doi:10.1016/j.jmb.2004.07.048 J. Mol. Biol. (2004) 343, 627–638

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Relative Flexibility of DNA and RNA: a Molecular Dynamics Study

Agnes Noy[1], Alberto Pérez[1], Filip Lankas[2], F. Javier Luque[3] and Modesto Orozco[1,4]*

[1]*Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028 Spain*

[2]*Bernoulli Institute for Mathematics, EPFL (Swiss Federal Polytechnical Institute) 1015 Lausanne, Switzerland*

[3]*Departament de Fisicoquímica Facultat de Farmàcia Universitat de Barcelona Avgda Diagonal 643, Barcelona 08028, Spain*

[4]*Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1 Barcelona 08028, Spain*

**Corresponding author*

State of the art molecular dynamics simulations are used to study the structure, dynamics, molecular interaction properties and flexibility of DNA and RNA duplexes in aqueous solution. Special attention is paid to the deformability of both types of structures, revisiting concepts on the relative flexibility of DNA and RNA duplexes. Our simulations strongly suggest that the concepts of flexibility, rigidity and deformability are much more complex than usually believed, and that it is not always true that DNA is more flexible than RNA.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: DNA; RNA; flexibility; molecular dynamics; similarity indexes

Introduction

Under physiological conditions DNA exists as an elongated helix known as *B*-form, while if possible RNA forms a double helix more compact than that of DNA,[1–3] which is named the *A*-form. The different puckering of sugars in *B* (South to South-East) and *A* (North) forms alters the structure of the grooves,[1–4] which changes completely the ability of the nucleic acids to interact with other molecules, particularly with proteins.[1–3] It seems the different helical structure of DNA and RNA might determine their different role in the cell. However, structure itself is not able to explain the incredible ability of proteins to distinguish between nucleic acid structures. For example, only a fraction of very similar structural distortions of DNA are recognized by the UvABC excision repair mechanism.[1–3,5] Indeed, it is known that RNase H recognizes and degrades DNA–RNA hybrids, but it is inactive against RNA–RNA duplexes and other RNA-hybrid molecules, despite the fact that all these duplexes pertain to the *A*-family.[6,7] More surprisingly, the complex mechanism involving gene silencing by small interference RNAs is activated by duplex RNA, but not by other RNA hybrid duplexes of similar structure.[8,9] Clearly, besides the general helical structure, other properties must be exploited by nature to distinguish between helical structures.

RNA adopts very unique conformations in the cell, like those of ribozymes, transfer or ribosomal RNAs. However, it is believed that this is mostly due to the fact that cellular RNA is single-stranded and intramolecular RNA duplexes contain unpaired bases and mismatches which favour strong twists and kinks in the helix. However, it is generally accepted in the scientific community that the DNA double helix is more flexible than the RNA one. This assumption is supported by indirect low resolution experimental data[10,11] like that derived

Abbreviations used: MD, molecular dynamics; RMSD, root mean square deviations.

E-mail address of the corresponding author: modesto@mmb.pcb.ub.es

0022-2836/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.

from electron micrography, gel electrophoresis, hydrodynamic measurements or fiber diffraction.[12] ^{31}P NMR experiments of highly oriented fibers also suggest that *B*-DNA is more flexible than *A*-RNA.[13,14] Protein NMR experiments also suggest that the RNA duplex is more rigid than the DNA one for equivalent sequences,[13,14] a finding further supported from the comparison of experimental NMR *J*-coupling constants with the values derived from molecular dynamics (MD) simulations.[15] Caution is, however, necessary, since we cannot ignore that NMR data provide information about local fluctuations in DNA microenvironments, it being less suitable to ascertain global structural changes of the duplex.

Crystal data deposited in the Protein Data Bank have been traditionally used to support the greater flexibility of *B*-DNA compared to *A*-RNA. Thus, depending on sequence, solvent composition or the presence of certain ions different *B*-like conformations (*C*-, *D*- or *T*-forms) can exist,[1–3] an effect that has not been detected for *A*-RNA. The larger occurrence of lattice distortions in DNA crystals[16,17] and the ability of a given DNA structure to crystallize in different space groups[15] have also been used to support the higher flexibility of *B*-DNA compared to *A*-RNA. However, we cannot ignore that similar lattice-dependence effects are detected for *A* forms.[18,19] Furthermore, a recent extensive comparison of crystal structures of *B*-DNA and *A*-RNA[20] have raised doubts on the hypothesis that DNA is always more flexible than RNA. Thus, not only the number of space groups found in RNA structures is very similar to that found in *B*-DNA, but no relevant differences are found in both thermal factors and resolution between *B*-DNA and *A*-RNA crystals. In summary, the analysis of the PDB structures does not support the view of *B*-DNA as an "intrinsically" more flexible entity than *A*-RNA.

Different theoretical calculations have explored the structure and flexibility of DNA and RNA duplexes. The landmark work by Cheatham and Kollman[15] provided an exhaustive MD comparison of DNA and RNA duplexes in aqueous solution. The analysis of the torsions around α, χ and γ dihedral angles and of sugar puckerings during 2 ns trajectories led the authors to conclude that *A*-RNA was more rigid than *B*-DNA.[15] The authors demonstrated the complex nature of the interactions rigidifying the RNA *versus* the DNA, suggesting that perhaps high ordered water molecules around the RNA reduce the flexibility of RNA. Similar calculations were later performed by Auffinger and Westhof,[21,22] who compared $d(CG)_6 \cdot d(CG)_6$ and $d(TA)_6 \cdot d(TA)_6$ DNAs with the corresponding RNA duplexes, $r(CG)_6 \cdot r(CG)_6$ and $r(UA)_6 \cdot r(UA)_6$. From the analysis of dihedral changes along the sequence and their time fluctuations, DNA was found to be more flexible than RNA. Only one MD simulation, performed recently by MacKerell's group has raised doubts on the dominant hypothesis that DNA is more flexible than RNA.[23]

In summary, despite the lack of definitive evidence, there is a general consensus that DNA duplex is more flexible than the RNA one under the same conditions. However, what does flexibility mean? Conceptually, it denotes the ability of a given structure to be deformed as a result of an external perturbation. For highly anisotropic macromolecules such as nucleic acids, such a definition is perhaps too ambiguous, as it would encompass a wide variety of structural deformations. For example, a molecule can be locally flexible but globally rigid when flexibility motions in different regions cancel. Here, we used state of art MD simulations to reinvestigate the relative flexibility/rigidity of *B*-DNA and *A*-RNA using a rigorous definition of these two concepts. Our purpose is then not only to obtain a qualitative picture of how "rigid" are DNA and RNA, but to describe for the first time the very complex scenario of the structural flexibility of nucleic acids.

Results and Discussion

Global structural descriptors

As it has been largely discussed,[24,25] current force-fields provide stable trajectories for both DNA and RNA duplexes. This trend is noted in the small root mean square deviations (RMSD) determined with regard to the MD-averaged structures (see Table 1), indicating that despite local oscillations (in the sub-nanosecond time scale) the two structures sample well defined regions of the conformational space centred on their respective averaged structures. In turn, this supports the validity of simple pseudoharmonic approaches in the analysis of fluctuations detected in MD simulations (see below).

