JCTC Journal of Chemical Theory and Computation

Exploring the Essential Dynamics of B-DNA

Alberto Pérez,[†,‡] José Ramón Blas,[†] Manuel Rueda,[†,§] Jose María López-Bes,[‡] Xavier de la Cruz,[†,||] and Modesto Orozco*[,†,§,⊥]

Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain, Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avgda Diagonal 643, Barcelona 08028, Spain, Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain, Institució Catalana per la Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08018 Barcelona, Spain, and Structure and Modeling Node, Instituto Nacional de Bioinformática, Spain

Received March 2, 2005

Abstract: The essential dynamics of different normal and chemically modified DNA duplexes pertaining to the B family have been extensively explored from molecular dynamics simulations using powerful data mining techniques. Some of them, which are presented here for the first time, might become standard, powerful tools to characterize the dynamical behavior of complex biomolecular structures such as nucleic acids. Their potential impact is illustrated by examining the extended trajectories sampled for the set of DNA duplexes considered in this study, which allows us to discuss the degree of conservation of the natural flexibility pattern of the different DNAs, which in specific cases contain severe chemical modifications.

Introduction

Though the first molecular dynamics (MD) simulation of a nucleic acid (NA) was done in the 1980s,[1,2] technical problems limited the systematic use of this powerful simulation technique to explore NAs in solution until the middle of the 1990s. The implementation of methods to treat long-range electrostatic forces, such as the Particle Mesh Ewald method,[3] and the development of more accurate force fields[4−8] are responsible for the huge explosion of MD simulations of NAs during the past decade. Current state-of-the-art simulations cover several nanoseconds of fully solvated 10−14-mer duplexes,[9−13] and simulations approaching the 0.1 μs range have been published by different groups (for example, see refs 14−16), showing that reasonable convergence can be obtained for many structural and enegetic properties in the sub-microsecond simulation time, at least for normal B-type DNA duplexes. Very recently, the Ascona B-DNA Consortium (ABC) published its first conclusions of B-DNA structure and flexibility obtained from the analysis of all of the 136 unique tetranucleotide sequences of DNA duplexes.[17] Clearly, the field of MD simulations of NAs has reached maturity and well-defined standards for simulation conditions and protocols are being established. Unfortunately, the capacity to run longer and more stable trajectories has not been accompanied by a similar increase in our ability to extract all the information included in such trajectories, and a large amount of useful information present in the trajectory is escaping our analysis.

The study of MD trajectories has traditionally pursued the verification of whether the trajectory was converged and the characterization of the average structural and energetic details of the molecular system.[9−13] Convergence has been typically checked by looking at the time evolution of the positional root-mean-square deviation (rmsd), or alternative global geometrical properties. The average structural features of molecules have been analyzed with a variety of geometrical

* Corresponding author e-mail: modesto@mmb.pcb.ub.es.
† Institut de Recerca Biomèdica.
‡ Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona.
§ Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona.
|| Institució Catalana per la Recerca i Estudis Avançats (ICREA).
⊥ Instituto Nacional de Bioinformática.

10.1021/ct050051s CCC: $30.25
Published on Web 06/30/2005

measurements, such as the helical parameters for NAs. Interaction properties have been explored by integrating the solvent population around the molecule, or by computing interaction potentials. Dynamical properties of the solvent distribution around the solute can be characterized by parameters such as the maximum or average residence time.[14−18] Recently, the ensemble of structures collected in the MD simulation has been used to estimate energetic properties by combining the average intramolecular energy with estimates of the solvation free energy derived from continuum[12,13,19−23] or discrete[24,25] linear-response calculations. This strategy has been very powerful in determining the relative stability of different DNA (or RNA) helical conformations,[12,23,26] but some caution is needed when used to discriminate between states very close in energy.

Convergence problems precluded, for many years, a careful analysis on the dynamical properties of macromolecules. However, it seems that even short (2−20 ns) simulations, which can be obtained quite easily with current computational resources, can be enough to capture the essential dynamics of small polymers such as B-DNA duplexes,[14−16] which opens the possibility of using MD as a method to reproduce NA flexibility in physiological conditions. In this paper, we want to examine this point in depth. For this purpose, we will first discuss several approaches to describe the dynamical behavior of NAs, and particularly new measurement techniques will be presented. Second, the potential impact of these techniques will be used to analyze the essential deformation pathways of DNA duplexes by analyzing a local library of MD trajectories collected for normal, mutated, and chemically modified DNA duplexes.

Methodological Approach

The dynamic analysis of trajectories is a typical data mining problem, where the information of interest is buried in many gigabytes of noise. Particularly, the trajectory contains a large amount of fast and irrelevant movements, which hide the soft deformation modes that are important to understand the flexibility of DNA. Most approaches to extract information from the trajectory are based on the construction of covariance matrices. For DNA simulations, such matrices can be built considering all or part of the system and using three different coordinate models: (i) Cartesian coordinates ($\mathbf{C}_x$), (ii) mass-weigthed Cartesian coordinates ($\mathbf{C}_m$), and (iii) helical coordinates ($\mathbf{C}_h$). We will summarize, in the following, how to process these covariance matrices to derive dynamic information on NAs.

Essential Dynamics. Following the principal component analysis (PCA) method, diagonalization of the Cartesian covariance matrix ($\mathbf{C}_x$) yields a set of $3N - 6$ eigenvectors (where N is the number of atoms used to define $\mathbf{C}_x$) and their associated eigenvalues. The eigenvectors display a series of interesting properties: (i) they explain all the variance of the system, (ii) they are orthogonal, and (iii) the fraction of the total variance of the trajectory explained by a given eigenvector is given by the magnitude of its eigenvalue. Thus, the diagonalization of $\mathbf{C}_x$ allows us to characterize the most important deformation modes and to evaluate their relative weight in the global flexibility of the molecule.[12,13,27−32] PCA is especially powerful in the study of DNA trajectories, since this biopolymer has quite simple dynamics (compared to proteins), which can be represented by a small number of "essential movements" (see Results section).

The analysis of the eigenvalues obtained by diagonalization of $\mathbf{C}_x$ provides very direct information on the flexibility of DNA. For example, they can be used to derive harmonic force constants associated with the essential deformations of DNA (see eq 1; refs 12,13,30−32).

$$K_i = k_B T/\lambda_i \tag{1}$$

where k_B is Boltzmann's constant (in kcal/mol K), T is the absolute temperature, and λ_i is the eigenvalue associated with the essential movement i (in Å^2).

Note that once the force constant is known, the deformation energy along the essential mode i can be easily determined from eq 2.

$$E_i = \frac{K_i}{2}(\Delta X_i)^2 \tag{2}$$

where ΔX_i is a Cartesian deformation along the eigenvector i.

