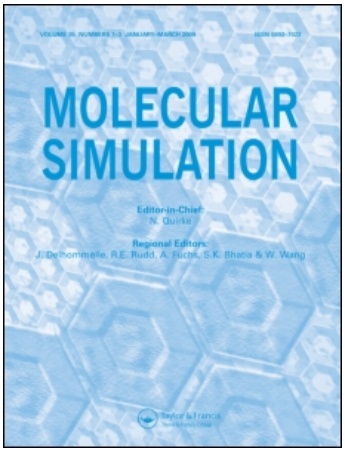

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Internal Structure and Dynamics of the Decamer D(ATGCAGTCAG) 2 In Li + -H 2 O Solution: A molecular Dynamics Simulation Study

Raimo A. Lohikoski[a]; Jussi Timonen[a]; Alexander P. Lyubartsev[b]; Aatto Laaksonen[b]
[a] Department of Physics, University of Jyväskylä, Jyväskylä, Finland [b] Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm, Sweden

Online publication date: 26 October 2010

To cite this Article Lohikoski, Raimo A. , Timonen, Jussi , Lyubartsev, Alexander P. and Laaksonen, Aatto(2003) 'Internal Structure and Dynamics of the Decamer D(ATGCAGTCAG) 2 In Li + -H 2 O Solution: A molecular Dynamics Simulation Study', Molecular Simulation, 29: 1, 47 — 62
To link to this Article: DOI: 10.1080/0892702031000065728
URL: http://dx.doi.org/10.1080/0892702031000065728

Molecular Simulation, 2003 Vol. 29 (1), pp. 47–62

Internal Structure and Dynamics of the Decamer D(ATGCAGTCAG)$_2$ in Li^+–H_2O Solution: A Molecular Dynamics Simulation Study

RAIMO A. LOHIKOSKI[a,*], JUSSI TIMONEN[a], ALEXANDER P. LYUBARTSEV[b] and AATTO LAAKSONEN[b]

[a]*Department of Physics, University of Jyväskylä, P.O. Box 35, FIN-40351, Jyväskylä, Finland;* [b]*Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, SE-10691, Stockholm, Sweden*

(Received March 2001; In final form April 2001)

Molecular dynamics simulation of the decamer d(ATGCAGTCAG)$_2$ in aqueous solution, electroneutralized by Li^+ ions has been carried out. Emphasis is on the verification of the equilibrium conditions and the related structural and dynamical properties. Applicability of the kinetic part of Boltzmann's *H* function as a measure of thermodynamic equilibrium is tested. Overall structural stability has been confirmed by different RMSDs. Conformational and helicoidal parameters have been analyzed statistically and dynamically. Dynamical analysis reveals the existence of dynamical sub-states, which typically appear as abrupt changes from a mean level to another in the value of parameter. In statistical analysis these sub-states cannot often be detected. Correlation analysis shows concerted motions among many nearby structural parameters.

Keywords: DNA; Molecular dynamics; Dynamical stability; Boltzmann's *H* function; Structural transitions; Equal time correlations

1 INTRODUCTION

In addition to the three-dimensional structure of large biomolecules, such as proteins and nucleic acids, the relationship between their internal structure and dynamics, and the functional properties of these molecules is also a fundamental part of information in molecular biophysics. It is important to understand, e.g. which kind of molecular motions and excitations play a significant role in the behavior of these molecules in physiologically relevant solution conditions. This becomes particularly crucial in problems like drug binding [1–3] and opening [4–6] of the double-stranded helical structure of DNA. Unfortunately, no experimental techniques are yet available to reveal the whole dynamical structure of these molecules in a solution environment. Therefore, molecular dynamics (MD) simulations have become a valuable tool in this and many related problems [7]. They can also be used to increase the information that can be extracted from crystallography and NMR spectroscopy [8].

MD simulations have been used over the last 15 years to determine the structure of various DNA oligonucleotides (see the Introduction of Ref. [20] and references therein). In the beginning, the force fields used were rather crude and the simulation times covered only a few tens of picoseconds [9]. There has been much progress over the last years in the refinement of the force fields [10–15] and the simulation methods (see e.g. Ref. [16] and references therein) as well as in the availability of more powerful computing resources. Simulations in the nanosecond time scale are now possible with increasingly realistic dynamic MD models of DNA. Different sequences such as, e.g. d(CCAACGTTGG)$_2$ and d(CGCGAATTCGCG)$_2$ have been considered as prototype DNAs with different force fields, and their conformational forms and solvent effects have been analyzed. Very recent advances in the field of computational DNA dynamics are reviewed in Refs. [17,18].

Recently, yet another decamer, d(ATGCAGTCAG)$_2$, has been considered for its hydration structure and counter-ion coordination [19–21]. In the present study we concentrate on analyzing the structural properties of this decamer and its stability in a Li^+–H_2O solution. In this analysis we use

*Corresponding author. E-mail: raimo.lohikoski@phys.jyu.fi

ISSN 0892-7022 print/ISSN 1029-0435 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0892702031000065728

the all-atom force field CHARMM [11]. The computational details are given in the following section. One of the goals here is to compare different methods that have been used for stability analysis. We also introduce a not very often used method for thermodynamical stability based on Boltzmann's H function. In this way one can also analyze in some detail the structural features that have been used to identify different conformational forms of DNA.

A closely related area of interest is the dynamics of fluctuations at equilibrium of the decamer considered. These fluctuations may be of importance when, e.g. the conformational form of a DNA is determined in some experimental situation. It will become evident in this work that in some of the structural and helicoidal parameters there indeed appear rather significant dynamic changes in the nanosecond time scale. Also, possible cross correlations of these dynamic fluctuations are analyzed.

2 DNA MODEL AND COMPUTATIONAL METHODS

A full periodic turn of a double-helix DNA (10 base pairs), surrounded by water and counter-ions was simulated in a rectangular box with periodic boundary conditions imposed in all directions. Full details of the simulated system are given in our previous work where the same DNA system was analyzed for its hydration structure and the counter-ion coordination based on spatial distribution functions [20]. The numbering scheme of the sequence, to which we refer later, is given in Fig. 1. The emphasis of the present work is on the internal structure, equilibrium conditions and dynamical fluctuations of the DNA itself.

The all-atom force field CHARMM [11] was used for the DNA. The Li^+ ions were treated within the force-field scheme as charged Lennard–Jones spheres with parameters defined by Dang [22]: $\sigma = 1.506$ Å and $\epsilon = 0.69$ kJ/M. The number of Li^+ counter-ions was 20, to fully neutralize the phosphate charge ($-20|e|$ for each turn of DNA). For water we used the flexible SPC water model by Toukan and Rahman [23]. All cross-interaction parameters were determined by using the Lorentz–Berthelot combination rules [24]. A constant temperature–constant pressure MD algorithm was employed [25] at the temperature of 298 K and at the pressure of 1 atm such that the long-range electrostatic interactions were taken into account by means of the Ewald method [24]. The number of water molecules was 500 and the average box size at the simulated temperature and pressure was $24.5 \times 24.5 \times 33.8$ Å^3.

As the initial configuration, the helix axis of DNA coincided with the z-axis of the simulation cell, and the DNA coordinates were chosen from the canonical B-form. The water molecules were arranged initially in a cubic lattice with random orientations and equilibrated as liquid water. Thereafter all water molecules in a cylinder of radius 11 Å, co-axial to the DNA helix axis, were displaced from the simulation cell. The ions were randomly distributed among the water molecules. During the first 100 ps of the simulation, the DNA was kept rigid, allowing for equilibration of the water molecules and the Li^+-ions. A constant-temperature algorithm with constant volume (NVT ensemble) was used at this stage. After 100 ps, all atoms of the DNA were allowed to move. Also, the constant temperature–constant pressure algorithm [25] was switched on. The system was further equilibrated during the next 100 ps, whereafter the production runs were started. The Nóse–Hoover thermostat parameters were set to 30 fs for the thermal bath and 700 ps for the pressure bath (volume fluctuations).

Due to the periodic boundary conditions together with the Ewald summation of the electrostatic interactions, the simulated system represented an infinite array of DNA molecules, packed in parallel. The average distance between the helix axes was equal to the box length in the x and y directions. This corresponds to a structure of packed cellular DNA. This kind of structure is experimentally relevant, as ordered DNA samples (fibers) are often used in the experimental studies. Also, the average distance between the DNA helices of about 24.5 Å corresponds to a relative humidity of fibers of about 95% [26]. At these conditions, DNA is normally found in the B-form. It is known, that at somewhat lower humidity, the Li-DNA is transformed into a C-form [27]. Thus our simulation parameters correspond to a "pre-transitional" region from the B-form to the C-form.

	1	2	3	4	5	6	7	8	9	10
Strand 1:	A	T	G	C	A	G	T	C	A	G
Strand 2:	T	A	C	G	T	C	A	G	T	C
	20	19	18	17	16	15	14	13	12	11

FIGURE 1 Numbering scheme of the decamer.

3 STABILITY OF THE SIMULATED DNA SYSTEM

3.1 Thermodynamical Stability

The basic assumption in MD simulations, as to thermodynamic equilibrium, is the equipartition principle, namely that the average kinetic energy of each degree of freedom of a molecule should be equal to $(1/2)k_BT$ [28,29]. In our simulations, the temperature of the system was computed based on the equipartition principle from all the atoms in the system. It was found to be stable and fluctuate slightly around the pre-defined value of 300 K.