The RMSD with respect to the corresponding canonical fiber[26] and crystal structures (PDB entries 1BNA and 157D for DNA and RNA duplexes, respectively) are also reasonably small (see Table 1), which indicates that the sampled structures are close to the experimental ones. Interestingly, the RMSD between RNA and the *A*-form fiber structure is clearly smaller (see Table 1) than that obtained between DNA and the *B*-form fiber one. This finding, which demonstrates the suitability of Parm99 to reproduce RNA duplexes, suggests that the structure of RNA in the fiber and solution does not change remarkably, while a larger change is expected for DNA. However, care must taken to conclude that DNA is more flexible than RNA, because we must stress that the standard deviation of the RMSD between the sampled conformations in DNA and RNA trajectories and the corresponding MD-averaged structures are similar (see Table 1), a result that does not support the assumption that RNA is more rigid than DNA.

The distance (C1′–C1′) between contiguous

Table 1. Average root mean square deviation (RMSD; in Å) between DNA/RNA trajectories and reference structures

	B-form[a]	*A*-form[a]	Crystal[b]	MD-averaged[c]
DNA	2.22(0.34)	4.43(0.44)	2.05(0.39)	1.16(0.29)
RNA	5.16(0.55)	1.63(0.36)	3.54(0.22)	1.23(0.33)

Standard deviations (in Å) are given in parentheses.
[a] Canonical structures from Arnott's data.[26]
[b] Crystal structures of DNA (PDB 1BNA) and RNA (157D). Two mismatches of 157D were not considered to compute the RMSD.
[c] Structures obtained by averaging the snapshots collected during the last ns of the corresponding MD simulations.

nucleotides is a simple, yet accurate general descriptor of the helical structure.[27] For each collected snapshot, the sequence-averaged C1′–C1′ distance and its associated standard deviation were computed. This calculation was then repeated along the entire trajectory to obtain time-averaged mean C1′–C1′ distances, and the associated standard deviations. As suggested by Hunter and co-workers[27] C1′–C1′ discriminates very well between DNA and RNA (see Figure 1(top)), since RNA average distances are always larger than those of DNA, and no overlap exists between DNA and RNA C1′–C1′ distance distributions. The oscillations (see Figure 1(top)) of the mean C1′–C1′ distance are slightly larger for RNA (0.07 Å SD) than for DNA (0.05 Å SD). This suggests that RNA is slightly more flexible than DNA. However, the oscillations in the standard deviations associated with the sequence-averaged C1–C1′ distance in each snapshot are larger in DNA (0.36 Å SD) than in RNA (0.27 Å SD), suggesting the opposite, i.e. that DNA is more flexible than RNA (see Figure 1(bottom)). Thus, the analysis of the C1′–C1′ distances and their fluctuations do not provide conclusive evidence about the relative flexibility between DNA and RNA. The same lack of conclusions appears when the distribution of the groove widths for DNA and RNA is examined (see Figure 2), since the minor groove fluctuates more in DNA than in RNA; but the reverse trend is found concerning the fluctuations in the major groove.

In summary, the analysis of the global structural descriptors presented in this section does not convincingly support the supposed higher flexibility of DNA compared to RNA. More precise analysis needs to be made to analyze the nature of the differences in flexibility of the nucleic acids.

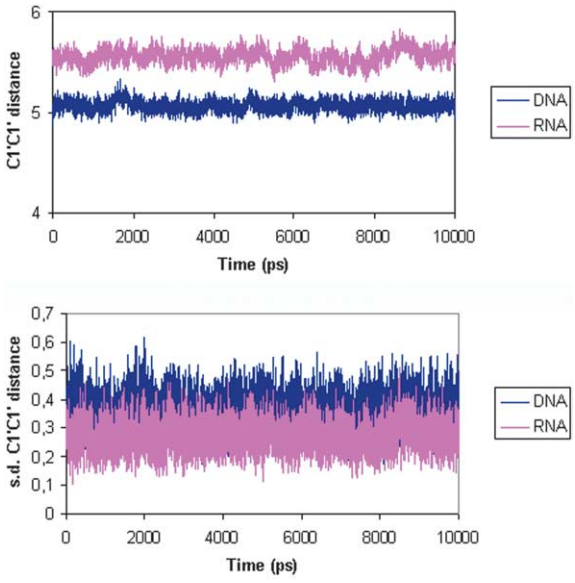

Figure 1. Top: Variation of the sequence averaged C1′–C1′ distance (in Å) along DNA and RNA trajectories. Bottom: Variation of the standard deviations (in Å) associated to the sequence-averaged C1′–C1′ distance (in Å) along DNA and RNA trajectories.

Entropy calculations

The intramolecular entropy can be used as a direct estimate of the global flexibility of DNA and RNA. As noted above, MD-derived entropy estimates depend on the length of the trajectory. However, the values obtained for 10 ns trajectories and those extrapolated to infinite time[28] are very close (see Table 2), which suggests a good convergence in the calculations, as expressed in the small

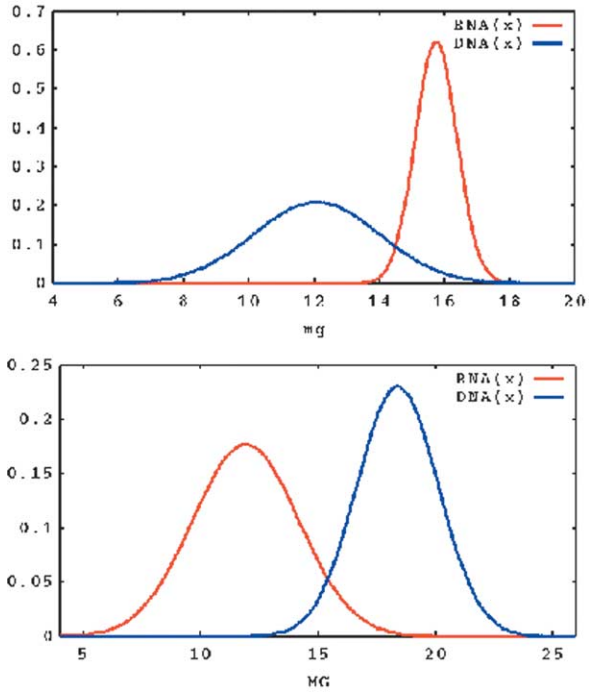

Figure 2. Distributions (fitted to normalized Gaussians) of the width (in Å) of the minor (top) and major (bottom) grooves of DNA and RNA in the MD simulations.