Assuming that the dynamical behavior of two molecules can be described by means of a limited set of t-harmonic oscillators, the accessible volume for a system at temperature T can be defined as shown in see eq 3a (see refs 31−33). Note that the practical use of the configurational volume (Vol) as a descriptor of flexibility faces a serious problem: Vol increases rapidly with the number of eigenvectors until a maximum is found; the introduction of additional eigenvectors (with $\lambda < 1$) decreases the volume, and a value of zero is reached when t becomes close to m. This behavior reflects the fact that the real dimensionality of the DNA is not equal to the total number of eigenvectors and that the stiffer eigenvectors do not contribute significantly to an explanation of trajectory variance. Several approaches can be used to overcome the intrinsic limitations of the volume: (i) the definition of an arbitrary number of important eigenvectors (t) for the systems studied, (ii) the use of an ad-hoc shifting value (see eq 3b) as, in fact, is performed in the quasiharmonic analysis of entropy (see below), and (iii) the use of the number of eigenvectors for which the maximum of Vol is obtained as an estimate of the dimensionality of the configurational space. Note that the dimensionality will roughly correspond to the order number of the eigenvector whose eigenvalue is closer to 1 (see eq 3a) in a given unit measuring system.

$$\mathrm{Vol} = \prod_{i=1}^{t}(\lambda_i)^{1/2} \tag{3a}$$

$$\mathrm{Vol}_g = \prod_{i=1}^{t}(1 + \lambda_i)^{1/2} \tag{3b}$$

where, here, t is the number of eigenvalues considered.

The selection of a common threshold for the definition of Vol might be acceptable when very similar molecules are compared, but for this study, it is not recommended. The

generalized volume shown in eq 3b provides information somehow redundant to that derived from quasiharmonic entropy measures (see below). Thus, in this paper, we decide to use the dimensionality as a measure of the configurational volume accessible to the DNA. We should note that the concept of dimensionality is close to physical intuition, since it can be interpreted as the limit above which the addition of additional dimensions does not improve our representation of the systems. For example, a sheet of paper is well represented by just two dimension, and the third one is irrelevant [i.e., the three-dimensional volume of a sheet of paper is close to zero, whereas its two-dimensional volume (surface) is not]. Note, however, that the value of the dimensionality depends on the units used to measure variance. In our case, we consider that the first irrelevant mode is that along which, at temperature T, the expected displacement of the molecule is less than 1 Å (a classical heteroatom−hydrogen distance).

The analysis of the eigenvectors projected on the Cartesian space gives very useful information on the nature of the essential movements of the DNA. Unfortunately, manipulation of this information is, again, difficult because of its high dimensionality, which forces us to focus on a reduced set of "important" n eigenvectors ($\{\mu_i\}_0$), defined as those needed to explain a certain degree of variance (for example, 80 or 90%, which, for normal DNA dodecamers, implies around 10 modes). Then, we can quantify the similarity in the pattern of deformation sampled in two trajectories (A and B) by comparing the corresponding sets of eigenvectors ($\{\mu\}^{\mathrm{A}}_{i0}$ and $\{\mu_i\}^{\mathrm{B}}_0$). This can be done by using eq 4,[12,13,30−32,34−36] which implicitly assumes that all the n eigenvectors have the same weight in defining the variance of the trajectory.

$$\gamma_{\mathrm{AB}} = \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}(\nu_i^{\mathrm{A}}\nu_j^{\mathrm{B}})^2 \tag{4}$$

where n is the minimum number of eigenvectors needed to define the set of "important" eigenvectors $\{\mu_{i}\}^{\mathrm{A}}_0$ and $\{\mu_i\}^{\mathrm{B}}_0$ and where ν_i^X stands for the i-unitary eigenvector of trajectory X (i.e., $\nu_i^X = \mu_i^X/|\mu_i^X|$). Note that γ_{AB} takes values from 0 (null similarity) to 1 (identical essential movements).

The absolute similarity index γ_{AB} is useful to verify[12,13,30−32,34−36] the convergence of a single trajectory by defining A and B as two distant portions of the same trajectory. Indeed, it can be modified (see eq 5) to determine how well the essential dynamics of a given molecule follow a given conformational transition.

$$\gamma_{\mathrm{A}}^R = \sum_{i=1}^{n}(\nu_i^{\mathrm{A}}R)^2 \tag{5}$$

where R is the unitary transition vector defining the transition between two conformations.

The most powerful use of γ_{AB}, however, is as a measure of the similarity between the essential dynamics of two molecules for which Cartesian covariance matrices of the same dimensionality can be created (this might imply, in some cases, a restriction of the number of atoms considered in the calculation of $\mathbf{C}_x$). As noted elsewhere,[12,13] eq 4 assumes that (i) trajectories are very long and (ii) a small number of eigenvectors are included in $\{\mu_i\}^{\mathrm{A}}_0$ and $\{\mu_i\}^{\mathrm{B}}_0$. The second requirement is not a real problem for NAs (see above), but clearly, current MD simulations are too short and part of the dissimilarities between the trajectories of two molecules can be simply due to the limited configurational sampling. Thus, it is often convenient to use relative similarity indexes (see eqs 6 or 7), where the noise in the original trajectories is largely canceled by normalizing cross similarities ($X-Y$) with self similarities ($X-X$ and $Y-Y$).

$$\kappa_{\mathrm{AB}} = 2\frac{\gamma_{\mathrm{AB}}}{(\gamma_{\mathrm{AA}}^T + \gamma_{\mathrm{BB}}^T)} \tag{6}$$

$$\kappa_{\mathrm{AB}} = \frac{\gamma_{\mathrm{AB}}}{(\gamma_{\mathrm{AA}}^T\,\gamma_{\mathrm{BB}}^T)^{1/2}} \tag{7}$$

where γ_{XX}^T is the absolute self-similarity index (eq 4) for trajectory X obtained by comparing the first and second halves of the trajectory.

As noted above, eqs 4−7 implicitly assume that all eigenvectors in the "important space" have the same weight in the dynamical behavior of the molecule. This means that, for the similarity index, there is a good overlap between the first eigenvectors (which might explain 40% of the variance) that is equivalent with that of higher modes, which might explain no variance. Furthermore, the method is insensitive to permutations in the ordering of the eigenvectors, which is quite counterintuitive since the same similarity index will be found if the first eigenvector of trajectory A matches the first of trajectory B, or if the matching happens between eigenvectors 1(A) and 1000(B).

Clearly, the use of a small set of "important eigenvectors" in the comparison ($n \leq 10$; see above) reduces but does not eliminate this intrinsic problem of indexes γ and κ, since cases can exist where the ordering of eigenvectors or the associated eigenvalues are quite different in the two trajectories. Here, we propose a new strategy to solve this problem. The method starts with the assumption that the molecule moves sampling states defined by a common displacement (Δx) along the different eigenvectors. Thus, the weight of each eigenvector in defining the flexibility space will be given by its Boltzman factor computed from the harmonic energy penalty given by eq 2 (see Figure 1 and eq 9). Within this approach, the relative importance of two eigenvectors (i and j) in defining the flexibility space is given by the ratio between their respective Boltzmann's factors (see eq 9). Accordingly, the relative importance of an eigenvector in the set of essential movements is given by eq 9, where the index z runs under either the "important" ($z = n$) or entire ($z = m$) set of eigenvectors.

$$w_i = \frac{e^{-(\Delta x)^2/\lambda_i}}{\sum_{j=1}^{z} e^{-(\Delta x)^2/\lambda_j}} \tag{8}$$

$$W_{ij} = \frac{e^{-(\Delta x)^2/\lambda_i}}{e^{-(\Delta x)^2/\lambda_j}} \tag{9}$$

