Intrinsic Conformational Properties of Deoxyribonucleosides: Implicated Role for Cytosine in the Equilibrium Among the A, B, and Z Forms of DNA

Nicolas Foloppe and Alexander D. MacKerell, Jr.
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201 USA

ABSTRACT Structural properties of biomolecules are dictated by their intrinsic conformational energetics in combination with environmental contributions. Calculations using high-level ab initio methods on the deoxyribonucleosides have been performed to investigate the influence of base on the intrinsic conformational energetics of nucleosides. Energy minima in the north and south ranges of the deoxyribose pseudorotation surfaces have been located, allowing characterization of the influence of base on the structures and energy differences between those minima. With all bases, χ values associated with the south energy minimum are lower than in canonical B-DNA, while χ values associated with the north energy minimum are close to those in canonical A-DNA. In deoxycytidine, χ adopts an A-DNA conformation in both the north and south energy minima. Energy differences between the A and B conformations of the nucleosides are $<$0.5 kcal/mol in the present calculations, except with deoxycytidine, where the A form is favored by 2.3 kcal/mol, leading the intrinsic conformational energetics of GC basepairs to favor the A form of DNA by 1.5 kcal/mol as compared with AT pairs. This indicates that the intrinsic conformational properties of cytosine at the nucleoside level contribute to the A form of DNA containing predominately GC-rich sequences. In the context of a B versus Z DNA equilibrium, deoxycytidine favors the Z form over the B form by 1.6 kcal/mol as compared with deoxythymidine, suggesting that the intrinsic conformational properties of cytosine also contribute to GC-rich sequences occurring in Z DNA with a higher frequency than AT-rich sequences. Results show that the east pseudorotation energy barrier involves a decrease in the furanose amplitude and is systematically lower than the inversion barrier, with the energy differences influenced by the base. Energy barriers going from the south (B form) sugar pucker to the east pseudorotation barrier are lower in pyrimidines as compared with purines, indicating that the intrinsic conformational properties associated with base may also influence the sugar pseudorotational population distribution seen in DNA crystal structures and the kinetics of B to A transitions. The present work provides evidence that base composition, in addition to base sequence, can influence DNA conformation.

INTRODUCTION

DNA assumes a variety of conformations, with the equilibrium between those conformations depending on base composition, base sequence, and the environmental conditions (Franklin and Gosling, 1953; Pohl and Jovin, 1972; Leslie et al., 1980; Dickerson et al., 1982; Saenger, 1984; Guzikevich-Guerstein and Shakked, 1996). The best-characterized forms of DNA double helices are the A, B, and Z forms. Although the B form is prevalent under physiological conditions, the A form is also involved in biological processes (Wang et al., 1982; Ivanov and Minchenkova, 1995), while a biological role for the Z form is still elusive (Herbert and Rich, 1996). Many of the factors stabilizing the different forms of DNA have been identified from a phenomenological point of view (Saenger, 1984), but their detailed molecular basis is still intensively researched. In particular, the base dependency of DNA structure has aroused considerable interest.

Base dependency of DNA structure can manifest itself locally by causing conformational heterogeneities in a given form of DNA or globally by affecting the relative stabilities of the various forms of DNA. The latter was documented by studies using x-ray diffraction of DNA fibers (Davies and Baldwin, 1963; Langridge, 1969; Arnott et al., 1974, 1980; Arnott and Selsing, 1974a,b; Leslie et al., 1980) and a variety of spectroscopies (Bram, 1971; Bram and Tougard, 1972; Pilet and Brahms, 1972; Pohl and Jovin, 1972; Wang et al., 1989). The base dependence of DNA structure can a priori be related to base composition, base sequence, or both. Interpretation of experimental data alone, however, does not generally allow for a clearcut distinction between the respective contributions of base composition and base sequence on the relative stabilities of different forms of DNA. That base composition can influence the overall structure of DNA was explicitly concluded from some early studies (Bram, 1971; Pilet and Brahms, 1972), but these conclusions were called into question (Leslie et al., 1980), shifting the emphasis of subsequent studies to the role of base sequence (Calladine and Drew, 1984; Peticolas et al., 1988; Mazur et al., 1989; Hunter, 1993).

Detailed structural information obtained on oligonucleotides by single-crystal x-ray diffraction studies has moti-

Received for publication 9 December 1998 and in final form 23 March 1999.

Address reprint requests to Dr. Alexander D. MacKerell, Jr., School of Pharmacy, Room 554, University of Maryland, 20 N. Pine St., Baltimore, MD 21201. Tel.: 410-706-7442; Fax: 410-706-0346; E-mail: alex@mmiris.ab.umd.edu.

Nicolas Foloppe's current address is Center for Structural Biology, Department of Bioscience, Karolinska Institutet, S-141 57, Huddinge, Sweden.

Abbreviations used: A, adenine; C, cytosine; G, guanine; HF, Hartree-Fock; MP2, second-order Møller-Plesset; T, thymine.

0006-3495/99/06/3206/13 $2.00

vated the analysis of how some base sequences may favor one form of DNA or another. For instance, a "spine of hydration" has been observed bridging the purine N3 and pyrimidine O2 atoms in adjacent basepairs in the minor groove of some B DNA crystal structures (Drew and Dickerson, 1981; Kopka et al., 1983; Privé et al., 1991). These crystal structures suggest that this spine of hydration is disrupted by the guanine N2 amino group, which is consistent with NMR studies (Liepinsh et al., 1992). It has been proposed that these differences in hydration may account for the increased stabilization of the DNA B form by AT versus GC basepairs (Drew and Dickerson, 1981; Kopka et al., 1983). Another factor that may contribute to the particular stability of the B form with poly (dA) · poly(dT) is the formation of additional non-Watson-Crick hydrogen bonds between successive basepairs (Nelson et al., 1987). Differences in base stacking interactions have also been a major theme in trying to rationalize aspects of the sequence dependency on DNA conformation (Calladine and Drew, 1984; Mazur et al., 1989; Hunter, 1993; Alhambra et al., 1997). From these studies it is evident that the base sequence can influence DNA conformation via both environmental contributions, in the form of the spine of hydration, and intrinsic conformational properties associated with base stacking and intramolecular hydrogen bonds.

Based on the direct relationship between nucleoside conformation and overall DNA structure (Dickerson et al., 1982; Hartmann and Lavery, 1996), it is possible that base composition, in addition to base sequence, may affect the relative stabilities of different DNA forms. Such contributions may occur via the base influencing the conformational energetics of the individual nucleosides in DNA. At the nucleoside level, the various forms of DNA differ mainly by the deoxyribose pseudorotation angle and the glycosyl torsion. It is well-established that the deoxyribose furanose populates two ranges of pseudorotation angles in DNA and its components, commonly referred to as the north and south ranges (Altona and Sundaralingam, 1972; Davies, 1978; Dickerson et al., 1982; Gelbin et al., 1996; Hartmann and Lavery, 1996). The A and B forms of DNA are predominantly associated with the north and south ranges, respectively, and both ranges are equally populated in the Z form of DNA.

Values of the glycosyl linkage between the sugar and the base are highly correlated with the pseudorotation angle, with details of that correlation depending on the DNA structural family (Dickerson et al., 1982; Hartmann and Lavery, 1996). The glycosyl torsion populates different conformational ranges in the A and B forms of DNA, both corresponding to an *anti* orientation of the base relative to the sugar. Only the purines in Z DNA are systematically *syn* relative to the deoxyribose. The prevalence of the *anti* orientation of the base in DNA reflects the same trend in free nucleosides, as documented by experimental studies (Altona and Sundaralingam, 1972, 1973; Davies, 1978; Gelbin et al., 1996). These experimental observations also indicate that the energy difference between the deoxyribose north and south conformations should be small enough to account for the existence of both conformations. Condensed phase information from experimental approaches, however, includes possible contributions from solvent effects, crystal packing interactions, or additional degrees of freedom in the molecules, which can influence the properties of interest. It is thus difficult to derive the intrinsic conformational properties of the nucleosides solely from experimental studies, including statistical analysis of condensed phase data.

Theoretical studies can complement the wealth of information obtained from experiment by providing insights into the intrinsic conformational properties of the nucleosides, independent of condensed phase effects. To date, theoretical studies of standard nucleosides have been limited to calculations using hard-sphere models (Sundaralingam, 1969) or semiempirical quantum mechanical methods (Berthod and Pullman, 1971a, b; Saran et al., 1972, 1973). A variety of nucleic acid building blocks structurally related to nucleosides were also studied using hard-sphere models (Lakshminarayanan and Sasisekharan, 1970; Olson and Flory, 1972a) or empirical potential energies (Lakshminarayanan and Sasisekharan, 1969; Olson and Flory, 1972b; Olson, 1973; Yathindra and Sundaralingam, 1973). Although these studies have provided valuable insights into the gross conformational properties of DNA constituents, such coarse physical models cannot yield accurate intrinsic conformational properties. More recently, the structural and energetic properties of a nucleoside analog, where the base was mimicked by an imidazole group, were calculated ab initio at the MP2/6-31G* level of theory, and compared to a large body of experimental data (Foloppe and MacKerell, 1998). This comparison showed the MP2/6-31G* level of theory to yield conformational properties in satisfactory agreement with experiment for these types of compounds.