Table 2. Intramolecular entropies (in kcal/mol K) computed using Schlitter's (roman font) and Andreocci-Karplus's methods (italics) for DNA and RNA

	$S(t=9$ ns)	$S(t=\infty)$	$S(3)$	$S(10)$
DNA (all atoms)	2.0067	2.14(0.05)	0.0328	0.0988
	1.8321	*1.93(0.05)*	*0.0327*	*0.0983*
RNA (all atoms)	1.8228	1.90(0.03)	0.0342	0.0976
	1.6516	*1.71(0.03)*	*0.0340*	*0.0971*
DNA (nucleobases)	0.8946	0.91(0.01)	0.0286	0.0849
	0.8205	*0.84(0.01)*	*0.0285*	*0.0845*
RNA (nucleobases)	0.8893	0.91(0.02)	0.0300	0.0860
	0.8145	*0.83(0.01)*	*0.0299*	*0.0858*
DNA (backbone)	1.3424	1.43(0.07)	0.0320	0.0962
	1.2395	*1.34(0.10)*	*0.0319*	*0.0956*
RNA (backbone)	1.1429	1.18(0.03)	0.0333	0.0944
	1.0950	*1.09(0.03)*	*0.0331*	*0.0940*

Total intramolecular entropies were determined considering all equivalent atoms. Partial entropies were computed considering only nucleobases or backbone atoms. In all the cases, values were determined considering all the principal components, as well as only the first three and tenth ones. For the total entropy, not only the value directly obtained from the 10 ns trajectory, but also that extrapolated at infinite simulation time is displayed (see Methods for details).

standard error associated to S_∞. This finding, in conjunction with the excellent agreement between the two methods used for the calculation of entropy, gives strong confidence on the quality of the configurational entropies determined for DNA and RNA.

The intramolecular entropy of DNA is 11–13% larger (around 0.24 kcal/mol K; see Table 2) than those of RNA. Therefore, DNA is globally more disordered and flexible that RNA, which supports the generally accepted picture of the relative DNA/RNA flexibility. However, a partitioning analysis of the entropy contributions shows that the concept of flexibility is more complex than expected. Thus, when the entropy is computed considering only the first ten principal components, the entropy of DNA is only around 1% larger than that of RNA, and when only the first three principal components are considered (see Table 2) the RNA appears 4% more disordered than DNA. This finding suggests that disorder in RNA stems from a relatively small number of "very soft" motions, while for DNA is the result of many "soft" motions.

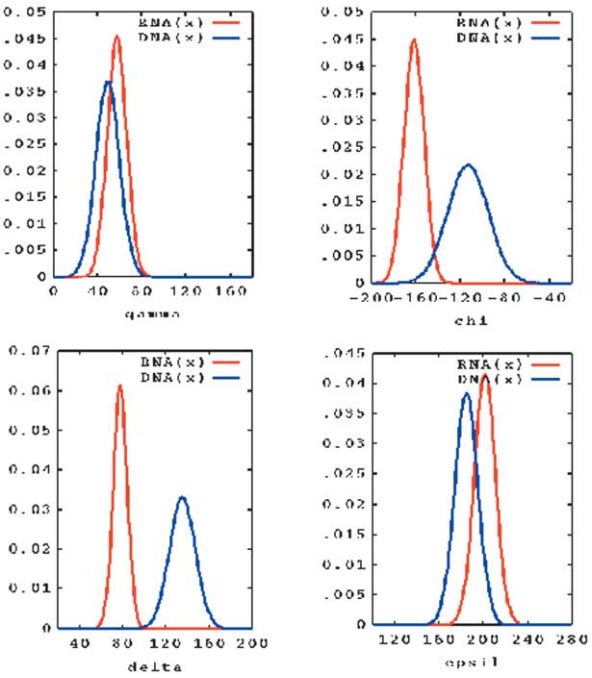

Figure 3. Distributions (fitted to normalized Gaussians) of selected backbone angles (γ,χ,δ,ε) in DNA and RNA trajectories.

Diagonalization of mass-weighted covariance matrices built up by considering only certain groups of atoms allows us to examine their contribution to the global disorder in DNA and RNA. For our purposes here, we have analyzed separately the contributions of atoms in the backbone and in the nucleobases. The results in Table 2 clearly demonstrate that the larger intramolecular entropy of DNA is due to the greater disorder in the backbone. This finding is further supported from inspection of the distributions of selected backbone angles sampled along the trajectory (see Figure 3). However, the fluctuations in the backbone do not affect the nucleobases, whose entropy in DNA and RNA is not very different. Finally, it is worth noting that the addition of backbone and nucleobase entropies approaches the total entropy, confirming the relatively small coupling between backbone and nucleobases movements.

Essential dynamics analysis

Similarity analysis (see equations (1) and (2)) shows that, in general, the easiest essential motions in DNA and RNA are similar, they being associated with twisting and bending of the polymers. When similarity is computed using 500 essential movements, absolute all-atoms similarity indexes $\gamma_{DNA/DNA}$ and $\gamma_{RNA/RNA}$ are around 0.87, which are close to similarity indexes obtained considering only the backbones (see Table 3). This demonstrates the larger complexity of the essential movements of backbones compared to those of the nucleobases. Furthermore, the all-atom cross similarity index $\gamma_{RNA/DNA}$ is around 0.74 (yielding a relative

Table 3. Absolute (γ) and relative (κ) similarity indexes for DNA and RNA essential movements

	All atoms	Nucleobases	Backbone
Ten eigenvectors			
$\gamma_{RNA/DNA}$	0.564	0.616	0.566
$\gamma_{RNA/RNA}$	0.860	0.926	0.865
$\gamma_{DNA/DNA}$	0.794	0.829	0.770
$\kappa_{RNA/DNA}$	0.682	0.701	0.692
500 eigenvectors			
$\gamma_{RNA/DNA}$	0.744	0.938	0.799
$\gamma_{RNA/RNA}$	0.873	0.967	0.919
$\gamma_{DNA/DNA}$	0.878	0.967	0.916
$\kappa_{RNA/DNA}$	0.850	0.970	0.870

Calculation is repeated considering the first ten and 500 essential movements.

$\kappa_{RNA/DNA}$ of 0.85; see Table 3), reflecting the different nature of the essential movements in DNA and RNA backbones ($\gamma_{RNA/DNA}=0.799$ and $\kappa_{RNA/DNA}=0.87$ in the backbone; see Table 3).

Absolute self-similarity indexes obtained considering only the first ten eigenvectors are much higher for RNA than for DNA (see Table 3). This finding demonstrates that more essential movements are necessary to explain the flexibility of DNA than that of RNA, which reflects the higher complexity of DNA flexibility compared to that of RNA.

In summary, despite a general similarity differences exist in the dynamics of DNA and RNA. These differences are mostly related to the higher complexity of movements in DNA, and to the nature of the essential movements of DNA and RNA backbones.