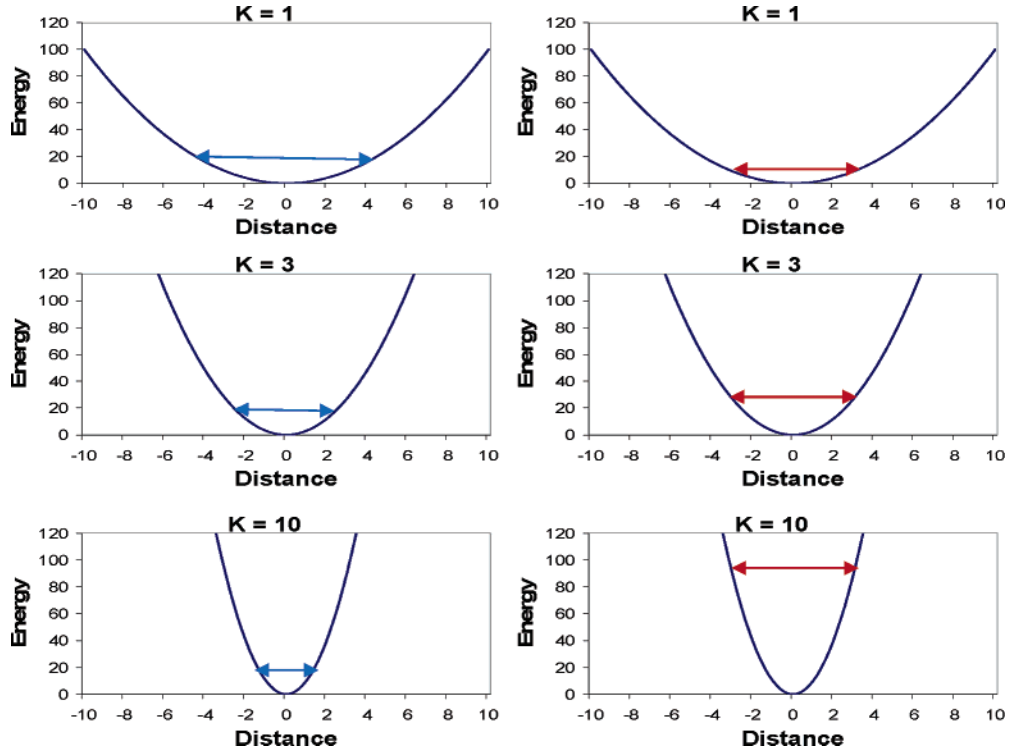

Figure 1. Scheme of the movements along three eigenvectors (softer to stiffer from top to bottom). Panels at the left indicate the sampling obtained for a common energy (equipartition principle). Panels at the right indicate the different energy levels needed to sample each node for a common displacement.

When comparing two sets of eigenvectors $\{\mu_i\}^A$ and $\{\mu_i\}^B$, the contribution to the global similarity of two individual eigenvectors μ_i^A and μ_j^B is assumed to be equal to the product of the relative probabilities of each eigenvector. Then, the extension of this concept to all (or the "important" set of) eigenvectors and the subsequent normalization considering the energy distribution of the sets of eigenventors- $\{\mu_i\}^A$ and $\{\mu_i\}^B$ yields to eq 10, which is equivalent to eq 4, but sensitive to the relative importance of the different eigenvectors in explaining trajectory variance.

$$\xi_{AB} = \frac{2\sum_{i=1}^{i=z}\sum_{j=1}^{j=z}\left\{(\nu_i^A\nu_j^B)\dfrac{\exp\left[-\dfrac{(\Delta x)^2}{\lambda_i^A}-\dfrac{(\Delta x)^2}{\lambda_j^B}\right]}{\sum_{i=1}^{i=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_i^A}\right]\sum_{j=1}^{j=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_j^B}\right]}\right\}^2}{\sum_{i=1}^{i=z}\left\{\dfrac{\exp\left[-2\dfrac{(\Delta x)^2}{\lambda_i^A}\right]}{\left(\sum_{i=1}^{i=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_i^A}\right]\right)^2}\right\}^2+\sum_{j=1}^{j=z}\left\{\dfrac{\exp\left[-2\dfrac{(\Delta x)^2}{\lambda_j^B}\right]}{\left(\sum_{j=1}^{j=z}\exp\left[-\dfrac{(\Delta x)^2}{\lambda_j^B}\right]\right)^2}\right\}^2} \quad (10)$$

where λ_i is the eigenvalue (in Å^2) associated with eigenvector μ_i, whose unitary vector is ν_i. The sum can be extended to all ($z = m$) or the "important set" ($z = n$) of the eigenvectors.

By analogy to the κ_{AB} index, the new absolute index ξ_{AB} can be manipulated to obtain relative similarity measures (δ_{AB}; see eq 11). The derivation of transition indexes equivalent to eq 5 is also straightforward.

$$\delta_{AB} = 2\frac{\xi_{AB}}{(\xi_{AA}^T + \xi_{BB}^T)} \quad (11)$$

Note that a key issue in the calculation of eq 10 is the selection of the common displacement (Δx). Large values of Δx will make it so that only the first modes will have importance in the comparison (see eqs 8 and 9), whereas small values will increase the weight in the comparison of higher essential modes, some of them with small impact in the molecule variance [note that for $\Delta x = 0$, eqs 10 and 11 converge to eqs 4 and 6, which means that all the modes (in the important space) will have the same impact in defining the similarity index]. As a compromise, the common distortion level (Δx) can be defined (i) by determining the number of eigenvectors (n') required to capture a given degree (for example 90%) of the variance and (ii) determine the smallest Δx for which the contribution of the $n' + 1$ mode to the similarity index is less than 1%. This process guarantees that no weight will be given to irrelevant modes and that, within the "important space", all eigenvectors will have some impact (depending on their relative Boltzman's factors) in the determination of the similarity function.

The different behavior of the two similarity indexes can be illustrated in the comparison of the essential movements of the central eight-mer portion of two duplexes of different lengths but identical sequence: $d(A)_{11}$ and $d(A)_{15}$ (trajectories

Table 1. DNA Duplexes Studied by MD Simulations[a]

code	sequence
trajectory 1	*d*(GATTAATTAATTAATC)
trajectory 2	*d*(GATTAATT***A***ATTAATC)
trajectory 3	*d*(GGCCGGCCGGCCGGCC)
trajectory 4	*d*(GGCCGGCC***G***GCCGGCC)
trajectory 5	*d*(AAAAAAAAAAAAAAAA)
trajectory 6	*d*(CGCGAATTCGCG)
trajectory 7	*d*(CTTTTC***F***TTCTT)
trajectory 8	*d*(CTTTTCTTTCTT)
trajectory 9	*d*(CTTTTC***T***TTCTT)
trajectory 10	*d*(CTTTTC***T***TTCTT)
trajectory 11	*d*(AAAAAAAAAAA)
trajectory 12	*d*(GAAGGAGGAGA)
trajectory 13	*d*(CCAAGCTTGG)
trajectory 14	*d*(CCAA***G***CTTGG)

[a] Bases in bold correspond to special cases: trajectories 2 and 4 have *S*-methylphosphonate substituting the charged group after the marked base; 7, 8, and 9 contain apolar surrogates of bases (T, Q, and Z respectively; see ref 51). Finally, trajectory 14 contains thymine paired to the marked guanine.