There is also another method to estimate the thermodynamical stability of the simulated system, although it has not been used very often. In this method, the statistical evolution of the entire simulated system from its initial state (assumed far from equilibrium) towards thermodynamical equilibrium, is followed by computing the kinetic part of Boltzmann's H function [28] (which we denote by H). In the literature it is traditionally considered as a tool for deciding when an equilibrium is reached. Here we consider its usefulness to monitor the thermodynamic stability of an atomistic simulation system in general, and the modeled DNA in particular, during the production phase of the MD simulations. (A more complete testing will be published elsewhere [30]).

The H function is often defined [28,31] as

$$H(t) = \int_{-\infty}^{\infty} f(\vec{v})\ln(f(\vec{v}))\,\mathrm{d}^3v, \tag{1}$$

where t is time, $\vec{v}$ is an instantaneous atomic velocity and $f(\vec{v})$ is the three-dimensional (3D) probability distribution function of these velocities. A shortcoming of the definition Eq. (1) is that it is not dimensionless. Therefore we have chosen to modify the definition of H so that

$$H(t) = \int_{-\infty}^{\infty} f(\vec{v})\ln(v_0^3 f(\vec{v}))\,\mathrm{d}^3v, \tag{2}$$

where v_0 is simply a constant to ensure that $v_0^3 f$ in Eq. (2) is dimensionless (we discuss below how v_0 is determined). Introduction of this constant corresponds to adding a constant contribution to the H function. When the thermodynamic equilibrium is reached, the atomic velocities should satisfy the Maxwell distribution. This distribution gives a value for H which should be obtained near the end of the equilibration phase of the simulation. If we assume a Gaussian form (such as the Maxwell distribution) for the three-dimensional velocity distribution, H can be expressed as a sum of three individual components,

$$H = \sum_{j=x,y,z} H_j, \tag{3}$$

where

$$H_j = \int_{-\infty}^{\infty} f(v_j)\ln(v_0 f(v_j))\,\mathrm{d}v_j. \tag{4}$$

We now choose v_0 such that $H = 0$ for the Maxwell distribution, i.e.

$$v_0\left(\frac{2\pi ekT}{m}\right)^{1/2}. \tag{5}$$

In discretized form the H_j can be expressed in the form

$$H_j = \frac{1}{2} + \sum_i \frac{N_i}{N}\ln\left(\frac{N_i}{N_0}\right), \tag{6}$$

where $N = \sum_i N_i$, with N_i the number of counts or the frequency in the velocity interval i (of width Δv) of the discrete velocity distribution, and N_0 is the number of atoms with zero velocity.

We first consider the effect of the number of velocity classes (the size of Δv in the discretized distribution) on H. For a Gaussian distribution (in one dimension)

$$f(v) = \frac{1}{\sigma(2\pi)^{1/2}}\exp\left[\frac{-(v-\mu)^2}{2\delta^2}\right], \tag{7}$$

ranging from -1500 to 1500 (v in arbitrary units), with zero mean and a standard deviation of 125, we find that $H = 0$ until Δv is so large that there are only three values of N_i that have clearly nonzero values. In this case $\Delta v = 250$ and $H = 0.07$. For the atom velocity distributions emerging from a MD simulation with Gaussian form, the behavior of H as a function of Δv is more complicated. When Δv is small, H is clearly negative. When Δv is increased, H eventually becomes positive, and approaches zero only when the fluctuations in the distribution die out. Finally, H begins to increase again when there are only three nonzero classes in the distribution.

We also tested the effect of a velocity cut-off (i.e. cutting off the tails of the distribution) on the value of H. Experimental distributions are necessarily finite and velocities that have a small probability do not appear. Using the Gaussian distribution defined above, Eq. (7), and measuring the size of the cut-off through the remaining area (RA) under the finite distribution curve, we find that H differs from zero already for RA = 98%, and $H = 0.20$ for RA = 90%. This shows that a truncated Gaussian form can be sensitively recognized from the value of H.

We can conclude that H is close to zero for a Gaussian distribution if the discrete distribution is qualitatively good, i.e. there are no strong fluctuations in the distribution curve and the cut-off velocity is high enough. From an experimental point of view (either real or computer experiments) the statistics has to be good, i.e. the number of events in

the distribution must be large enough. This indicates that H is not suitable for a continuous monitoring of simulations because the amount of atoms is usually too low for good statistics. On the other hand, monitoring is sensible if it is determined over a long enough time interval. We have computed H for all atom types of our model DNA using non-overlapping time intervals of length 10 ps. Considered as functions of these discrete time units, the H functions steadily fluctuate around a mean value near $H = 0$. This demonstrates the dynamical stability of the simulations.

3.2 Structural Stability

A few years ago, when enough computer power became available to carry out long (≈ ns and more) MD simulations even for large molecules, it was found that the conformational dynamics of large (bio)molecules were not stable in long simulations [32]. This gave one reason for the force-field developers to refine codes, and several major force-fields were subsequently revised, including CHARMM [11,13,33] and AMBER [10,34]. Together with other significant developments in the methodology (e.g. the introduction of the particle-mesh Ewald method [16,35,36]), it now seems that the simulated DNA structures stay near the experimentally observed structures during even very long simulations (several nanoseconds) [17,18,37]. In the MD simulations of macromolecules, the overall structural stability is in the first place measured as root mean square deviations (RMSD) of the constituent atoms,

$$\mathrm{RMSD} = \left(\sum_i^N w_i d_i^2 / \sum_j^N w_j \right)^{1/2}, \qquad (8)$$

where d_i is the deviation of atom i from its reference position, and the w_i are suitable weighting factors. This is, however, by no means a unique way. One can take into account all atoms of the molecule [38] or only the heavy atoms [39–41]. Also, the atomic masses are sometimes used as the weights [38]. If the structural stability is the only purpose, the RMSDs can be computed with respect to the initial structure (of the thermodynamical equilibrium part of the simulation) [38]. In some cases, the RMSDs are computed with respect to the average structure of the system [42] in order to reveal the largest fluctuations in a convenient way. The average structure is not necessarily a characteristic structure of the system, especially if fluctuations are strong. This again may lead to incorrect conclusions concerning the structural evolution. McConnell *et al.* [43] use two-dimensional RMSDs for more detailed information. In the case of DNA, RMSDs can also be used to measure the structural family of the simulated DNA by computing it with respect to the corresponding canonical structures A-DNA, B-DNA and Z-DNA [37,38,40].

In Fig. 2, we have used RMSDs to obtain a stability profile of the simulated DNA (more detailed local structural properties are studied later using other means). Figure 2 shows the RMSDs of all atoms of the DNA as a function of time. Two reference structures are shown together with the RMSD along various coordinate directions. All curves indicate that, after a certain rise time, the RMSDs reach a mean constant value around which they fluctuate in the course of time.

RMSDs with respect to various reference structures show differences. In Fig. 2a, when the RMSD is computed with respect to the initial structure of the production part, it shows a clear tendency to increase with a rise time of about 500 ps before its mean constant level, 2.9 Å, is obtained. When the RMSD is computed with respect to the final structure of the simulation, Fig. 2b, the corresponding "decay time" is roughly 1800 ps and the structural separation does not seem to be continuous. We can conclude that in case (a) the overall structure is still developing towards a structural equilibrium, while in case (b) the DNA structure fluctuates around some mean structure. RMSDs in various coordinate directions are quantitatively very similar, which also indicates structural stability. However, much longer simulations would be needed to completely ensure the stability of the system.

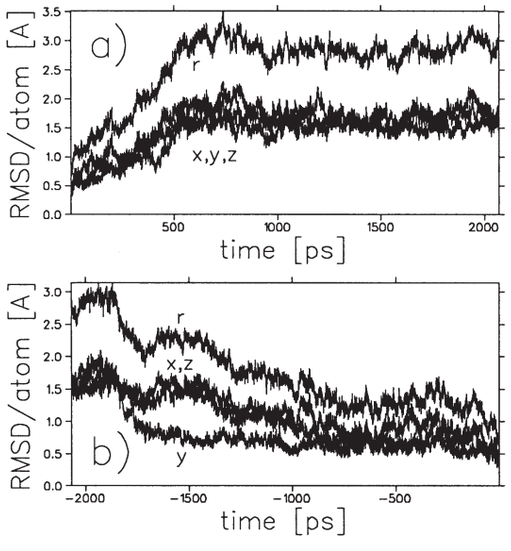

FIGURE 2 Unweighted atomic RMSDs as a function of time. In (a) comparison is made with respect to the initial and in (b) to the final structure of the simulation. Symbols x, y, and z denote RMSDs in the cartesian coordinate directions, and r is the total RMSD.

4 STRUCTURAL ANALYSIS

4.1 Conformation

In Fig. 4 we show the distributions of the conformational parameters by following each local parameter separately (for definition of the parameters see Fig. 3 and Appendix). In this way we can show the differences observed in the parameters measured at different sites. In Fig. 4b we show the total distributions. These correspond to distributions that are possible to observe experimentally. In order to make it convenient to compare the simulated values with experimental data, we have included in the Fig. 4b some values based on experiments on A-DNA and B-DNA sequences. The reference value are taken from Saenger [44]. Notice that the angle values given by the CURVES program [45] are between -180 and 180° (except for P: $0 \leq P \leq 360°$) whereas in Fig. 4 these values are between 0 and 360° ($-90° \leq P \leq 270°$). We have done these shifts only to get continuous distributions, which makes their inspection easier.