Stiffness analysis

The stiffness analysis provides useful information on the force necessary to deform a molecule along a given geometrical variable. Thus, the analysis performed using the eigenvectors obtained by diagonalization of the covariance matrix as deformation variables can be used to measure the deformability of DNA/RNA along the axis defined by their essential motions. Elastic force constants associated with the first five essential movements of

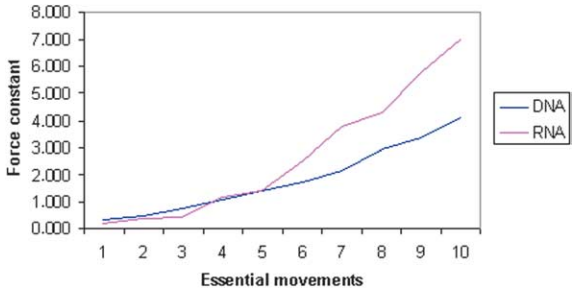

Figure 4. Force constants (in cal/mol Å^2) associated to different essential movements of DNA and RNA. Essential movements are labelled in decreasing order of impact in the explanation of trajectory variance.

RNA are similar or even weaker than those of DNA, but the situation dramatically changes when further essential movements are considered (see Figure 4), and the RNA becomes stiffer than DNA. Therefore, RNA is very deformable along a small set of essential motions, but very rigid along all the others, whereas DNA has a more degenerated pattern of deformability (see Figure 4). Note that this agrees with the picture of DNA as a molecule with a much more complex flexibility pattern than RNA.

The stiffness analysis performed using essential motions as deformation variables provides information on the natural deformability of nucleic acids. However, it is difficult to interpret in term of canonical motions (see Methods). Thus, we extended the stiffness analysis using helical parameters as deformation variables. First, we used a reduced set of global parameters (tilt, roll, twist and stretch) for the 10mer central portion of DNA and RNA duplexes. The global parameters are defined essentially,[29] and also we describe the geometry of the fragment using the X3DNA analyzer to maintain consistency with other parts of this study. The global stretch is the sum of distances between neighbouring base-pairs along the fragment, global twist is the sum of base-pair step local twists. We define a middle frame by taking its z-axis as the average of the normal vectors of the first and last base-pair in the fragment, and its x-axis as the projection of the x-axis of the central base-pair onto the plane perpendicular to the z-axis. The y-axis complements the set to form a right-handed triad. The global roll and tilt are then defined exactly as roll and tilt in the X3DNA algorithm, using the middle frame just described. Since the x-axis of a base-pair (as defined by X3DNA) is close to the dyadic vector pointing to the major groove, our global roll measures the bending of the fragment towards the major groove in its center, and global tilt describes the bending in the perpendicular direction. If the fragment has even number of pairs, we first average the x-axes of the two central base-pairs.

The stiffness at the global level can be described by the stretch modulus and by twisting and bending persistence lengths. Since we measure bending in two perpendicular directions, we have two bending persistence lengths corresponding to the global roll (bending towards the grooves) and global tilt. In this way, we can capture a possible anisotropy in the elastic behaviour. For longer fragments (several helical turns) the anisotropy is in general supposed to disappear and the bending stiffness is then fully described by one, "isotropic" bending persistence length, which can be estimated as the harmonic average of the two anisotropic ones.[29]

Table 4 indicates that deformation of the global twist of DNA is easier than that of RNA, but the stiffness of DNA and RNA with respect to roll is similar, and quite surprisingly DNA is stiffer than RNA in terms of global tilt and stretch. Caution is

Table 4. Elastic force constants for deformations along a reduced set of global helical parameters (angular force constants in kcal/mol degree^2 and displacement force constants in kcal/mol Å^2) computed for the central 10mer portion of DNA and RNA duplexes (top); RNA and DNA length independent persistence lengths for tilt, roll and twist deformations (in nanometers) and stretch modulus (in picoNewtons) (bottom)

	Global tilt[a]	Global roll[a]	Global twist[a]	Global stretch[b]
Top				
DNA	0.0043	0.0059	0.0056	1.50
RNA	0.0030	0.0058	0.0129	0.80
Bottom				
DNA	73.85	101.85	96.15	3269
RNA	56.01	107.80	242.00	1891

[a] In kcal/mol deg^2.
[b] In kcal/mol Å^2.

then necessary when general global concepts like flexibility or rigidity are applied to DNA and RNA.

Stiffness analysis performed using local helical parameters (tilt, roll, twist, shift, slide and rise) computed using X3DNA[30] provides a direct measure of the average local deformability of each base-pair in DNA and RNA. Fortunately for our purposes, the local helical values sampled in our MD simulations follow a normal distribution ($r^2 > 0.998$ between the real distribution and the Gaussian's predicted one (equation (7)), which supports the goodness of the harmonic analysis described below). RNA is around 190%, 75%, 35% and 12% stiffer than DNA regarding twist, slide, shift and rise deformations (see Table 5). On the contrary, DNA is stiffer than RNA in terms of roll and tilt deformations, but the differences are small (between 10% and 20% increment in force-constants; see Table 5). When cross-terms are considered (equation (8)), qualitative results remain mostly unaltered (see Table 6), reflecting the fact that stiffness matrix is dominated by diagonal terms. The only significant difference between force constants in Table 5 and diagonal force-constants in Table 6 appears for rise, as a consequence of the strong (especially for RNA) positive coupling of rise with slide and the negative one with twist and roll.

As a whole, results in Tables 5 and 6 show that for most local motions DNA is more flexible than RNA, but it is difficult to evaluate the implication of the different force constants in the local deformability. To obtain additional information in this point we computed the local deformation energy (see equation (9)) of DNA and RNA in front of random perturbation of helical parameters (to obtain meaningful results random changes are limited to a maximum of two times the largest standard

Table 5. Elastic force constants for deformations along local helical parameters of DNA and RNA

	Tilt[a]	Roll[a]	Twist[a]	Shift[b]	Slide[b]	Rise[b]
DNA	0.031	0.016	0.016	1.094	1.564	5.174
RNA	0.028	0.013	0.046	1.484	2.709	5.814

Couplings between different parameters are neglected.
[a] In kcal/mol deg^2.
[b] In kcal/mol Å^2.

Table 6. Stiffness matrix for deformations along local helical parameters of DNA and RNA

	Tilt	Roll	Twist	Shift	Slide	Rise
DNA						
Tilt	0.031	0.001	0.000	−0.043	−0.001	0.004
Roll		0.023	0.004	−0.016	−0.011	−0.021
Twist			0.014	−0.003	−0.060	−0.186
Shift				1.228	0.044	−0.001
Slide					1.886	0.951
Rise						7.217
RNA						
Tilt	0.026	0.000	−0.001	−0.180	−0.007	0.004
Roll		0.015	0.001	0.001	−0.007	−0.111
Twist			0.052	−0.020	−0.141	−0.156
Shift				1.369	−0.030	−0.013
Slide					3.193	1.077
Rise						6.183

All couplings (in kcal/mol deg. Å) are included. Since the Tables are symmetric, only the upper part is shown.

