J Mol Model (2004) 10:185–197
DOI 10.1007/s00894-004-0182-0

ORIGINAL PAPER

Konstantinos Sfyrakis · Astero Provata · David C. Povey · Brendan J. Howlin

Local sequential minimization of double stranded B-DNA using Monte Carlo annealing

Received: 20 February 2003 / Accepted: 15 December 2003 / Published online: 24 March 2004

Abstract A software algorithm has been developed to investigate the folding process in B-DNA structures in vacuum under a simple and accurate force field. This algorithm models linear double stranded B-DNA sequences based on a local, sequential minimization procedure. The original B-DNA structures were modeled using initial nucleotide structures taken from the Brookhaven database. The models contain information at the atomic level allowing one to investigate as accurately as possible the structure and characteristics of the resulting DNA structures. A variety of DNA sequences and sizes were investigated containing coding and non-coding, random and real, homogeneous or heterogeneous sequences in the range of 2 to 40 base pairs. The force field contains terms such as angle bend, Lennard-Jones, electrostatic interactions and hydrogen bonding which are set up using the Dreiding II force field and defined to account for the helical parameters such as twist, tilt and rise. A close comparison was made between this local minimization algorithm and a global one (previously published) in order to find out advantages and disadvantages of the different methods. From the comparison, this algorithm gives better and faster results than the previous method, allowing one to minimize larger DNA segments. DNA segments with a length of 40 bases need approximately 4 h, while 2.5 weeks are needed with the previous method. After each minimization the angles between phosphate–oxygen-carbon A_1, the oxygen–phosphate–oxygen A_2 and the average helical twists were calculated. From the generated fragments it was found that the bond angles are A_1=150°±2°and A_2=130°±10°, while the helical twist is 36.6°±2° in the A strand and A_1=150°±6° and A_2=130±6° with helical twist 39.6°±2° in the B strand for the DNA segment with the same sequence as the Dickerson dodecamer.

Keywords Double strand · B-DNA · Monte Carlo · Annealing · Coding/non-coding DNA

K. Sfyrakis · D. C. Povey · B. J. Howlin
School of Biomedical and Life Sciences, Chemistry,
University of Surrey,
Guildford, GU2 7XH, UK

K. Sfyrakis · A. Provata (✉)
Institute of Physical Chemistry,
National Research Center "Demokritos",
15310 Athens, Greece
e-mail: aprovata@limnos.chem.demokritos.gr

Introduction

Today due to the fast increase in computational power it is possible to investigate the dynamics of the conformational properties of large double helical DNA. This is very interesting because the overall three-dimensional structure and the local flexibility of DNA is thought to play a very important role in the recognition of binding sites by proteins and other small or large, biologically important, molecules. In order to do this, an accurate force field is needed to predict the energy fluctuations during the folding process.

Many studies of DNA folding have been carried out previously, starting from the first ones using Molecular Dynamics (MD) in 1983, [1, 2] 5 years after the first paper on protein dynamics, as quoted in Rappe et al. (1997). [3] Most of the calculations have been carried out in vacuum or the effect of the solvents have been approximated by reducing the charges on the phosphate groups, or a distance-dependent dielectric parameter has been used due to the enormous computational requirements. Recent numerical works include Molecular Dynamics (MD) and Monte Carlo (MC) methods that contain B-DNA strands with water or other solvents. [4, 5] Most of the MD methods investigate the internal dynamics of DNA using atomic or mesoscopic scales [6] and others are based on sequence, characteristics and functions of DNA. [7, 8, 9, 10, 11, 12, 13, 14, 15, 16]

Some of the MC methods investigate large double DNA segments concerning helix deformations upon stretching using internal coordinates giving results that lead to force curves which exhibit a plateau as the conformational transition occurs. [8] Other Monte Carlo

methods use an internal coordinate model associated with pseudorotational representation of sugar repuckering. [17] The molecular dynamics simulation of the Dickerson dodecamer using the particle mesh Ewald sum method by Yong Duan [18] shows very good agreement with the X-ray structure (1.1 Å).

Molecular mechanics studies of double stranded DNA give results with strong sequence effects on the position of the unpaired base and on the overall curvature induced by the "abasic lesion". [19] Such algorithms are based on the "multi-copy approach enabling us to determine base sequences that favor given structural changes or interactions via a single energy minimization". [20] Besides simulation methods, systems of different equations have been used to describe DNA supercoiling and thus treat sequences containing thousands of base pairs. [21] Other methods investigate large supercoiled DNA using Langevin and Euler molecular dynamics [5, 22, 23] obtaining trefoil knotting results. [24] K. Vlahovicek and S. Pongor use a different construction method to form long DNA chains that use constraint molecular dynamics simulations in the final structure to find the lowest energy molecular conformation. [10] Wilma K. Olson and Victor B. Zhurkin developed an algorithm that models double-helical DNA at four levels of three-dimensional structure, showing that conformational changes are critical to the functioning of the helix, including its packaging in the close confines of the cell, the mutual fit of DNA and protein in nucleoprotein complexes, and the effective recognition of base pairs in recombination and transcription. [9] Throughout these methods it is very common to compare the modeled structures and dynamics with the Dickerson dodecamer. [25, 26, 27]

The aim of this work is to compare this local, sequential minimization algorithm with the "full strand minimization method" [28] and to compare the structural and coiling characteristics of the double stranded B-DNA segments originating (a) from coding regions of DNA, (b) from homogeneous DNA strands found frequently in non-coding regions and (c) from artificial random DNA sequences. The two algorithms use a different method during the folding process while each force field is composed of similar energy potentials such as Coulombic and bend angles. Each one of these interactions has the same parameters but a different range of influences.

This study develops a method for modeling three-dimensional structures of B-DNA using a simple but accurate force field in order to examine the overall structure and the characteristics such as helical twist and curvature in the modeled DNA structures. The influence on the folding process of the major forces used in the force field were investigated and analyzed. All the calculations during the minimization were made in vacuum where a minimal complexity model was adopted for every structure.

The optimum energy minimization process of any organic molecule of any size requires all the bonds, angles and torsional angles of the molecule to be included in the minimization procedure. This would have the result of an exponential increase in the computational power required to search for the energy minimum depending on the size of the molecule. However, in the last few years computational power has increased significantly, allowing for minimizing large macromolecules such as DNA and proteins with a small but significant number of degrees of freedom. In many cases, orbital theories have proved that specific groups of atoms (such as sugar rings) are very stable during many chemical or structural processes [29, 30] allowing us to treat them as geometrically stable during the minimization procedure. In this way less conformational changes are required for the study of macromolecules such as B-DNA. In this study, the three-dimensional structure of DNA was determined by seven degrees of freedom (torsional angles) for the sugar phosphate backbone (T_1, T_2,..., T_7) and two angles (A_1, A_2) (see Fig. 1). In this way it is possible to construct highly flexible three-dimensional double stranded DNA structures [31] that could be used for energy minimization models. Later in this study, extra hydrogen bond and angle degrees of freedom will be discussed, in order to monitor other structural changes in the model.