Two of the backbone torsional angles, δ and χ, have well separated experimental values for the A- and B-forms of DNA. The simulated δ distribution shows that roughly half of the δ values belong to a range typical of the A-form and the other half are close to the B-form values. In the simulated χ distribution most of the χ angles are near the experimental range for the A-form. A noticeable part of the distribution is however between the experimental A and B values, and a minor part close to the B values. This distribution information thus leads to the conclusion that the DNA of our simulations is initially in the A-form but after a while transforms into the B-form (or *vice versa*), or it is a mixture of coexisting structural entities of both types. Inspection of the dynamical trajectories of these parameters reveals that the latter possibility seems more probable in the present simulations.

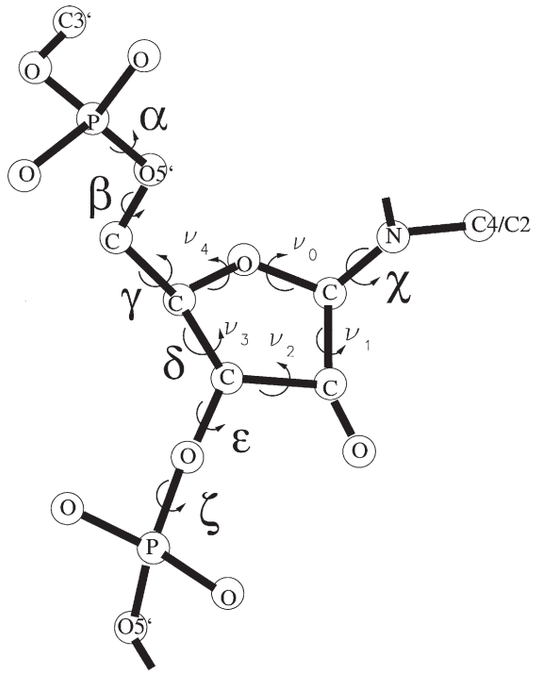

FIGURE 3 Definition of the conformational parameters.

Another striking feature in Fig. 4 is that our simulations produce α and γ angles far from the observed values. For the α angle the simulated values range from 60 to 180°, and for γ from 150 to 210°. These angle values are sterically possible (see Fig. 4 of the paper by Young *et al.* [38]). Possibly the experimental conditions used have restricted some of the observables. On the other hand, conformational angles, neither in our simulation, nor in the 5 ns simulation by Young *et al.* [38], cover all the sterically allowed ranges of angle values. In the steady state or in a "sleepy" phase, these parameters probably occupy only some subspace of the allowed values, and other parts may be visited during an "active" phase, e.g. when interacting with proteins, mRNA or other molecules.

The last boxes of Fig. 4a,b contain the distributions of the pseudorotation phase angle P. In the canonical A-form structure P is in the range from 0 to 36° (C3′-endo) and in the canonical B-form between 144 and 190° (C2′-endo). The plot of P in Fig. 4b shows two preferred ranges of P values, from 100 to 200° and from 320 to 40°. In 11 cases out of the twenty the maximum of the P distribution occurs in the C2′-endo region. In six cases they are in the C3′-endo region, and three maxima are in the C2′-exo region. As the plots in Fig. 4a indicate, in living systems the widths of the P distributions may be quite broad. Evolution trajectories show that each P changes its value rapidly between the two main regions at least once during the simulation.

4.2 Helicoidal Properties

The distributions of the axis-base pair, intra-base pair and inter-base pair parameters are shown in Fig. 5 (definitions and grouping of these parameters are given in Appendix). These are formed by summing up the individual distributions. Table I quotes the mean values with standard deviations of these distributions along with some reference data. Columns of the canonical and crystal structure data of B-DNA and A-DNA together with the Y-MD data, are all taken from recent simulation results of Young *et al.* [38]. The B-DNA crystal structure data are based on the 92 B-form oligonucleotide crystal structures, and the A-DNA data is composed of 42 A-form structures. The Y-MD results are from a 5 ns

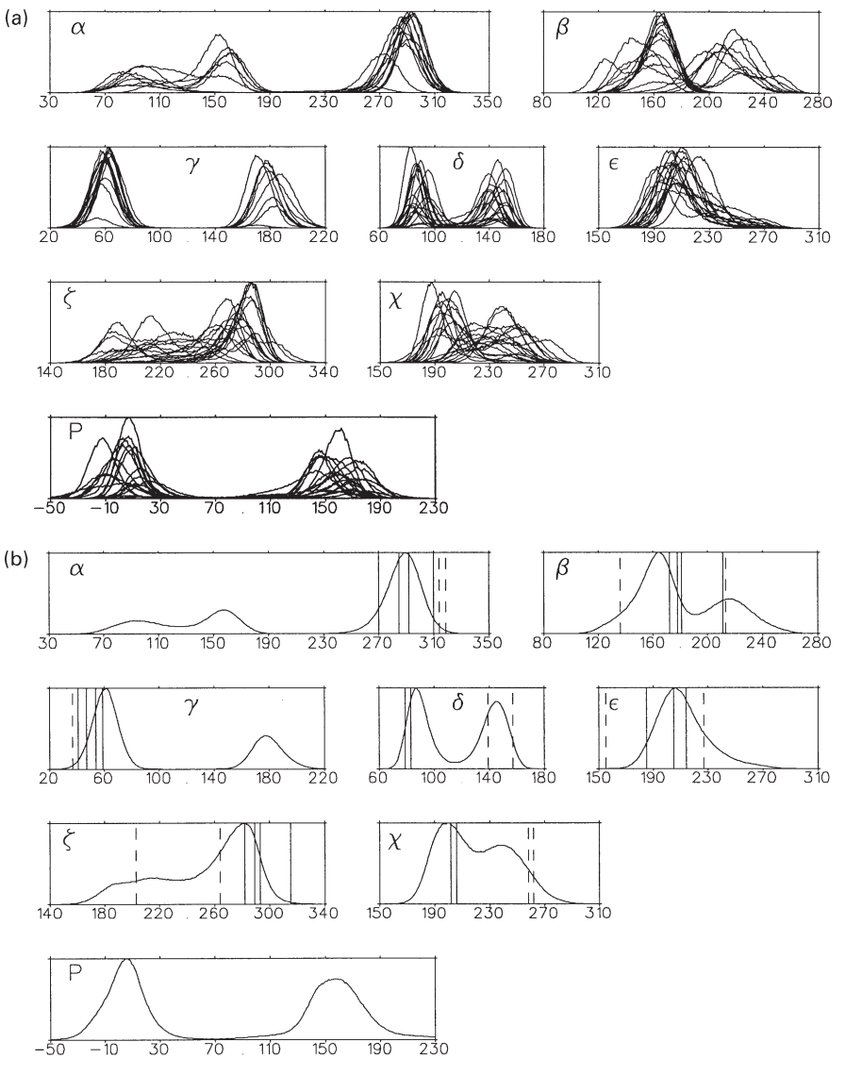

FIGURE 4 Distributions of the conformational parameters. Plots in (a) show all individual distributions of the simulation, and plots in (b) show their sum distributions. Experimental reference values, shown by bars, are taken from Saenger [44]. Solid bars correspond to A-DNA and dashed bars to B-DNA.

MCI/PME simulation of the sequence d(CGCGAA-TTCGCG)$_2$. The values of the canonical parameters XDP, INC, RIS and TWS can be used to identify the structural family with respect to the sequence. Table I shows that the XDP of our simulation is slightly more shifted towards the XDP of the A-form than the XDP of Y-MD. One reason for this can be seen by inspecting the individual distributions. One of the XDP distributions has a second peak at -3.1 Å. The overall mean, -1.4 Å, is however far from the canonical value for A-DNA, -5.3 Å. Compared with the crystal structure, the XDP of this work is between the A- and B-form values, slightly closer those of the B-form. The mean INC is near the observed value for B-DNA. It is slightly smaller but by not nearly so much as the INC of Y-MD. The mean RIS of our simulation, likewise that of the simulation of Young *et al.* [38], is the same as the canonical and crystal-structure value in the B-form of DNA. Our mean TWS is also very near the canonical TWS value of B-DNA, whereas the TWS of Y-MD differs from that by nearly 5°, having roughly the mean value of the crystalline A-form. Notice, however, that in our simulation the system is periodic (also in the helix direction), which probably forces the TWS angle to be near 36°. In Y-MD the helix was not periodic, and it had the freedom to settle in any conformation allowed by the force field.

If the crystal structure data are also considered as valid reference data with respect to the structural identity of the simulated DNA, then TIP, PRP, TLT and ROL may also be used in the comparison. The mean TIP of our simulation was positive while the A- and B-form both have negative values, the TIP of the A-form being closer to our result. Also the mean PRP and TLT of our simulation were closer to

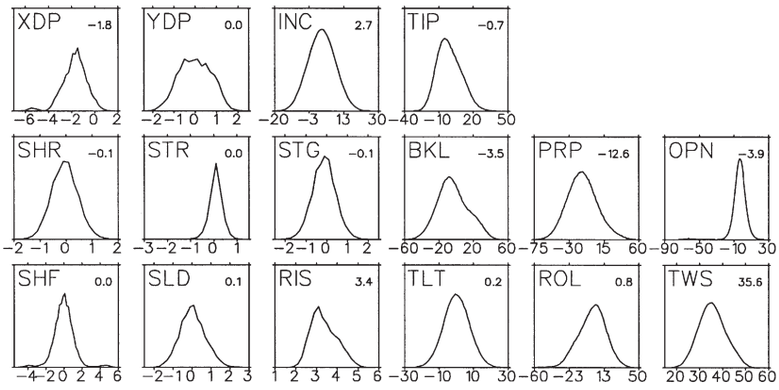

FIGURE 5 Distributions of the helicoidal parameters.

the observed A-form values. Only ROL had a mean value, which is identical to that of the B-form. On the other hand, although the DNA structure in Y-MD is interpreted to represent B-DNA, the parameters of the Y-MD simulation do not show "better" agreement.