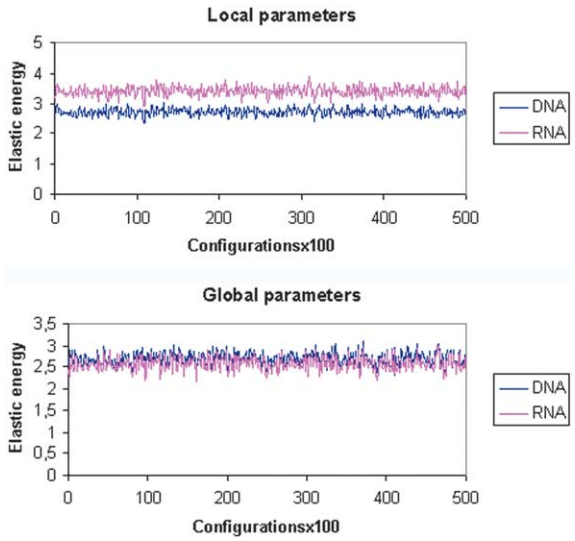

Figure 5. Elastic energy (in kcal/mol) associated to local (top) or global (bottom) distortions of DNA and RNA. In order to reduce noise, values represented here are averaged in blocks of 100 configurations.

deviation for each helical parameter). The elastic energy profiles (see Figure 5) clearly demonstrate that in terms of local deformations RNA is stiffer than DNA, despite that for a few types of special deformations the opposite situation is found (see also Tables 5 and 6). On the contrary, when the same analysis is performed using global elastic constants, RNA is not found to be more rigid than DNA (see Figure 5), in line with the results found in essential dynamics and entropy analysis.

In summary, stiffness analysis using either essential motions or helical coordinates demonstrates that the concepts of flexibility/rigidity in DNA and RNA are complex, and that the ability of a polymer to be perturbed by an external force depends on the type of deformation. Overall, DNA is more susceptible than RNA to local deformations, but the resulting local geometrical changes are more efficiently propagated to the global level in RNA.

Analysis of interaction profiles

As noted by other authors,[15,21,22] the relative flexibility of DNA and RNA is probably not only a direct consequence of its covalent structure, but also of its solvent environment (water and counterions). Solvation maps (see Figure 6) are clearly different for DNA and RNA because of the different groove distribution and the presence of an extra hydroxyl group, which alters the electrostatic potential around the polymer (see Figure 7). The total number of water molecules that solvate the RNA duplex is slightly larger (by around three per

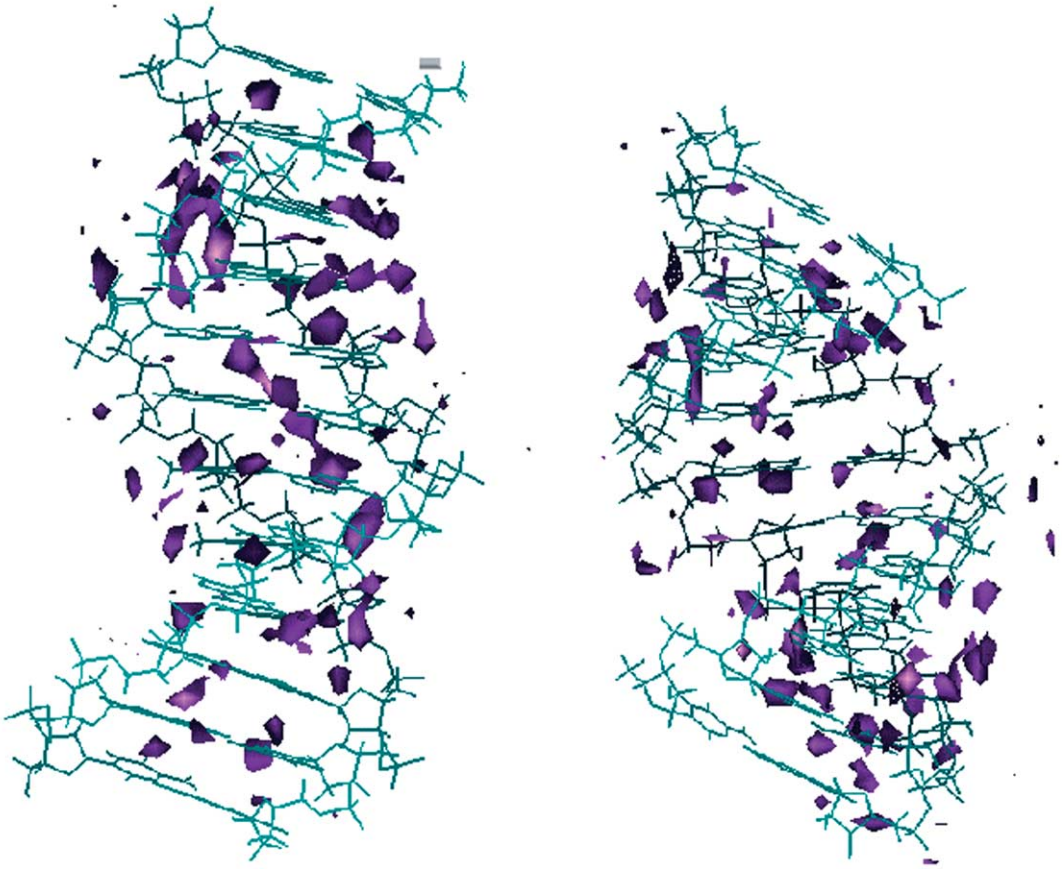

Figure 6. Solvation maps for DNA (left) and RNA (right). Water density contours correspond to an apparent density of 2 g/ml.

