

# Quaternion modeling of the helical path for analysis of the shape of the DNA molecule

A.F. Muterko

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

The three-dimensional shape of a DNA molecule is a key property influencing its functional specificity and the nature of its molecular interactions. The characteristic shape into which a DNA molecule folds under certain conditions is a manifestation of its micromechanical and structural features, which are sequence-dependent. DNA shape-related properties can therefore be determined in a predictable manner. A number of models have been designed to describe intrinsic DNA curvature, incorporating a set of helical parameters which can be applied to operative three-dimensional reconstruction of the DNA structures. Alternatively, desired base pair parameters can be computed based on publicly available information about atomic DNA structures. Further, taking the base pairs as rigid bodies, their relative location in space can be estimated based on these parameters. Matrices are a common method to implement any rigid body transformations and are widely used in the modeling of DNA structures. Quaternions are the more straightforward and robust alternative for matrices. Unit quaternions can represent only a rotation, whereas dual quaternions combine rotation and translation into a single state. In the present guide, the algebra of unit and dual quaternions is applied for the first time to modeling of the DNA helical path, based on conformational parameters of the base pair steps. Although dual quaternions are preferable for modeling of DNA structure in detail, the use of unit quaternions is sufficient to predict the DNA trajectory and all calculations of DNA shape features. In order to analyze DNA shape and chain statistics, and predict the micromechanical properties of DNA molecules based on coordinates of the helical path, the widely used as well as original algorithms for computing DNA curvature, radius of gyration, persistence length and phasing of DNA bends are described. Taken together, these algorithms will be useful both in the *in silico* analysis of relatively short DNA fragments as well as in topological mapping of whole genomes.

Key words: quaternion modeling; DNA structure; curvature; radius of gyration; persistence length; Fourier transform; phasing; bending.

## Кватернионное моделирование траектории спирали для анализа формы молекулы ДНК

А.Ф. Мутерко

Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

Пространственная организация, форма молекулы ДНК являются ключевой характеристикой, определяющей ее функциональную специфичность и природу межмолекулярных взаимодействий. Специфическая форма, которую молекула ДНК принимает при определенных условиях, обусловлена ее микромеханическими и структурными особенностями, зависящими от последовательности нуклеотидов. Следовательно, отдельные характеристики формы ДНК могут быть прогнозированы. Предложен ряд моделей для описания внутренней кривизны ДНК, включающий набор геометрических параметров двойной спирали, применяемых при компьютерной реконструкции пространственных структур. С другой стороны, необходимые параметры пар оснований можно рассчитать исходя из общедоступной информации атомной структуры ДНК. Принимая пары оснований как твердые тела, их относительное расположение в пространстве можно оценить по полученным параметрам. Матрицы являются наиболее распространенным способом реализации преобразований твердого тела и широко используются в моделировании формы ДНК. Более простая и надежная альтернатива матрицам – кватернионы. Единичные кватернионы представляют только поворот, тогда как двойные кватернионы объединяют в себе и поворот, и смещение. В настоящем руководстве алгебра единичных и двойных кватернионов впервые применена для моделирования траектории молекулы ДНК исходя из конформационных параметров динуклеотидных шагов. Хотя использование двойных кватернионов оптимально для детального моделирования структуры, единичные кватернионы достаточны для прогнозирования траектории двойной спирали и последующих расчетов ее пространственных характеристик. Обсуждаются широко используемые, а также оригинальные алгоритмы вычисления кривизны, радиуса гирации, персистентной длины и фазирования статических изгибов для анализа формы молекулы, вычисления статистики полимерной цепи и прогнозирования микромеханических свойств на основе координат траектории ДНК-спирали. Приведенные алгоритмы будут полезны как в ходе *in silico* анализа относительно коротких фрагментов ДНК, так и в топологическом картировании полных геномов.

Ключевые слова: кватернионное моделирование; структура ДНК; кривизна; радиус гирации; персистентная длина; преобразование Фурье; фазирование; изгибаемость.

### HOW TO CITE THIS ARTICLE:

Muterko A.F. Quaternion modeling of the helical path for analysis of the shape of the DNA molecule. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2017;21(8):878-886. DOI 10.18699/VJ17.308

Received 12.09.2017  
Accepted for publication 29.11.2017  
© AUTHOR, 2017

The shape of DNA is sequence-dependent and determined by the intrinsic curvature and flexibility of the molecule, which is driven by external forces. DNA bending and curving play a crucial role in many important biological processes, including recombination, replication and the excision of damaged nucleotides. They are also important for transcriptional regulation of numerous prokaryotic and eukaryotic genes, as well as DNA-protein recognition and interactions. By now it is definitively known that intrinsic DNA curvature is a major determinant of nucleosome organization and positioning, facilitating chromosome folding and DNA packaging in the nucleoid, promoting the appropriate mode of supercoiling and protecting the prokaryotic genome from phage integration (see References in the Supplementary material)<sup>1</sup>.

