its contribution to the opening and melting in DNA. Analysis of opening pathways using a simplified backbone geometry has appeared [9]. The low frequency collective motions relevant to opening also have been probed [10]. Despite these efforts, microscopic details on the pathways of base pair opening remain elusive.

Imino proton exchange experiments show that base pairs open one at a time [2]. A single base pair opening event was detected in a 10 base pair duplex. Therefore it is likely that individual opening events in a long DNA sequence occur at least this distance apart, which can thus be considered approximately as an isolated event. Comparison between the observed [3,4] and our computed opening enthalpy, based on a widely used set of force field parameters [11], suggests that the low energy pathway likely involves base rotation towards the grooves. This rotation is facilitated by collective motions of relevant torsion angles at the opening site. Although motions in the complementary bases and associated backbones in a base pair opening event are yet to be determined experimentally, our study is useful in probing whether a strictly low energy single base displacement is sufficient to facilitate imino proton exchange and thus correlates with observed base pair opening processes. Such a correlation will be discussed in this paper.

THEORETICAL METHODS

Models of DNA duplexes

Our models of DNA duplexes are atomic-level models with individual atoms connected by chemical bonds. These models are from crystal structures of B-form DNA duplexes deposited in the Nucleic Acids Database [12]. For comparison the canonical B-form DNA, constructed from the fiber diffraction data [13], are also used in this study. A total of 32 B-form DNA structures have been considered in this work.

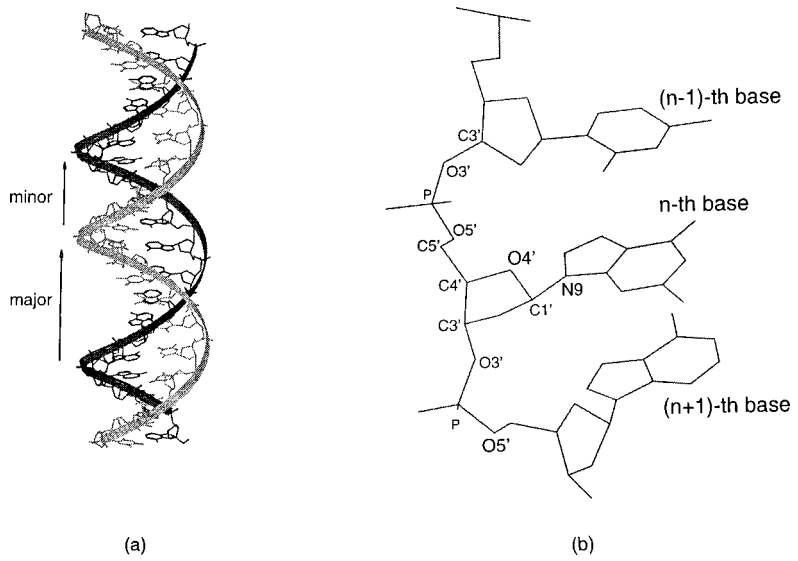

FIG. 1. (a) Major and minor grooves in a duplex DNA are indicated. (b) Torsion angles of the rotatable bonds involved in the opening of the nth base in a DNA duplex. The torsion angle, for instance, ζ of the four atom chain O5′-P$\overset{\zeta}{-}$O3′-C3′ is defined as the angle between the O5′-P-O3′ plane and the P-O3′-C3′ plane. The standard notations for the backbone torsion angles are: $(n+1)$O5′-P$\overset{\zeta_n}{-}$O3′$\overset{\epsilon_n}{-}$C3′ $\overset{\delta_n}{-}$C4′$\overset{\gamma_n}{-}$C5′$\overset{\beta_n}{-}$O5′$\overset{\alpha_n}{-}$P$\overset{\zeta_{n-1}}{-}$O3′-C3′$(n-1)$. The backbone angles involved in the opening of the nth base are the principle torsion angle ζ_n and adjustable torsion angles α_n, β_n, γ_n, and ζ_{n-1}. The glycosidic bond torsion angle χ_n(O4′-C1′$\overset{\chi_n}{-}$N9-C8) is also an adjustable variable that keeps the base in plane.

The atomic motions can lead to the stretch, angle bending (change of the angle between two bonds), and torsion angle deformation in the rotatable bonds. The stretch and angle bending motion do not contribute to the low energy opening process, as large deviation in these variables involves significantly higher energy than the observed <30 kcal/mol base pair opening enthalpy [3,4,14]. Therefore bond length and angle bending are assumed fixed in this work. Because our focus is on the opening of a single base, one can further assume that only the torsion angles associated with the opening in one base are allowed to change, the rest of the helix including the base aromatic ring is held rigid. These rotatable torsion angles are illustrated in Fig. 1.