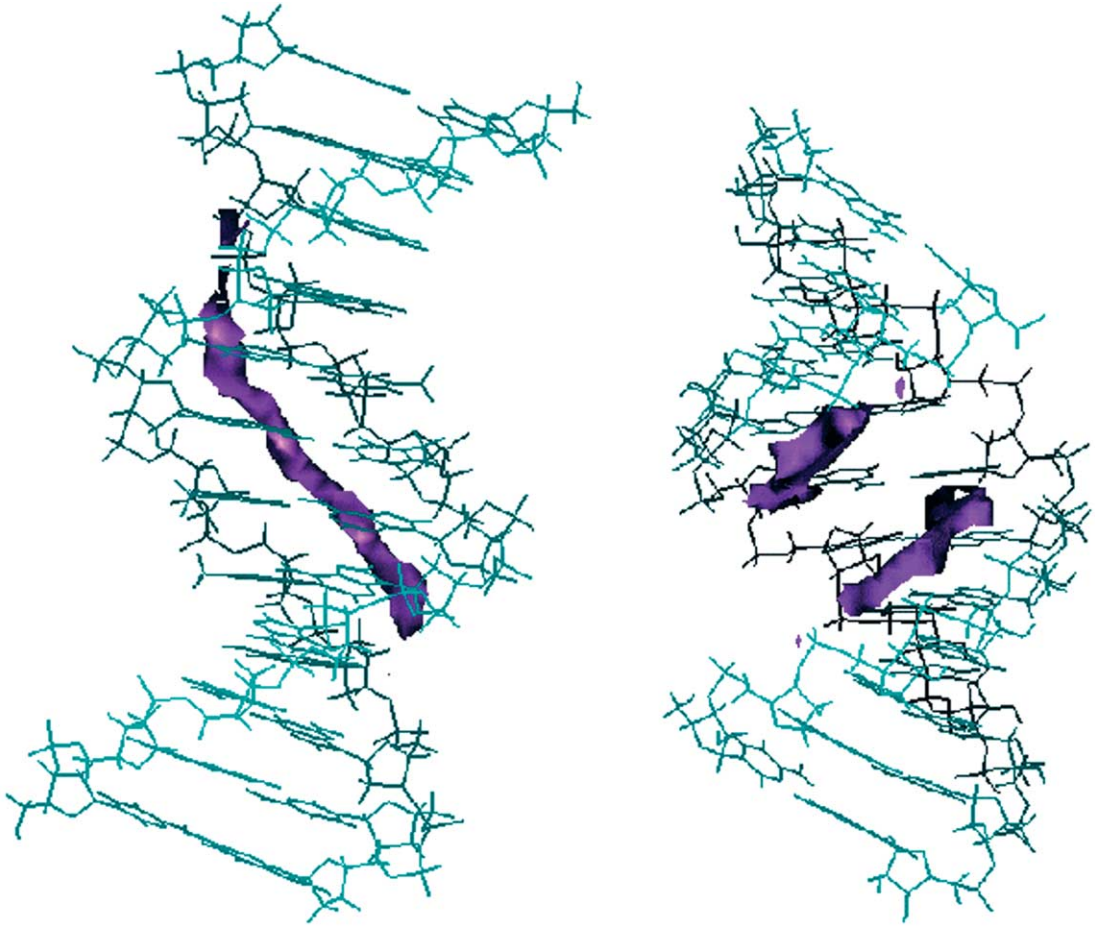

Figure 7. Classical molecular interaction potentials (in kcal/mol) for DNA (left) and RNA (right). Contour plots correspond to −4 kcal/mol for RNA and −3 kcal/mol for DNA.

base-pair in average; see Table 7) than for the DNA, the difference being mostly located in the backbone (see Table 7). However, longer residence times are found in general for DNA than for RNA, including those in the backbone (see Table 8), which is more flexible for DNA than for RNA (see above). These results strongly suggest that polymer flexibility is not closely related to specific solvation, and that the DNA/RNA motions are slow enough to allow a fast reorganization of the solvent.

The different electrostatic potential around DNA and RNA generates a different pattern of interaction with Na^+ (see Figure 7), thus justifying a different counterion distribution around DNA and RNA. Not only are Na^+ located closer to RNA than to DNA, but they have larger residence time for RNA when bound to the nucleic acid structure (up to 2 ns residence times are found for RNA in our simulations; see Table 7). Most of these "extra" bound-Na^+ are located in the major groove of the RNA, which seems to be a better cationic binding site than the minor groove of DNA. It is worth to note that similar results were previously obtained by Westhof and co-workers[21,22] and Cheatham & Kollman.[15] Thus, though much caution is necessary in the analysis of counterion distribution due to the slow Na^+ equilibration,[25] MD simulations suggest that the ability of RNA to tightly capture Na^+ is larger than that of DNA. However, we must emphasize that, as shown in Figure 2, the major groove is more flexible in RNA than in DNA, which indicates that Na^+ binding has not a direct relationship with differential flexibility.

Conclusions

Essential dynamics analysis strongly suggests

Table 7. Average number of water molecules located at each base-pair of DNA and RNA

	Backbone	m-groove	M-groove	Grooves (both)	Total
DNA	10.4	6.5	8.4	14.9	25.3
RNA	12.8	6.1	9.3	15.4	28.2

A contact is detected when there is a water oxygen at less than 3.5 Å from a heteroatom of DNA or RNA.

Table 8. Maximum residence times (in ps) for water molecules (roman font) or sodium (in italics) bound to polar groups of DNA/RNA

Atom	DNA	RNA
O2′	–	244
O4′	617	171
O3′	270	113
O1P	518/*668*	234/*723*
O2P	410/*298*	510/*600*
O5′	266	388
Cyt O2	600/*353*	200/*273*
Gua N3	438	145
Gua N2	534	244
Thy/Ura O2	521	245
Ade/Ura N3	571	175
Cyt N4	297	332
Gua O6	537	544
Gua N7	352/*121*	465/*655*
Thy/Ura O4	402/*351*	374/*449*
Ade N6	563	442
Ade N7	558/*415*	500/*2006*

First block, backbone atoms; second block, atoms in the minor groove; third block, atoms in the major groove.

that the pattern of flexibility of DNA and RNA duplexes is different. In the first case there are many possible deformations of low energetic cost, while for RNA there are a few very soft deformations, but the others are very stiff.

Entropy and stiffness analysis suggest that overall the DNA duplex is more disordered and can fluctuate more than the RNA one. However, this greater flexibility of DNA is mostly local and due to backbone fluctuations that do not introduce large global changes in the helix.

Stiffness analysis using local helical parameters shows that DNA is much more flexible than RNA in terms of local twist, slide and shift and slightly more flexible for local rise. RNA is slightly more flexible in terms of local roll and tilt deformations. Considering global helical deformations the DNA is more deformable in terms of global twist, but stiffer for global tilt and stretch.

Overall, molecular dynamics simulations demonstrate that the concept of flexibility needs to be applied with caution to duplexes of RNA and DNA, since one duplex can be more flexible for some perturbations and more rigid for others. Clearly, a new, more complex view of nucleic acids dynamics seems necessary.

Methods

Molecular dynamics simulations

Two dodecamers, d(CGCGAATTCGCG)$_2$ and r(CGCGAAUUCGCG)$_2$, were built up using standard helical parameters[26] for *B*-DNA and *A*-RNA conformations. The structures were surrounded by around 4400 water molecules and 22 Na^+ placed in the regions of more electronegative potential. The hydrated systems were then optimized, heated and pre-equilibrated using our standard multistage protocol,[31,32] followed by 1 ns unrestrained MD simulation for equilibration. Finally, a 10 ns trajectory was collected for each structure. All MD simulations were performed in the isothermic-isobaric ensemble (1 atm, 298 K). Periodic boundary conditions and the Particle Mesh Ewald (PME[33]) technique were used to account for long-range effects. All bonds were constrained using SHAKE,[34] which allowed us to use 2 fs time step for integration of Newton's equations. Parm99[24,35] and TIP3P[36] force-fields were used to describe molecular interactions. Global translations were removed every 0.5 ns to remove erroneous partitions of the kinetic energy of the system. The analysis (see below) was performed using the last 9 ns of each trajectory and removing the base-pairs at both 5′ and 3′ ends. All the trajectories were obtained using the SANDER module of AMBER6.1 computer program.