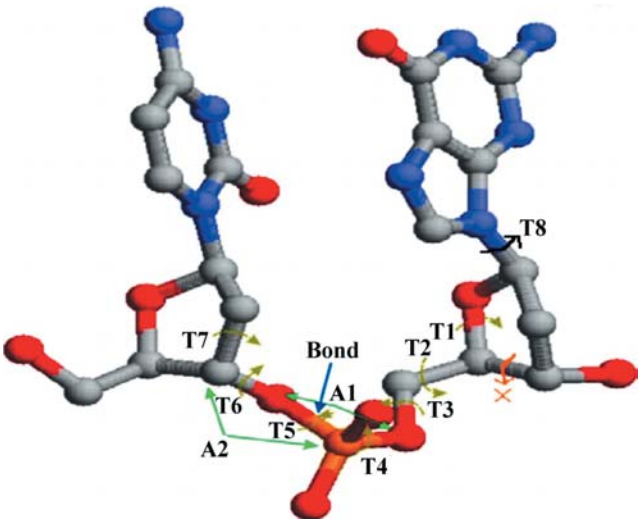

Fig. 1 Two DNA bases and sugars connected with a P–O bond. The DNA strand is rotated via the O–P–O and P–O–C angle as shown. Different types of bonds, angles and torsions used in the simulation are shown

Many of the structural rules that determine the path through space followed by the central axis of the DNA macromolecule are not yet understood. In the case of linear DNA free in solution, the phenomenon of sequence-dependent bending has received considerable attention. [32] In the past, continuum elastic models have been used to investigate the trajectories of linear molecules. [33, 34, 35, 36] These models treat the DNA macromolecules as isotropic, homogeneous, elastic materials without incorporating a good level of detail. Thus, they cannot predict accurately enough local sequence-dependent effects or conformational transitions requiring rearrangements of the secondary structure. More accurate models have been developed by describing the structure

by pseudo-atomic models. [37] These models, although they are less detailed than all-atom models, have the advantages of having smaller computational demands and give reasonable estimations of DNA backbone bending and DNA folding at the same time.

It is well known that the DNA of living organisms is composed of coding and non-coding regions. Although both of these regions are constructed as "random" sequences of base pairs, there are many important statistical differences in their construction. [38, 39] An unexpected feature of the non-coding regions may be reflected in their conformations and functionality, especially in the coiling and super-coiling of the helix. In this work, the structural conformations of three different types of double DNA sequences were studied: (a) segments of coding sequences obtained from real organisms (lambda virus); (b) segments of DNA consisting of the same base-pairs as the ones which are found in non-coding regions of real organisms; (c) random, artificial sequences (the coding regions of DNA have the statistical characteristics of random sequences).

This algorithm minimizes double DNA linear segments of any size and sequence to double folded B-DNA structures using the Monte Carlo Annealing [40] method. It uses a very accurate representation of all the atoms for each base (including the hydrogen atoms) taken from B-DNA structures found from the Brookhaven database. The hydrogen atoms are needed to calculate the hydrogen bond interactions within each base pair. Between each two neighbor bases nine degrees of freedom in total (two angles and seven torsions) were used to rotate each chain in space. The two initial chains are linear and connected by hydrogen bonds only at the first base pair. During the minimization two neighboring base pairs were minimized sequentially starting from the original connected hydrogen bonded pair to the last one in the chain. The phosphate–oxygen bonds between neighboring bases remain constant to prohibit the bases from escaping into space. The force field includes one bonded and three non-bonded interactions whose parameters are found from the Dreiding force field. [41] During the minimization the atoms are treated as hard spheres. This algorithm is similar to those of molecular mechanics and molecular dynamics, where the conformational space is examined to find the global minimum in a wide range of temperatures. During the minimization, information about the energy, atomic distances, angles and torsions are saved together with the corresponding coordinates of each new structure generated. This allows one to determine the variations of the helical twist and bending for each base pair generated. Programs such as Rasmol, [42] Web Lab Viewer and Pdbfit [42] were used for displaying and for calculating the Root Mean Square (RMS) deviation of the structures. The RMS values produced by the sequential Monte Carlo algorithm were compared with the RMS values obtained from the crystallographic B-DNA data.

In the section Model and method, the model of B-DNA is described with its most important characteristics, followed by a description of the initial structures and the different types of sequence used. The method for producing random DNA structures, the different parts and parameters of the force field and the basic description of this algorithm are explained. In the section Double stranded B-DNA, the results of the different type of sequences minimized are presented and analyzed and the average helical twist, the hydrogen bonds, the bond angles and the torsion angles of each structure are calculated. Results are presented for (a) sequences originating from coding DNA of real organisms (b) for homogeneous segments, which are frequently met in non-coding DNA and (c) for artificial random sequences. In the same section Double stranded B-DNA, the differences and similarities in the spatial conformations of the coding and non-coding DNA segments are studied by comparing the spatial characteristics of the produced structure. In the section Differences and similarities between local sequential and full strand minimization procedures, the advantages and disadvantages found between this local sequential minimization algorithm and the global length from the "full strand minimization method" [28] are discussed by comparing the accuracy and the efficiency of the two methods. Finally, in the Conclusion the main conclusions of this work are summarized and future applications and modifications are proposed.

Model and method

In the current section the most important structural characteristics found in the B-DNA form are described including a detailed description of the initial DNA structure used to predict the double B-DNA strands. The parameters needed to develop the force field for this simulation are also described as a function of a set of the internal coordinates. The methods of Monte Carlo and simulated annealing, which are used in order to determine the lowest energy structure of the diverse sequences are also described. The program is written in such a way as to accept from the command line the DNA sequence that will be minimized. During the sequential minimization, for each interval the molecule is stored into a different file and all the different parts of the energy force field for each step of the process are printed into a single file.

For the generation of the minimum-energy structures, a multiprocessor server containing four CPUs running at 150 MHz was used, running on the IRIX64 operating system. The simulation was written in the C++ programming language using the "CC" compiler, creating a class for each part of the algorithm. The most important parts of the algorithm include the generation of the new structures, the calculation of the energy–force field and the initialization and storing of the molecule coordinates during the minimization.

Structural characteristics of B-DNA

In the current work, the investigation of B-DNA is chosen since it is the most common DNA structure found in nature. Also, it is regarded as the native form because its X-ray pattern resembles that of the DNA in intact sperm heads. [31] However, the general methodology of this research should be similar for minimization of the other forms of DNA. It should only be necessary to change some of the basic parameters of this minimization procedure.

B-DNA consists of two polynucleotide strands that wind about a common axis with a right-handed twist to form a ~20-Å diameter double helix. [43, 44] The two strands are anti-parallel (run in opposite directions) and wrap around each other in such a way that they cannot be separated without unwinding the helix. The natural B-DNA helix has ten base pairs (bp) per turn (a helical twist of 36° per bp), [38, 39] a pitch (rise per turn) of 34 Å and the planes of the bases are nearly perpendicular to the helical axis. [43] Later, these basic theoretical structural characteristics found from the X-ray crystallographic data will be compared with our modeled structures.

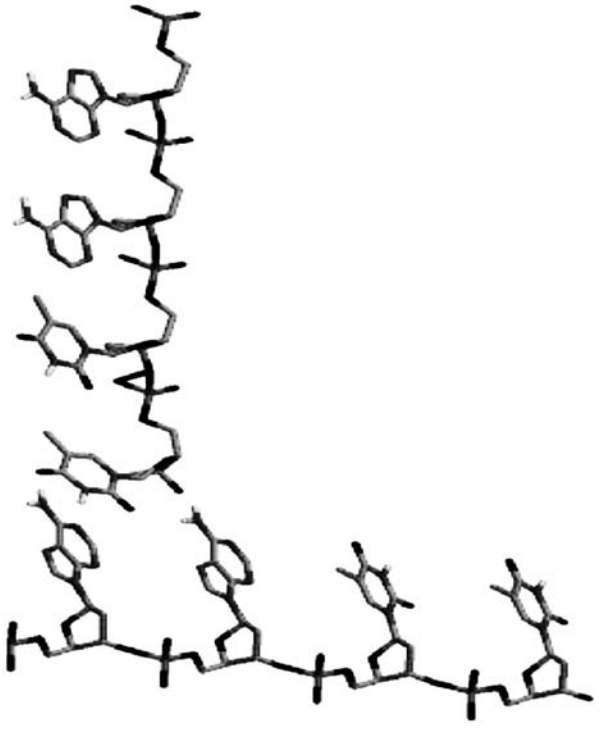

Fig. 2 Example of an initial linear B-DNA structure used in the sequential minimization simulation

General description of the algorithm

The algorithm starts with an initial state of very high energy. The choice of this initial state is described in the section Initialization of structure. For the construction of the new structures, the random generator method is used, described in the section Random generator of new structures. The energy of every new molecular structure is calculated using a general force field described in the section Force field, whereas the methodology and the parameters used to obtain the conformational energy minimum of the structures are presented in the section Monte Carlo annealing simulation.