The present work systematically extends and refines the conclusions drawn from the study on the imidazole nucleoside analog by using ab initio calculations to examine the intrinsic conformational properties of nucleosides with the usual nucleic acid bases. The investigated conformational space has been selected to be relevant to that occurring in the A, B, and Z forms of DNA. Results indicate that the base influences the intrinsic structural and energetic properties of the corresponding nucleoside, with the largest effects occurring with deoxycytidine. These results indicate that cytosine contributes to GC-rich duplexes favoring the A and Z forms of DNA as compared to AT-rich duplexes. Furthermore, the present results indicate that base composition may also influence the kinetics of B-to-A transitions. Results also suggest that the nucleoside moieties in B DNA adopts a conformation that significantly deviates from the closest energy minimum, in contrast to A DNA.

METHODS

The atom names and dihedral angle nomenclature are as in Saenger (1984). Deoxyribose puckering pseudorotation angles (P) and amplitudes (τ) have been determined following Altona and Sundaralingam (1972) using the

same reference state for $P = 0.0°$. The pseudorotation space is divided into four equally sized quadrants centered around $P = 0.0°$, $P = 90.0°$, $P = 180.0°$, and $P = 270.0°$ that are referred to as the north, east, south, and west quadrants, respectively. The pseudorotation angles and amplitudes were extracted from the energy-minimized structures.

Quantum mechanical calculations were carried out with the GAUSSIAN 94 program (Frisch et al., 1996) using the 6-31G* basis set. All energy minimizations were first performed at the restricted HF level of theory. The HF energy-minimized structures were used as initial geometries for energy minimization at the second-order MP2 level of theory. Energy minimizations were performed to the default tolerances in the GAUSSIAN program. All degrees of freedom, other than those specified as being fixed, including all bond lengths and angles, were allowed to relax during the energy minimizations.

In the calculations aimed at locating the pseudorotation north and south energy minima, the initial furanose conformations were C3′endo and C2′endo, respectively. In these calculations the dihedral angles β, γ, ϵ, and χ were initially positioned in their conformational ranges observed in the A and B forms of DNA (t, g^+, t, and *anti*, respectively), where they remained during the energy minimizations without being constrained. The structures corresponding to the north and south energy minima have been used as starting structures for the energy minimizations where χ was fixed to an A DNA-like ($\chi = 201.1°$) or B DNA-like ($\chi = 258.1°$) conformation. These values of χ were taken from Table 2 in Hartmann and Lavery (1996) and correspond to the representation of the ideal A and B forms of DNA arrived at with the program CURVES (Lavery and Sklenar, 1989). Given the wide spread of B DNA χ values, it is important to note that the value of 258.1° is close to a mean value of 256° obtained from high-resolution crystal structures (Heinemann et al., 1994). It is also close to another mean value of 257° calculated from different set of B DNA crystal structures (Table 5 in Dickerson, 1992).

Energy minimizations of the nucleosides with the furanose in the O4′endo (east) conformation were carried out by fixing the dihedral angle C1′-C2′-C3′-C4′ to 0.0°. Energy minimizations of the nucleosides with a planar furanose (inversion barrier) were carried out with all the furanose endocyclic dihedral angles fixed to 0.0°. Normal mode frequency calculations were performed on the O4′endo HF/6-31G* optimized structures. In all cases a single negative frequency was observed, indicating the structures to represent true transition states.

Various potential energy differences were calculated to understand the relation of conformation to energetics in the studied nucleosides. $\Delta E_{\text{n-s}}$ is the energy of the north minimum minus the energy of the south minimum. $\Delta E_{\text{A-n}}$ is the energy of the A DNA-like conformation minus the energy of the north energy minimum. $\Delta E_{\text{B-s}}$ is the energy of the B DNA-like conformation minus the energy of the south energy minimum. $\Delta E_{\text{B-A}}$ is the energy of the B DNA-like conformation minus the energy of the A DNA-like conformation. *Be* is defined here as the energy of the O4′endo conformation minus that of the lowest in energy of either the north or the south energy minimum. *Bi* is defined here as the energy of the planar furanose conformation minus that of the lowest in energy of either the north or the south energy minimum. Be^{B} is the energy of the O4′endo conformation minus that of the B DNA-like conformation ($\chi = 258.1°$), and Be^{A} is the energy of the O4′endo conformation minus that of the A DNA-like conformation ($\chi = 201.1°$).

Distributions of the glycosyl torsions in crystal structures of DNA duplexes were obtained from the Nucleic Acids Database (Berman et al., 1992) as of March 1998. Structures excluded from the distributions were those containing nonstandard DNA components, bound drugs, or proteins. The results are presented as probability distributions and have been obtained separately for the A, B, and Z DNA families by sorting the data into 2° bins. The mean and standard deviation of each bell-shaped part of the distribution were obtained by best fit to a Gaussian function.

RESULTS AND DISCUSSION

The present calculations were performed on the nucleosides of the four standard DNA bases (deoxyadenosine, deoxycytidine, deoxyguanosine, and deoxythymidine). Calculations were also performed with deoxyuridine, for the sake of completeness and because it has been studied experimentally (Altona and Sundaralingam, 1973; Guschlbauer and Jankowski, 1980). Because the base moiety is relatively rigid and its structure unaffected by the sugar conformation (Clowney et al., 1996), the structural properties discussed in the following are only those of the deoxyribose moiety and the glycosyl linkage, although the base geometries were allowed to relax in all energy minimizations. Since many of the general aspects of the conformational properties of deoxyribose and the glycosyl linkage that are relevant to DNA structure have already been discussed (Altona and Sundaralingam, 1972; Davies, 1978; Westhof and Sundaralingam, 1980; Olson and Sussman, 1982; Olson, 1982; Gelbin et al., 1996; Foloppe and MacKerell, 1998) the present discussion specifically emphasizes the influence of the base on these properties.

Calculations performed in the present study provide potential energies in vacuo and not free energies in the condensed phase. The merits and limitations of this approach have already been assessed in a similar context (Foloppe and MacKerell, 1998), showing that this type of calculation can provide insights into the intrinsic properties of the nucleosides and their counterparts in DNA. Some results from the previous study with an imidazole nucleoside analog are listed alongside the present results on the standard nucleosides to facilitate their comparison.

The first step of the present study was aimed at locating the intrinsic energy minima of the deoxyribonucleosides and investigating the influence of base on these energy minima. In the second step, the contribution of the nucleosides to the intrinsic conformational properties of the A and B forms of DNA was investigated by constraining the glycosyl torsion to characteristic canonical values. In the third step, calculations to investigate possible differences between deoxycytidine and deoxythymidine with respect to the stabilization of Z DNA were performed. In a fourth step, the nucleosides served as a basis to identify and characterize the energy barriers separating the deoxyribose north and south conformations.

Influence of the base on the nucleoside intrinsic energy minima

Differences between the calculated and crystal deoxyribose bond lengths and valence angles, in both the north and south energy minima, are summarized in Table 1. These differences are all small, confirming that calculations at the MP2/6-31G* level yield structures in satisfactory agreement with experiment for this type of compound, as previously discussed (Foloppe and MacKerell, 1998). The individual deoxyribose bond lengths are given in the Appendices (Table A1, north energy minimum; Table A2, south energy minimum), as well as the individual valence angles (Tables A3, north energy minimum; Table A4, south energy mini-

TABLE 1 Average differences between calculated and crystal deoxyribose bond lengths and valence angles

	Bonds (Å)		Angles (deg.)	
Base	North*	South*	North*	South*
Cytosine	0.005	0.005	1.4	1.1
Thymine	0.005	0.007	1.3	1.0
Uracil	0.005	0.007	1.3	1.0
Adenine	0.006	0.007	1.2	1.0
Guanine	0.007	0.007	1.1	0.9
Imidazole[#]	0.007	0.008	1.0	1.0

The differences are averages of the individual differences between the calculated bond lengths (or valence angles) and their mean crystal counterpart (Gelbin et al., 1996).

*North and south refer to the deoxyribose conformation.

[#]Imidazole refers to the nucleoside analog used in a previous study, where the base was mimicked by an imidazole moiety (Foloppe and MacKerell, 1998).

mum). The variations in the individual bond lengths and valence angles associated with different bases are small, making them poor descriptors of the influence of the base on the deoxyribose conformation.

Results on the influence of the base on the deoxyribose pseudorotation angle and amplitude are presented in Table 2. The north (P_n) and south (P_s) pseudorotation energy minima fall within the expected ranges (Altona and Sundaralingam, 1972; Dickerson et al., 1982; Gelbin et al., 1996), but their exact location is influenced by the base. P_s is located at higher values for the purines than for the pyrimidines, which is compatible with the corresponding shift observed in the pseudorotation angle distributions obtained from crystal structures of DNA (Drew et al., 1981; Dickerson et al., 1982). P_n spans a range of values (7.0–12.4°) with variations similar to that observed with P_s (162.1–168.6°). The differences in P_n values, however, do not follow any trend in terms of purines versus pyrimidines, in contrast to P_s.

The north amplitudes of puckering (τ_n, Table 2) are systematically higher than the south ones (τ_s). Although this trend has already been noted by examining crystal structures of nucleosides and nucleotides, the amplitudes in these crystal structures are distributed over a broad range of values (Gelbin et al., 1996). It is therefore difficult to conclude solely from crystal structures that the observed trend reflects an intrinsic property of the molecules in question. The present calculations confirm such an interpretation.