Structural and energetic analysis

The helical parameters of the structures sampled along the 10 ns MD simulations were analyzed using the X3DNA program.[30] Additional geometrical parameters were derived with analysis modules in AMBER and with in house software. Classical molecular interaction potentials were computed using the CMIP program[37] with Na^+ as probe particle, and the electrostatic potentials were computed by solving numerically the Poisson-Boltzman equation.[38]. Water densities around DNA and RNA duplexes were determined by integrating water population around polar atoms in DNA/RNA (cutoff distance of 3.5 Å). Maximum water residence times were computed by tracing all water molecules around polar groups of the nucleic acids.

Essential dynamics

This powerful analysis was used to determine which type of motions represents the most important structural fluctuations in DNA and RNA.[25,37,38] Essential motions were determined from a principal component analysis. To this end, covariance matrices for common atoms of DNA and RNA (i.e., by excluding 5-methyl/H groups of T/U and 2′-OH/H groups in sugar moieties) were built and diagonalized. The resulting eigenvectors define the type of essential motions, and the associated eigenvalues determine how much of variance in the trajectory is explained by each eigenvector. Note that for two molecules of the same size the number of eigenvectors necessary to explain a given positional variance indicates the complexity of the molecular motions: the larger the number of essential motions, the greater the complexity.

The similarity between the essential motions of two molecules (defined by covariance matrices of the same size) can be computed using absolute (γ) and relative (κ) similarity indexes,[25,41,42] as shown in equations (1) and (2):

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

where ν_i^A stands for the unit eigenvector i of molecule A, n is the minimum number of essential motions that account for a given variance in the trajectory and the dot denotes a scalar product:

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

where the self-similarity indexes γ^{T}_{AA} are calculated by comparing eigenvectors obtained with the first and second parts of the same trajectory.

Entropy calculation

Since the intramolecular entropy is the best single parameter to describe the intrinsic order of a molecule, it provides a good estimate of the overall flexibility of DNA and RNA. The entropy of DNA and RNA was determined using Schiltter's[43] and Andreocci-Karplus[44] methods. The two models are based on the diagonalization of the mass-weighted covariance matrix, and on the use of a simple harmonic oscillator (hybrid classical-quantum (*S*) or purely quantum (*A–K*)) to link eigenvalues and entropy (see equations (3) and (4)). As noted before, intramolecular entropies were computed considering only atoms common to DNA and RNA:

$$S \approx 0.5k \sum_i \ln\left(1 + \frac{e^2}{\alpha^2}\right) \tag{3}$$

$$S = k \sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(1 - e^{-\alpha_i}) \tag{4}$$

where $\alpha_i = \hbar\omega_i/kT$, ω denotes the eigenvalues obtained by diagonalization of the mass-weighted covariance matrix, and the sum extends to all the non-trivial vibrations.

The entropy estimates obtained by equations (3) and (4) depend on the length of the trajectory (the larger the simulation, the greater the number of visited microstates), which generates uncertainties in the calculation. To accelerate convergence, we used the extrapolation method,[28] which allows us to estimate the entropy at infinite simulation times from the time-evolution of the entropy estimated along the trajectory.[25,28]

Stiffness analysis

The positional fluctuations of atoms along the trajectory were used to derive force-constants to describe the stiffness of DNA or RNA with respect to certain types of elastic deformations. This analysis provides then a unique tool for the description of the flexibility of nucleic acids.[25,29,45–47]

We first computed the stiffness associated with the deformation of DNA and RNA along their essential modes. This analysis provides quantitative information on the deformability of both helices along the easiest motional pathways. Taking advantage of the orthogonality of eigenvectors and assuming a harmonic behaviour, force constants are easily determined from the eigenvalues by using equation (5):

$$K_i = kT/\lambda_i \tag{5}$$

where λ_i is the eigenvalue (in Å^2) associated with the essential movement ν_i determined by diagonalization of the covariance matrix, *T* is the temperature (in Kelvin), *k* is the Boltzman constant and K_i is a force constant associated to the essential motion.

Stiffness analysis applied to essential motions provides useful 3-D information on the deformability of nucleic acids. Unfortunately, it is not always simple to interpret the nature of eigenvectors, thus making it desirable to compute stiffness in a helical coordinate system closer to chemical intuition.[45,47] Under the assumptions that the distribution of values adopted by a given variable X is fully Gaussian and that the deformation variables are orthogonal (independent), harmonic (elastic) force constants K_X can be easily derived from the variance of the variable X during the trajectory using equation (6):

$$K_X = kT \frac{1}{\langle (X - X_o)^2 \rangle} \tag{6}$$

where X_o is the equilibrium value of variable *X* and the brackets mean that fluctuations are averaged over thermodynamic ensemble.

An alternative way to compute force constants for a given deformable variable can be obtained by fitting the normalized histogram of values sampled for a given parameter along the trajectory, $\mathfrak{I}(X)$, to a Gaussian function (equation (7)). The exponent of the fitted Gaussian provides directly the force constant, and the goodness of the fitting (determined for example by the χ^2 test) gives an indication of the harmonicity of the deformation:

$$\mathfrak{I}(X) - \left[\frac{1}{\sqrt{2\pi}} \exp \frac{K_X}{2kT}(X - X_0)^2\right] - \text{error}_X = 0 \tag{7}$$

The non-orthogonality of the helical parameters can be taken specifically into account by means of coupling terms, which requires the inclusion of force constants, K_{ij}, to account for contributions to the deformation energy arising from coupling interactions. After simple matrix algebra, equation (7) can be generalized by defining the stiffness matrix *K* with entries K_{ij}, which can be easily derived by inverting the covariance matrix *C* with entries $C_{ij} = \langle (X_i - X_{i0})(X_j - X_{j0}) \rangle$ as shown in equation (8). Diagonal elements in *K* represent contributions to deformation arising from individual helical variables, while off-diagonal components account for coupling terms:

$$K = kTC^{-1} \tag{8}$$

Note that when the stiffness matrix is known, the deformation energy of a given configuration can be calculated by equation (9), where the subindex 0 stands for the equilibrium value:

$$E_{def} = \sum_i \frac{K_{ii}}{2}(X_i - X_{i0})^2 + \sum_{i \neq j} \frac{K_{ij}}{2}(X_i - X_{i0})(X_j - X_{j0}) \tag{9}$$

Acknowledgements

This work has been supported by the Spanish Ministry of Science and Technology (BIO2003-06848 and SAF2002-04282), the Fundación BBVA and the VIth Framework Program of the European Union (UE project QLG1-CT-1999-00008). The computer facilities of the CIRI-CEPBA computer centre are also acknowledged. F.L. thanks for support from the Swiss National Science Foundation.