Initialization of structure

The initial coordinates of the four DNA bases are copied from a high-resolution pdb file "166D.pdb", [45] from the Brookhaven database.

This polynucleotide strand is known by the name "Dickerson–Drew dodecamer" and is widely discussed in the literature. [25, 26, 27] Some of the studies in the literature involve comparison with its own X-ray structure and that of the NMR structure of the native counterpart, [46] others involve MD including water and counterions [4, 18, 47] and others investigate the stability and the conformation of the dodecamer by inducing sugar puckers or binding to other molecules. [48, 49]

The 166D.pdb file contains a B-DNA polynucleotide strand (deoxyribonucleic acid), and a docked gamma-oxapentamidine. It is formed by 12 base pairs with the sequence in strand A (cgcgaattcgcg) and is resolved at a resolution of 2.2 Å (Appendix A, Tables 6 and 7).

In order to produce the initial building blocks of the simulated B-DNA strands, the pdb file was decomposed into its constituent base pairs with attached side-chains. Each pair of bases, i.e. AT, TA, GC and CG from the pdb file were then superimposed on the side-chains and the sugar part using the program pdbfit. [42] The average coordinates of the backbone were then generated and used to build the starting model for each different type of base for each chain. The hydrogen atoms have been added in order to calculate the angles for the hydrogen bond interactions. To add these hydrogen atoms in the initial building model, every previously defined base was superimposed with the corresponding one from the 166D.pdb file found in the Brookhaven database. [50] The new hydrogen atom coordinates are saved and transferred to the initial models.

The initial structure of the double stranded B-DNA was made by connecting DNA bases forming a linear strand. These bases remain unchangeable at the phosphate–oxygen bonds (~1.6 Å) where the hydrogen bonds between the conjugate bases were not formed in the initial structure (see Fig. 2). In that way the system is made computationally simpler, both strands are allowed to fit their own conformations without bias and the bonds are allowed to form naturally from the chosen conformation. The current version of the program can handle only double DNA strands assuming that they exist in a vacuum (Coulombic interactions are included). Minimization in water or another medium could be achieved by changing the Coulombic parameters accordingly. The initial state

used here is chosen in order to show how the single chain curls under the influence of the force field. Other authors use different initial configurations as starting states. [15]

For the generation of the minimum energy structures, various double DNA strands are used. Some of them are homogeneous or artificial (random) and others are extracted from the DNA of test organisms. The real DNA segments are taken from a coding region of DNA, of the Bacteriophage Lambda virus, whose entire genome can be found in Genbank. The Lambda virus is the most commonly used test organism as it consists of almost pure coding DNA. In order to understand how DNA strands coil, sequences with variable number of bases (2, 5, 10, ... 40 base pairs) are used for minimization.

In each run four different structures are minimized simultaneously in order to decrease the total time of the minimizations for all the runs. In this simulation, the energy of the system and the new conformations generated after each interval neither depend on the number of base pairs minimized nor on the different energy-components forming the force field, which are the same for every DNA sequence. A 20-base, double DNA strand needs approximately 2 h for minimization, and a 40-base strand needs approximately 4.5 h. The time to complete the minimization process increases slightly with the number of bases because the algorithm has to translate more atoms during each step.

Random generator of new structures

An important part of this molecular-energy minimization simulator, is the random generator of new structures of DNA. Here the base pairs were chosen sequentially, minimizing only two base pairs each time, where during each minimization the algorithm rotates the second of the two chosen base pairs in three-dimensional space. In order to find a new lower-energy conformation, the following steps are taken:

- Move to the next base pair.
- Choice of two randomly generated angles for each chain corresponding to the angles A_1 and A_2, (see Fig. 1), with a defined step of ±1° respectively (see Table 1).

Table 1 The range of all randomly generated angles in degrees for B-DNA structures. The *a* and *b* are integer numbers corresponding to the base counting with *b* equal to *a*+1

Symbol	Angles	Strand	Range
A1	O(a)–P(a)–O(b)	A, B	0.0° ↔ 360.0°
A2	P(a)–O(b)–C(b)	A, B	0.0° ↔ 360.0°

- Choice of 14 torsion angles, seven defined for the first strand and seven for the second (see Fig. 1), with a defined step for all of them of ±5°. The T8 torsion angle is not used in this method in order to reduce the size of the conformational space (see Fig. 1 and Table 2).

The angles defined in Fig. 1 and Table 1 do not correspond completely to the standard nucleic acid torsion notation [51] but they are chosen here as more appropriate for the needs of the current algorithm. In the minimization, seven torsion angles are used for both strands. All other angles and lengths shown in Fig. 1 are kept constant during the simulations. In more detailed calculations, variations of T8 must also be taken into account, but in the current study we fix T8 in order to reduce computational time. More precisely, the sugar ring geometry and the bases (A. T, C and G) remain undistorted. Moves of the O–P–O and P–O–C valence angles are allowed to be as general as possible (0–180°) in order to give the molecule maximum freedom to reach the minimum-energy state.

The prediction of the step for both angle and torsion is based on experience to avoid very large or small conformation changes. The range for the angles is defined to give maximum rotation in space, whereas the range for the torsion angles is defined from the file 166D.pdb and is based on the type of DNA that is minimized.

Force field

The molecular-energy minimization scheme is developed with a force field suitable for fast and accurate predictions of the energy of the DNA structure treating the atoms of the same type identically. It is assumed that the potential energy of a molecule with arbitrary geometry is expressed

Table 2 The range of all randomly generated torsion angles in degrees based on pdb file 166D.pdb. Again *a* and *b* are integer numbers corresponding to the base counting with *b* equal to *a*+1

Symbol	Torsion angles	Strand	Range (strand A)	Range (strand B)
T1	C(a)–C(a)–O(a)–C(a)	A, B	110.0° ↔ 130.0°	110.0° ↔ 130.0°
T2	O(a)–C(a)–C(a)–O(a)	A, B	−50.0° ↔ −90.0°	−50.0° ↔ −100.0°
			150.0° ↔ 160.0°	
			−140.0° ↔ −160.0°	
T3	P(a)–O(a)–C(a)–C(a)	A, B	140.0° ↔ 180.0°	140.0° ↔ 180.0°
			−150.0° ↔ −180.0°	−150.0° ↔ −180.0°
T4	O(a)—P(b)–O(b)–C(b)	A, B	0.0 ↔ 360.0°	−50.0° ↔ −80.0°
T5	C(a)–O(a)—P(b)–O(b)	A, B	−60.0° ↔ −180.0°	−90.0° ↔ −170.0°
				160.0° ↔ 162.0°
T6	C(a)–C(a)–O(a)—P(b)	A, B	60.0° ↔ 120.0°	60.0° ↔ 150.0°
T7	C(b)–C(b)–C(b)–O(b)	A, B	60.0° ↔ 110.0°	60.0° ↔ 110.0°

Table 3 The energy of real double DNA strands calculated by the force field used in the current algorithm. The equilibrium average energy values for each structure are recorded