Low energy opening pathway

As shown in Fig. 1, in order for a base to move freely out of the stacked helix, its effective rotation axis must be perpendicular to the base plane. Our analysis on both single and multiple torsion angle displacements indicates that only ζ(C3′-O3′-P-O5′) and δ(C5′-C4′-C3′-O3′) torsion angles satisfy this criterion. However, the rotation of δ causes angle bending distortion in C5′-C4′-O4′ or O3′-C3′-C2′. Hence ζ is the only energetically feasible torsion angle to drive base rotation. The variation of ζ along with fixed δ and ϵ(C4′-C3′-O3′-P) enables the base to rotate toward the major or minor grooves. The base can be kept in plane along this pathway by further adjusting the glycosidic bond torsion angle χ(O4′-C1′-N9-C8 for for purine and O4′-C1′-N1-C6 for pyrimidine). Because of its principal role in determining the pathway, ζ can be considered as the reaction coordinate for the single base opening described in this work.

Without any adjustment of the other backbone torsion angles, the change in ζ causes unphysical displacement of the backbone. In this work, the helix is kept rigid except for the section between the nth C3′ and $(n-1)$th C3′ (Fig. 1). Displacement of ζ in the nth base causes the above section to rotate. This results in the shift of the $(n-1)$th C3′ away from its original position, causing a drastic distortion in the $(n-1)$th sugar ring. Hence the adjustment of the other torsion angles along this section is necessary so as to move the $(n-1)$th C3′ back, close to its original position. This gives rise to a geometric requirement that guides the adjustment of rest of the torsion angles along the opening pathway. These angles, together with χ, can be regarded as adjustable variables controlled by the helix restoring forces. They can be determined by a search in the multidimensional torsion angle space to find a set of values that gives the least deviation in the $(n-1)$th C3′ coordinates. No energy minimization has been performed along the pathway. These angles have been found to change comparably with ζ.

An angular grid of 0.5° spacing is used to describe the torsion angle search space. Although the use of a finite grid size gives rise to a finite displacement in the computed new position of the C3′ atom, the grid size employed here seems to be optimum to yield converged results. A reduction of 50% in grid size results in $<10\%$ change in the computed opening extent $\Delta\zeta_{\max}$. The tolerance or the upper bound in the error can be estimated as follows: The maximum error in the computed position of the end atom in each of the torsion set is the average length of a covalent bond multiplied by the angular grid spacing $(=1.5\ \text{Å}\times 0.5°\times\pi/180°)$. There are seven torsion linkages connecting the two C3′ atoms of the nth and $(n-1)$th nucleotides. The maximum value for the

cumulative error or tolerance is then 7×0.014 Å$=0.1$ Å. Because of the partial cancellation among these errors, the computed deviation is expected to be smaller than this tolerance.

The backbone geometric restraint sets a limit on the extent of ζ rotation, $\Delta\zeta_{\max}$, and thus the opening extent. $\Delta\zeta_{\max}$ is determined from the maximum allowed deviation of the $(n-1)$th C3′ position due to the angular grid used in this study. A natural choice for the maximum allowed deviation is the tolerance derived above. The energy barrier along the pathway may also limit the extent of base opening. An upper limit of observed energy of 30 kcal/mol is used along with the geometric restraint to determine $\Delta\zeta_{\max}$.

Calculation of opening energy

The energy profile for the opening is given by the empirical functional form:

$$V=\sum_{\text{torsions}}\frac{V_n}{2}[1+\cos(n\phi-\gamma)]+\sum_{\text{H bonds}}[V_0(1-e^{-a(r-r_0)})^2-V_0]+\sum_{\text{nonbonded}}\left[\frac{A}{r_{ij}^{12}}-\frac{B}{r_{ij}^{6}}+\frac{q_iq_j}{\epsilon r_{ij}}\right], \quad (1)$$

where ϕ denotes a torsion angle, V_n, n, and γ are torsion potential parameters; r is the H-bond donor-acceptor distance, and V_0, a, and r_0 are H-bond potential parameters; A and B are nonbonded van der Waals parameters; ϵ is the dielectric constant, q_i and q_j are the charges of the ith and jth atoms, and r_{ij} is the distance between them. AMBER force field [11] parameters have been used for potential terms other than the H-bond term. Bond stretch and angle bending terms are excluded as they do not contribute to the low energy opening process. The torsion terms only include those backbone and glycosidic bond torsions that are associated with the opening of a single base. The stacking interactions are implicitly included in the nonbonded van der Waals and electrostatic energy terms. A distance-dependent dielectric constant [1] is used. Our results are relatively insensitive to the choice of dielectric constant.