Predictions of sequence-dependent mechanical properties of DNA are important for understanding many biological processes associated with the various DNA-protein interactions, including the phenomenon of “indirect readout”. The micromechanical characteristics of the DNA molecule can be described by a variety of models, including the remarkably effective worm-like chain (WLC) model for semiflexible polymers (Kratky, Porod, 1949). In this model, the bending stiffness of a molecule is described by its persistence length. However, the WLC model is valid only for isotropic, intrinsically straight, homogeneous polymers and cannot be applied in its original form to DNA molecules, which contain non-Gaussian deformations, such as the intrinsic bends (Schellman, Harvey, 1995; Rivetti et al., 1998).

The intrinsic curvature of DNA is mainly determined by the length and localization of adenine tracts (A-tract,  $A_nT_m$ ,  $n+m \geq 4$ ). The helical phasing of A-tracts and other bent DNA-related sequences for a long time formed the basis of the central concept of intrinsic DNA curvature. More than 30 years ago it was demonstrated how localization of local bends affects the global curvature of a DNA molecule (Wu, Crothers, 1984; Hagerman, 1985; Koo et al., 1986). In particular, it has been shown that macroscopic DNA curvature is strongly affected by the phasing of local bends. Since the bends produced by A-tracts have directional preference, their systematic alternation in phase with the helix screw add coherently and significantly increases the macroscopic DNA curvature, whilst for “straighter” DNA molecules the systematic bends are nearly exactly out of phase (Koo et al., 1986). The Discrete Fourier Transform (DFT) is used to detect the periodicity of any property along a sequence, through calculation of the Fourier transform power spectrum. However, this method is valid only for planar systematic bends, which are oriented in the same direction or, often, the bends, which are determined by repeats of a sequence motif. It is important therefore to predict the macroscopic DNA curvature based on the phase of local bends, regardless of their origin.

Since DNA conformation has been shown to be sequence dependent, this has led to the development of series of models for *in silico* prediction of DNA shape based on sequence-dependent parameters of base pair steps in di-, three- and tetranucleotide contexts. These base pair steps (or often wedge) parameters were chosen through many different methods, such as circularization, gel-mobility, DNase I digestion and

nucleosome positioning data, X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, conformational energy minimization and computer simulation (see References in Supplementary material). Nonetheless, there is no complete consensus on sequence-dependent parameters that are optimal for accurate prediction of DNA shape.

Base pair orientation in the DNA helical structures is described by a set of intra and inters parameters, each including of three rotational and three translational parameters. Internal parameters determine the conformation of bases in the local coordinate frame of a base pair, hence, they do not affect the trajectory of the DNA helical path. The set of inter base pair geometrical parameters consists of the three rotations by *tilt* ( $\tau$ ), *roll* ( $\rho$ ) and *twist* ( $\Omega$ ), and three translations *shift* ( $Dx$ ), *slide* ( $Dy$ ) and *rise* ( $Dz$ ) about  $X$ ,  $Y$  and  $Z$  axes, respectively (Fig. 1), in the standard coordinate frame approved by the Cambridge convention (Dickerson, 1989). In this manner, the inter base pair parameters are used to construct low resolution (coarse-grained) three-dimensional DNA structures. Such translational parameters as *shift* and *slide* define the local displacement between adjacent base pairs ( $\sqrt{\text{slide}^2 + \text{shift}^2}$ ). The magnitude of *twist* characterizes the degree of torsional twisting of the double helix. Only the parameters of *roll* and *tilt* make the main contribution to bending of the DNA molecule, herewith the total local bend of each dinucleotide step can be represented as  $\sqrt{\text{tilt}^2 + \text{roll}^2}$ . Thus, the DNA trajectory, modeled in this way is a spatial curve, the shape of which is mainly determined by the accumulation of local bends (of specific magnitude and direction), local displacements, translation by *rise* and rotation by *twist*. In a recent study, the genome-scale computational analysis of DNA curvature based on three-dimensional trajectories of DNA molecules calculated from a set of inter base pair conformational parameters was carried out for *Arabidopsis* and rice (Masoudi-Nejad et al., 2011).