TABLE 2 Descriptors related to the structure and energetics of the nucleosides with the deoxyribose in the north and south energy minima

Base	P_n	P_s	τ_n	τ_s	χ_n	χ_s	ΔE_{n-s}
Cytosine	8.8	162.1	39.4	37.4	194.8	207.0	−0.3
Thymine	12.4	162.7	39.3	37.7	197.9	230.6	0.9
Uracil	12.4	162.9	39.2	37.7	197.4	230.2	0.8
Adenine	7.0	168.3	39.7	36.7	192.3	230.2	0.4
Guanine	9.6	168.6	39.4	36.7	198.2	232.8	0.7
Imidazole*	13.1	167.2	38.6	37.0	205.2	234.8	1.2

P (deg.), τ (deg.), and χ (deg.) refer to the furanose pseudorotation angle, the furanose amplitude, and the glycosyl torsion, respectively. Subscripts n and s refer to the north and south energy minima, respectively. ΔE_{n-s} (kcal/mol) is the energy of the north minimum minus the energy of the south minimum.

*Imidazole refers to the nucleoside analog used in a previous study, where the base was mimicked by an imidazole moiety (Foloppe and MacKerell, 1998).

The calculated values of the glycosyl torsion in the north energy minima (χ_n, Table 2) are in excellent agreement with their counterparts obtained from crystal structures of nucleosides and nucleotides (193.3 ± 14.0° for purines, 14.0 is the standard deviation; 195.7 ± 6.6° for pyrimidines) (Gelbin et al., 1996). This agreement is better than when χ was calculated using the imidazole nucleoside analog. For the south sugar conformation the glycosyl torsions, χ_s, for deoxyadenosine, deoxyguanosine, deoxythymidine, and deoxyuridine fall within the experimental ranges (237.0 ± 24.3° for purines and 229.8 ± 18.4° for pyrimidines), the largest departure being 6.8° for deoxyadenosine. The calculated χ_s value with deoxycytidine, however, is ~20° lower than χ_s calculated with the other bases and is more similar to the values associated with the A form (201.1°) versus the B form (258.1°) of DNA (Hartmann and Lavery, 1996). This property of deoxycytidine suggested that the intrinsic conformational properties of cytosine may favor the A form of DNA, stimulating additional calculations to investigate this possibility (see below).

Differences in the potential energy between the north and south energy minima (ΔE_{n-s}, Table 2) were found to be modulated by the base, but are $<\pm 1.0$ kcal/mol with all bases. ΔE_{n-s} favors the south energy minimum with adenine, guanine, thymine, and uracil, but not cytosine (Fig. 1 *a*). This mirrors the experimental populations obtained by NMR for deoxyribonucleosides in solution (Altona and Sundaralingam, 1973; Davies, 1978; Uesugi et al., 1979; Guschlbauer and Jankowski, 1980), which indicate that the deoxyribose south conformer is generally more favored than the north. A recent statistical survey of the crystal structures of deoxyribo-based monomers (Gelbin et al., 1996) also reported a majority of them to have a south pucker. This survey, however, did not distinguish between nucleosides and nucleotides or the different bases. The present calculations are also compatible with the previous observation (Altona and Sundaralingam, 1973) that cytosine may favor the north pucker significantly more than uracil.

The calculated values of ΔE_{n-s}, in Table 2, however, do not follow any trend in terms of purine versus pyrimidine, which is in contrast with previous analysis of NMR data in solution (Altona and Sundaralingam, 1973; Davies, 1978; Uesugi et al., 1979; Guschlbauer and Jankowski, 1980). Those NMR studies indicated that purines favor the south pucker more than pyrimidines, although only by a small margin. This discrepancy may reflect the limitations of the present calculations, difficulties in interpreting the NMR data, or a combination of both. It must be reiterated that the present calculations yield potential energies in vacuo and, therefore, cannot be rigorously compared with experimental distributions that reflect free energies in the condensed

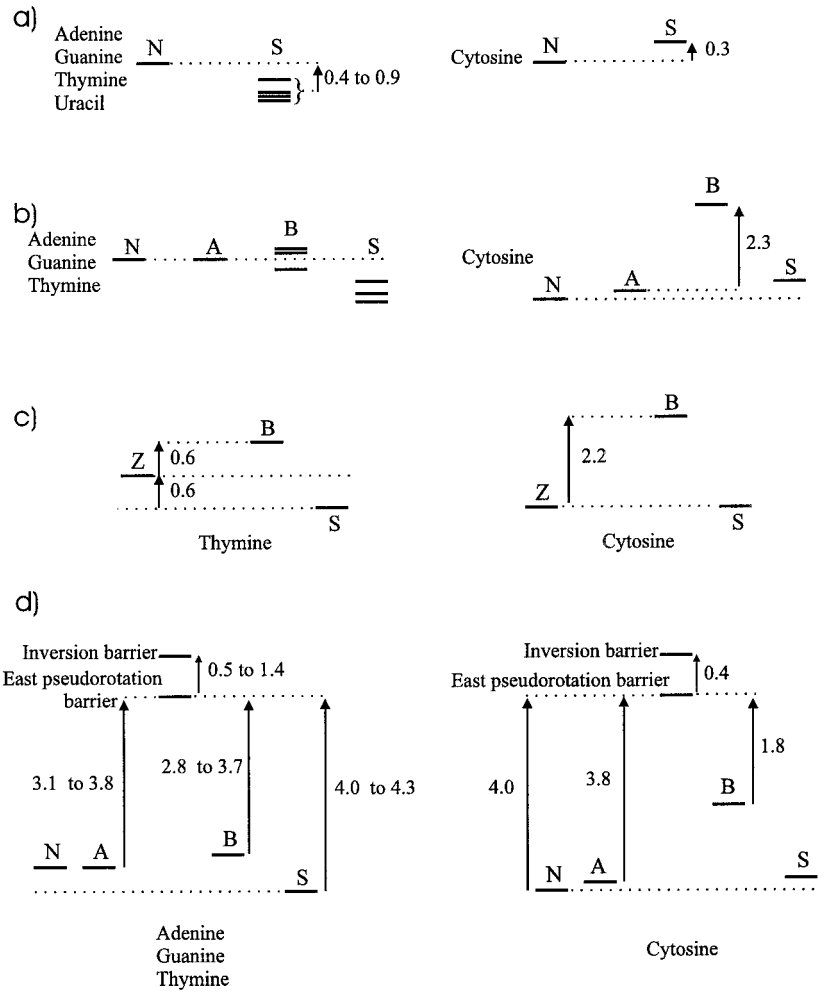

FIGURE 1 Schematic representation of selected relative potential energies (kcal/mol) associated with various conformations of the deoxyribonucleosides. The right-hand diagrams correspond to deoxycytidine, and the left-hand diagrams correspond to the other nucleosides, as noted. The conformations represented are the north minimum (N), the south minimum (S), the A DNA-like conformation (A), the B DNA-like conformation (B), the Z DNA-like conformation (Z), the east pseudorotation barrier, and the inversion barrier. (*a*) Energy of the south minimum relative to that of the north minimum (the various energy levels in the south minimum represent different bases, see Results and Discussion and Table 2). (*b*) Energies of the A and B DNA-like conformations relative to that of the north and south energy minima (the various energy levels in the B DNA-like, and south minimum represent different bases, see Results and Discussion and Table 3). (*c*) Energy of the Z DNA-like conformation relative to that of the B DNA-like conformation, and south energy minimum (see Results and Discussion). (*d*) Energy of the east pseudorotation barrier, and inversion barrier relative to each other, and relative to the global energy minimum, and the A and B DNA-like conformations (see Results and Discussion and Table 4).

phase. It is worth noting that NMR data represent averages over a large number of different conformers, reflecting the flexibility of all the furanose substituents. NMR-based conformational analysis of the deoxyribose in nucleosides is still an area of active research (Bandyopadhyay et al., 1997).

The results presented in Table 2 show the conformations of the nucleosides in their north and south energy minima to be in the vicinity of the conformations that occur in A and B DNA, respectively. Some systematic differences, however, do exist between the glycosyl torsion in the nucleosides south energy minima and in the B form of DNA (see next section). This implied that $\Delta E_{n\text{-}s}$ values cannot be directly used to understand intrinsic conformational contributions to the A versus B forms of DNA. To obtain a more realistic estimate of the intrinsic energy contributions of the nucleoside moieties, they were energy-minimized with constraints on the glycosyl torsion, as described in the next section.

Contribution of the nucleoside moieties to the intrinsic conformational energy of the A and B forms of DNA

The results shown in Table 2 and discussed in the previous section indicate that χ_n values are significantly closer to their A DNA counterpart as compared to χ_s values and their B DNA counterparts. For the glycosyl linkage $\chi_A = 201.1°$ and $\chi_B = 258.1°$, values have been proposed as being characteristic of the A and B forms of DNA, respectively (Hartmann and Lavery, 1996). These χ_A and χ_B values are compatible with the corresponding χ distributions for A and B DNA crystal structures obtained from the nucleic acids database as of March 1998 (Fig. 2). Differences between χ_n values in Table 2 and χ_A range from 2.9° to 8.8°. Excluding cytosine, the differences between χ_s values in Table 2 and χ_B are significantly larger, from 25.3° to 27.9°. The difference between χ_s and χ_B for cytosine is 51.1°. The χ distribution in crystal structures of nucleosides and nucleotides is also shifted to lower values (Gelbin et al., 1996) than in B DNA (Fig. 2). To investigate these differences and their possible contribution on the energetics of the B form of DNA, the nucleosides have been energy-minimized with $\chi_B = 258.1°$ and the furanose in the south range. To allow for a systematic comparison between the different nucleosides in an A DNA-like conformation, they have also been energy-minimized with χ constrained to $\chi_A = 201.1°$ and the furanose in the north range. Selected properties of the nucleosides when energy minimized in the A and B DNA-like conformations are given in Table 3.