The H-bond terms are based on Morse potential with implicit hydrogen atom. The Morse parameters in these terms are those used to predict H-bond premelting and melting in DNA [8]. In principle, the hydrogen bond energy can be computed by means of the explicit hydrogen atom models. However, in practice this is complicated by the difficulty in solving for the dynamics of the hydrogen atoms along base pair opening pathway and the requirement for more accurate force parameters and the positions of hydrogen atoms. This difficulty can be circumvented by using an empirical implicit hydrogen atom potential that gives reasonable energy for the hydrogen bonds. This potential, based on a Morse function of the donor and acceptor separation, has been shown to fit well with the potential energy obtained from *ab initio* calculations [15].

RESULTS AND DISCUSSION

The potential curves for the opening of bases in one of the B-DNA duplexes studied [16] is shown in Fig. 2. The opening pathway of a base in this duplex is illustrated in Fig. 3. Similar profiles have been found for all the other B-DNA systems. The energy barrier along the major groove of all the bases has been found to be less than 30 kcal/mol. Hence the major groove opening extent $\Delta\zeta_{\max}^{\text{major}}$ is exclusively determined by the geometric restraint. Further opening exceeds this constraint and thus most likely involves the cooperative displacement of neighboring nucleotides, which is consistent with the observed cooperative nature in DNA melting.

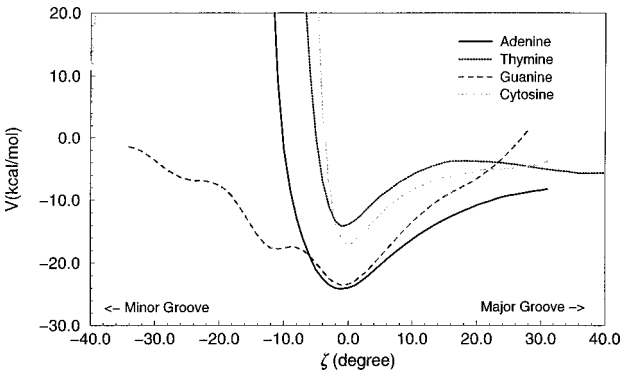

FIG. 2. Energy barrier for the opening of adenine, thymine, guanine, and cytosine bases in the crystal B-DNA duplex $d(\text{CGCGAATTCGCG})_2$ [16]. The potential curves are for $A5$, $T8$, $G2$, and $C3$ bases respectively and terminate at the maximum extent dictated by the geometric restraint.

In contrast, a large potential barrier is found along the minor groove of all bases except for guanines. Thus the minor groove opening extent, $\Delta\zeta_{\max}^{\text{minor}}$, for these bases is primarily determined by the energetic constraint. The high energy barrier arises from a steric clash between each of these bases and its complementary base on the opposite strand, which limits $\Delta\zeta_{\max}^{\text{minor}}$ of these bases to $4-10°$, as compared to that of $\sim45°$ given by the geometric restraint.

Guanines are exceptional because their O6-C6 bond is oriented in such a way that it avoids the close contact with the complementary base in the minor groove. Thus $\Delta\zeta_{\max}^{\text{minor}}$ of guanines is primarily determined by the geometric restraint.

Although the positions of neighboring nucleotides are fixed, surprisingly large amplitude ζ displacement and thus base opening can be found for majority of bases in B-DNA duplexes. Our results show that the average value of $\Delta\zeta_{\max}^{\text{major}}$ for all bases in the crystal B-DNA entries in Nucleic Acids

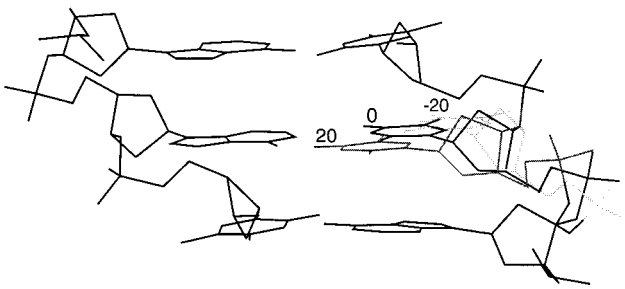

FIG. 3. Displacement of a base in a crystal B-DNA duplex with respect to the rotation of its ζ angle. The numbers in the figure indicate the rotation angles in degrees. The positive (negative) values correspond to rotation towards the major (minor) groove.

TABLE I. Energy barrier and maximum opening extent $\Delta\zeta_{max}^{major}$ for single base opening towards the major groove in DNA. For comparison the observed values are also included.