DNA sequences	Average energies (arbitrary units)			
	Total	VDW	Q	ANG
cg, gc	3.362	0.768	0.751	1.853
aa, tt	2.482	0.882	0.733	0.867
at, ta	2.534	0.990	0.800	0.745
ag, ga, tc, ct	2.176	0.800	0.741	0.635
aatt, ttaa	7.089	2.727	2.215	2.712
attcg, taagc	10.027	3.141	2.957	3.930
gcgaattc, cgcttaag	17.182	5.461	5.172	6.522

as a superposition of valence (or bonded) interactions (E_{val}) that depend on the specific connections (bonds) of the structure and non-bonded interactions (E_{nb}) that depend only on the distance between the atoms. The valence interactions consist of the bond-angle bend (E_{ANG}, three-body), whereas the non-bonded interactions contribute the van der Waals or dispersion potential (E_{VDW}), the Coulombic (E_{Q}), and the hydrogen bond (E_{HB}) potentials.

$$E_{\mathrm{total}} = E_{\mathrm{val}} + E_{\mathrm{nb}} = 10^{-7}E_{\mathrm{ANG}} + 10^{-4}E_{\mathrm{VDW}} + 10^{-5}E_{\mathrm{Q}} + E_{\mathrm{HB}} \quad (1)$$

In the current version of the model, the atoms are allowed to move by three- and four-body interactions, which are formed by two and three bonds, respectively. In this simulator the force field does not need to be adjusted for each base because it is independent of the length of the DNA strand. To determine the accuracy and the global energy minimum (defined at 1 arbitrary unit), the energy of known DNA structures is calculated. Table 3 shows these structures and their energies. The constants before each potential term (in Eq. 1) have been computed by calculating the energy from known B-DNA structures found in the Brookhaven database. They allow the different potentials to be scaled correctly.

There are two types of three-body interaction taken into account in the force field. The first takes place between two oxygen atoms and the phosphate atom and the second between the oxygen, the phosphate and the carbon atom of two neighboring bases, which are allowed to rotate in three-dimensional space. The energy of every three-body interaction is calculated from the formula:

$$E_{\mathrm{ANG}} = \sum_{n=0}^{n=\mathrm{totalbases}} K \sin(\vartheta - \vartheta_0)^2 \quad (2)$$

where under the usual approximation $(\vartheta - \vartheta_0) \ll 1 \Rightarrow \sin(\vartheta - \vartheta_0) \cong (\vartheta - \vartheta_0)$ the E_{ANG} becomes:

$$E_{\mathrm{ANG}} = \sum_{n=0}^{n=\mathrm{totalbases}} K(\vartheta - \vartheta_0)^2,$$

$$K = 100\,(\mathrm{kcal\ mol^{-1}})/\mathrm{degrees}^2 \quad (3)$$

In Eqs. (2) and (3), K is the force constant for all angle bend interactions; ϑ, in degrees, is the new angle between the two bonds and ϑ_0 is the equilibrium angle. The equilibrium angle ϑ_0 was found from a high resolution X-ray file, from the Brookhaven database [41] and is 104° for the first and 121° for the second interaction. However, both strands in the DNA molecule are also allowed to move by four-body interactions; the energies of these interactions were not included in the force field. This is due to the large number of different torsion angles found in the real DNA structures and to the difficulty and uncertainty in predicting the constants for the various torsion interactions.

In the non-bonded energies, the hydrogen bond is incorporated due to the importance of structural changes between the opposite base pairs of different chains. Some force fields replace the Lennard-Jones 6–12 term between hydrogen-bonding atoms by an explicit hydrogen-bonding term, which is often described using a 10–12 Lennard-Jones potential. Other force fields, such as Dreiding, improve the accuracy that the geometry of the hydrogen system adopts by using the 10–12 term and include the angle between the hydrogen donor (D), the hydrogen (H), and the hydrogen acceptor (A). In this simulator, both previous potential terms were tested, but both of them force the geometry of the two DNA strands to unfold, pushing the opposite bases as far apart as possible. For this reason, a new hydrogen bond potential is used to describe the hydrogen bonding. It contains a bonding factor to bring the bases closer and an angle factor to make them as planar as possible.

$$E_{\mathrm{HB}} = \sum_{i=0}^{i=\mathrm{totalhydrogenbonds}} \mathrm{abs}[100(R_{\mathrm{hb}} - R_{\mathrm{DA}})^2 + (180 - \mathrm{abs}(\theta))] \quad (4)$$

In Eq. (4), R_{hb} is the equilibrium distance (2.75 Å) between the donor and acceptor, R_{DA} is the new distance between the donor and acceptor, θ is the angle between donor, hydrogen and acceptor atom and abs means the absolute value. The value of 100 is used to give extra weighting to the bonding rather than to the angle factor.

In a DNA strand, the phosphate groups in the backbone are heavily charged, making the electrostatic interactions significant in the current model. It is not possible to avoid the electrostatic potential due to the nature of the charges, which are balanced by the presence of magnesium ions associated with the DNA. However in this simulation there are no heavy metals present, so it would be of interest to investigate how they would effect the current DNA models. The electrostatic interactions are calculated as:

$$E_{\mathrm{Q}} = \sum_{i=0}^{i=\mathrm{totalinteraction}} [(322.0637 Q_i Q_j)/(\varepsilon R_{ij})] \quad (5)$$

where Q_i and Q_j are charges in electron units, R_{ij} is the distance between charges i and j in Å, ε is the dielectric constant (ε=1 for vacuum) and the factor 332.0637 converts E_{Q} to kcal mol^{-1}. The charges are positioned on each individual atom and the net charge on each atom for the

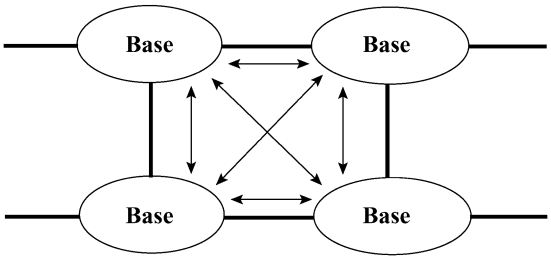

Fig. 3 The Lennard-Jones interactions between two adjacent base pairs

Table 4 The van der Waals parameters used to calculate the potential energy

Atom	r_o (Å)	C (kcal mol^{-1})
H	3.195	0.0152
C	3.8983	0.0951
N	3.6621	0.0774
O	3.4046	0.0957
P	4.1500	0.3200

nucleic acids in units of electron charge are taken from [52]. In this simulation, only interactions between atoms of opposite or neighboring bases and not within the same bases are included in order to make the algorithm computationally simpler.

The third non-bonded interaction is the van der Waals interaction (see Fig. 3), which is described by the Lennard-Jones expression. This expression describes the potential energy of two non-bonded molecules or atoms. For short distances, the nuclear and electronic repulsions and the rising kinetic energy begin to dominate the attractive forces. The repulsions increase steeply with decreasing separation in a way that can be deduced only by very extensive, complicated molecular-structure calculations. The sum of the repulsive and the attractive interactions is here approximated by the Lennard-Jones (n,6) potential:

$$V = \frac{C_n}{r^n} - \frac{C_6}{r^6} \tag{6}$$

As usual, it is convenient to select n=12 for mathematical reasons. The new equation can be written as:

$$V_{\mathrm{vdw}} = \sum_{n=0}^{n=\mathrm{totalbases}} C\left\{\left(\frac{r_0}{r_n}\right)^{12} - \left(\frac{r_0}{r_n}\right)^{6}\right\} \tag{7}$$

where C is a constant parameter, r_0 is the equilibrium distance between the two atoms or molecules, and r_n is the new distance between them, Table 4. To truncate large numbers of Lennard-Jones calculations, the energy between neighboring base pairs is calculated and a cut-off distance of 10 Å was used, to avoid calculations of the energy for distances larger than the cut-off.