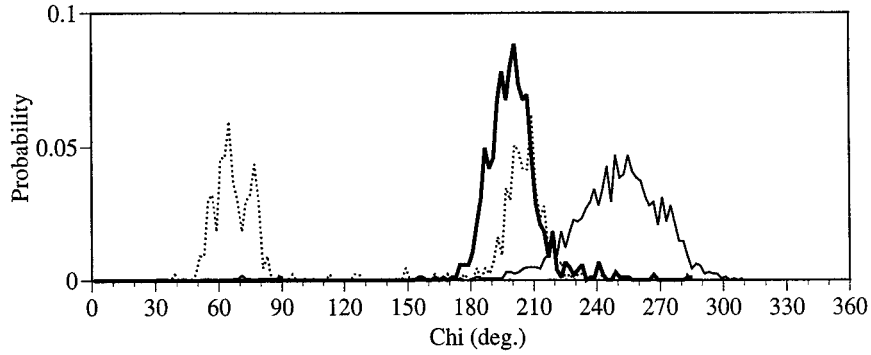

FIGURE 2 Probability distribution of the glycosyl torsion χ (chi) in crystal structures of A DNA (*thick line*), B DNA (*thin line*), and Z DNA (*dotted line*). The mean (m), and standard deviation (σ) for each distribution are: for A DNA, m = 199.°, σ = 9°; for B DNA, m = 252°, σ = 19°; for Z DNA anti, m = 206°, σ = 8°; for Z DNA syn, m = 66°, σ = 9°. Sample sizes were 1053, 1648, and 436 for A DNA, B DNA, and Z DNA, respectively.

The furanose pseudorotation angles in the B DNA conformation (P_B, Table 3) are systematically higher than the corresponding P_s values (Table 2) by an average of 5.3° for both the purines and pyrimidines. Differences between the pseudorotation angle in the A DNA conformation (P_A, Table 3) and the corresponding P_n values (Table 2) are all 3.9° or less, with P_A systematically closer to the canonical C3′endo conformation (p = 18.0°) than P_n. The furanose amplitudes in the A and B DNA-like conformations (τ_A and τ_B, respectively; Table 3) remain very close to τ_n and τ_s values (Table 2), respectively.

Constraining χ to a B DNA-like value increases the energy of the nucleosides relative to their south energy minimum (see $\Delta E_{B\text{-}s}$, Table 3). $\Delta E_{B\text{-}s}$ is 1.2 kcal/mol or less with adenine, guanine, and thymine, but is 2.2 kcal/mol with cytosine. Constraining χ to an A DNA-like value leads to small increases of 0.2 kcal/mol or less in the energy of the nucleosides relative to their north energy minimum (see $\Delta E_{A\text{-}n}$, Table 3). The $\Delta E_{A\text{-}n}$ values being less than the corresponding $\Delta E_{B\text{-}s}$ values is consistent with the better agreement of χ_n values with the canonical A form values as compared to χ_s and the canonical B form values.

For each nucleoside, the energy differences between the B and A DNA-like conformations ($\Delta E_{B\text{-}A}$) are also given in Table 3. With deoxyadenosine, deoxyguanosine, and deoxythymidine, $\Delta E_{B\text{-}A}$ values are small (<0.5 kcal/mol), suggesting that the intrinsic conformational energies of these DNA building blocks do not significantly favor either the A or B forms of DNA. In contrast, the value of $\Delta E_{B\text{-}A}$ for deoxycytidine (2.3 kcal/mol) suggests that the intrinsic conformational energetics of cytosine favors the A form of DNA over the B form. These results are summarized graphically in Fig. 1 *b*. It has recently been suggested that the nucleoside moiety in DNA may stabilize the B form relative to the A form by ~1.0 kcal/mol, based on the relative energies of a nucleoside analog in its north and south energy minima (Foloppe and MacKerell, 1998). Given the present results, this previous conclusion must be revised and replaced by the proposal that the nucleosides intrinsic conformational energy does not significantly favor either the A or the B form of DNA, except in the case of deoxycytidine, which favors the A form.

TABLE 3 Descriptors related to the structure and energetics of the nucleosides with χ in A and B DNA-like conformations

Base	P_A	P_B	τ_A	τ_B	ΔE_{A-n}	ΔE_{B-s}	ΔE_{B-A}
Cytosine	12.7	166.9	39.1	37.3	0.2	2.2	2.3
Thymine	12.9	167.9	39.0	37.2	<0.1	1.2	0.3
Adenine	9.3	173.8	39.2	36.2	<0.1	0.8	0.4
Guanine	10.1	174.5	39.1	36.2	<0.1	0.6	−0.1

P (deg.) and τ (deg.) refer to the furanose pseudorotation angle and amplitude, respectively. Subscripts A and B refer to the conformations with the A DNA-like and B DNA-like glycosyl torsions, respectively. ΔE_{A-n} (kcal/mol) is the energy of the A DNA-like conformation minus the energy of the north energy minimum. ΔE_{B-s} (kcal/mol) is the energy of the B DNA-like conformation minus the energy of the south energy minimum. ΔE_{B-A} (kcal/mol) is the energy of the B DNA-like conformation minus the energy of the A DNA-like conformation.

Combining $\Delta E_{B\text{-}A}$ values for the standard DNA basepairing suggests that the internal conformational energy of a GC basepair ($\Delta E^{GC}{}_{B\text{-}A}$ = 2.2 kcal/mol) may stabilize the A form of DNA more than AT ($\Delta E^{AT}{}_{B\text{-}A}$ = 0.7 kcal/mol) by 1.5 kcal/mol. This energy difference would be even more significant in the context of GC-rich DNA duplexes. These results are compatible with a number of experimental studies that indicate that GC-rich DNA duplexes are more amenable to the A form than AT-rich duplexes (Bram, 1971; Bram and Tougard, 1972; Pilet and Brahms, 1972; Arnott and Selsing, 1974a, b; Arnott et al., 1974; Leslie et al., 1980; Peticolas et al., 1988; Wang et al., 1989; Ivanov and Minchenkova, 1995).

A factor commonly invoked to rationalize this phenomenon is the presence of a "spine of hydration" in the minor groove of AT sequences in B DNA (Drew and Dickerson, 1981; Kopka et al., 1983). Experiments where guanine is replaced by inosine have supported the hypothesis that the disruption of the spine of hydration by the guanine N2 amino group facilitates transitions from the B form of DNA to the A form (Langridge, 1969; Leslie et al., 1980). The energetics associated with a spine of hydration in the DNA minor groove, however, remains unclear. It has been argued that the spine of hydration stabilizes the B form of some, but not all, AT sequences (Chuprina, 1985, 1987). While the spine of hydration most likely contributes to the conformational equilibrium between the A and B forms of DNA, the present calculations indicate that the deoxyribonucleosides'

intrinsic conformational energy also makes a significant contribution to this equilibrium.

Cytosine versus thymine in Z DNA

Z DNA typically occurs in alternating purine-pyrimidine sequences dominated by GC basepairs (Pohl and Jovin, 1972; Wang et al., 1979; Arnott et al., 1980; Drew et al., 1980; Saenger, 1984). Z DNA can contain AT basepairs (Arnott et al., 1980; Wang et al., 1987a; Vorlickova et al., 1982; Zimmer et al., 1982); however, the ability to assume the Z form becomes increasingly difficult as the GC content decreases (Pohl and Jovin, 1972; Ho et al., 1986; Peticolas et al., 1988; Wang et al., 1984, 1987b). Why GC basepairs accommodate the Z form of DNA better than AT basepairs remains unclear. There is evidence that guanine has a higher propensity than adenine to adopt a *syn* orientation relative to the sugar in 5′-nucleotides (Son et al., 1972; Yathindra and Sundaralingam, 1973; Olson, 1973), and the relationship between this difference and the characteristic *syn* orientation of the purines in Z DNA has been noted (Saenger, 1984). The present calculations suggest that the intrinsic conformational properties of deoxycytidine as compared to deoxythymidine contribute to GC pairs being more amenable than AT pairs with respect to the Z form of DNA.

In Z DNA, the pyrimidine nucleosides combine a south sugar with a glycosyl torsion much lower than in B DNA, populating a range overlapping that observed in A DNA (Fig. 2). Values of the glycosyl torsion in pyrimidines range from ~199 (Z_I) to 213° (Z_{II}) (Hartmann and Lavery, 1996). These values correspond closely to the structural properties obtained in the present calculations for deoxycytidine in its south energy minimum (Table 2), where $\chi_s = 207.0°$, while with deoxythymidine $\chi_s = 230.6°$. This suggests that deoxycytidine is intrinsically more prone than deoxythymidine to adopt the conformation typical of pyrimidines in Z DNA. When deoxythymidine is energy minimized with a south pucker ($P = 161.5°$, unconstrained) and χ constrained to the value obtained for deoxycytidine, 207.0°, its energy is 0.6 kcal/mol above its intrinsic south energy minimum.