Table 2. Example of the Behavior of Absolute Similarity Indexes (γ and ξ; See Eqs 5 and 10) in the Comparison of the Essential Dynamics of the Eight-mer Portion of Trajectories 5 and 11 in Normal Conditions and Forcing Some Eigenvector Rotations (See Text for Details)

	γ	ξ
normal	0.77	0.87
1↔10 rotation (traj11)	0.77	0.63
1↔10 rotation (traj5)	0.77	0.64
1↔10 rotation (both)	0.77	0.62
1↔456 rotation (both)	0.66	0.53
10↔456 rotation (both)	0.66	0.86

5 and 11, see Table 1). By defining an important space of 10 eigenvectors, absolute similarity indexes obtained from the indexes γ (eq 5) and ξ (eq 10) are large, which shows the similarity in the nature of the movement of both duplexes (see Table 2). When eigenvectors are permuted in the essential space, that is, by interchanging the position of vectors 1 and 10, the γ index remains unaltered whereas the ξ index detects the smaller similarity between the two trajectories (see Table 2). The substitution of the first eigenvector by an irrelevant one (taken here as the 456th eigenvector) decreases the γ index at a level similar to that obtained when such a replacement affects the 10th eigenvector despite the difference in importance of these two eigenvectors (see Table 2). On the contrary, the new ξ index properly captures the difference in impact of the substitution of the 1st and 10th eigenvectors (see Table 2). In summary, although, for normal cases, the behavior of both similarity indexes is similar, the new similarity ξ index is more powerful in analyzing anomalous situations.

As noted above, the new indexes (10 and 11) are dependent on the common displacement (Δx). However, in practical use, for the series of the trajectories considered here, where no large permutation of eigenvectors exists, the dependence of the new indexes on Δx is very small, as seen for selected cases in Table 3. Note that, in general, the

Table 3. Absolute Similarity Index (ξ; Eq 10) between Trajectories 1−4 for Different Values of the Common Displacement (Δx)[a]

	traj1	traj2	traj3	traj4
traj1		0.73	0.65	0.69
		0.75	0.67	0.71
		0.82	0.73	0.77
		0.83	0.74	0.77
traj2			0.63	0.65
			0.66	0.68
			0.72	0.74
			0.74	0.75
traj3				0.69
				0.70
				0.74
				0.72

[a] Values in each cell correspond, top to bottom, to Δx= 0.0, 1.0, 5.0, and 15.0 Å.

similarity indexes increase when the displacements do until large Δx values are reached (for which only the first mode contributes to the similarity), indicating that, for the series of trajectories studied, the first eigenvectors are the better conserved (see Results section).

Entropy Calculation. The fluctuations in the root-mean square deviation or mass-weighted root-mean-square deviation (rmsd or rmsd_w) along the trajectory have been traditionally used as an indicator of the flexibility of a molecule. The rmsd is defined as the average deviation of a conformation with respect to a reference structure (see eq 12).

$$\text{rmsd}_k = [1/N \sum_{l=1}^{3N}(x_{kl} - x_l)^2]^{1/2} \tag{12}$$

where x_{kl} stands for the l coordinate in structure k and x_l is the value of the l coordinate in the reference structure.

The average deviation ($\langle \text{rmsd}^2 \rangle_t$) can be defined in terms of the time-averaged position of each Cartesian or mass-weigthed coordinate ($\langle x_l \rangle_t$; see eq 13). It can be easily shown that the average deviation is equal to the variance of the trajectory and the deviation between the average and the reference structures (see eq 14). Note that the variance of the trajectory equals the sum of the eigenvalues, which for most covariance matrices is closely related to the product of the eigenvalues, which as shown below is closely related to the entropy of the system. The connection between rmsd fluctuation and entropy that has been implicitly assumed in many MD studies is then clear.

$$\langle \text{rmsd}^2 \rangle_t = \frac{1}{Z}\sum_{k=1}^{Z}\left[\frac{1}{N}\sum_{l=1}^{3N}(x_{kl} - x_l)^2\right] = \frac{1}{N}\sum_{l=1}^{3N}\left[\frac{1}{Z}\sum_{k=1}^{Z}(x_{kl} - \langle x_l \rangle_t - x_l + \langle x_l \rangle_t)^2\right] \tag{13}$$

$$\langle \text{rmsd}^2 \rangle_t = \text{rmsd}_{\text{ref}}{}^2 + \frac{1}{N}\sum_{l=1}^{3N}\frac{1}{Z}\sum_{k=1}^{Z}(x_{kl} - \langle x_l \rangle_t)^2 = \text{rmsd}_{\text{ref}}{}^2 + \frac{1}{N}\sum_{l=1}^{3N}\text{var}_l \tag{14}$$

where Z stands for the number of snapshots collected during the trajectory.

The conformational freedom of a molecule is then qualitatively captured in the oscillation of the mean square deviation, in the generalized configurational volume, and in the dimensionality of the space (see Methods section). However, the best thermodynamic property to describe the conformational freedom is entropy. In principle, for any system, entropy can be computed using basic thermodynamic principles (see eq 15). However, in general, the direct use of eq 15 is not advisable since the definition of microstates is difficult and somehow arbitrary for DNA and convergence problems might be severe.

$$S \propto \sum_k P_k \log_2 P_k \tag{15}$$

where P_k is the probability of microstate k and the sum extends to all the possible conformational space.

More useful in DNA simulations is the evaluation of molecular entropy from the harmonic oscillator model using either Schlitter's (ref 37; see eq 16) or Andricioaei and Karplus' (ref 38; see eq 17) methods. The two approaches rely on the same principles: the quantum oscillator and the concept of "generalized configurational volume". Both require the diagonalization of the mass-weighted covariance matrix to obtain the frequencies (derived directly from the eigenvalues) associated with the essential deformations. Note that the sum of the eigenvalues of a $\mathbf{C}_x$ (or $\mathbf{C}_w$) in displacement (or mass-weighted displacement) units equals the variance, connecting, then, the concepts of entropy and rmsd fluctuations (see above).

In practice, Schlitter's and Andricioaei and Karplus' methods provide very close results and share the same intrinsic shortcomings derived from (i) the assumption that the molecule moves from the equilibrium geometry according to harmonic modes and (ii) the time dependence of the entropy estimate. Thus, both methods can be used in trajectories near the equilibrium but not in those that sample irreversible conformational transitions. The time dependence of both methods is intrinsic to any simulation-based estimate of entropy, since as the length of the trajectory increases the number of microstates visited and, then, the entropy become larger. The time-dependence problem can be alleviated by using an exponential correction formula developed by Harris et al. (see eq 18 and ref 39), which using partial entropy estimates obtained for different simulation time windows (snapshots always collected every 1 ps) allows us to estimate the entropy at infinite simulation time.

$$S \approx 0.5k\sum_i \ln\left(1 + \frac{e^2}{\alpha^2}\right) \tag{16}$$

$$S = k\sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \tag{17}$$

where $\alpha_i = \hbar\omega_i/kT$, ω being the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix ($\mathbf{C}_w$), and the sum extends to all the nontrivial vibrations.

$$S(t) = S_\infty - \frac{\alpha}{t^\beta} \tag{18}$$

where α and β are fitted parameters and t is the simulation time (in picoseconds) used to obtain the entropy estimate.