Monte Carlo annealing simulation

The current sequential minimization algorithm combines Metropolis Monte Carlo sampling with a simulated annealing procedure in order to identify the global energy minimum of the linear B-DNA strands (in arbitrary energy units). The starting conformation exists in vacuum and during the minimization only four types of interaction are taken into account, one bonded and three non-bonded, as explained in the section Force field. During each step of the minimization, the algorithm searches for the minimum energy structure at a given temperature. In order to minimize the energy, at each temperature, a large enough number of N attempts is undertaken, called a "stage" or "step". The total number of these stages for each temperature is constant and is called one "interval". Each interval does not depend on the size of the sequence and is normally taken as N=400. Each minimization usually finishes after 120 intervals (or temperature gradings) while during the minimization processes only approximately 50% of the structures generated are accepted.

In each step the strand is allowed to move randomly in the configuration space by changing angles or torsions of the minimized base-base pair in the range of ±5° and ±1°, respectively. During this process, the new structure and the values of the energies, angles and torsions of the new strands are saved at every 20 stages, while during each step the algorithm keeps track of two particular strand conformations, the current and the trial. The most favorable conformation is then calculated and saved. The lower-energy structures are automatically accepted while those of higher energy are accepted on the basis of the Boltzman factor of the energy increase. The new conformations are accepted with probability p given by:

$$p = \begin{cases} 1 & \text{if} \Delta E < 0 \\ \mathrm{e}^{-(\Delta E \times 10)/kT)} & \text{if} \Delta E > 0. \end{cases} \tag{8}$$

The new structures are accepted if ΔE<0 or if the random number between 1 and 0 is smaller than the quality $\mathrm{e}^{-(\Delta E \times 10)/kT)}$, where T is the current temperature of the system, k the Boltzman factor and ΔE is the energy difference E_{new}–E_{old}. The value 10 in the expression guarantees that 40%–60% of the moves among all trials are accepted. The constant k (Boltzman factor) is equal to 1 because it is absorbed in the temperature T in order to simplify the calculations. This value works only with the current energy scale and changing the scale of the force field would change the acceptance ratio.

The system is allowed only to approach an equilibrium distribution at a given starting temperature T_0=200, which is chosen by experiment. This temperature is reduced by an accelerated cooling procedure that is lowered by a factor of 97% during the 120 intervals. The cooling procedure is only allowed to continue if the energy E_{old} is larger or equal to 1 unit (see Appendix B, Table 8) or the temperature T is lower than 1. Otherwise, the minimization process is terminated, saving the energy, torsion, angle and bond values and the coordinates of the equilibrated structure at this point.

Double stranded B-DNA

During the minimization, double stranded DNA chains were used in order to achieve the doubly folded B-DNA structures. The two chains are connected originally at their first bases by hydrogen bonds, which remain unchanged during the minimization, while the neighboring hydrogen bonds are allowed to move. However, the phosphate–oxygen bonds between the bases in the same strand remain constant (1.6 Å) allowing the two neighboring bases that are minimized to equilibrate, thus avoiding the bases escaping into space, or into the solution if the DNA is surrounded by a medium.

This type of minimization is always local and is between neighboring base pairs. This procedure is an alternative to the full-length minimization introduced previously [28] and uses global sequence minimization in order to achieve the energy minimum. Every minimized base pair is chosen sequentially by taking the last minimized base pair and the next one in the chain.

Figure 4 shows a real double stranded B-DNA structure, with the same nucleotide sequence as the Dickerson dodecamer (cgcg*aattcgcg*). This structure was minimized starting from two completely linear chains with a very high energy. During the whole process, every structure generated was saved after each base-pair minimization was completed successfully. In the initial structure, all the hydrogen bonds were completely disconnected (see Fig. 2) except the ones in the first base pair. The starting hydrogen potential was made stronger than the other potentials, forcing the two chains to come close enough to form the hydrogen bond interactions. Making this potential weaker will result in bringing the chains together more slowly, making the minimization process more conformation-search demanding, but probably giving slightly more accurate results. Several DNA sequences were generated giving similar results, including homo- or heterogeneous and real or random base sequences.

The temperature of the system during each local minimization was decreased after each interval, allowing the structure to relax to the equilibrium position, thus avoiding trapping in a local minimum. In total, 20 different small or larger strands were minimized, composed of 2 to 40 base pairs. It would be possible to minimize much larger DNA segments (due to the fast algorithm) but it was not the purpose of this research at this stage. Again, in order to compare the current results, the B-DNA structure from the166D.pdb file was compared with the same segment generated with the same sequence as the Dickerson dodecamer, by superimposing them in order to calculate the RMS deviation (Figs. 4 and Fig. 5). The results show that the structure generated has RMS 1.9 Å compared with the 1.6 Å from the "full strand minimization method". The new structure generated is very similar to the real one. This is due to the limited range of the force field used that does not take into account wider range interactions inside the double helix.

This method gives relatively accurate predictions of the Dickerson dodecamer, including the angles (see Fig. 6) the torsions (see Fig. 7) and the hydrogen bonds (see Fig. 8) and thus can predict larger homogeneous or heterogeneous segments of B-DNA chains. The new structures generated have fluctuations between the equilibrium hydrogen bond angles (180°) and distances (2.75 Å) which are in the range 150–180° and 2.75±1 Å respectively (see Fig. 8) and the bond angles are at the range of A_1=130±10°and A_2=150±5° (see Fig. 6). All the

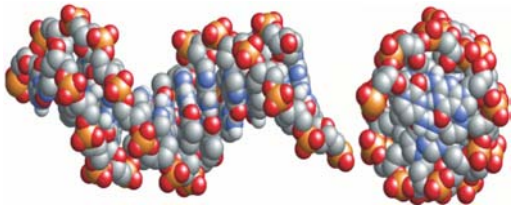

Fig. 4 The final minimized DNA segment of the Dickerson dodecamer sequence represented by ball drawings and viewed (*left*) perpendicular and (*right*) down the helical axis

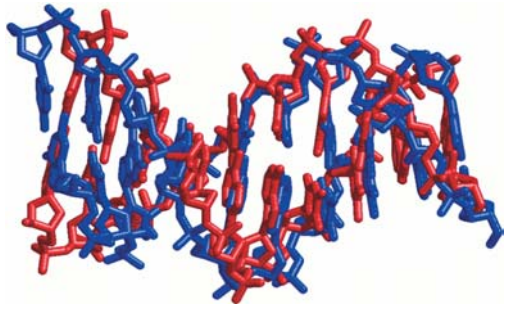

Fig. 5 The superposition of the Dickerson dodecamer sequence minimized using the local Monte Carlo minimization algorithm and the one found in Brookhaven database. The *red* is the minimized one and the *blue* is the real taken from the 166D.pdb" file

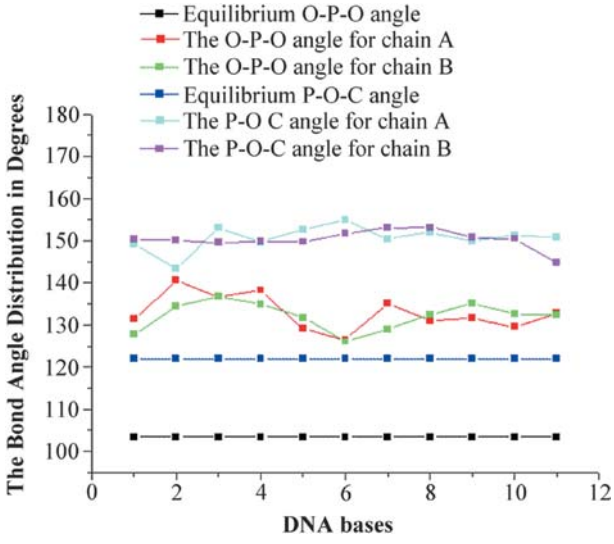

Fig. 6 The values of the angles A1=O–P–O and A2=P–O–C (in degrees) from sequence "cgcgaattcg". For comparison the values of the equilibrium angle are depicted