A more appropriate comparison of the energetics of deoxycytidine and deoxythymidine would be in the context of a B versus Z DNA equilibrium. The energy of the B conformation (see above) minus that of the Z conformation ($\chi = 207.0°$) is 2.2 kcal/mol for deoxycytidine (identical to $\Delta E_{B\text{-}s}$ in Table 2) and 0.6 kcal/mol for deoxythymidine. This suggests that the intrinsic conformational energies of both deoxycytidine and deoxythymidine favor the Z form of DNA over the B form, with deoxycytidine favoring the Z form by 1.6 kcal/mol more than deoxythymidine (Fig. 1 *c*). Thus, the intrinsic conformational energetics of cytosine at the nucleoside level will be more amenable to the Z form than thymine, contributing to the composition of Z DNA being dominated by GC basepairs.

That cytosine intrinsically favors the Z DNA conformation more than thymine raises further questions concerning the effect of 5-methylation on the intrinsic energetics of 5-methyl-deoxycytidine, as compared to deoxycytidine. Indeed, methylation of cytosine at position 5 induces the B to Z DNA transition at lower salt concentration than for poly-(dGdC) (Behe and Felsenfeld, 1981). Proposed mechanisms for the promotion of the B to Z DNA transition by 5-methylation of cytosine have focused on base stacking interactions and solvation properties (Wang and Kool, 1995, and references therein). It is, therefore, interesting to test whether the nucleoside intrinsic energetics contribute to this effect, i.e., is the energy difference between the B and Z conformations higher in 5-methyl-deoxycytidine than in deoxycytidine? To investigate this possibility, 5-methyl-deoxycytidine was energy-minimized in the south minimum (corresponding to a Z DNA conformation for deoxycytidine) and the B DNA-like conformation. The south minimum ($P_s = 162.4°$) of 5-methyl-deoxycytidine is associated with a χ value of 208.7°, very similar to that observed with deoxycytidine. This confirms the special character of cytosine with respect to the preferred glycosyl torsion when associated with a south deoxyribose. As with deoxycytidine, the south energy minimum of 5-methyl-deoxycytidine corresponds to a Z DNA conformation. In 5-methyl-deoxycytidine, the energy difference between the south energy minimum and the B DNA-like conformation is 2.5 kcal/mol, marginally higher than the corresponding 2.2 kcal/mol calculated for deoxycytidine. It is interesting to note that, although 5-methylation renders cytosine structurally more analogous to thymine, it does not bring $\Delta E_{B\text{-}s}$ (see above) closer to the deoxythymidine value (1.2 kcal/mol), but rather slightly increases $\Delta E_{B\text{-}s}$ as compared to deoxycytidine. Although this increase may contribute to the effect of cytosine 5 methylation on the promotion of the B to Z DNA transition, it is not significant enough to fully rationalize it.

Energy barriers between the deoxyribose north and south puckering ranges

Ab initio calculations provide a powerful tool to probe the energy barriers between the deoxyribose north and south conformational ranges. Previous calculations indicate that the west pseudorotation energy barrier is very unlikely to be the path of lowest energy between the deoxyribose north and south puckering ranges, particularly in DNA (Saran et al., 1973; Olson, 1982; Foloppe and MacKerell, 1998). Therefore, the pseudorotation west energy barrier is not discussed further in the present work. In previously reported calculations using the imidazole nucleoside analog, the O4′endo conformation of the furanose was found to be a good approximation of the east pseudorotation barrier (*Be*), and the inversion barrier (*Bi*) was probed by constraining the furanose to a planar conformation (Foloppe and MacKerell, 1998). These approximations have also been used in the present study. With the imidazole nucleoside analog, *Be* was found to be only 0.6 kcal/mol lower than *Bi*, which prompted the calculation of both quantities for the four DNA nucleosides and deoxyuridine in the present work.

Results on the two barriers to pseudorotation are presented in Table 4 for the five nucleosides. *Bi* is higher in energy than *Be* for all the deoxyribonucleosides (Fig. 1 *d*). The energy difference between *Bi* and *Be*, however, varies with the base, the smallest difference being 0.4 kcal/mol (cytosine), and the highest being 1.4 kcal/mol (thymine). This further documents that cytosine and thymine impart different intrinsic properties at the nucleoside level. The present calculations confirm that the path of lowest energy between the north and south puckers is likely to be through the east pseudorotation barrier. In view of the relatively small differences between *Bi* and *Be* it cannot be concluded that the inversion mechanism is systematically forbidden as a pathway between the deoxyribose north and south puckering ranges.

Calculations on the imidazole nucleoside analog suggested that the furanose amplitude at the east energy barrier (τ_e, Table 4) may be significantly lower than in the north and south energy minima (Foloppe and MacKerell, 1998). The present calculations confirm that the O4′endo conformation in standard deoxyribonucleosides is also associated with a significant flattening of the furanose ring and suggest that this flattening is more pronounced with purines than with an imidazole analog (Table 4). τ_e is ~4.0° lower with adenine and guanine than when these purines are mimicked by an imidazole. τ_e is 5.0 to 6.0° higher with the pyrimidines than with the purines. These results illustrate further the need to explicitly study the sugar with the actual nucleic acids bases when details of the sugar structural features are to be characterized.

Given the differences in τ_e as a function of base, it is surprising that *Be* shows very little dependence on the base (Table 4), being ~4.0 kcal/mol with all bases. This suggests that there is no simple relationship between the degree of flattening of the furanose ring and the height of the east pseudorotation barrier. The present calculations are compatible with an experimental estimate of 4.7 ± 0.5 kcal/mol for the activation energy separating the north and south furanose puckers in purine ribonucleosides (Röder et al., 1975). The same experimental work, however, also led to the proposal that this activation energy should be higher with pyrimidines than with purines, which is not supported by the results of the present calculations. Those experimental studies, however, were carried out with ribonucleosides, while the present results have been obtained on deoxyribonucleosides.

The glycosyl torsion in the east pseudorotation barrier shows minimal base-dependent properties, except with cytosine, and the glycosyl torsion for the inversion barriers, χ_i, are higher with the purines than the pyrimidines, as shown in Table 4. In the east quadrant, the χ_e values are intermediate to the values seen in the north and south quadrants (Table 2). This suggests that the furanose pseudorotation angle and χ may change conformations in a concerted fashion during a B to A DNA transition. Notable is the low value of χ_e with cytosine, consistent with the low value of χ in the south conformation for this nucleoside.

In the context of a B to A DNA transition, the east pseudorotation energy barrier is different from *Be* due to the energy of the nucleosides in a B DNA-like conformation being higher than in their south energy minimum (see above). The east pseudorotation energy barriers relative to the energy of the B DNA-like conformations (Be^B), and relative to the energy of the A DNA-like conformations (Be^A) are listed in Table 4. The base dependence of Be^B values parallels the $\Delta E_{B\text{-}s}$ values (Table 3). The heterogeneity in Be^B values contrasts the relative homogeneity of *Be* values, and may be of importance for the base dependency of DNA dynamics. Interestingly, the lowest Be^B value is for deoxycytidine, suggesting that cytosine may not only stabilize A DNA versus B DNA, but may also kinetically facilitate B to A DNA transitions. Also of note is that the Be^B values of the pyrimidines are lower than for the purines, consistent with the pyrimidine pseudorotation angle distribution extending more into the east quadrant than with purines in B DNA (Hartmann and Lavery, 1996; Drew et al., 1981; Foloppe and MacKerell, 1998).

TABLE 4 Structural descriptors and energetics of the nucleosides in O4′endo and planar furanose conformations

Base	τ_e	χ_e	χ_i	*Be*	*Bi*	Be^B	Be^A
Cytosine	19.8	200.5	191.8	4.0	4.4	1.8	3.8
Thymine	20.5	223.6	194.3	4.0	5.4	2.8	3.1
Uracil	20.4	222.2	193.8	4.0	5.2	na	na
Adenine	14.3	219.6	225.7	4.2	4.7	3.4	3.8
Guanine	14.4	221.5	231.8	4.3	4.8	3.7	3.6
Imidazole	18.1	220.6	238.3	4.4	5.0	na	na

τ (deg.) and χ (deg.) refer to the furanose amplitude and the glycosyl torsion, respectively. Subscripts e and i refer to the east pseudorotation energy barrier and the inversion energy barrier. *Be* (kcal/mol) and *Bi* (kcal/mol) are the east pseudorotation and the inversion energy barriers, respectively, relative the global (north or south) energy minimum. Be^B (kcal/mol) and Be^A (kcal/mol) are the east pseudorotation energy barriers relative to the B DNA-like and A DNA-like conformations, respectively. Be^B and Be^A are not available (na) for deoxyuridine and the imidazole nucleoside analog used in a previous study (Foloppe and MacKerell, 1998).

CONCLUSIONS

This is the first report presenting high-level ab initio calculations on all the standard deoxyribonucleosides. Only a fraction of the conformational space accessible to these molecules has been probed, but the investigated conformations have been selected to be relevant to a DNA context. The results document how the intrinsic conformational properties of deoxyribonucleosides are influenced by the base and identify base composition as an important factor contributing to the overall structure of DNA. The overall contribution of these base-dependent conformational properties is difficult to calibrate considering the different factors influencing DNA structure; however, these contributions could be significant, especially in the context of accumulating information suggesting that DNA sequence-dependent features result from a combination of rather

subtle effects (McCall et al., 1985; Nelson et al., 1987; Privé et al., 1991; Poncin et al., 1992; Dickerson et al., 1994; Grzeskowiak et al., 1991; Subirana and Faria, 1997).