Helical Stiffness. Standard double helical NAs such as physiological DNA are often represented by means of helical coordinates,[40−42] which implies a moderate loss of information, but offers two major advantages: (i) it is very close to chemical intuition, and (ii) it dramatically simplifies the definition of the DNA conformational space. Thus, for canonical base pairs, the conformational space of DNA can be reasonably represented by only $3(K - 1)$ rotational and $3(K - 1)$ translational degrees of freedom, where K is the number of base pair steps in the duplex. For large and regular polymers, the coordinate systems can be even more simplified by using a reduced set of global polymer parameters (like global twist, bending, or stretch; see ref 43).

Covariance matrices defined in the helical reference system ($\mathbf{C}_h$) can be manipulated to determine the flexibility of DNA with respect to perturbations along helical coordinates. As shown by Olson and co-workers[44,45] and Lankas and co-workers,[43,46] this can be obtained by inversion of $\mathbf{C}_h$, which yields (see eq 19) a stiffness matrix $\mathbf{F} = F(K_{ij})$, whose diagonal elements represent contributions due to deformations arising from pure helical variables, whereas off-diagonal components account for coupling terms. Once the stiffness matrix is known, the mesoscopic calculation of the elastic energy of a DNA helix can be easily done even for very large DNA fragments (see eq 20).

$$\mathbf{F} = kT\mathbf{C}_h^{-1} \tag{19}$$

$$E = \sum_a \frac{K_a}{2}(X_a - X_a^0)^2 + \sum_{a,b} \frac{K_{ab}}{2}(X_a - X_a^0)(X_b - X_b^0) \tag{20}$$

where X_a and X_b stand for two different helical coordinates ($a \neq b$), K_a stands for a diagonal force constant (for variable a), and K_{ab} represents out-of-diagonal force constants.

Note that the analyses of the stiffness matrix and of the elastic energy provide complete information on the global isotropic and anisotropic deformability of any NA, provided its geometry can be well-represented in the helical space.[12,13,30−32,43−46]

Computational Details. Trajectories for 14 B-DNA duplexes of length 10−16-mer (see Table 1) were obtained from isothermal−isobaric simulations ($T = 298$ K; $P = 1$ atm). The duplexes considered here involve different sequences and, in some cases, include point alterations, such as mismatched pairs, unusual bases, or the substitution of a phosphate by a neutral phosphonate. The PARM-98 AMBER force field[8] was used in conjunction with the TIP3P water model[47] to describe standard nucleotides and water. When needed, force-field parameters were developed to describe unusual bases using the RESP/HF-6-31G(d)[48] and PARMQMD[49] methodologies. SHAKE[50] was used in all simulations to constrain all bonds at equilibrium lengths, which

Table 4. Percentage of Variance Explained for Different Trajectories (Central Eight-mer) When Different Number of Eigenvectors Are Considered[a]

trajectory	1	5	10	50	100
traj1	25,1	63,2	78,2	95,9	98,5
traj2	28,0	61,4	76,6	95,4	98,3
traj3	21,9	60,2	76,6	96,1	98,4
traj4	30,0	66,2	79,4	96,5	98,6
traj5	31,7	66,8	79,3	95,9	98,5
traj6	16,2	52,4	70,0	93,8	97,5
traj7	24,5	64,3	76,9	95,1	98,1
traj8	31,0	66,7	79,0	95,5	98,3
traj9	21,7	57,0	72,7	94,4	97,8
traj10	20,1	57,6	73,4	94,9	98,0
traj11	23,9	62,4	74,4	94,6	97,9
traj12	18,0	56,5	73,6	94,8	98,0
traj13	21,9	58,7	74,0	94,8	97,9
traj14	17,5	57,8	72,9	94,6	97,8
average	23.7 ± 4.8	60.8 ± 4.8	75.5 ± 2.8	95.2 ± 0.7	98.1 ± 0.3
traj5[b]	29.7	57.3	72.5	94.5	97.7

[a] The average values with standard deviations are also shown. [b] This value corresponds to trajectory 5 considering the central 13-mer.

allowed us to use a 2 fs time step for integration of Newton's equations. All studied systems were neutralized by adding a suitable amount of Na^+ and hydrated by immersion in suitable rectangular boxes of TIP3P waters.[47] The PME strategy[3] and periodic boundary conditions were used to account for long-range effects. Some of the simulations were partially or totally obtained from our local trajectory library, whereas others were especially performed for this paper (see Table 1). All the trajectories were extended for at least 5 ns, and the trajectory comprised between the 1st and 5th nanosecond was used for analysis.

Some analyses were performed considering the entire helices, whereas others considered only the common eight-mer portion. Since the sequences were different, we were often forced to use simplified covariance matrices to facilitate comparison between different duplexes (purines were represented by atoms N9, C8, H8, N7, C5, C6, N6/O6, N1, C2, N3, and C4 and pyrimidines by N1, C6, H6, C5, C4, N4/O4, N3, C2, and O2; i.e., 20 atoms × base dimer). For other studies (for example, to study conformational transitions), a further reduced coordinate system was used by taking only four atoms for each base (C1′−C4−N9−C8 for purines and C1′−C6−N1−C2 for pyrimidines). Finally, a last set of analyses (for example, to compare essential movements between trajectories) was performed considering only backbone atoms (up to N1 or N9). In all cases, comparisons were performed always considering covariance matrices of the same dimensionality.

Results and Discussion

Essential Dynamics. Around 100 eigenvectors are able to explain nearly all the variance of the eight-mer DNA duplexes irrespective of its composition and of the presence of chemical alterations in the structure (see Table 4). Interestingly, not many more eigenvectors are necessary to explain all the variance of longer duplexes (see Table 4). Thus, 100 eigenvectors explain between 97.9% (traj11) and 98.5% (traj5) of the variance in $d(A)_8$, and the same number of eigenvectors represent 97.7% of the variance of $d(A)_{13}$.

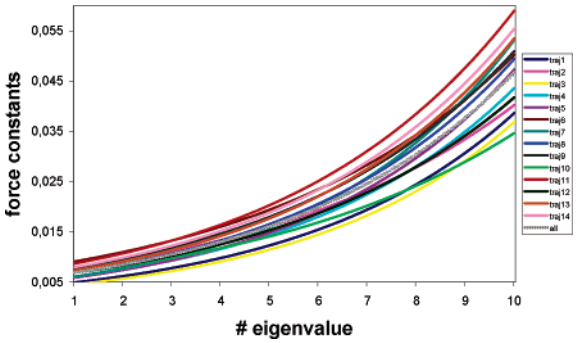

Figure 2. Force constants (in kcal/mol Å^2) associated with the first 10 essential deformations of different DNA duplexes. Line in gray labeled "all" corresponds to the exponential fit [$K_i = 0.005\,572 \exp(0.2126i)$] performed considering data for all the duplexes. Force constants were determined considering only the central eight-mer portion of the different DNA duplexes.