The forms of the distributions can also be used to distinguish the structural features. According to the crystal-structure data collected by Young *et al.* [38], certain structural parameters seem to have different distribution profiles in A-DNA and in B-DNA. If these profiles are assumed to appear also in the time evolutions of the various forms of DNA, then comparison of the distribution profiles of Figs. 4 and 5 of Young *et al.* [38] will shed some light into the nature of the DNA structure of our simulation. According to the crystal-structure data the distributions of XDP, STR and SLD are sharper in B-DNA than in A-DNA. Similarly the distributions of TIP, SHR, TLT and ROL are sharper in A-DNA than in B-DNA. Because Young *et al.* [38] have not given the widths of these distributions, we can only do a visual comparison of the two data. Our TIP seems rather broad and STR quite narrow indicating B-form distributions. A similar indication can be drawn from the broadness of the TLT and ROL distributions. The other distributions do not seem to provide conclusive evidence either way.

5 DYNAMICS

5.1 Conformational Parameters

The dynamic behavior of parameter α generally indicated stability of the structure. Ten of the 20 torsion angles oscillated around their mean values near 290°. These mean values seemed to have temporal variations resembling restricted Brownian motion. At some sites the αs however showed distinct jumps from one average level to another. These were CYT 4, where α rapidly changed from 100 to 250° at 800 ps, GUA 13 with a jump from 260 to 100° at 250 ps, and CYT 18 with a jump from 260 to 110° at 1450 ps. After these jumps a seemingly stable

TABLE I Helicoidal parameters. First column shows the mean values of this work. Subsequent columns are reference values taken from Ref. [38]. Superscript *can* refers to the canonical structure, superscript *cry* refers to the crystal structures and Y-MD is the results of the MCI/PME simulation of the sequence d(CGCGAATTCGCG)$_2$

	Mean (std)	*B-DNA* ^{can}	*A-DNA* ^{can}	*B-DNA* ^{cry}	*A-DNA* ^{cry}	*Y-MD*
XDP	−1.4 (0.5)	0.0	−5.3	−0.2	−3.8	−0.9
YDP	−0.2 (0.4)	0.0	0.0	0.1	−0.1	0.1
INC	0.4 (6)	1.5	20.7	0.7	4.9	−2.2
TIP	0.7 (6)	0.0	0.0	−2.6	−0.3	−1.4
SHR	0.0 (0.4)	0.0	0.0	0.0	0.0	0.0
STR	0.0 (0.2)	0.0	0.0	0.0	0.2	0.2
STG	−0.1 (0.5)	0.0	0.0	0.0	0.0	0.0
BKL	−3.8 (18)	0.0	0.0	0.7	−0.1	0.3
PRP	−11.8 (13)	−7.5	−7.5	−13.9	−12.5	−12.5
OPN	−3.5 (10)	0.0	0.0	2.6	2.8	1.9
SHF	0.0 (1.0)	0.0	0.0	0.0	0.0	0.1
SLD	0.0 (0.4)	0.0	0.0	0.0	−0.1	−0.2
RIS	3.4 (0.7)	3.4	2.6	3.4	3.1	3.4
TLT	1.2 (13)	0.0	0.0	−1.2	0.2	−0.2
ROL	−0.5 (12)	0.0	0.0	−0.4	3.6	5.4
TWS	35.7 (9)	36.0	32.7	33.8	31.6	31.1

evolution continued. There were also sites, which showed continuous switching between two values. This kind of behavior was seen in GUA 3 whose extremum values were 50 and 180°, in ADE 9 where variation occurred between 80 and 180°, and in ADE 19 where the torsion angle varied between 80 and 180° during the first nanosecond of the simulation, and then stayed near 150° during the next nanosecond. The other sites showed dynamics where quite stable (dynamically stable) trajectories were interrupted by quick jumps into clearly distinct trajectories around different mean values, where they stayed for a while until the next similar jumps occurred.

The time evolution of parameter β was qualitatively similar to that of α. None of the individual trajectories had a well-defined mean value but all of them seemed to continuously drift over a large range of values. Only at few sites and at few times during the simulation did βs show rapid jumps away from their previous values (jumps by 50° on the average).

Parameter γ was the most stable of the conformational parameters. Its steady evolution was broken only by some very rapid (jump-like) changes. These jumps were observed at ADE 1 as a shift from 60 to 180° at 2 ns, at ADE 5 from 180 to 60° at 800 ps, at GUA 13 from 60 to 180° at 25 ps, at ADE 14 from 60 to 180° at 250 ps and at ADE 19 from 60 to 180° at 1450 ps. Four of the five γs linked to the ADE groups showed noticeable change during the simulation.

Parameter δ was bistable. Its distributions had two well-separated peaks centered near 85 and 145°. All individual dynamical trajectories included abrupt jumps between these two values. Lifetimes at either level could be only few picoseconds though transitions did not occur very often, five times on the average in each nucleotide unit during the 2.07 ns simulation time.

Visually the basic features of parameter ϵ was a steady trajectory broken by many random jumps to slightly larger values and back. The "upper levels" were so short lived that no secondary distribution peaks were formed except in few single cases.

Typical of parameter ζ were relatively long stable periods of evolution (of the order of nanosecond) broken by rapid rather big jumps, with a stochastic evolution within the jump periods. As the distribution curves show, the range of ζ values was quite large. No correlation between the ζ values and a nucleotide type could be found.

Parameter χ was quite stable. Again no unique mean value for any single χ angle could be found, and the time evolution was similar to that of parameter β apart from the abrupt big changes, which did not occur for χ.

The plot of distributions in Fig. 4 shows two main ranges for the values of pseudorotation phase angle *P*. The evolution trajectories reveal that every single sugar have *P*s in both of these ranges. The behavior of *P* seems dynamically stable but however includes sudden jumps to the other main value. These kinds of transition occur a few times in each sugar during the whole simulation. There also appeared quick transitions of a duration of a few picoseconds back and forth between the two main values.

Combined transitions in nearest-neighbor parameters could also be seen. The amount of coupled movements depended on the site and parameters. Strong coupling was found among the $\epsilon - \zeta$ pairs, for which many sites showed coupled behavior during the simulation, and also among the $\alpha - \beta$ pairs.

5.2 Helicoidal Parameters

The distributions of the axis-base pair parameters shown in Fig. 4 result from trajectories evolving in a random-walk like fashion, i.e. parts of the distributions are caused by noise but other parts arise from drifting over large ranges of values. These trajectories reveal some structural deformations. The XDP in base pair THY 7-ADE 14 shifted from -3 Å to -1 Å after 400 ps of the simulation time. At the same time the same base pair showed a long-lived (≈ 0.5 ns) shift in INC, from 10 to -10°. Other base pairs also showed changes of similar size, but they did not last as long.

SHR, STR and STG of the intra-base pair parameters seemed to oscillate around their global mean values while BKL, PRP and OPN showed similar behavior as the axis-base pair parameters.

Some of the inter-base pair parameter distributions were skew. This was due to drifting of the local mean value. This drifting occurred so slowly that, in the time scale of this simulation, distributions did not have enough statistics, i.e. complete ergodicity was not realized. The overall distribution of SHF had wings. These are due to some structural change in base pair THY 7-ADE 14. The SHF of GUA 6/THY 7 began at -4 Å, was 1 Å after 0.5 ns, and drifted to near 0 Å towards the end of the simulation. In THY 7/CYT 8, SHF began at 5 Å, was 0 Å after 0.5 ns, and stayed there during the rest of the simulation. Otherwise the dynamical trajectories did not differ much from a random walk like behavior.

5.3 Correlation Analysis

In order to find some quantitative information of the dynamics of the helicoidal and conformational parameters of the DNA we carried out a partial correlation analysis. The full dynamical correlation matrix was computed between all the helicoidal and conformational structure parameters used in this work. Technically this analysis was based on 1 ns long time series. The first variable of the pair always

had a fixed time origin at the beginning of the production run of the simulation. Dynamical correlation functions were obtained using a time delay of 200 fs between the consecutive zero points of the second variable. Our interest was in the concerted motions in DNA. We therefore used cross-correlation data to reveal statistically relevant co-operative motions. First between all the structure parameters and second between some parameters which have experimentally proved to be correlated.