This work confirms the main conclusions previously obtained based on an imidazole nucleoside analog, but complements and refine them by taking the influence of base into account. Some of the deoxyribose structural properties depend on the base being a purine, versus a pyrimidine. These structural properties are the furanose pseudorotation angle in the south range (P_s and P_B) being higher with purines than with pyrimidines, and the furanose amplitude of puckering in the O4′endo conformation (τ_e) being higher with pyrimidines than with purines. On the other hand, the furanose pseudorotation angle in the north range (P_n and P_A) shows some dependence on the base, but without any trend in terms of purine versus pyrimidines. Some other deoxyribose structural properties show little, if any, influence of the base. These properties include the furanose amplitudes in the north (τ_n and τ_A) and south (τ_s and τ_B) ranges and the glycosyl torsion associated with the north energy minimum (χ_n). The glycosyl torsions associated with the south energy minimum (χ_s) and with the east pseudorotation energy barrier (χ_e) also show little dependence on the base, except with cytosine.

Deoxycytidine strikingly differs from the other nucleosides by its χ_s value, which falls in the range populated by the glycosyl torsion in A and Z DNA. Deoxycytidine is also the only nucleoside for which the north minimum is lower in energy than the south. This strongly suggests that deoxycytidine is intrinsically more prone to accommodate the A form of DNA as compared to the other deoxyribonucleosides. With the other nucleosides, χ_s values fall in a range ($\sim$230.0°) which is intermediate between the χ values characteristic of the canonical A and B forms of DNA, and the south energy minimum is slightly more stable than the north, by $<$1.0 kcal/mol.

When the deoxynucleosides, except deoxycytidine, are energy-minimized in a B DNA-like conformation by constraining χ accordingly, their internal energy increases by $\sim$0.5 to 1.0 kcal/mol relative to the corresponding south energy minima. This indicates that the internal energy of deoxyadenosine, deoxyguanosine, and deoxythymidine is approximately the same in the A and B forms of DNA. In contrast, the intrinsic conformational energetics of deoxycytidine favors the A form of DNA over the B form by 2.3 kcal/mol. At the basepair level, the present calculations suggest that the intrinsic conformational energetics of a GC basepair stabilize A DNA by $\sim$1.5 kcal/mol relative to B DNA. This result is consistent with GC basepairs favoring the equilibrium toward the A form of DNA, as compared to AT basepairs. Therefore, the present results indicate that the intrinsic conformational properties at the nucleoside level, and thus base composition, influence the equilibrium between the A versus B form of DNA. It should be emphasized that this contribution is but one of many effects, including but not limited to base stacking, screening of phosphate-phosphate electrostatic repulsion and minor groove hydration, that contribute to the equilibrium between the A and B forms of DNA.

The present work also suggests that the south energy minimum conformation of deoxycytidine can more readily accommodate a Z DNA pyrimidine conformation than deoxythymidine. Deoxycytidine internal energy favors the Z form of DNA over the B form by 1.6 kcal/mol as compared to deoxythymidine. This difference may contribute to the formation of Z DNA being favored by a high-GC basepair content in alternating purine-pyrimidine sequences.

The different intrinsic energetics of deoxycytidine as compared to the other nucleosides needs not to be limited to the canonical A, B, and Z DNA conformations. This difference could also extend to other conformations, occurring, for instance, under specific binding of a protein or a drug. If so, the nucleoside intrinsic energetics would be an additional indirect readout factor contributing to the selective binding of a protein to its cognate sequence. That the B DNA conformation is destabilized by cytosine, at the nucleoside level, relative to other conformations, may facilitate the deformation of some sequences upon protein or drug binding.

Information about the influence of base on the east pseudorotation and inversion energy barriers between the deoxyribose north and south conformational ranges has been presented. The east pseudorotation barrier is found lower in energy than the inversion barrier in all cases, with the extent of the difference being base-dependent. When the east pseudorotation barrier is calculated relative to the global energy minimum, this barrier shows little dependence on the base, and is in agreement with a previous estimate of $\sim$4.0 kcal/mol (Foloppe and MacKerell, 1998). However, the magnitude of the east energy barrier decreases and shows base dependency when it is calculated relative to the B DNA conformation, ranging from 1.8 to 3.7 kcal/mol. The lowest of these values is obtained for deoxycytidine, suggesting that this nucleoside may kinetically facilitate B to A DNA transitions. Furthermore, the east pseudorotation barrier for deoxythymidine, 2.8 kcal/mol, is also lower than that of the purines. The lower east pseudorotation energies relative to the B DNA conformation of the pyrimidines versus the purines correlates with the higher population of pseudorotation values in the C1′exo region in pyrimidines as compared to purines in B DNA (Hartmann and Lavery, 1996; Drew et al., 1981; Foloppe and MacKerell, 1998).

The present results confirm that the conformational properties of thymine and uracil at the nucleoside level are nearly identical, although differing significantly from cytosine. Structural and energetic properties of deoxythymidine and deoxyuridine minimum energy structures are similar (see Table 2). Furthermore, the glycosyl torsion and energetics associated with the east pseudorotation barrier are nearly identical for the two compounds (Table 4). χ_e in deoxycytidine is $\sim$22.0° lower than in the two other pyrimidine deoxynucleosides. While the similarity of the uracil and thymine nucleosides is expected, the inclusion of the uracil data can be considered to be an additional control to

TABLE A1 Bond lengths (Å) for the deoxyribose moiety in the nucleosides in the north energy minimum

Bond	l_{crys}	σ_{crys}	l_{Ade}	l_{Cyt}	l_{Gua}	l_{Thy}	l_{Ura}
C1′–C2′	1.519	0.010	1.524	1.527	1.526	1.527	1.527
C2′–C3′	1.518	0.012	1.526	1.525	1.527	1.526	1.526
C3′–C4′	1.521	0.010	1.523	1.521	1.522	1.521	1.521
C4′–O4′	1.449	0.009	1.443	1.440	1.441	1.441	1.441
O4′–C1′	1.418	0.012	1.420	1.423	1.421	1.420	1.420
C3′–O3′	1.419	0.006	1.423	1.424	1.423	1.423	1.423
C4′–C5′	1.509	0.011	1.508	1.508	1.509	1.508	1.508
C1′–N1/N9	1.488	0.013	1.465	1.478	1.463	1.480	1.481
C5′–O5′	1.423	0.011	1.427	1.429	1.427	1.429	1.428

l_{crys} refers to the average values obtained from statistical analysis (standard deviation σ_{crys}) of crystal structures of nucleosides and nucleotides (Gelbin et al., 1996). l_{Ade}, l_{Cyt}, l_{Gua}, l_{Thy}, and l_{Ura} refer to the calculated bond lengths for deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine, and deoxyuridine, respectively.

verify the unique influence of cytosine on the intrinsic conformational energetics at the nucleoside level.

An important application of the present analysis will be an improved calibration of DNA molecular mechanics force fields. Recent tests of some currently available force fields have demonstrated that such improvements are necessary (Feig and Pettitt, 1998; MacKerell, 1998). Based in part on the present data, a revised version of the CHARMM all-atom force field for nucleic acids has been developed (Foloppe and MacKerell, 1999, submitted for publication; MacKerell and Banavali, 1999, submitted for publication).

Supporting information available

The following appendices contain four tables presenting the calculated (MP2/6-31G*) bond lengths and valence angles for the deoxyribonucleosides in their north and south energy minima, and in the O4′endo conformation, along with 15 tables presenting the calculated (MP2/6-31G*) Cartesian coordinates of the deoxyribonucleosides in these conformations.

APPENDIX A

Individual nucleoside bond lengths and valence angles calculated at the MP2/6-31G* level of theory and their crystal counterpart.

APPENDIX B

Cartesian coordinates (X, Y, Z) calculated at the MP2/6-31G* level of theory for the nucleosides may be obtained from the web page of ADM, Jr. at URL www.pharmacy.ab.umd.edu/~alex/research.html. See Methods for the definitions of atom names.

We thank the Pittsburgh Supercomputing Center and NCI's Frederick Biomedical Supercomputing Center for providing computational resources. We also thank C. Zardecki of the Nucleic Acid Database for providing us with data on the glycosyl torsions in the DNA crystal structures.

This work has been financially supported by National Institutes of Health Grant GM51501.

REFERENCES

Alhambra, C., F. J. Luque, F. Gago, and M. Orozco. 1997. Ab initio study of stacking interactions in A- and B-DNA. *J. Phys. Chem. B.* 101: 3846–3853.

Altona, C., and M. Sundaralingam. 1972. Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.* 94:8205–8212.

Altona, C., and M. Sundaralingam. 1973. Conformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants. *J. Am. Chem. Soc.* 95:2333–2344.

Arnott, S., R. Chandrasekaran, D. L. Birdsall, A. G. W. Leslie, and R. L. Ratcliff. 1980. Left-handed DNA helices. *Nature.* 282:743–745.