References

1. Bloomfield, V. A., Tinoco, I., Hearst, J. E. & Crothers, D. M. (2000). *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, Calif.
2. Gait, M. J. & Blackburn, G. M. (1990). *Nucleic Acids in Chemistry and Biology*. IRL Press at Oxford University Press, Oxford.
3. Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer, New York.
4. Soliva, R., Luque, F. J., Alhambra, C. & Orozco, M. (1999). Role of sugar re-puckering in the transition of A and B forms of DNA in solution. A molecular dynamics study. *J. Biomol. Struct. Dynam.* **17**, 89–99.
5. Sancar, A. (1996). DNA excision repair. *Annu. Rev. Biochem.* **65**, 43–81.
6. Pradeepkumar, P. I., Amirkhanov, N. V. & Chattopadhyaya, J. (2003). Antisense oligonuclotides with oxetane-constrained cytidine enhance hetero-duplex stability, and elicit satisfactory RNase H response as well as showing improved resistance to both exo and endonucleases. *Org. Biomol. Chem.* **1**, 81–92.
7. Verbeure, B., Lescrinier, E., Wang, J. & Herdewijn, P. (2001). RNase H mediated cleavage of RNA by cyclohexene nucleic acid (CeNA). *Nucl. Acids Res.* **29**, 4941–4947.
8. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
9. Hammond, S. M., Caudy, A. A. & Hannon, G. J. (2001). Post-transcriptional gene silencing by double-stranded RNA. *Nature Rev. Genet.* **2**, 110–119.
10. Hagerman, P. J. (1988). Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265–286.
11. Hagerman, P. J. (1997). Flexibility of RNA. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 139–156.
12. Thomas, G. J., Jr, Benevides, J. M., Overman, S. A., Ueda, T., Ushizawa, K., Saitoh, M. & Tsuboi, M. (1995). Polarized Raman spectra of oriented fibers of A DNA and B DNA: anisotropic and isotropic local Raman tensors of base and backbone vibrations. *Biophys. J.* **68**, 1073–1088.
13. Fujiwara, T. & Shindo, H. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 2. Molecular motions in hydrated DNA. *Biochemistry*, **24**, 896–902.
14. Shindo, H., Fujiwara, T., Akutsu, H., Matsumoto, U. & Kyogoku, Y. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 1. Static geometry of DNA double helices. *Biochemistry*, **24**, 887–895.
15. Cheatham, T. E. & Kollman, P. A. (1997). Molecular dynamics simulations highlight the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* **119**, 4805–4825.
16. Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). "…the tyranny of the lattice…". *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.
17. Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J. & Dickerson, R. E. (1993). Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: conformational flexibility of B-DNA. *Biochemistry*, **32**, 1373–1389.
18. Portmann, S., Usman, N. & Egli, M. (1995). The crystal structure of r(CCCCGGGG) in two distinct lattices. *Biochemistry*, **34**, 7569–7575.
19. Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. (1989). The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature*, **342**, 456–460.
20. Perez, A., Noy, A., Lankas, F., Luque, F. J. & Orozco, M. (2004). The relative flexibiliy of DNA and RNA a database analysis. *Nucl. Acid Res.* In the press.
21. Auffinger, P. & Westhof, E. (2000). Water and ion binding around RNA and DNA (C,G) oligomers. *J. Mol. Biol.* **300**, 1113–1131.
22. Auffinger, P. & Westhof, E. (2001). Water and ion binding around r(UpA)12 and d(TpA)12 oligomers—comparison with RNA and DNA (CpG)12 duplexes. *J. Mol. Biol.* **305**, 1057–1072.
23. Pan, Y. & MacKerell, A. D., Jr (2003). Altered structural fluctuations in duplex RNA *versus* DNA: a conformational switch involving base pair opening. *Nucl. Acids Res.* **31**, 7131–7140.
24. Cheatham, T. E., Cieplak, P. & Kollman, P. A. (1999). A modified version of the Cornell *et al*. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dynam.* **16**, 845–862.
25. Orozco, M., Perez, A., Noy, A. & Luque, F. J. (2003). Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.* **32**, 350–364.
26. Arnott, S., Bond, P. J., Selsing, E. & Smith, P. J. (1976). Models of triple-stranded polynucleotides with optimised stereochemistry. *Nucl. Acids Res.* **3**, 2459–2470.
27. Packer, M. J. & Hunter, C. A. (1998). Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J. Mol. Biol.* **280**, 407–420.
28. Harris, S. A., Gavathiotis, E., Searle, M. S., Orozco, M. & Laughton, C. A. (2001). Cooperativity in drug-DNA recognition: a molecular dynamics study. *J. Am. Chem. Soc.* **123**, 12658–12663.
29. Lankas, F., Sponer, J., Hobza, P. & Langowski, J. (2000). Sequence-dependent elastic properties of DNA. *J. Mol. Biol.* **299**, 695–709.
30. Lu, X. J., Shakked, Z. & Olson, W. K. (2000). *A*-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840.
31. Shields, G. C., Laughton, C. A. & Orozco, M. (1997). Molecular dynamics simulations of the d(T center dot A center dot T) triple helix. *J. Am. Chem. Soc.* **119**, 7463–7469.
32. Shields, G. C., Laughton, C. A. & Orozco, M. (1998). Molecular dynamics simulation of a PNA center dot DNA center dot PNA triple helix in aqueous solution. *J. Am. Chem. Soc.* **120**, 5895–5904.
33. Darden, T., York, D. & Pedersen, L. (1993). Particle Mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
34. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical-integration of cartesian equations of motion of a system with constraints—molecular-dynamics of *N*-alkanes. *J. Comput. Phys.* **23**, 327–341.
35. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Ferguson, D. M. *et al*. (1995). A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
36. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
37. Gelpi, J. L., Kalko, S. G., Barril, X., Cirera, J., de La Cruz, X., Luque, F. J. & Orozco, M. (2001). Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. *Proteins: Struct. Funct. Genet.* **45**, 428–437.

38. Orozco, M. & Luque, F. J. (2000). Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.* **100**, 4187–4226.
39. Sherer, E. C., Harris, S. A., Soliva, R., Orozco, H. & Laughton, C. A. (1999). Molecular dynamics studies of DNA A-tract structure and flexibility. *J. Am. Chem. Soc.* **121**, 5981–5991.
40. Wlodek, S. T., Clark, T. W., Scott, L. R. & McCammon, J. A. (1997). Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. *J. Am. Chem. Soc.* **119**, 9513–9522.
41. Cubero, E., Abrescia, N. G. A., Subirana, J. A., Luque, F. J. & Orozco, M. (2003). Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. *J. Am. Chem. Soc.* **125**, 14603–14612.
42. Rueda, M., Kalko, S. G., Luque, F. J. & Orozco, M. (2003). The structure and dynamics of DNA in the gas phase. *J. Am. Chem. Soc.* **125**, 8007–8014.
43. Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance-matrix. *Chem. Phys. Letters*, **215**, 617–621.
44. Andricioaei, I. & Karplus, M. (2001). On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* **115**, 6289–6292.
45. Lankas, F., Sponer, J., Langowski, J. & Cheatham, T. E. (2003). DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* **85**, 2872–2883.
46. Matsumoto, A. & Olson, W. K. (2002). Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. *Biophys. J.* **83**, 22–41.
47. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl Acad. Sci. USA*, **95**, 11163–11168.

Edited by P. J. Hagerman

(Received 8 March 2004; received in revised form 30 June 2004; accepted 6 July 2004)