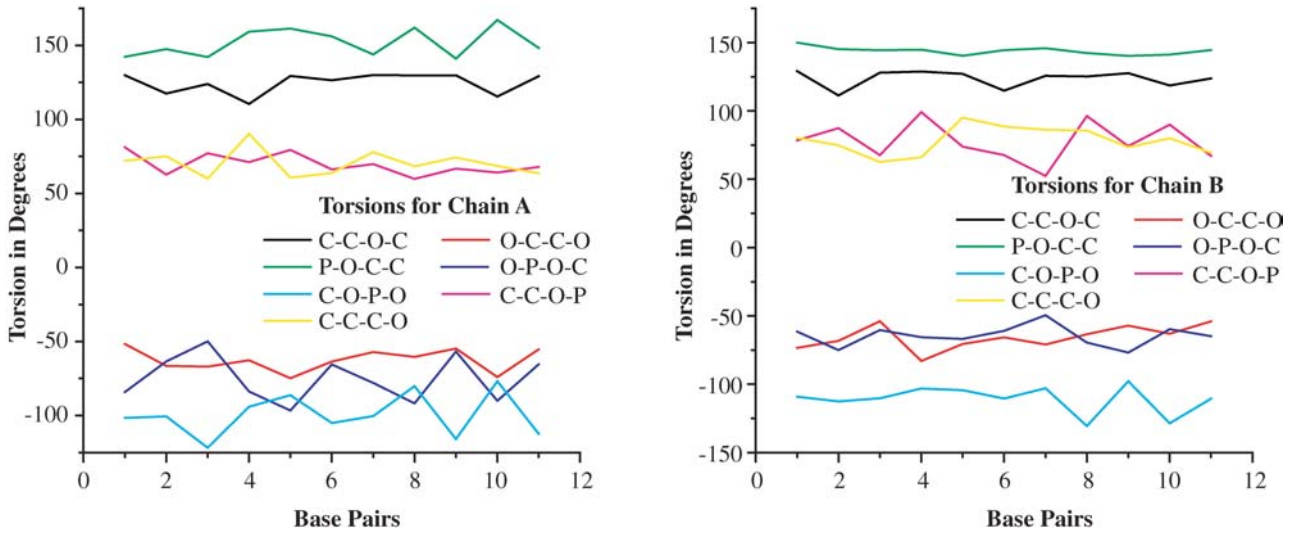

Fig. 7 The distribution of the seven different torsion angles (in degrees), moving across the sequence. *Left* for chain A and *right* for chain B. The angles are taken from the final generated DNA segments after the local sequential minimization

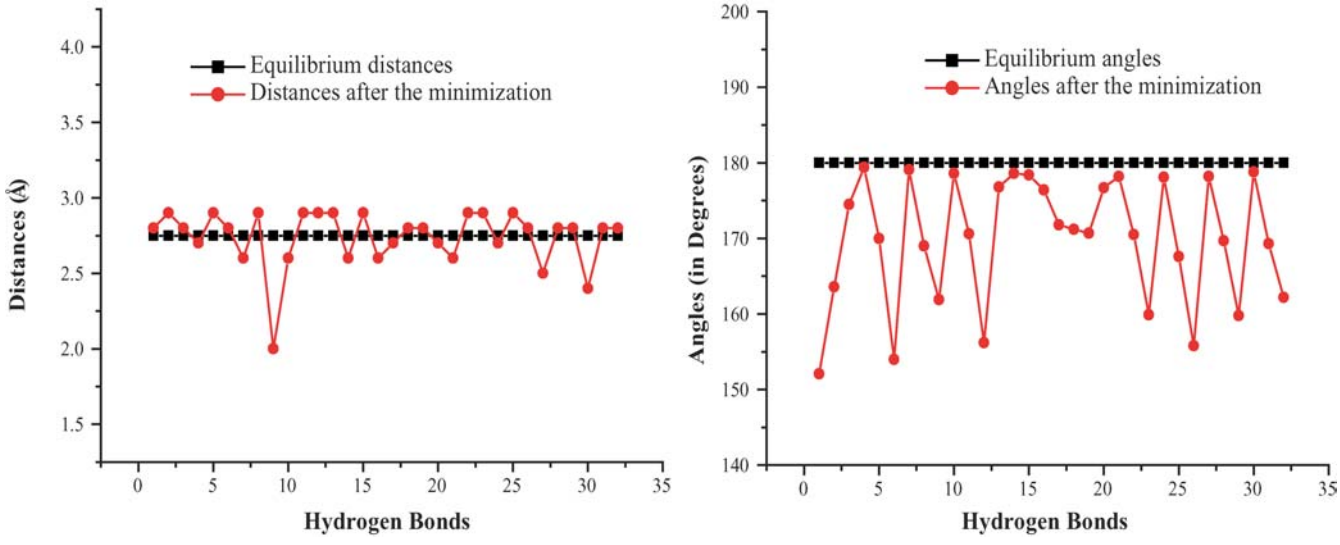

Fig. 8 The distribution of the hydrogen bond distances *left* and the hydrogen bond angles *right*. They are taken from the minimized Dickerson dodecamer sequence. For comparison the values of the equilibrium angle are displayed

Table 5 The average helical twist of both chains, for the structures Dickerson dodecamer, dC_{40}, and tgagaacgaa aagctgcgcc gggaggttga agaactgcgg

Structure	cgcgaattcgcg	dC_{40}	tgagaacgaa aagctgcgcc gggaggttga agaactgcgg
Strand A	36.6°±1°	33.9°±1°	37.2°±1°
Strand B	39.6°±1°	41.1°±1°	40.1°±1°

torsion angles are found in the angle space that the algorithm generated (see Table 2). In order to find out how close the generated structures are to the real crystallographic B-DNA, we test the helical twist values of the generated Dickerson dodecamer against the helical twist values measured by X-ray crystallography for the same structure (see Table 5). In the same table the helical twist of a few other generated sequences is also reported for comparison. The structures generated from the runs using dC40 and "tgagaacgaaagctgcgcgggaggttgaagaactgcgg" are shown as Fig. 9. It was found that the real structures were very close to the numerically predicted ones with small differences between the neighboring chains. A factor preventing exact overlap of the two structures is the organic molecule intercalated into the structure in the pdb file. This is not taken into account in the simulation, al-

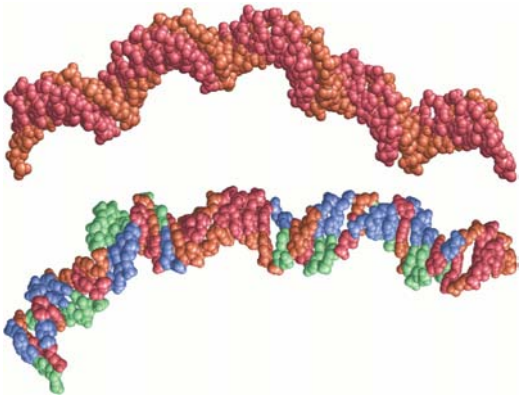

Fig. 9 Two of the final minimized DNA segments generated by this algorithm. The top is the homogeneous dC_{40} base pairs while the bottom sequence is "tgagaacga aagctgcgc gggaggttga agaactgcgg"

though organic molecules (drugs) in the minor groove typically have little effect on the helical conformation.