This suggests that the dynamics of the DNA is quite simple (see below). For qualitative comparison, when the same analysis is performed with a set of trajectories taken from our local database of small protein trajectories (PDB entries 1AA2, 1AA3, 1ACF, 1ACX, 1AG4, 1ARK, 1ATA, 1B40, 1BM8, 1CDZ, 1CK2, 1EXG, 1FOW, and 1HY8), the average variance explained by 100 eigenvectors is 88%.

The softer deformation modes have associated elastic force constants below 10 cal/mol Å^2 for all the central eight-mer segments (see Figure 2), and even for the 10th mode, force constants are, in general, below 50 cal/mol Å^2. This implies very low frequencies in the range of 13−44 (1st−10th) cm^{-1} for the studied systems, values that are similar to those obtained for proteins (range between 9 and 36 cm^{-1}) and to those associated with weak rotations in small molecules. The low force constants associated with the essential modes indicate that thermal energy can introduce important fluctuations in the DNA structure and that the DNA duplex is an extremely flexible entity irrespective of the sequence or the presence of chemical alterations. Thus, the rmsd's between the snapshots that have explored more distant regions along

Table 5. Dimensionality (see Methods Section) of the Configurational Space of the Different Trajectories[a]

trajectory	dimensionality
traj1	48
traj2	44
traj3	45
traj4	42
traj5	45
traj6	42
traj7	50
traj8	46
traj9	45
traj10	48
traj11	46
traj12	46
traj13	42
traj14	42

[a] In all cases, values are computed considering the central eight-mer fragments.

the first eigenvector are in the range 3.5−4.6 Å (these structures are selected from the values of the projection of the trajectory along the corresponding first eigenvector), a quite large value considering that they are computed for the central eight-mer and considering the deformation only along one eigenvector.

The increase in the value of the elastic constants with the eigenvectors follows well an exponential law (see Figure 1). In fact, the force constants (in kcal/mol Å^2) associated with the ith eigenvector can be reasonably determined (at least for eight-mer duplexes) using the equation $K_i = 0.005\,572 \exp(0.2126i)$, obtained by analyzing all the force constants of the central eight-mer. Trajectories leading to force constants far from these consensus values must be analyzed with special care. Also interestingly, the different lines in Figure 2 are very close for the fist eigenvector, whereas they diverge as higher modes are considered. This indicates that (in absolute terms) force constants associated with the first modes are less dependent on the nature of the oligonucleotides than those associated with higher modes.

The average dimensionality of the configurational space of the different eight-mer duplexes considered here is 45 ± 3 (see Table 5), which suggests that the configurational space of eight-mer DNAs can be represented in a small space of 42−50 dimensions corresponding to the important essential movements (see also Table 4). This small number sharply contrasts with the thousands of degrees of freedom of the studied duplexes and confirms the relative simplicity of the accessible configurational space of the DNA duplex. The difference in dimensionality due to changes in sequence, presence, or chemical alterations and sequence environments is rather small (see Table 5) and does not seem to be clearly related to the chemical nature of the duplex.

As expected, the dimensionality of the system increases with the length of the duplex (see Figure 3). An analysis of the data shows that, at least within the interval studied, we can predict the dimensionality of the system from the linear equation dim = 4.71 × length + 8.9 ($r^2 > 0.999$). This relationship suggests that, on average, each base pair adds only around five more dimensions to the DNA configurational space. These empirical equations can then be used to estimate the complexity of accessible space of large DNA duplexes, defining then suitable simulation protocols, since simulations that are able to explore a space of small dimensionality will not be large enough to reproduce that of larger spaces well.

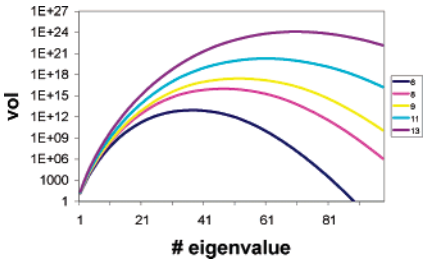

Figure 3. Variation in the configurational volume (Vol in Å^t; where t is the number of eigenvalues, logarithmic scale) for a common sequence [d(A)$_n$] with the length of the duplex and the number of eigenvalues.

There is a close similarity in the nature of movements sampled by the eight-mer duplexes in the different trajectories. Absolute similarity indexes (γ, eq 4 and ξ, eq 10) range typically between 0.6 and 0.8. The ξ values (eq 10) are generally larger than the γ ones (eq 4), showing that the similarity is greater for the most important deformation modes (see Table 6), whereas it decreases as higher modes are considered. Clearly, the new index seems to be more powerful than traditional Hess metrics[36] to properly weight the importance of low-frequency modes to draw the essential dynamics of DNA.

As expected, the best similarity is found when two similar sequences are compared (see, for example, similarity indexes for trajectories 8, 9, and 10 or for 5 and 11), and the worst are found when duplexes of different sequences and containing chemical alterations are compared (see Table 6). When the entire set of trajectories is combined into a single one, a consensus covariance matrix can be generated ($\mathbf{C}_x^{\text{cons}}$), containing not only data on the variability of DNA conformation due to natural deformations in the Cartesian space but also information on the deformation of the DNA structure due to change in the sequence or the presence of unusual modifications in the chemical structure. The eigenvectors obtained upon diagonalization of $\mathbf{C}_x^{\text{cons}}$ provide, then, information on both types of deformations (sequence-derived and due to thermal fluctuations). It is worth noting that these two types of conformational changes could, in principle, be very different, but a comparison of eigenvectors obtained from $\mathbf{C}_x^{\text{cons}}$ with those obtained by diagonalization of the different individual covariance matrices ($\mathbf{C}_x$) yields very large similarity indexes (in the range 0.7−0.8; see Table 3). This finding strongly supports the idea that there is an essential deformation pattern for the DNA, which can explain 70−80% of the normal flexibility of any DNA duplex. Furthermore, our results also suggest that the conformational changes in the DNA induced by alterations in its covalent structure happen, in most cases, along the natural deformation modes of the duplex.[31]

Entropy calculations. Entropies determined from Schlitter's[37] and Andricioaei and Karplus'[38] methods provide a