Method	System	Residue	$\Delta\zeta_{max}^{major}$ (deg)	Energy barrier (kcal/mol)		Ref.
				V_{total}	V_{VdW}	
This work	$d(CGCGAATTCGCG)_2$	G	28	22.2	17.3	
		C	31	14.3	9.5	
		A	29	11.1	10.3	
		T	52	10.0	14.0	
Fluorescence	$d(CTGAA^*TTCAG)_2$	A^*		17		[3]
NMR	$d(CGCAGATCTGCG)_2$	GC pair		17-26	13-17	[4]
		AT pair		17-21	14	[4]
NMR	Oligonucleotides	Average	18-29			[17]

Database (NDB) [12] is 35°, while that of the canonical B-DNA is 54°. For bases with an equilibrium ζ of $\sim -100°$, this corresponds to a displacement of 3.8 and 5.9 Å, respectively. This is sufficient to expose the imino proton to the solvent, which should facilitate imino proton exchange. The displacement completely reduces the stacking interactions of the base with one of its two neighbors. As shown in Table I, the calculated opening energy barriers are comparable to the observed enthalpies from NMR [4] and fluorescence [3] experiments.

The amplitude of the base angular motion in DNA has been estimated from the data obtained from NMR and optical experiments. The estimated spatial restriction for the base angular motion in several oligonucleotides is 29° based on the diffusion in a cone model and 18° based on the overdamped libration model [17]. These are comparable to our calculated opening extent. Thus our results on both the opening extent and energy barrier indicate that the single base opening described here correlates, at least partially, with the observed premelting base pair opening.

As shown in Fig. 2, because of similar energy barrier, guanines can be opened towards both the major and the minor groove. However, from Table I the energy barrier for guanines is substantially higher than that of other bases. Moreover three out of four bases can only open toward the major groove. Hence there is an overall preference for major groove opening.

A relatively small difference can be found in the potential parameters of different force fields, as these parameters are refined based on similar set of experimental data and quantum chemistry calculations. Therefore the effect of different force fields on the calculated energy landscape and the potential barrier is expected to be relatively small. In the minor groove, $\Delta\zeta_{max}^{minor}$ is determined by structural clash with complementary base. A small change in potential parameters thus has a small effect on its value. In the major groove, $\Delta\zeta_{max}^{major}$ is primarily determined by the geometric restraint, which is force field independent. Hence $\Delta\zeta_{max}^{major}$ and $\Delta\zeta_{max}^{minor}$ are expected to be relatively insensitive to the choice of potential parameters.

CONCLUSION

Our study reveals the existence of a principal torsion angle as the reaction coordinate for single base opening. The identification of this coordinate along with the accompanying geometric requirement on other torsions makes it possible to determine the microscopic pathway and energetics of this collective motion. Motions along this pathway may be further examined by other methods such as normal mode analysis. Investigation on multibase motions is also needed in order to fully explore base pair opening pathways. The proposed hypothesis is likely to find wider applications for various collective motions in other systems and further tests are warranted.

[1] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, New York, 1991); J. A. McCammon and M. Karplus, Nature (London) **268**, 765 (1977).

[2] M. Gueron, M. Kochoyan, and J.-L. Leroy, Nature (London) **328**, 89 (1987).

[3] D. Xu, K. O. Evans, and T. M. Nordlund, Biochemistry **33**, 9592 (1994).

[4] E. Folta-Stogniew and I. M. Russu, Biochemistry **33**, 11 016 (1994).

[5] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).

[6] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).

[7] Y. Z. Chen, W. Zhuang, and E. W. Prohofsky, Biopolymers **31**, 1273 (1991).

[8] Y. Z. Chen and E. W. Prohofsky, Phys. Rev. E **49**, 873 (1994).

[9] J. Ramstein and R. Lavery, J. Biomol. Struct. Dyn. **7**, 915 (1990).

[10] A. E. Garcia, C.-S. Tung, and J. A. Krumhansl, Biophys. J. **49**, 1239 (1986); G. Edwards and C. Liu, Phys. Rev. A **44**, 2709 (1991); S. M. Lindsay and J. W. Powell, in *Structure and Dynamics: Nucleic Acids and Proteins*, edited by E. Clementi

and R.H. Sarma (Adenine Press, New York, 1983), pp. 241–259.

[11] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. M. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[12] H. M. Berman, W. K. Olson, D. L. Beveridge, J. W. A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider, Biophys. J. **63**, 751 (1992).

[13] R. Chandrasekaran, and S. Arnott, in *Crystallographic and Structural Data II*, edited by W. Saenger, Landolt-Börnstein, New Series, Group VII, Vol. 1, Pt. b (Springer-Verlag, Berlin, 1989), pp. 31–170.

[14] J. G. Moe and I. M. Russu, Biochemistry **31**, 8421 (1992).

[15] N. C. Baird, Int. J. Quantum Chem. Quantum Biol. Symp. **1**, 49 (1974).

[16] R. E. Dickerson and H. R. Drew, J. Mol. Biol. **149**, 761 (1981).

[17] W. Eimer, J. R. Williamson, S. G. Boxer, and R. Pecora, Biochemistry **29**, 799 (1990).