5.3.1 Coincident Motions

For all cross-correlation functions we first determined the correlation minima and maxima and the related time delays. From these data we selected the local parameter pairs whose correlation minima or maxima occurred at zero delay time. Notice that each (global) parameter pair appears in many places locally in the decamer sequence analyzed. From the resulting set of parameter pairs we then chose the pairs for which at least a half of the local pairs were present in the zero delay-time set. This was used as the criterion for statistical relevance. For the final consideration we then selected from the set of statistically relevant pairs those, which showed features of interest. The correlation information for this final set of pairs is shown in Tables II–V. (The full correlation data is available on request). These tables give the selected cross-correlation pairs and the mean values for the minimal or maximal correlation coefficients, or both if anti-correlated and correlated cases both appeared. Numbers in the parenthesis are the numbers of local pairs that appear, for each (global) parameter pair, and were thus used to determine the mean value of the correlation coefficient quoted. The purpose of the selection routine was to find those parameter pairs, which typically show non-delayed coupled motions in the helicoidal double-stranded nucleotide sequence. The mean value of the correlation coefficient indicates the strength of the coupling, although it does not provide a quantitative measure. When two trajectories seem to follow each other quite closely, its absolute value is around 0.6. The sign of the correlation coefficient shows the relative phase of the two correlated trajectories. However, in the following we are not interested in the nature of correlation, although Tables II–VI also give this information, and use the term correlation independent of whether we refer to correlation or anticorrelation.

Table II shows that the base–sugar torsion angle χ has a clear correlation with the INC angle of the base pair. Obviously this causes correlation also between χ and BKL. One must also note the χ-STG and χ-TLT correlations. First of these two couplings shows that staggering of the bases of a base pair is caused at least partly by rotation of the base plane with respect to the plane of the attached sugar ring, and not only by a shift of the base in the helix direction. The last coupling indicates that the tilt between base pairs mainly originates from a rotation of the bases and not from a rotation of the sugar–base complex. Of the correlations between the backbone torsion angle χ and the other backbone related parameters, the most striking is the $\chi - \delta$ correlation. This coupling indicates a rotation of the furanose ring, which causes changes in torsion angles χ and δ residing on the opposite sides of the ring. Correlation between the pseudorotation phase angle P and χ can be understood through the $\chi - \delta$ correlation [44]. The backbone torsion angle γ is generally nearly independent of the other parameters. Only the χ and ζ angles in the second strand of the double helix show some degree of coupling with it. These correlations refer to conformational changes of the sugar group. The torsion angle δ is clearly correlated with the PRP of the base pair and the backbone torsion angle χ in both strands of the helix, more strongly with the latter. This correlation can be understood as a pendulum motion of the glycosyl group between the backbone and the base. The δ-PRP correlation is indicative of the motion of the whole base–sugar unit during the propeller motion. The $\delta - \zeta$ correlation is most obviously due to their common atom, O_3^*. Similarly, common atoms cause observed correlations both between δ and ϵ and between δ and β. Coupled motion of δ and P can be seen in the definition of P [44]. Notice that contribution to the mean correlation coefficient between δ and P comes from six out of the 10 possible sites. Of the correlations ϵ has with the other parameters only that with α is significant. These torsion angles do not have any common atom. Moreover, their coupled motion is observed only in one of the two strands. The torsion angle ζ shows correlated motion with many structural parameters. In strand one it is coupled with many helical parameters. This is noteworthy because ζ is spatially well separated from these. Based on the other correlations it has with the backbone conformational parameters, one can infer that the dynamics of the backbone is quite concerted without long delays. Only γ does not show correlation with ζ within our time window of 200 fs. β behaves similarly to ζ, and does not correlate with γ even though this is one of its neighbors. The pseudorotation phase angle P correlates with some helicoidal and conformational parameters, the strongest correlation is with δ.

Of the axis base-pair parameters the displacements XDP and YDP show quite small correlation with the other parameters (see Table III). XDP is coupled with STR and OPN, often also with SHF, and in many cases with one or both of the ζ-angles.

TABLE II Statistically relevant zero-lag cross-correlation coefficients of the conformational parameters. Subscript in the parameter symbols denotes the strand of the double helix

Parameter 1	*Parameter* 2	*crac**	N^{crac}†	*crc*‡	N^{crc}¶
χ_1	INC			0.52	(9)
	TIP	−0.24	(5)		
	STR			0.28	(6)
	STG	−0.28	(6)		
	BKL			0.40	(9)
	TLT	−0.45	(7)		
	δ_1			0.66	(7)
	ζ_1	−0.59	(5)		
	P_1			0.73	(6)
χ_2	INC			0.47	(10)
	STG	−0.24	(6)		
	BKL	−0.48	(9)		
	RIS	−0.52	(5)		
	TLT	−0.46	(7)		
	γ_2	−0.31	(6)		
	δ_2			0.54	(9)
	ζ_2	−0.65	(6)		
	P_2			0.60	(7)
γ_1	TWS			0.18	(5)
γ_2	TIP			0.29	(5)
	χ_2	−0.36	(7)		
	δ_2	−0.29	(5)		
	ζ_2			0.40	(6)
δ_1	PRP			0.30	(7)
	ROL	−0.44	(6)		
	χ_1			0.66	(7)
	ϵ_1			0.34	(5)
	ζ_1	−0.71	(8)		
	β_1	−0.33	(4)	0.27	(2)
	P_1			0.92	(6)
δ_2	TIP	−0.45	(8)		
	PRP			0.46	(8)
	χ_2			0.62	(9)
	ζ_2	−0.64	(9)		
	P_2			0.86	(6)
ϵ_1	ζ_1	−0.62	(7)		
	α_1	−0.70	(5)	0.27	(3)
	β_1	−0.44	(5)	0.64	(3)
ϵ_2	ζ_2	−0.63	(8)		
	β_2	−0.56	(5)	0.46	(2)
ζ_1	XDP	−0.59	(5)		
	TIP	−0.40	(7)		
	PRP	−0.37	(6)		
	SHF			0.42	(6)
	χ_1	−0.66	(4)	0.17	(1)
	δ_1	−0.71	(8)		
	ϵ_1	−0.72	(6)		
	α_1	−0.24	(2)	0.85	(3)
	β_1	−0.75	(3)	0.59	(4)
	P_1	−0.64	(6)	0.35	(1)
ζ_2	χ_2	−0.75	(5)		
	δ_2	−0.65	(7)		
	ϵ_2	−0.63	(8)		
	α_2	−0.11	(3)	0.89	(2)
	β_2	−0.64	(2)	0.65	(5)
	P_2			0.61	(5)
α_1	ϵ_1	−0.70	(5)	0.27	(3)
	ζ_1	−0.17	(3)	0.83	(4)
	β_1	−0.54	(8)		
α_2	ϵ_2	−0.49	(4)	0.24	(2)
	β_2	−0.52	(7)		
β_1	ROL1	−0.35	(2)	0.19	(3)
	δ_1	−0.39	(3)	0.26	(3)
	ϵ_1	−0.44	(5)	0.64	(3)
	ζ_1	−0.75	(3)	0.54	(5)
	α_1			0.54	(8)
β_2	χ_2	−0.50	(4)	0.43	(2)
	ϵ_2	−0.56	(5)	0.46	(2)
	ζ_2	−0.64	(2)	0.65	(5)
	α_2	−0.52	(7)		
P_1	TIP	−0.57	(5)		
	STG	−0.24	(5)		
	ROL	−0.49	(6)		
	χ_1			0.57	(8)
	δ_1			0.83	(8)
	ζ_1	−0.65	(7)		
P_2	PRP			0.41	(5)
	δ_2			0.84	(7)
	β_2	−0.36	(3)	0.37	(3)

* Mean value of the anti-correlated cases. † Number of observations in crac. ‡ Mean value of the correlated cases. ¶ Number of observations in crc.

Correlation with ζ shows that motion of a base or a base pair often originates from the backbone, while the sugar–base pair behaves as rigid body. YDP shows coupling with SLD and TWS. INC and TIP are correlated with many structure parameters. Most interesting is perhaps the coupling with the sugar buckering parameter *P*. Orientation of the base is coupled with the orientation of the sugar.

As shown in Table IV the intra base-pair parameters excluding OPN are correlated with many other parameters. SHR is correlated with STR. This means that a breathing motion of the bases also involves a shear motion. SHR often seems to be correlated with the twisting, TWS, occurring between base pairs. An interesting observation is coupling between the motions of STR and STG. An obvious reason for this behavior is the parrying motion resulting in a layered structure from

TABLE III Statistically relevant zero-lag cross-correlation coefficients of the axis-base pair parameters. Notation is the same as in Table II

Parameter 1	*Parameter* 2	*crac*	N^{crac}	*crc*	N^{crc}
XDP	STR			0.39	(8)
	OPN			0.35	(9)
	SHF	−0.59	(6)		
	ζ	−0.51	(10)		
YDP	SLD	−0.66	(6)		
	TWS	−0.39	(5)		
	ζ_1	−0.61	(1)	0.23	(4)
INC	STR			0.26	(6)
	STG	−0.41	(10)		
	BKL	−0.34	(4)	0.36	(3)
	PRP	−0.36	(7)		
	TLT	−0.54	(9)		
	ROL	−0.41	(5)		
	χ			0.25	(20)
	P_1			0.39	(6)
TIP	INC	−0.42	(1)	0.30	(4)
	STR	−0.15	(1)	0.32	(4)
	BKL	−0.32	(9)		
	ROL	−0.63	(9)		
	γ	−0.19	(2)	0.29	(6)
	δ_1			0.45	(8)
	δ_2	0.51	(7)		
	ϵ	−0.25	(4)	0.30	(4)
	ζ_2			0.49	(5)
	P_1			0.43	(8)

TABLE IV Statistically relevant zero-lag cross-correlation coefficients of the intra-base pair parameters. Notation is the same as in Table II