TABLE A2 Bond lengths (Å) for the deoxyribose moiety in the nucleosides in the south energy minimum

Bond	l_{crys}	σ_{crys}	l_{Ade}	l_{Cyt}	l_{Gua}	l_{Thy}	l_{Ura}
C1′–C2′	1.518	0.010	1.527	1.524	1.527	1.525	1.525
C2′–C3′	1.516	0.008	1.520	1.519	1.521	1.521	1.521
C3′–C4′	1.529	0.010	1.530	1.532	1.530	1.532	1.532
C4′–O4′	1.446	0.010	1.439	1.441	1.439	1.439	1.439
O4′–C1′	1.420	0.011	1.426	1.426	1.426	1.429	1.429
C3′–O3′	1.435	0.013	1.430	1.431	1.431	1.430	1.430
C4′–C5′	1.512	0.007	1.513	1.512	1.513	1.512	1.512
C1′–N1/N9	1.468	0.014	1.446	1.461	1.446	1.455	1.456
C5′–O5′	1.418	0.025	1.430	1.429	1.430	1.431	1.431

l_{crys} refers to the average values obtained from statistical analysis (standard deviation σ_{crys}) of crystal structures of nucleosides and nucleotides (Gelbin et al., 1996). l_{Ade}, l_{Cyt}, l_{Gua}, l_{Thy}, and l_{Ura} refer to the calculated bond lengths for deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine, and deoxyuridine, respectively.

TABLE A3 Valence angles (deg.) for the deoxyribose moiety in the nucleosides in the north energy minimum

Angle	θ_{crys}	σ_{crys}	θ_{Ade}	θ_{Cyt}	θ_{Gua}	θ_{Thy}	θ_{Ura}
C1′–C2′–C3′	102.4	0.8	101.6	102.1	101.9	102.2	102.2
C2′–C3′–C4′	102.2	0.7	100.9	100.9	100.9	100.9	100.9
C3′–C4′–O4′	104.5	0.4	105.4	105.2	105.3	105.2	105.2
C4′–O4′–C1′	110.3	0.7	109.6	109.7	109.7	109.8	109.8
O4′–C1′–C2′	106.8	0.5	106.7	106.6	106.6	106.6	106.6
C2′–C3′–O3′	112.6	3.3	115.3	115.5	115.2	115.4	115.4
C4′–C3′–O3′	112.3	2.0	107.7	107.7	107.8	107.7	107.7
C5′–C4′–C3′	115.7	1.2	115.9	116.2	115.9	116.2	116.1
C5′–C4′–O4′	109.8	1.1	109.9	109.9	109.8	109.9	109.8
O4′–C1′–N1/N9	108.3	0.3	108.4	109.1	108.3	108.8	108.8
C2′–C1′–N1/N9	112.6	1.9	111.4	111.7	111.9	112.1	112.1
O5′–C5′–C4′	111.0	2.5	108.6	108.8	108.7	108.8	108.6
C1′–N9–C4	123.9	1.0	124.4	—	124.5	—	—
C1′–N1–C2	117.5	1.4	—	114.9	—	115.0	114.9

θ_{crys} refers to the average values obtained from statistical analysis (standard deviation σ_{crys}) of crystal structures of nucleosides and nucleotides (Gelbin et al., 1996). θ_{Ade}, θ_{Cyt}, θ_{Gua}, θ_{Thy}, and θ_{Ura} refer to the calculated valence angles for deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine, and deoxyuridine, respectively.

TABLE A4 Valence angles (deg.) for the deoxyribose moiety in the nucleosides in the south energy minimum

Angle	θ_{crys}	σ_{crys}	θ_{Ade}	θ_{Cyt}	θ_{Gua}	θ_{Thy}	θ_{Ura}
C1′–C2′–C3′	102.5	1.2	101.8	101.6	101.8	101.6	101.6
C2′–C3′–C4′	103.1	0.9	102.7	102.8	102.7	102.7	102.7
C3′–C4′–O4′	106.0	0.6	106.3	106.5	106.3	106.5	106.5
C4′–O4′–C1′	110.1	1.0	110.1	109.7	110.1	109.8	109.8
O4′–C1′–C2′	105.9	0.8	105.8	105.7	105.8	105.3	105.3
C2′–C3′–O3′	109.4	2.5	111.6	111.8	111.6	111.6	111.6
C4′–C3′–O3′	109.7	2.5	105.5	105.4	105.6	105.4	105.4
C5′–C4′–C3′	114.1	1.8	114.7	114.8	114.6	114.5	114.5
C5′–C4′–O4′	109.3	1.9	109.2	109.0	109.2	109.1	109.0
O4′–C1′–N1/N9	108.0	0.7	107.7	108.2	107.8	107.6	107.5
C2′–C1′–N1/N9	114.3	1.4	113.7	113.5	113.7	114.3	114.2
O5′–C5′–C4′	110.9	1.7	108.4	108.2	108.4	108.4	108.3
C1′–N9–C4	126.3	1.2	126.7	—	126.7	—	—
C1′–N1–C2	117.8	1.1	—	116.5	—	118.5	118.5

θ_{crys} refers to the average values obtained from statistical analysis (standard deviation σ_{crys}) of crystal structures of nucleosides and nucleotides (Gelbin et al., 1996). θ_{Ade}, θ_{Cyt}, θ_{Gua}, θ_{Thy}, and θ_{Ura} refer to the calculated valence angles for deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine, and deoxyuridine, respectively.

Arnott, S., R. Chandrasekaran, D. W. L. Hukins, P. J. C. Smith, and L. Watts. 1974. Structural details of a double-helix observed for DNAs containing alternating purine and pyrimidine sequences. *J. Mol. Biol.* 88:523–533.

Arnott, S., and E. Selsing. 1974a. The structure of polydeoxyguanilic acid · polydeoxycytidylic acid. *J. Mol. Biol.* 88:551–552.

Arnott, S., and E. Selsing. 1974b. Structures for the polynucleotide complexes poly(dA) · poly(dT) and poly(dA) · poly(dT) · poly(dT). *J. Mol. Biol.* 88:509–521.

Bandyopadhyay, T., J. Wu, W. A. Stripe, I. Carmichael, and A. S. Serianni. 1997. ^{13}C-^{1}H and ^{13}C-^{13}C spin couplings in [2′-^{13}C]2′-deoxyribonucleosides: correlations with molecular structure. *J. Am. Chem. Soc.* 119:1737–1744.

Behe, M., and G. Felsenfeld. 1981. Effects of methylation on a synthetic polynucleotide: the B–Z transition in poly(dG-m5dC) · poly(dG-m5dC). *Proc. Natl. Acad. Sci. U.S.A.* 78:1619–1623.

Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751–759.

Berthod, H., and B. Pullman. 1971a. Molecular orbital calculations of the conformation of nucleic acids and their constituents. I. Conformational energies of β-nucleosides with C(3′)- and C(2′)-endo sugars. *Biochim. Biophys. Acta.* 232:595–606.

Berthod, H., and B. Pullman. 1971b. Molecular orbital calculations of the conformation of nucleic acids and their constituents. II. Conformational energies of β-nucleosides with C(3′)- and C(2′)-exo sugars. *Biochim. Biophys. Acta.* 246:359–364.

Bram, S. 1971. Secondary structure of DNA depends on base composition. *Nat. New Biol.* 232:174–176.

Bram, S., and P. Tougard. 1972. Polymorphism of natural DNA. *Nat. New Biol.* 239:128–131.

Calladine, C. R., and H. R. Drew. 1984. A base-centered explanation of the B-to-A transition in DNA. *J. Mol. Biol.* 178:773–782.

Chuprina, V. P. 1985. Regularities in formation of the spine of hydration in the DNA minor groove and its influence on the DNA structure. *FEBS Lett.* 186:98–102.

Chuprina, V. P. 1987. Anomalous structure and properties of poly(dA) · poly(dT). Computer simulation of the polynucleotide structure with the spine of hydration in the minor groove. *Nucleic Acids Res.* 15:293–311.

Clowney, L., S. C. Jain, A. R. Srinivasan, J. Westbrook, W. K. Olson, and H. M. Berman. 1996. Geometric parameters in nucleic acids: nitrogenous bases. *J. Am. Chem. Soc.* 118:509–518.

Davies, D. B. 1978. Conformations of nucleosides and nucleotides. *Prog. Nucl. Magn. Reson. Spectrosc.* 12:135–226.

Davies, D. R., and R. L. Baldwin. 1963. X-ray studies of two synthetic DNA copolymers. *J. Mol. Biol.* 6:251–255.

Dickerson, R. E. 1992. DNA structure from A to Z. *Methods Enzymol.* 211:67–111.

Dickerson, R. E., H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka. 1982. The anatomy of A-, B-, and Z-DNA. *Science.* 216:475–485.

Dickerson, R. E., D. S. Goodsell, and S. Neidle. 1994. ". . . the tyranny of the lattice. . . ". *Proc. Natl. Acad. Sci. USA.* 91:3579–3583.

Drew, H. R., and R. E. Dickerson. 1981. Structure of a B-DNA dodecamer III. Geometry of hydration. *J. Mol. Biol.* 151:535–556.

Drew, H., T. Takano, S. Tanaka, K. Itakura, and R. E. Dickerson. 1980. High-salt d(CpGpCpG), a left-handed Z′-DNA double helix. *Nature.* 286:567–573.

Drew, H. R., R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. S. Dickerson. 1981. Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA.* 78:2179–2183.

Feig, M., and B. M. Pettitt. 1998. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* 75:134–149.

Foloppe, N., and A. D. MacKerell, Jr. 1998. Conformational properties of the deoxyribose and ribose moieties of nucleic acids: a quantum mechanical study. *J. Phys. Chem. B.* 102:6669–6678.

Franklin, R. E., and R. G. Gosling. 1953. Molecular configuration in sodium thymonucleate. *Nature.* 171:740–741.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople. 1996. Gaussian 94. Gaussian, Inc., Pittsburgh, PA.