The most important contribution to the force field iteration during the first intervals are the hydrogen bond interactions, whereas at the end of the minimization the other energy components become more important. The electrostatic interactions contribute less than the other forces, so they are not very important during the minimization. In all cases, the energy is reduced rapidly in the first intervals and then falls logarithmically with large fluctuations during the process of avoiding trapping in a local minimum. In the current simulations, larger organic molecules attached to the double chain were not taken into account. To include such molecules, one must introduce additional degrees of freedom to account for the various components of the attached molecules. Alternatively, solvents could be added into the system easily by changing the parameters of the force field.

Differences and similarities between local sequential and full strand minimization procedures

Using local minimization techniques rather than full-length techniques has many advantages and disadvantages. From the local minimization, the results show that it is faster and easiest to find out the correct conformations required for the prediction of the folded DNA structure. A smaller number of trials is also used to achieve folding and thus less time is needed for the total minimization to be completed. B-DNA sequences containing 40 base pairs need approximately 4 h using this method and 2.5 weeks using the full-length minimization method, which is global. It is obvious that the local minimization is more efficient than the global.

As was reported in the previous section, using the local minimization method, the generated DNA strands are not as accurate as expected because of the very restricted force field used for calculating the energy. This force field does not take into account the wider forces existing in these small segments of DNA. In a global minimization one has to use a wider range force field to take into account all the changes in the structure. To eliminate this problem in a local minimization algorithm, it is possible to increase the range of the force field in the molecule until it gives more accurate results. It is possible to take into account wider changes of the DNA strands, calculating the energy forces from the already predicted DNA segments.

Here the algorithm uses a very narrow force field that calculates the energy between two successive base pairs without taking into account any structural information of the neighboring base pairs. In the near future, the authors plan to increase the range of the force field in the algorithm and to include further base pairs. However, the results are accurate enough even at this level of approximation.

Conclusions

Using the proposed sequential minimization approach, our aim was to understand and investigate how DNA actually folds in cells. Many biologists believe that DNA folds by adding each base to an already built (linear or folded) single or unfolded DNA strand, allowing each newly added base to fold independently from the rest. This investigation tries to apply a similar approach by using a linear starting strand and folding each base sequentially until reaching the last one. The results show that it is possible to fold accurately and rapidly large DNA segments, by using the Monte Carlo annealing method and a force field with only a minimum number of energy terms. Using a local minimization method gives the advantage of high-speed performance and results that are nearly as accurate as the global minimization methods.

With the sequential minimization approach, in order to achieve new structural conformations, each chain was moved independently from the others, using two angles and seven torsions for each strand, defined in Tables 1 and 2. To bring the two strands together a simple type of hydrogen bond potential was applied between the two strands. In the algorithm there are no bond interactions because it is assumed that the bond between the bases for each strand remain constant (1.6 Å) and the total energy is independent of the number of base pairs contained in the DNA chain.

Only double stranded homogeneous or heterogeneous real or double B-DNA molecules were calculated. However, A- and Z- type DNA molecules can be minimized by changing the restriction table for the torsion angles. The results show that the linear initial conformations could be folded quickly into recognizable helical shaped DNA fragments. For the minimization, the Monte Carlo annealing method was used and the force field was built by taking the parameters from the Dreiding force field. During the construction of the algorithm many assumptions were taken in order to reduce the time of the minimization process and to make the model simpler to handle. These assumptions may be relaxed in more de-

tailed models, which will allow the exploration of larger conformational space and produce more accurate results. Additional angle and torsion rotations could be added to the system allowing the system to equilibrate better. Other force fields, such as the Amber and Charmm force fields should also be tested and compared both with the Dreiding force field results and with crystallographic data.

The average helical twist was calculated for several real and theoretical or random structures (see Table 5) in order to determine how close these structures are to the real structures and to the numerical values predicted by the local sequential minimization algorithm. The results show that the neighboring chains are not identical. This conclusion together with the much faster performance of the sequential algorithm shows a large improvement over the full strand minimization algorithm. Further studies could investigate the folding process in a solvent (such as water) with addition of magnesium atoms and of organic molecules attached to the DNA chain.

A question, which was posed at the beginning of this study, was whether it is possible to determine quantitative structural/conformational differences between coding, non-coding, homogeneous and random DNA sequences, i.e. to determine whether or not the structural characteristics relate to the functional role of the sequence. The current algorithm has tested chains of length up to 40 base pairs. For these chain sizes, and although the chains differ in their detailed structures, the overall conformational characteristics, such as bond angles or helical twists, do not present any consistent features which could allow us to determine the functional role (coding, non-coding) of each chain. To test this hypothesis, longer chains, of sizes up to few thousand base pairs, need to be tested in detail and this will be possible with (a) further amelioration of the algorithm performance and with (b)incorporation of the "full strand minimization" scheme in critical steps of the minimization process.

Appendix A

Initial structures are shown in Tables 6 and 7.

Table 6 The values for the bonds, and angles and torsion angles for the B-DNA strand from the pdb file 166D.pdb

Pair sequence	Bonds (Å)	Angles (degrees)			
	Strand B	Strand A		Strand B	
		A1	A2	A1	A2
cg	1.556	99.431	119.633	100.851	121.693
gc	1.608	91.351	121.134	95.327	119.034
ag	1.624	93.019	122.008	95.400	122.455
aa	1.607	98.744	118.724	99.589	125.903
ta	1.593	96.731	118.459	99.808	120.381
tt	1.609	100.776	117.534	100.214	116.611
ct	1.569	103.802	125.117	103.272	125.592
gc	1.621	97.097	126.144	102.447	123.011
cg	1.553	96.549	117.381	95.536	108.223
gc	1.663	100.903	122.330	100.351	127.826

Table 7 The values for the torsion angles for the B-DNA strand from the pdb file 166D.pdb

Pair sequence	Torsions (degrees)													
	Strand A							Strand B						
	T1	T2	T3	T4	T5	T6	T7	T1	T2	T3	T4	T5	T6	T7
Cg	116.8	−77.9	173.9	−55.5	−119.6	78.4	89.9	126.0	−70.4	−157.6	−106.2	−149.2	90.9	81.7
Gc	114.8	−96.4	163.8	−65.5	160.3	149.2	80.1	106.4	−62.4	150.8	−64.3	−101.2	92.7	74.0
Ag	115.2	−65.3	175.8	−73.0	−117.2	88.2	86.4	110.5	−78.5	149.9	−30.9	−122.7	88.2	106.5
Aa	126.9	−77.0	179.6	−59.2	−111.4	83.1	76.0	114.8	−88.4	171.6	−44.9	−101.1	81.6	74.5
Ta	125.7	−53.5	173.9	−69.4	−102.4	91.2	75.0	124.7	−56.2	168.9	−61.1	−101.9	72.0	71.8
tt	120.3	−58.9	164.8	−69.9	−112.8	83.7	74.7	124.9	−57.7	170.2	−65.6	−105.6	78.4	78.5
ct	117.7	−68.6	179.7	−60.4	−98.1	62.4	74.4	121.0	−54.1	173.9	−72.2	−130.4	82.2	72.0
gc	108.7	−65.0	−162.3	−67.6	−106.9	68.8	71.6	113.3	−60.6	168.6	−63.7	−137.1	85.4	81.4
cg	114.9	−56.8	145.1	−54.6	−152.7	102.1	87.2	108.9	−60.8	177.7	−61.3	−179.0	92.2	68.8
gc	121.7	−50.0	173.3	−73.7	−95.8	70.9	80.0	107.7	156.9	−161.1	52.7	−66.7	113.3	71.8

Appendix B

The average starting and final energy values of the minimized double B-DNA strands are shown in Table 8.