Parameter 1	*Parameter* 2	*crac*	N^{crac}	*crc*	N^{crc}
SHR	STR	−0.28	(2)	0.20	(5)
	OPN	0.15	(3)	0.26	(2)
	SLD	−0.15	(6)		
	TWS	−0.37	(7)		
	ζ_2	−0.06	(2)	0.13	(4)
STR	XDP			0.40	(9)
	INC			0.24	(7)
	SHR	−0.28	(3)	0.22	(4)
	STG	−0.23	(8)		
	BKL	−0.26	(5)	0.31	(2)
	PRP	−0.21	(5)		
	OPN			0.68	(10)
	SHF	−0.34	(7)		
	χ			0.20	(10)
STG	INC	−0.41	(10)		
	STR	−0.23	(8)		
	BKL	−0.39	(3)	0.41	(3)
	OPN	−0.28	(3)	0.21	(2)
	RIS	−0.46	(4)	0.16	(2)
	TLT			0.42	(9)
	χ	−0.29	(9)		
BKL	INC	−0.34	(4)	0.52	(1)
	TIP			0.40	(6)
	SHR	−0.20	(2)	0.28	(3)
	STR	−0.26	(5)	0.23	(4)
	STG	−0.31	(5)	0.39	(3)
	RIS			0.46	(8)
	χ_1			0.46	(8)
	χ_2	−0.44	(9)		
	P			0.26	(5)
PRP	INC	−0.37	(7)		
	STR	−0.21	(6)		
	BKL	−0.49	(2)	0.32	(3)
	TLT			0.20	(6)
	χ_1	−0.34	(8)		
	δ_1			0.33	(6)
	δ_2			0.45	(7)
	ζ	−0.47	(10)		
	P			0.36	(8)
OPN	XDP			0.38	(8)
	SHR	−0.16	(4)	0.26	(2)
	STR			0.68	(10)
	SHF	−0.41	(7)		

TABLE V Statistically relevant zero-lag cross-correlation coefficients of the inter-base pair parameters. Notation is the same as in Table II

Parameter 1	*Parameter* 2	*crac*	N^{crac}	*crc*	N^{crc}
SHF	XDP	−0.52	(8)		
	YDP			0.60	(5)
	STR	−0.34	(7)		
	OPN	−0.38	(8)		
	ϵ	−0.23	(8)		
	ζ			0.43	(10)
SLD	YDP	−0.57	(8)		
	TWS			0.44	(6)
	ϵ_1			0.30	(6)
	ϵ_2	−0.21	(2)		
RIS	STR	−0.20	(3)	0.26	(2)
	BKL			0.49	(8)
	TWS			0.28	(5)
	χ_2	−0.39	(9)		
	ζ			0.30	(5)
	P	−0.28	(5)		
TLT	INC	−0.58	(8)		
	STG			0.42	(9)
	χ_1	−0.46	(7)		
	χ_2	−0.40	(8)		
ROL	TIP	−0.66	(8)		
	TWS	−0.40	(5)		
	δ_1	−0.45	(7)		
	δ_2			0.37	(4)
	ϵ_1	−0.30	(3)	0.28	(2)
	ζ_1			0.44	(7)
	β_1	−0.35	(2)	0.23	(3)
	P_1	−0.49	(6)		
TWS	YDP			0.26	(5)
	SHR	−0.32	(9)		
	SLD			0.36	(8)

the molecular force fields. STR and OPN are correlated at all possible sites. This is natural because the opening of a base pair involves a stretching like motion and Curves is not able to distinguish the opening motion from a stretching motion. STR also seems to have an accompanying change in the glycosyl angle χ. This shows that in stretching the motions of a base plane (or the mean plane of a base pair) and the related sugar are coupled. This is understandable because the sugar groups are parts of the backbone, which thus restricts their motion. Staggering of the bases of a base pair, STG, seems to be coupled with the tilt, TLT, between the neighboring base pairs. In a half of the possible sites there is also correlation with χ. This relates back to the stretching motion. BKL shows correlations with STR, STG, RIS and χ with which it is naturally linked. The most notable correlation of the propeller, PRP, is obviously with either or both of the two backbone torsion angles ζ. This shows that propeller motion originates from the backbone or has effects to this distance. Moreover, one should note its correlation with the sugar bucker phase angle P.

The correlation data for the inter-base pair parameters is given in Table V. Of these parameters SHF shows zero time-delay correlation with XDP eight times out of 10. This is due to their similar nature. Both are displacement type motions and in the same direction. SHF also seems to follow the motion of STR, obviously due to the few structural changes occurring in the backbone (correlations with ϵ and ζ). RIS correlates only with the χ of the second strand. No reason for this could be found. TLT naturally correlates with χ on both sides. ROL has significant correlations with TWS and δ. The ROL-δ correlation indicates that the sugar–base groups function as stiff units with hinge points in the backbone. ROL is often coupled with the other backbone torsion angles (ϵ, ζ and β).

5.3.2 Comparison With Experimental Data

Some experimental correlation data can be found in the literature [44] for the structural parameters of DNA. Most of these data are for the backbone and glycosyl torsion angles. For comparison we have again used the minimum–maximum data of our

correlation matrix from which we have chosen those parameter pairs whose correlation data contains at least one backbone or glycosyl torsion angle, or *P*. In this case, we allow the time delay to be different from zero if it is small (under 1 ps), and demand that the absolute value of the correlation coefficient is at least 0.3. Again we define a parameter pair to be correlated if roughly a half of all possible pairs fulfills the correlation criteria. The results of these analyses are shown in Table VI. For each parameter pair the mean absolute value of the correlation coefficient at its maximum or minimum, and the number of sites included in the mean value, divided into correlated (+) and anti-correlated (−) columns, are given. Notice that the mean correlation coefficients indicate quite strong correlations although the limit used was quite low.

According to the literature [44] all backbone torsion angles of the helical polynucleotides are correlated. Correlation coefficients are observed in the range from 0.65 to 0.71. Table VI shows that in our simulations this is not quite the case. Especially the torsion angle γ lacks correlations. The torsion angle χ seems to be coupled with most of the backbone torsion angles which shows that backbone deformations extend to the core of the double helix. Table VI supports the observation on B-DNA that sugar conformations are strongly correlated with χ. On the other hand our simulations do not show correlation between α and γ as observed for the A-DNA. Moreover, in A-DNA the motion of the bases of a base-pair is related to the angles α, γ and χ, while in the simulations STR-α, STR-γ and STR-χ had correlation coefficients below 0.2, 0.2 and 0.3, respectively (not shown in Table VI). These values indicate only weak correlation. Our simulations thus support a B-form in favor of an A-form for the structure of the DNA of this study.

TABLE VI Mean absolute values of the cross-correlation coefficients (cc) of statistically relevant correlations between the conformational parameters

Parameter pair	*cc*	N^{crac}*	N^{crc}†
$\chi - \delta$	0.70	0	16
$\chi - \epsilon$	0.45	5	6
$\chi - \zeta$	0.64	16	0
$\chi - \beta$	0.45	7	2
$\chi - P$	0.68	1	14
$\delta - \zeta$	0.73	16	0
$\delta - P$	0.84	1	15
$\epsilon - \zeta$	0.66	15	0
$\epsilon - \beta$	0.52	10	5
$\zeta - \beta$	0.63	7	8
$\zeta - P$	0.70	10	1
$\alpha - \beta$	0.58	13	0

* number of the anti-correlated cases. † number of the correlated cases.

6 DISCUSSION

In this work we studied the decamer d(ATGCAGTCAG)$_2$ embedded in ionic water solution at the relative humidity of 95% and at 300 K temperature. In the first part of the study we concentrated on establishing the equilibrium state of the decamer. To this end we also have studied the applicability of the kinetic part of Boltzmann's H function to the verification of thermodynamic equilibrium of the molecular system. In the latter part of the study we studied both the statistical and dynamic properties of the decamer using conformational and helicoidal parameters produced by the methodology of the Curves program.

In the thermodynamically stable state the velocity distribution of the atoms is the Maxwell–Boltzmann distribution, i.e. it has a Gaussian form. The kinetic part of Boltzmann's H function, H, is defined such that in this state $H = 0$. This work shows that $H = 0$ can be obtained if the distribution really has a Gaussian form, is smooth enough, and if it is not truncated at a small velocity. Both in experimental research and in simulations smoothness of the probability density functions is not necessarily guaranteed due to the finite number of constituents in the system. This holds especially true in molecular systems. According to this work the usefulness of H in deciding thermodynamical stability is better in cases where there are quite many atoms of the same type, like in bulk matter. In this kind of systems the adequate statistics can easily be obtained. If the original requirement of H as a continuously monitored quantity is lifted, then better statistics can be obtained for distributions over non-zero time intervals, and this method could also be used in molecular systems. Applied in this way, Boltzmann's H function shows our DNA sequence is at thermodynamical equilibrium.

The statistical structural stability of the decamer was studied using the conventional RMSD method. RMSDs show stable behavior in the same way as in recent similar MD studies using the latest force fields.