Gelbin, A., B. Schneider, L. Clowney, S.-H. Hsieh, W. K. Olsen, and H. M. Berman. 1996. Geometric parameters in nucleic acids: sugar and phosphate constituents. *J. Am. Chem. Soc.* 118:519–529.

Grzeskowiak, K., K. Yanagi, G. G. Prive, and R. E. Dickerson. 1991. The structure of B-helical CGATCGATCG and comparison with CCAACGTTGG. The effect of basepair reversals. *J. Biol. Chem.* 266: 8861–8883.

Guschlbauer, W., and K. Jankowski. 1980. Nucleoside conformation as determined by the sugar substituent. *Nucleic Acids Res.* 8:1421–1433.

Guzikevich-Guerstein, G., and Z. Shakked. 1996. A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nat. Struct. Biol.* 3:32–37.

Hartmann, B., and R. Lavery. 1996. DNA structural forms. *Q. Rev. Biophys.* 29:309–368.

Heinemann, U., C. Alings, and M. Hahn. 1994. Crystallographic studies of DNA helix structure. *Biophys. Chem.* 50:157–167.

Herbert, A., and A. Rich. 1996. The biology of left-handed Z-DNA. *J. Biol. Chem.* 271:11595–11598.

Ho, P. S., M. J. Ellison, G. J. Quigley, and A. Rich. 1986. A computer-aided thermodynamic approach for predicting the formation of Z-DNA in naturally occurring sequences. *EMBO J.* 5:2737–2744.

Hunter, C. A. 1993. Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* 230:1025–1054.

Ivanov, V. I., and L. E. Minchenkova. 1995. The A-form of DNA: in search of biological role (a review). *Mol. Biol. (Engl. ed.).* 28:780–788.

Kopka, M. L., A. V. Fratini, H. R. Drew, and R. E. Dickerson. 1983. Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* 163:129–146.

Lakshminarayanan, A. V., and V. Sasisekharan. 1969. Stereochemistry of nucleic acids and polynucleotides. IV. Conformational energy of base-sugar units. *Biopolymers.* 8:475–488.

Lakshminarayanan, A. V., and V. Sasisekharan. 1970. Stereochemistry of nucleic acids and polynucleotides. II. Allowed conformations of the monomer unit for different ribose puckerings. *Biochim. Biophys. Acta.* 204:49–59.

Langridge, R. 1969. Nucleic acids and polynucleotides. *J. Cell Physiol.* 74(Suppl. 1):1–20.

Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: convention and principles. *J. Biomol. Struct. & Dyn.* 6:655–667.

Leslie, A. G. W., S. Arnott, R. Chandrasekaran, and R. L. Ratliff. 1980. Polymorphism of DNA double helices. *J. Mol. Biol.* 143:49–72.

Liepinsh, E., G. Otting, and K. Wütrich. 1992. NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. *Nucleic Acids Res.* 20:6549–6553.

MacKerell, A. D., Jr. 1998. Observations on the A versus B equilibrium in molecular dynamics simulations of duplex DNA and RNA. *In* Molecular Modeling of Nucleic Acids. N. B. Leontis and J. SantaLucia, Jr., editors. American Chemical Society, Washington, DC. 304–311.

Mazur, J., A. Sarai, and R. L. Jernigan. 1989. Sequence dependence of the B-A conformational transition of DNA. *Biopolymers.* 28:1223–1233.

McCall, M., T. Brown, and O. Kennard. 1985. The crystal structure of d(GGGGCCCC). A Model for Poly(dG) · poly(dC). *J. Mol. Biol.* 183: 385–396.

Nelson, H. C. M., J. T. Finch, B. F. Luisi, and A. Klug. 1987. The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature.* 330:221–226.

Olson, W. K. 1973. Syn-anti effects on the spatial configuration of polynucleotide chains. *Biopolymers.* 12:1787–1814.

Olson, W. K. 1982. How flexible is the furanose ring? 2. An updated potential energy estimate. *J. Am. Chem. Soc.* 104:278–286.

Olson, W. K., and P. J. Flory. 1972a. Spatial configuration of polynucleotide chains. I. Steric interactions in polyribonucleotides: a virtual bond model. *Biopolymers.* 11:1–23.

Olson, W. K., and P. J. Flory. 1972b. Spatial configuration of polynucleotide chains. II. Conformational energies and the average dimensions of polyribonucleotides. *Biopolymers.* 11:25–56.

Olson, W. K., and J. L. Sussman. 1982. How flexible is the furanose ring? 1. A comparison of experimental and theoretical studies. *J. Am. Chem. Soc.* 104:270–278.

Peticolas, W. L., Y. Wang, and G. A. Thomas. 1988. Some rules for predicting the base-sequence dependence of DNA conformation. *Proc. Natl. Acad. Sci. U.S.A.* 85:2579–2583.

Pilet, J., and J. Brahms. 1972. Dependence of B-A conformational change in DNA on base composition. *Nat. New Biol.* 236:99–100.

Pohl, F. M., and T. M. Jovin. 1972. Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.* 67:375–396.

Poncin, M., B. Hartmann, and R. Lavery. 1992. Conformational substates in B-DNA. *J. Mol. Biol.* 226:775–794.

Privé, G. G., K. Yanagi, and R. E. Dickerson. 1991. Structure of the B-DNA decamer CCAACGTTGG and comparison with isomorphous decamers CCAAGATTGG and CCAGGCCTGG. *J. Mol. Biol.* 217: 177–199.

Röder, O., H.-D. Lüdemann, and E. von Goldammer. 1975. Determination of the activation energy for pseudorotation of the furanose ring in nucleosides by ^{13}C nuclear-magnetic-resonance relaxation. *Eur. J. Biochem.* 53:517–524.

Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York.

Saran, A., D. Perahia, and B. Pullman. 1973. Molecular orbital calculations of the conformation of nucleic acids and their constituents. VII. Conformation of the sugar ring in β-nucleosides: the pseudorotational representation. *Theoret. Chim. Acta.* 30:31–44.

Saran, A., B. Pullman, and D. Perahia. 1972. Molecular orbital calculations on the conformation of nucleic acids and their constituents. IV. Conformations about the exocyclic C(4′)-(C5′) bond. *Biochim. Biophys. Acta.* 287:211–231.

Son, T., W. Guschlbauer, and M. Guéron. 1972. Flexibility and conformations of guanosine monophosphates by the Overhauser effects. *J. Am. Chem. Soc.* 94:7903–7911.

Subirana, J. A., and T. Faria. 1997. Influence of sequence on the conformation of the B-DNA helix. *Biophys. J.* 73:333–338.

Sundaralingam, M. 1969. Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleosides mono-, di-, tetraphosphates, nucleic acids, and polynucleotides. *Biopolymers.* 7:821–860.

Uesugi, S., H. Miki, M. Ikehara, H. Iwahashi, and Y. Kyogoku. 1979. A linear relationship between electronegativity of 2′-substituents and conformation of adenine nucleosides. *Tetrahedron Lett.* 42:4073–4076.

Vorlickova, M., J. Kypr, S. Stokrova, and J. Sponar. 1982. A Z-like form of poly(dA-dC) · poly(dG-dT) in solution? *Nucleic Acids Res.* 10: 1071–1080.

Wang, A. H.-J., S. Fujii, J. H. van Boom, G. A. van der Marel, S. A. A. van Boeckel, and A. Rich. 1982. Molecular structure of r(GCG)d(TATACGC): a DNA-RNA hybrid helix joined to double helical DNA. *Nature.* 299:601–604.

Wang, A. H.-J., T. Hakoshima, G. van der Marel, v. B. J. H., and A. Rich. 1984. AT base pairs are less stable than GC base pairs in Z-DNA: the crystal structure of d(m5CGTAm5CG). *Cell.* 37:321–331.

Wang, S., and E. T. Kool. 1995. Origins of the large differences in stability of DNA and RNA helices: C- 5 methyl and 2′-hydroxyl effects. *Biochemistry.* 34:4125–4132.

Wang, A. H.-J., G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich. 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature.* 282:680–686.

Wang, Y., G. A. Thomas, and W. L. Peticolas. 1987a. Sequence dependence of the B to Z transition in crystals and aqueous NaCl solutions for deoxyribonucleotides containing all four canonical DNA bases. *Biochemistry.* 26:5178–5186.

Wang, Y., G. A. Thomas, and W. L. Peticolas. 1987b. Sequence dependent conformations of oligomeric DNA's in aqueous solutions and in crystals. *J. Biomol. Struct. & Dyn.* 5:249–274.

Wang, Y., G. A. Thomas, and W. L. Peticolas. 1989. A duplex of the oligonucleotides d(GGGGGTTTTT) and d(AAAAACCCCC) forms an A to B conformational junction in concentrated salt solutions. *J. Biomol. Struct. & Dyn.* 6:1177–1187.

Westhof, E., and M. Sundaralingam. 1980. Interrelationships between the pseudorotation parameters P and τm and the geometry of the furanose ring. *J. Am. Chem. Soc.* 102:1493–1500.

Yathindra, N., and M. Sundaralingam. 1973. Conformational studies on guanosine nucleotides and polynucleotides. The effect of the base on the glycosyl and backbone conformations. *Biopolymers.* 12:2075–2082.

Zimmer, C., S. Tymen, C. Marck, and W. Guschlbauer. 1982. Conformational transitions of poly(dA-dC)poly(dG-dT) induced by high salt or in ethanolic solution. *Nucleic Acids Res.* 3:1081–1091.