Table 6. Absolute Similarity Indexes (Indexes γ, Eq 5, above the Diagonal in Roman Font) and Indexes ξ, below the Diagonal in Italics) between the Eigenvectors Associated the Central Eight-mer Portion of the Different Duplexes[a]

	traj1	traj2	traj3	traj4	traj5	traj6	traj7	traj8	traj9	traj10	traj11	traj12	traj13	traj14	**All**
traj1		0.73	0.65	0.69	0.62	0.55	0.68	0.66	0.69	0.67	0.66	0.64	0.57	0.56	**0.72**
traj2	*0.84*		0.63	0.65	0.59	0.55	0.63	0.65	0.62	0.56	0.61	0.59	0.55	0.59	**0.68**
traj3	*0.76*	*0.75*		0.69	0.62	0.62	0.60	0.63	0.60	0.57	0.63	0.65	0.61	0.54	**0.69**
traj4	*0.78*	*0.76*	*0.74*		0.59	0.52	0.58	0.63	0.67	0.59	0.60	0.61	0.52	0.55	**0.65**
traj5	*0.73*	*0.70*	*0.73*	*0.74*		0.61	0.65	0.64	0.63	0.62	0.78	0.68	0.60	0.59	**0.69**
traj6	*0.63*	*0.65*	*0.68*	*0.63*	*0.70*		0.61	0.66	0.56	0.57	0.59	0.64	0.79	0.67	**0.68**
traj7	*0.76*	*0.74*	*0.69*	*0.71*	*0.77*	*0.70*		0.82	0.70	0.73	0.65	0.65	0.62	0.58	**0.79**
traj8	*0.77*	*0.73*	*0.72*	*0.76*	*0.79*	*0.72*	*0.89*		0.70	0.73	0.65	0.65	0.61	0.59	**0.80**
traj9	*0.79*	*0.75*	*0.69*	*0.74*	*0.74*	*0.66*	*0.79*	*0.78*		0.74	0.66	0.65	0.55	0.54	**0.72**
traj10	*0.76*	*0.68*	*0.68*	*0.70*	*0.73*	*0.65*	*0.81*	*0.79*	*0.80*		0.63	0.64	0.56	0.51	**0.71**
traj11	*0.77*	*0.75*	*0.73*	*0.77*	*0.87*	*0.69*	*0.79*	*0.79*	*0.76*	*0.74*		0.70	0.60	0.57	**0.69**
traj12	*0.72*	*0.67*	*0.73*	*0.74*	*0.79*	*0.68*	*0.72*	*0.74*	*0.71*	*0.72*	*0.82*		0.63	0.59	**0.72**
traj13	*0.65*	*0.64*	*0.65*	*0.66*	*0.73*	*0.85*	*0.70*	*0.72*	*0.68*	*0.66*	*0.71*	*0.70*		0.70	**0.66**
traj14	*0.69*	*0.68*	*0.65*	*0.71*	*0.75*	*0.76*	*0.74*	*0.76*	*0.70*	*0.66*	*0.74*	*0.71*	*0.78*		**0.63**
ALL	**0.81**	**0.77**	**0.79**	**0.75**	**0.78**	**0.72**	**0.83**	**0.84**	**0.79**	**0.81**	**0.80**	**0.79**	**0.71**	**0.75**	

[a] Similarities of each trajectory with respect to a "consensus DNA trajectory" obtained by combining the 14 individual trajectories are also displayed, in bold.

Table 7. Intramolecular Entropies (in kcal/mol K) Computed for the Different Trajectories from 3.5 ns Samplings[a]

name	Schlitter	Andreociaei–Karplus	rmsd fluctuations
trajectory 1	1.57/*1.73*	1.45	1.23
trajectory 2	1.54/*1.68*	1.42	1.03
trajectory 3	1.58/*1.79*	1.46	1.31
trajectory 4	1.51/*1.69*	1.39	1.15
trajectory 5	1.56/*1.72*	1.43	1.32
trajectory 6	1.51/*1.63*	1.39	1.07
trajectory 7	1.53/*1.71*	1.41	1.53
trajectory 8	1.57/*1.74*	1.43	1.59
trajectory 9	1.55/*1.70*	1.42	1.24
trajectory 10	1.61/*1.81*	1.48	1.56
trajectory 11	1.55/*1.71*	1.43	1.26
trajectory 12	1.57/*1.74*	1.44	1.35
trajectory 13	1.53/*1.69*	1.41	1.33
trajectory 14	1.52/*1.65*	1.39	1.29

[a] Values in italics are those obtained by extrapolating to infinite simulation time. Total rmsd fluctuations ($\langle \text{rmsd}^2 \rangle_t - \text{rmsd}_{\text{ref}}^2$; see eqs 13–14) from the average are in Å^2. The central eight-mer portions of all duplexes were considered in the calculations, where the base pairs were represented by a common reference system comprising 20 atoms.

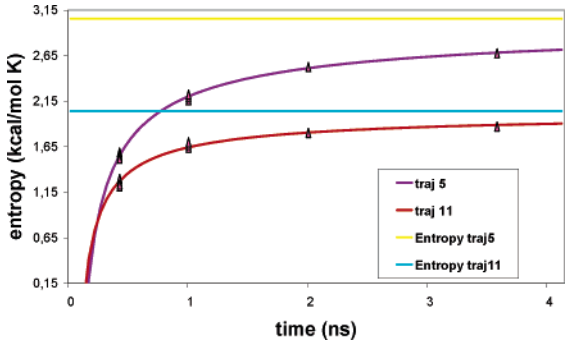

Figure 4. Change in (Schlitter) entropy with simulation time (snapshots collected every picosecond) for two duplexes of different lengths but the same composition: traj 5 is $d(A)_{15}$ and traj 11 is $d(A)_{11}$. Individual values used to fit the exponential profiles are shown, as well as the estimate of entropy at infinite simulation time (labeled as "Entropy traj*x*").

quantitative measure of the conformational freedom of a molecule moving in a quasi-harmonic regime. Andricioaei and Karplus' method provides systematically smaller values than the Schlitter approach, but in relative terms, both approaches give identical results (see Table 7). As noted previously (see Methods section), the entropy strongly depends on the extent of sampling used to create the mass-weighted covariance matrix. Not surprisingly, such dependence becomes more evident (see Figure 4) as the length of the duplexes increases, which warns against the quantitative value of entropy estimates of large duplexes obtained from 1 to 10 ns simulation times (note the difference between the last entropy estimate and the entropy at infinitive simulation time in Figure 4). In fact, a 5 ns simulation time is probably not enough to reproduce the real intramolecular entropy of even the shorter eight-mer duplexes.

Overall, entropies are only moderately dependent on the eight-mer duplex considered (a range of less than 10% variation in the total entropy, with a standard deviation in S_∞ of only 2%), confirming that global flexibility is mostly determined by the general polymeric properties of the duplex, and changes in sequence, presence of chemical alterations, or other moderate changes do not have a dramatic influence on the DNA entropy. Interestingly, the entropy differences between the DNA duplexes analyzed here are evident only for long simulation times, when the structure is allowed to explore less populated regions of the configurational space (Supporting Information, Figure S1). The fitting of entropy estimates at different simulation times to eq 18 yields similar α and β values for all the duplexes, and consensus values ($\alpha = 24.3 \pm 3.2$, $\beta = 0.62 \pm 0.03$; see eq 18) can be used to compute entropies at infinite simulation time from a unique estimate at 3500 ps with a reasonable error (rmsd = 0.02 kcal/mol K) for a variety of DNA sequences. Simulations yielding α and β values far from these consensus ones (for the central eight-mer) should be taken with caution, since

Table 8. Diagonal Stiffness Constants (Translations in kcal/mol Å^2 and Rotations in Kcal/Mol Degree^2) Associated with Harmonic Deformation of Local Helical Parameters for Some Selected Base Pair Steps[a]

steps	shift	slide	rise	tilt	roll	twist
AG/CT	1.36 (0.50)	2.14 (0.45)	6.34 (1.10)	0.029 (0.008)	0.024 (0.002)	0.018 (0.008)
GA/TC	1.54 (0.31)	2.04 (0.13)	8.30 (0.97)	0.040 (0.004)	0.025 (0.001)	0.032 (0.007)
GC/GC	2.07 (0.65)	3.61 (0.77)	10.46 (0.96)	0.045 (0.003)	0.030 (0.007)	0.041 (0.008)
GG/CC	1.19 (0.25)	1.78 (0.13)	8.37 (0.55)	0.042 (0.003)	0.024 (0.000)	0.030 (0.004)
TT/AA	1.66 (0.50)	2.72 (0.59)	8.31 (0.81)	0.035 (0.007)	0.025 (0.002)	0.032 (0.008)

[a] Values were averaged for all the base pair steps of a given type in the entire set of duplexes (the standard deviation is given in parentheses). Only values for which at least 10 examples were found are reported in this table.

they might be capturing conformational transitions moving out of the harmonic regime.