The conformational and helicoidal structure of the decamer was analyzed here both statistically and dynamically. Distributions of the conformational parameters show that the backbone of the DNA has locally a few, not only one preferred orientation. Evolution of the trajectories shows this even better. A typical dynamical behavior of a conformational parameter is that it fluctuates around a certain level for a time and then jumps into another mean level, from which it returns back or jumps to a third mean level after a while. The frequency of jumps of this kind depends on the parameter. The average distributions of the helicoidal parameters usually show only one peak. The broadness

of the distributions, however, varies a lot, and a detailed study of the evolution of the local individual parameters shows that these distributions hide different local behaviors. Parts of the helicoidal parameters have a similar jerky character as the conformational parameters. For the helicoidal parameters jumps happen however in a smaller scale and, moreover, one level of values seems to be preferred over the others. This kind of behavior cannot be found from the average distributions. It appears evident that a large molecule in aqueous solution typically has a family of possible structures, and not a single well defined arrangement of atoms. The concept of dynamical stability must therefore be interpreted accordingly.

One part of the dynamical analysis carried out was a partial correlation analysis. We found that there exist concerted motions in the decamer, and these are very localized. Structural parameters over even one nucleic acid pair do not show significant correlations. Within one site there also are parameters, which do not show co-operative motion even with their neighbors. Most striking of these is the conformational parameter γ. A typical feature is that conformational and helicoidal parameters are correlated with parameters, which belong to the same main class. Of the conformational parameters only χ has correlations with many helicoidal parameters on both strands, and with ζ in strand one. On the other hand only TIP, PRP and ROL generally have several correlations with conformational parameters. In all these cases a base pair or consecutive base pairs are winding around the axis parallel to the base pair. Correlations of these parameters with the conformational parameters mean that this winding affects the backbone or originates from changes occurring in the backbone. All these parameters are coupled with the position of the sugar (correlation with δ). Otherwise they are correlated with different conformational parameters. So, in principle this may offer a selective method to manipulate the orientation of base pairs. Comparison of the correlation results of this work with the experimentally known correlations show that, in the simulated decamer, the backbone torsional angle γ, contrary to the experimental results, is not coupled to the other conformational parameters.

As to the classification of the structure of the decamer studied, our sequence shows features of both the A-form and the B-form DNA. When we go individually through all the structural parameters for which there is available experimental data, we see that the values of χ, TIP, PRP and TLT of our decamer are nearer the A-DNA. The values of INC, RIS, TWS and ROL, and the shapes of the distributions for TIP, STR, TLT and ROL are similar to those of B-DNA. XDP is in between the experimental values for the A-form and the B-form, slightly nearer that for the B-form. δ-values are distributed equally between those for these two forms. Most of the χ-values are nearer that for the A-form although there are many values between those of these two main forms. The other parameters do not show any evidence towards either structural form. α and γ are far from the observed, and the forms of the XDP, SLD and SHR distributions do not have any characteristic shape. As for the knowledge about the dynamical behavior of the structural parameters in various structural forms of DNA, we only have some experimental data on the correlation behavior of the conformational parameters. Here we have not observed correlation between α and γ, and correlations STR-α, STR-γ and STR-χ are weak. This means that our decamer is not fluctuating in the A-form. Sugar conformation and χ show correlation which is typical of the B-form. So, as a summary of the attempts to find the structural category for the simulated piece of DNA, we can conclude that it clearly is not in the A-form. On the other hand, it is not purely in the B-form. It is quite possible that this simulation describes a decamer which is in a transitional state between the B- and C-forms [27]. Recent experimental observation [46] gives support for the existence of sequence dependent dynamically stable intermediate states of DNA. Future studies, both on the experimental side and simulations, will be needed to reveal the richness of the structural phase space.

We also plan to do similar simulations for counterions other than the Li^+ ions used here. These simulations will provide additional information about the conformational form of the model decamer. On the other hand, CHARMM force field is known [47] to slightly favor the A-form while the periodicity in the z direction used here, i.e. topological constraints; tend to prevent transition from the B-form to the A-form. The conclusions made so far based on the present simulation are in agreement with a competition between these two tendencies.

References

[1] Chen, Y.Z., Szabo, A., Schroeter, D.F., Powell, J.W., Lee, S.A. and Prohofsky, E.W. (1997) "Effect of drug-binding-induced deformation on the vibrational spectrum of a DNA-daunomycin complex", *Phys Rev. E* **55**, 7414.

[2] Lee, S.A., Rupprecht, A. and Chen, Y.Z. (1998) "Drug binding to DNA: observation of the drug-DNA hydrogen-bond-stretching modes of netropsin bound to DNA *via* Raman spectroscopy", *Phys Rev. Lett*. **80**, 2241.

[3] Szabo, A., Flowers, R.A., Carter, B.J. and Lee, S.A. (1998) "Observation of two-mode binding to DNA by bipyridyl-(ethylenediamine)platinum(II): isothermal titrational calorimetry and midinfrared absorption studies", *Phys Rev. E* **58**, 7754.

[4] Englander, S.W., Kallenbach, N.R., Heeger, A.J., Krumhansl, J.A. and Litwin, S. (1980) "Nature of the open state in long

polynucleotide double helices: possibility of soliton excitations", *Proc. Natl. Acad. Sci.* **77**, 7222.
[5] Peyrard, M. and Bishop, A.R. (1989) "Statistical mechanics of a nonlinear model for DNA denaturation", *Phys Rev. Lett.* **62**, 2755.
[6] Briki, F. and Genest, D. (1993) "Molecular dynamics study of the base pair opening process in the self-complementary octanucleotide d(CTGATCAG)", *J. Biomol. Struct. Dyn.* **11**, 43.
[7] Frenkel, A. and Smit, B. (1996) Understanding Molecular Simulation—From Algorithms to Applications (Academic Press, San Diego).
[8] Blackburn, G.M., Gait, M.J., (1996) In: *Nucleic Acids in Chemistry and Biology,* 2nd ed. (Oxford University Press, Oxford).
[9] Van Gunsteren, W.F., Berendsen, H.J.C., Geurtsen, R.G. and Zwinderman, H.R.J. (1986) "A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: comparison with experimental two-dimensional NMR data", In: Beveridge, D.L. and Jorgensen, W.L., eds, *Computer Simulation of Chemical and Biomolecular Systems* (Annals of the New York Academy of Sciences, New York), pp 287–303.
[10] Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, Jr, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules", *J. Am. Chem. Soc.* **117**, 5179.
[11] MacKerell, Jr., A.D., Wiorkiewicz-Kuczera, J. and Karplus, M. (1995) "An all-atom empirical energy function for the simulation of nucleic acids", *J. Am. Chem. Soc.* **117**, 11946.
[12] Cheatham, III, T.E., Cieplak, P. and 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. Dyn.* **16**, 845.
[13] Foloppe, N. and MacKerell, Jr., A.D. (2000) "All-atom empirical force field for nucleic acids: I. Parameter optimisation based on small molecule and condensed phase macromolecular target data", *J. Comp. Chem.* **21**, 86.
[14] Langley, D.R. (1998) "Molecular dynamic simulations of environment and sequence dependent DNA conformations: the development of the BMS nucleic acid force field and comparison with experimental results", *J. Biomol. Struct. Dyn.* **16**, 487.
[15] MacKerell, Jr., A.D. and Banavali, N.K. (2000) "All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution", *J. Comp. Chem.* **21**, 105.
[16] Sagui, C. and Darden, T.A. (1999) "Molecular dynamics simulations of biomolecules: long-range electrostatic effects", *Annu. Rev. Biophys Biomol. Struct.* **28**, 155.
[17] Beveridge, D.L. and McConnell, K.J. (2000) "Nucleic acids: theory and computer simulation, Y2K", *Curr. Opin. Struct. Biol.* **10**, 182.
[18] Cheatham, III, T.E. and Kollman, P.A. (2000) "Molecular dynamics simulation of nucleic acids", *Annu. Rev. Phys Chem.* **51**, 435.
[19] van Dam, L., Lyubartsev, A.P., Laaksonen, A. and Nordenskiöld, L. (1998) "Self-diffusion and association of Li^+, Cs^+, and H_2O in oriented DNA fibers. An NMR and MD simulation study", *J. Phys Chem. B* **102**, 10636.
[20] Lyubartsev, A.P. and Laaksonen, A. (1998) "Molecular dynamics simulations of DNA in solution with different counter-ions", *J. Biomol. Struct. Dyn.* **16**, 579.
[21] Lyubartsev, A.P. and Laaksonen, A. (1999) "Effective potentials for ion–DNA interactions", *J. Chem. Phys* **111**, 11207.
[22] Dang, L.X. (1992) "Development of nonadditive intermolecular potentials using molecular dynamics: solvation of Li^+ and F^- ions in polarizable water", *J. Chem. Phys* **96**, 6970.
[23] Toukan, K. and Rahman, A. (1985) "Molecular-dynamics study of atomic motions in water", *Phys Rev. B* **31**, 2643.
[24] Allen, M.P. and Tildesley, D.J. (1987) *Computer Simulations of Liquids* (Clarendon Press, Oxford).
[25] Melchionna, S., Ciccotti, G. and Holian, B.L. (1993) "Hoover NPT dynamics for systems varying in shape and size", *Mol. Phys* **78**, 533.
[26] Lee, S.A., Lindsay, S.M., Powell, J.W., Weidlich, T., Tao, N.J., Lewen, G.D. and Rupprecht, A. (1987) "A Brillouin-scattering study of the hydration of Li-DNA and Na-DNA films", *Biopolymers* **26**, 1637.
[27] Lorprete, D.M. and Hartman, K.A. (1993) "Conditions for the stability of the B-structural, C-structural, and Z-structural forms of poly(dG–dC) in the presence of lithium, potassium, magnesium, calcium, and zinc cations", *Biochemistry* **32**, 4077.
[28] Haile, J.M. (1992) Molecular Dynamics Simulation (Wiley, New York).
[29] Heermann, D.W (1986) Computer Simulation Methods in Theoretical Physics (Springer, New York).
[30] R. Lohikoski and J. Timonen, unpublished results.
[31] Reif, F. (1965) Fundamentals of Statistical And Thermal Physics (McGraw-Hill, Tokyo).
[32] Swaminathan, S., Ravishanker, G. and Beveridge, D.L. (1991) "Molecular dynamics of B-DNA including water and counter-ions: a 140-ps trajectory for d(CGCGAATTCGCG) based on the GROMOS force field", *J. Am. Chem. Soc.* **113**, 5027.
[33] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) "CHARMM: a program for macromolecular energy, minimization, and dynamics calculations", *J. Comp. Chem.* **4**, 187.
[34] Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, Jr, S. and Weiner, P. (1984) "A new force field for molecular mechanics simulation of nucleic acids and proteins", *J. Am. Chem. Soc.* **106**, 765.
[35] Darden, T.A., York, D. and Pedersen, L.G. (1993) "Particle mesh Ewald: an $N\log(N)$ method for Ewald sums in large systems", *J. Chem. Phys* **98**, 10089.
[36] Cheatham, III, T.E., Miller, J.L., Fox, T., Darden, T.A. and Kollman, P.A. (1995) "Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins", *J. Am. Chem. Soc.* **117**, 4193.
[37] Sprous, D., Young, M.A. and Beveridge, D.L. (1998) "Molecular dynamics studies of the conformational preferences of a DNA double helix in water and an ethanol/water mixture: theoretical considerations of the A ↔ B transition", *J. Phys Chem. B* **102**, 4658.
[38] Young, M.A., Ravishanker, G. and Beveridge, D.L. (1997) "A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation", *Biophys J.* **73**, 2313.
[39] Norberg, J. and Nilsson, L. (1995) "Constant pressure molecular dynamics simulations of the dodecamers: d(GCG-CGCGCGCGC)$_2$ and r(GCGCGCGCGCGC)$_2$", *J. Chem. Phys* **104**, 6052.
[40] Yang, L. and Pettitt, B.M. (1996) "B to A transition of DNA on the nanosecond time scale", *J. Phys Chem.* **100**, 2564.
[41] MacKerell, Jr., A.D. (1997) "Influence of magnesium ions on duplex DNA structural, dynamic, and solvation properties", *J. Phys Chem. B* **101**, 646.
[42] Cruzeiro-Hansson, L. and Goodfellow, J.M. (1994) "Comparison of DNA duplexes with and without O^4-methylthymine: nanosecond, molecular dynamics simulations", *J. Chem. Soc. Farad. Trans.* **90**, 1415.
[43] McConnell, K.J., Nirmala, R., Young, M.A., Ravishanker, G. and Beveridge, D.L. (1994) "A nanosecond molecular dynamics trajectory for a B-DNA double helix: evidence for substates", *J. Am. Chem. Soc.* **116**, 4461.
[44] Saenger, W. (1984) Principles of Nucleic Acid Structure (Springer, Heidelberg).
[45] Lavery, R. and Sklenar, H. (1996) *Curves 5.1—Helical Analysis of Irregular Nucleic Acids Laboratoire de Biochimie Theorique, CNRS URA 77,* (Institut de Biologie Physico-Chimique, Paris).
[46] Ng, H.-L., Kopka, M.L. and Dickerson, R.E. (2000) "The structure of a stable intermediate in the A ↔ B DNA helix transition", *Proc. Natl. Acad. Sci. USA* **97**, 2035.
[47] Feig, M. and Pettitt, B.M. (1998) "Structural equilibrium of DNA represented with different force fields", *Biophys J.* **75**, 134.
[48] IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN)(1983) "Symbols for specifying the conformation of