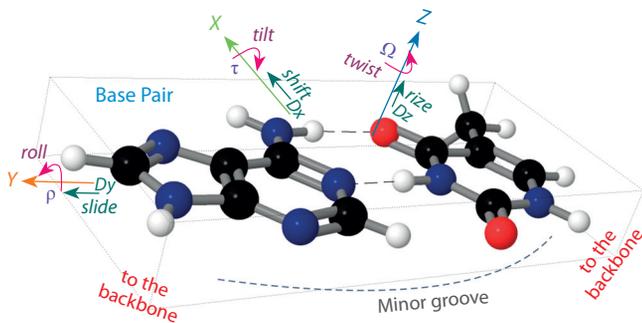
The rigid transform includes six degrees of freedom, consisting of three translational and three rotational components. The most popular method of storing and combining these transforms are matrices. Use of the matrix transformation technique is a standard method for calculation of the DNA path (De Santis et al., 1990; Bolshoy et al., 1991; Babcock et al., 1994; Liu, Beveridge, 2001). Quaternions are the more straightforward and robust alternative for matrices. They are more compact, reduce the volume of algebra and minimize computation. However, the unit quaternion can represent only a rotation without translation. Dual quaternions allow us to unify the translation and rotation into a single state.

In the present guide to DNA shape analysis the use of unit and dual quaternions to coarse-grained modeling of three-dimensional DNA structures is described in detail. Proven algorithms for efficient prediction and analysis of the shape of DNA molecules are described as follows. The discussed methods are fast and robust making them appropriate for operative prediction of curvature- and bendability-dependent DNA shape at any scale.

## 1. Obtaining base pair step parameters

The calculation of the DNA helix trajectory is performed based on the average values of local helical parameters. These parameters can be obtained from various sources, such as scientific papers, the base pair parameters database (DiProDB, [http://](http://www.bionet.nsc.ru/vogis/download/pict-2017-21/appx11.pdf)

<sup>1</sup> Supplementary materials are available in the online version of the paper: <http://www.bionet.nsc.ru/vogis/download/pict-2017-21/appx11.pdf>



**Fig. 1.** Base pair orientation parameters, following the Cambridge convention. Three rotations (*twist*, *roll*, *tilt*) and three translations (*rise*, *slide*, *shift*) of a base pair about the Z, Y and X axes are indicated.

diprod.b.leibniz-fl.de) or alternatively, they can be deduced from three-dimensional DNA structures. This last method is optimal, since it allows us to choose the most suitable or desirable to resolve specific problems the DNA structures (with specific nucleotide content or conformation, obtained in different conditions etc.) to deduce base pairs parameters. This significantly increases the flexibility of DNA molecular modeling and expands the possibilities for further investigations. In particular, specific sets of base pair parameters can be used for modeling and analysis of unusual DNA structures that are different from canonical B-DNA.

As the Protein Data Bank (PDB) is the largest current source of information about the three-dimensional structures of biological molecules (Berman et al., 2003), including nucleic acids, its will be enough to obtain primary data for further estimation of DNA base pair geometry. Currently, more than 1 700 DNA structures are publicly available in the PDB database. They are represented in different conformations obtained using a range of methods, such as X-ray crystallography (of various resolutions), solution NMR spectroscopy, neutron diffraction and others. However, many DNA molecules are chemically modified or represented in complex with ligands. It seems quite clear that DNA conformation is substantially affected by crystal packing, bases modification and interactions of the DNA with the other molecules or ions. Thus, the most suitable base pair conformational parameters for modeling of the DNA structures are provided by NMR in water only solutions. Previously it was shown, that distinct from the X-ray crystallographic analysis, helical parameters derived from NMR structures can correctly predict the curvature of DNA molecules (Gabrielian, Pongor, 1996). However, it should be noted that until the late 1990s the accuracy of the NMR method was not sufficient to evaluate small intrinsic bends of the DNA axis (Vermeulen et al., 2000). A full set of helical parameters from the selected DNA structures can be deduced using the Curves+ (Lavery et al., 2009) or 3DNA (Lu, Olson, 2003) programs.

As a rule, each DNA structure from PDB is represented by 5–10 conformers possessing the lowest energy. For this reason, base pair parameters deduced from individual structures should be averaged over all conformers, and the standard deviation for each parameter should be evaluated (further, the standard deviations can be used for estimation of the dynamic persistence length). To eliminate errors associated with “end

effects”, only the central base pairs should be used in computation. Since structural properties of dinucleotides in the DNA molecule depend upon the flanking base pairs, the parameters of base pair steps should wherever possible be applied in tetra-nucleotide context. (The base pair step parameters used here for reconstruction of the DNA trajectories are provided in the Supplementary material.)