Helical Stiffness. The helical elasticity of DNA has a moderate dependence on the local sequence (20−50% in force constants; see Table 8), with some steps being systematically more rigid than others. However, the data in Table 8 show that the general helical elastic properties of the DNA are marked by its general polymeric structure rather than by its sequence. When stiffness constants for a given (5′-XY-3′) step are computed in different chemical environments (for example, in different positions of the same duplex or in different duplexes), no negligible differences are found, as noted in the standard deviations associated with averages in Table 8. The same range of variation is evident when present force constants are compared with those obtained by Lankas and co-workers[43,46] using sequences different from those considered here. Overall, this indicates that the molecular environment surrounding the base pair step can play a major role in determining the local rigidity of DNA, suggesting then that fragments longer than two base pairs are needed to capture the deformability of DNA. Data in Table 8, combined with an analysis of flexibility in Cartesian space (see above), clearly show that the concept of flexibility is rather diffuse and should be linked to the nature of the perturbation introduced to the system. Some duplexes might be more flexible in terms of twist deformations but more rigid for bending distortions. In any case, large flexibility in the local helical space might yield, depending on the sequence, small flexibility in the Cartesian space. Caution is then necessary when playing with very anisotropic concepts such as flexibility or rigidity.

Conclusions

(1) MD trajectories contain a large amount of information on the flexibility of DNA duplexes, but hidden in gigabytes of noise. Data mining methods such as those discussed here can be used for the routine analysis of nanosecond-scale MD simulations.

(2) DNA is a very flexible entity, whose essential deformation movements are associated with very small force constants. However, despite this large flexibility, the pattern of DNA deformability is not very complex and can be described by a relatively small set of movements.

(3) The pattern of essential movements of DNA is largely conserved irrespective of length, sequence, or chemical alterations. It is possible to obtain a reduced set of consensus essential movements of DNA duplexes that are able to explain a large amount of variance in the MD trajectories. These movements are also informative in understanding how the DNA can change its structure as a result of small covalent changes.

(4) For a given length, the level of disorder and flexibility in the Cartesian space of the different DNAs studied is quite similar. However, different DNAs can react in different ways when they are perturbed along specific helical coordinates, indicating that concepts such as flexibility or rigidity are not universal, but must be tightly related to the coordinate frame used to measure the conformational space and on the nature of the deformation introduced to the structure.

Acknowledgment. We thank Prof. Luque and two anonymous reviewers for their valuable comments. This work has been supported by the Spanish Ministry of Education and Science (BIO2002-06848), the BBVA, and La Caixa Foundation. We also thank the support of the Instituto Nacional de Bioinformatica (INB—Genoma España) and the Centre de Supercomputacio de Catalunya (CESCA).

Supporting Information Available: Projection of some of the trajectories in the first eigenvectors (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Levitt, M. *Cold Spring Harbor Symp. Quant. Biol.* **1983**, *47* (Pt 1), 251−62.

(2) Tidor, B.; Irikura, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983**, *1*, 231.

(3) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98* (12), 10089−92.

(4) Cornell, W. D.; Cieplak, P.; Baily, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−97.

(5) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117* (48), 11946−75.

(6) Foloppe, N.; Mackerell, A. D. *J. Comput. Chem.* **2000**, *21* (2), 86−104.

(7) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16* (3), 487−509.

(8) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.

(9) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182−196.

(10) Cheatham, T. E.; Kollman, P. A. *Annu. Rev. Struct. Dyn.* **2000**, *51*, 435−471.

(11) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350−357.

(12) Orozco, M.; Pérez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* **2003**, *32*, 350−364.

(13) Orozco, M.; Rueda, M.; Blas, J. R.; Cubero, E.; Luque, F. J.; Laughton, C. A. *Encyclopedia of Computational Chemistry*; Wiley: New York, 2004 (published online April 15, 2004; doi: 10.1002/0470845015.cn0080).

(14) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (41), 14771−5.

(15) Rueda, M.; Cubero, E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2004**, *87* (2), 800−11.

(16) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32* (14), 4269−80.

(17) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. *Biophys. J.* **2004**, *87* (6), 3799−813.

(18) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2000**, *304*, 803−820.

(19) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.

(20) Srinavasanm, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc*. **1998**, *120*, 9401−9409.

(21) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127−9.

(22) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246* (1, 2), 122−9.

(23) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem*. **2002**, *23* (14), 1297−304.

(24) Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Proteins* **2004**, *57* (3), 458−67.

(25) Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J. L.; Luque, F. J.; Orozco, M. *Proteins* **2005**, *58* (1), 101−9.

(26) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 14603−14612.

(27) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121* (25), 5981−5991.

(28) Amadei, A.; Linssen, B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412−425.

(29) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9513−9522.

(30) Noy, A.; Perez, A.; Lankas, F.; Luque, F., J.; Orozco, M. *J. Mol. Biol.* **2004**, *343* (3), 627−38.

(31) Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2004**, *32* (20), 6144−51.

(32) Noy, A.; Pérez, A.; Márquez, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc*. **2005**, *127*, 4910−20.

(33) Go, M.; Go, N. *Biopolymers* **1976**, *15* (6), 1119−27.

(34) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125* (26), 8007−14.

(35) Rueda, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc*. **2005**, in press.

(36) Hess, B. *Phys. Rev.* **2000**, *62*, 8438.

(37) Schliter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.

(38) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.

(39) Harris, S.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **2001**, *123*, 12658.

(40) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300* (4), 819−40.

(41) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31* (17), 5108−21.

(42) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn*. **1989**, *6* (4), 655−67.

(43) Lankas, F.; Sponer, J.; Hobza, P.; Langowski, J. *J. Mol. Biol.* **2000**, *299*, 695−709.

(44) Olson, W. K.; Zhurkin, V. B. *Curr. Opin. Struct. Biol*. **2000**, *10*, 286−97.

(45) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163−8.

(46) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E., III. *Biophys. J.* **2003**, *85* (5), 2872−83.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(48) Bayly, C. E.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.

(49) Aleman, C.; Canela, E. I.; Franco, R.; Orozco, M. *J. Comput. Chem.* **1991**, *12*, 664−674.

(50) Rickaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1997**, *23*, 327−341.

(51) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323−2324.

CT050051S