Table 8 The average starting and final energy values of the minimized double B-DNA strands. The energy values are composed from the addition of all initials and final potentials of each base which composed the DNA fragments

Type	0–10–20	Average initial energy (arbitrary units)	Average final energy (arbitrary units)				
		Total	Total	VDW	ANG	HB	Q
Homogeneous	cc	30312.844	54.902	1.800	15.001	38.100	0.001
Homogeneous	aa	20274.687	24.309	0.901	16.202	7.201	0.005
Real	cg	32007.590	57.125	1.101	17.511	38.503	0.010
Real	ta	20274.714	20.903	0.002	13.300	7.600	0.001
Real	ag	32007.602	59.303	0.700	13.602	45.001	0.000
Real	aatt	321644.854	79.477	4.550	44.203	30.723	0.001
Homogeneous	ggggg	32007.615	240.204	3.001	56.301	180.900	0.002
Homogeneous	ccccc	1040603.302	215.518	3.703	55.512	156.301	0.002
Dyadic	cgcgcgcgcg	10401088.080	509.624	12.300	129.420	367.901	0.003
Dyadic	cgcgcgcgcg cgcgc	37478446.240	791.004	29.001	210.802	551.200	0.001
Real	cgcgaattcgcg	17057325.350	504.624	16.001	157.720	330.901	0.002
Real	cgaattcg	4419101.968	272.533	12.902	97.716	161.912	0.003
Homogeneous	ccccccccccc ccccccccccc	91485283.530	1064.640	15.989	297.742	750.901	0.008
Homogeneous	ccccccccccc ccccccccccc ccccc	181944801.890	1343.527	31.302	393.523	918.701	0.001
Homogeneous	ccccccccccc ccccccccccc cccccccccc	317932334.624	1588.74	26.123	421.112	1141.502	0.003
Homogeneous	ccccccccccc ccccccccccc cccccccccc ccccccccccc	762733113.839	2203.377	28.302	583.521	1591.543	0.011
Real-coding	tgagaacgaa aagctgcgcc	84181177.320	867.517	32.412	287.003	548.101	0.001
Random	tcaaatgggc cccgaaatcg	87404653.640	920.119	20.101	288.910	611.104	0.004
Real-coding	actgataccg gggttgctga	78174923.939	930.118	41.902	281.412	606.802	0.002
Real-coding	tgagaacgaa aagctgcgcc gggaggttga agaactgcgg	669256988.142	1780.151	59.734	575.710	1144.705	0.002

References

1. Tidor B, Irikura KK, Brooks BR, Karplus M (1983) J Biomol Struct Dyn 1:231–252
2. Levitt M (1983) Cold Spring Harbor Symp Quant Biol 45:251–262
3. Rappe AK, Cassewit CJ (1997) Molecular mechanics across chemistry. University Science Books
4. Swamminathan S, Ravishaanker G, Beveridge DL (1991) J Am Chem Soc 113:5027-5040
5. Drukker K, Schatz GC (2000) J Phys Chem B 26:6108–6111
6. Bruant N, Flatters D, Lavery R, Genest D (1999) Biophys J 5:2366–2376
7. Jian H, Vologodski AV (1997) J Comput Phys 136:168–179
8. (a) Lavery R, Lebroun A (1999) Genetica 106:1–2; (b) Lavery R, Lebroun A (1999) Genetica 106:75–84
9. Olson WK, Zhurkin VB (2000) Curr Opin Struct Biol 3:286–297
10. Vlahovicek K, Pongor S (2000) Bioinformatics 11:1044–1045
11. Treger M, Westhof E (1987) J Mol Graphics 5:178–183
12. Srinivasan AR, Olson WK (1987) J Biomol Struct Dyn 4:895–938
13. Schlick T (1995) Curr Opin Struct Biol 5:245–262
14. MacKerrell Jr AD, Banavali N, Foloppe N (2000) Biopolymers 56:257–265
15. Schneider B, Neidle S, Berman HM (1997) Biopolymers 42:113-124
16. Sherer EC, Harris SA, Soliva R, Orozco M, Laughton CA (1999) J Am Chem Soc 121:5981–5991
17. Gabb HA, Prevost C, Bertucat G, Robert CH, Lavery R (1997) J Comput Chem 16:2001-2011
18. Duan Y, Wilkosz P, Crowley M, Rosenberg JM (1997) J Mol Biol 4:553–572
19. Ayadi L, Coulombeau C, Lavery R (1999) Biophys J 6:3218–3226
20. Lafontaine L, Lavery R (2000) Biophys J 2:680–685
21. Westcott TP, Tobias I, Olson WK (1997) J Chem Phys 10:3967–3980
22. Ramachandran G, Scchlick T (1995) Phys Rev E 6:6188–6203
23. Schlick T, Olson WK (1992) J Mol Biol 223:1089–1119
24. Schlick T, Olson WK (1992) Science 257:1110–1115
25. Ezaz-Nikpay K, Verdine GL (1992) J Am Chem Soc 114:6562–6563
26. Tereshko V, Minasov G, Egli M (1999) J Am Chem Soc 121:470–471
27. Tereshko V, Minasov G, Egli M (1999) J Am Chem Soc 121:6970–6970
28. Sfyrakis K, Provata A, Povey DC, Howlin BJ (2003) Molecular Simulations 29:555–575
29. Creighton TE (1993) Structures and Molecular Properties, 2nd edn. Proteins, New York
30. Goodfellow JM, Moss DS (1992) Computer modeling of bimolecular processes. Ellis Horwood
31. Voet D, Voet JG (1995) Biochemistry, 2nd edn. Wiley, New York
32. Olson WK, Srinivasan AR, Hao HH, Nauss JL (1988) In: Olson WK, Sarma MH, Sarma RH, Sundaralingman M (eds) Structure and expression 3, DNA bending and curvature. Adenine Press, New York, pp 225–242
33. Schurr JM (1985) Biopolymers 24:1233–1246
34. Selepova P, Kypr J (1985) Biopolymers 24:867–882
35. Shimada J, Yamakawa H (1985) J Mol Biol 184:319–329
36. Tsuru H, Wadati M (1986) Biopolymers 25:2083–2096
37. Tan RK, Harvey SC (1989) J Mol Biol 205:573-591
38. Provata A, Tsakiroglou M, Almirantis Y (2000) Non Randomness and Non Linearities in DNA. Proceedings to 1st Interdisciplinary Symposium on Non-linear Problems, Athens
39. Almirantis Y, Provata A (1999) J Stat Phys 97:233–262
40. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical Recipes in Fortran. The art of scientific computing, 2nd edn. Cambridge University Press, Cambridge
41. Mayo SL, Olafson BD, Goddard WA (1990) J Phys Chem 94:8897–8909

42. Sayle R (1994) Rasmol home pagehttp://www.rasmol.com
43. Albert B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1994) Molecular biology of the cell, 3rd edn. Garland Publishing
44. Zimmerman SB (1982) Annu Rev Biochem 51:395–427
45. Nunn CM, Jenkins TC, Neidle SE (1994) Eur J Biochem 226:953–961
46. Denisov AY, Zamaratski EV, Maltseva TV, Sandstrom A, Bekiroglou S, Altmann K–H, Egli M, Chattopadhyaya J (1998) J Biomol Struct Dyn 3:547–568
47. Maltseva TV, Altmann K–H, Egli M, Chattopadhyaya J (1998) J Biomol Struct Dyn 3:569–578
48. Ikeda H, Fermandez R Wilk A, Barchi Jr JJ, Huang X, Marquez VE (1998) Nucleic Acid Res 9:2237–2244
49. Gelasco A, Lippard SJ (1998) Biochemistry 26:9230–9239
50. Protein Data Bank Contents Guide. Atomic co-ordinates entry format description,http://www.rcsb.org
51. IUPAC: International Union of Pure and Applied Chemistry,http://:www.chem.qmul.ac.uk/iupac
52. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) J Am Chem Soc 106:765–784