polysaccharide chains. Recommendations 1981", *Eur. J. Biochem.* **131**, 5–7.
[49] IUPAC–IUB Joint Commission on Biochemical Nomenclature (JCBN)(1983) "Abbreviations and symbols for the description of conformations of polynucleotide chains. Recommendations 1982", *Eur. J. Biochem.* **131**, 9–15.
[50] Dickerson, R.E. (1989) "Definitions and nomenclature of nucleic acid structure parameters", *J. Biol. Struct. Dyn.* **6**, 627.
[51] Lavery, R. and Sklenar, H. (1988) "The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids", *J. Biol. Struct. Dyn.* **6**, 63.
[52] Lavery, R. and Sklenar, H. (1989) "Defining the structure of irregular nucleic acids: conventions and principles", *J. Biol. Struct. Dyn.* **6**, 655.

APPENDIX: OF THE DEFINITIONS OF THE CONFORMATIONAL AND HELICOIDAL PARAMETERS

Conformational and helicoidal analysis of DNA typically includes a determination of certain structural parameters. The term conformational refers to the structure complex of the backbone including the sugar. Helicoidal is in this context reserved to the parameters originating merely from the bases, their ordering and relationships, because the helicity of the chain mostly affects the relative orientations of the bases.

Using IUPAC–IUB recommendations on the definitions and notations of the conformational parameters, the conformational parameter set of a nucleic acid involves backbone torsion angles, the exocyclic sugar–base torsion angle denoted by χ and the sugar pucker or pseudorotation phase angle P [44,48,49]. The backbone torsion angles are denoted by α, β, γ, δ, ϵ and ζ, respectively, when moved in the $5' \rightarrow 3'$ direction along the backbone from one phosphorous atom to the next. Figure 3 shows the definitions of these angles. The pseudorotation phase angle is defined by equation

$$\tan(P) = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin 36^\circ + \sin 72^\circ)}. \qquad \text{(A1)}$$

To our knowledge, the helicoidal parameters do not have any officially recommended names, and they may vary in different sources. We use here the set agreed upon in the EMBO Workshop on DNA Curvature and Bending in 1988 [50], together with the parameters defined by Lavery and Sklenar [51,52]. These two sets of parameters seem to largely overlap.

Helicoidal parameters include 20 parameters grouped into four sub-groups: the axis-base pair (4), intra-base pair (6), inter-base pair (6) and axis junction (4) parameters. Number in the parenthesis gives the number of parameters in each set [51,52]. The axis-base pair parameters describe the relation of a base pair with the axis of the helical structure. The intra-base pair parameters are related to the mutual relationship of the two bases of a base pair. The inter-base pair parameters give information about the spatial relationship of consecutive base pairs. The axis junction parameters describe the displacement and the orientation of the global helical axis of the system with respect to the straight axis of the corresponding canonical structure.

All definitions of the helicoidal parameter set are based on the orthogonal axis system fixed to the (flat) base. This provides four parameters through which the location of the nucleotide with respect to the global helical axis system can be determined. Two of these parameters are shifts, the y displacement (YDP) along the base-base axis and the x displacement (XDP) perpendicular to the y displacement. The other two are right-handed rotations, "tip" (TIP) is around the longitudinal base axis and "inclination" (INC) around the x displacement axis. These parameters constitute the axis-base pair group. When working with irregular nucleic acids, a practical problem is the determination of the axis. In the method of Lavery and Sklenar, rectangular axes are fixed on every single base and all local parameters of the axis-base group are used as optimization variables in finding an optimal helical axis. This is made by minimizing a function, which describes simultaneously the change in orientation between successive bases and the nonlinearity of the helical axis. When the minimization procedure is accomplished, a global axis is given for the whole double chain. This global axis is composed of the local axis systems whose orientation at each base pair is such that the x axis points towards the major groove, the y axis is towards the first strand and the z axis is normal to the average base-base plane pointing in the $5' \rightarrow 3'$ direction of the first strand.

The group of the intra-base pair parameters is composed of shear (SHR), stretch (STR), stagger (STG), buckle (BKL), propeller (PRP) and opening (OPN). Shear, stretch and stagger are mutual distances of the bases of a base pair measured in the local helicoidal axis system in the x, y and z direction, correspondingly. Buckle, propeller and opening are angles formed between the planes of the bases of a base pair when viewed along the local helicoidal x, y and z axis, correspondingly.

The inter-base parameters are called shift (SFH), slide (SLD), rise (RIS), tilt (TLT), roll (ROL) and twist (TWS). These parameters describe distances and angles between consecutive base pairs. Shift, slide and rise are the displacements of a base pair with respect to the previous base pair in the helical x, y and z directions. Tilt, roll and twist are the angles between the intermediate base pair planes of consecutive base pairs when viewed along the helicoidal x, y and z axis. A possible axis curvature is taken into account in the determination of these inter-base pair parameters.

The axis junction parameter set is composed of the axis x-displacement (AXD), axis y-displacement (AYD), axis inclination (AIN) and axis tip (ATP). The axis x-displacement is a shift of the helical axis in the x direction, and the axis y-displacement similarly in the y direction with respect to the corresponding canonical structure. Axis inclination (tip) is the rotation angle around the x axis (y axis) of the local helical xy plane with respect to the same canonical plane.

The exact mathematical definitions of all these parameters are given in Ref. [51]. The above description is only meant to give a conceptual idea of the terms, and of their origin, used to describe the behavior of the helical double stranded nucleic acid systems.