## 2. DNA path modeling

According to quaternion algebra, the vector function of the DNA sequence describing the position of the  $n$ -th base pair in the DNA path in the Cartesian coordinate system can be represented as:

$$P_{(n)} = \sum_{i=2}^n \left[ \prod_{k=2}^i \left[ Q_{\Omega_k \rho_k \tau_k} \right] \circ \vec{V}_{(x_i + Dx_i, y_i + Dy_i, z_i + Dz_i)} \circ \left\{ \prod_{k=2}^i \left[ Q_{\Omega_k \rho_k \tau_k} \right] \right\}^{-1} \right], \quad (1)$$

where  $\prod_{k=2}^i \left[ Q_{\Omega_k \rho_k \tau_k} \right]$  is the consistent quaternion multiplication of the quaternion  $Q_{\Omega_k \rho_k \tau_k}$  realizing rotation about the Z, Y and X axes consistently by angles, defined by the *twist* ( $\Omega$ ), *roll* ( $\rho$ ) and *tilt* ( $\tau$ ) parameters of the neighboring base pairs ( $k-1, k$ ):

$$Q_{\Omega_k \rho_k \tau_k} = \begin{bmatrix} \cos \frac{\Omega_k}{2} \cos \frac{\rho_k}{2} \cos \frac{\tau_k}{2} + \sin \frac{\Omega_k}{2} \sin \frac{\rho_k}{2} \sin \frac{\tau_k}{2} \\ \cos \frac{\Omega_k}{2} \cos \frac{\rho_k}{2} \sin \frac{\tau_k}{2} - \sin \frac{\Omega_k}{2} \sin \frac{\rho_k}{2} \cos \frac{\tau_k}{2} \\ \cos \frac{\Omega_k}{2} \sin \frac{\rho_k}{2} \cos \frac{\tau_k}{2} + \sin \frac{\Omega_k}{2} \cos \frac{\rho_k}{2} \sin \frac{\tau_k}{2} \\ -\cos \frac{\Omega_k}{2} \sin \frac{\rho_k}{2} \sin \frac{\tau_k}{2} + \sin \frac{\Omega_k}{2} \cos \frac{\rho_k}{2} \cos \frac{\tau_k}{2} \end{bmatrix}. \quad (2)$$

Vector  $\vec{V}_{(x_i + Dx_i, y_i + Dy_i, z_i + Dz_i)}$  contains the coordinates of the  $i$ -th element (atom, molecule, point or any structure) in the intrinsic coordinate frame of the dinucleotide step (for example, the coordinates of the idealized phosphate of the first strand are contained in the vector  $\vec{V}_{(Dx_i, 8.91 + Dy_i, 2.08 + Dz_i)}$ ). The quaternion of rotation and translational vector together calculate the rigid-body transformation between two successive base pair steps.

The initial displacement on *shift* ( $Dx$ ), *slide* ( $Dy$ ) and *rise* ( $Dz$ ) for each transformation can be simplistically represented using dual quaternions:

$$\begin{cases} P_{(n)} = \sum_{i=2}^n \left[ \hat{Q}_{TR_i} \circ \hat{Q}_{e_i} \circ \hat{Q}_{TR_i}^* \right] \\ \hat{Q}_{TR_i} = Q_{r_i} + \varepsilon \frac{1}{2} Q_{r_i} \circ Q_{d_i} \\ Q_{r_i} = \prod_{k=2}^i \left[ Q_{\Omega_k \rho_k \tau_k} \right] \\ Q_{d_i} = [0, Dx_i, Dy_i, Dz_i] \\ \hat{Q}_{TR_i}^* = Q_{r_i}^* - \left[ \varepsilon \frac{1}{2} Q_{r_i} \circ Q_{d_i} \right]^* \end{cases}. \quad (3)$$

The dual quaternion  $\hat{Q}_{e_i}$  contains the coordinate of any element, which is associated with the  $i$ -th base pair (for example, the coordinates of the base pair center are contained in the dual quaternion  $\hat{Q}_{e_i} = [1, 0, 0, 0][0, 0, 0, 0]$ ; coordinates of the idealized phosphate of the first strand in  $\hat{Q}_{e_i} = [1, 0, 0, 0][0, 0, 8.91, 2.08]$ ).

The use of dual quaternions is a fast, simple and robust way of molecular modeling. For example, to model base pair conformation, determined by such local (inter base pair) parameters as *shear* ( $Sx$ ), *stretch* ( $Sy$ ), *stagger* ( $Sz$ ), *buckle* ( $\kappa$ ),

*propeller* ( $\pi$ ) and *opening* ( $\sigma$ ) the dual quaternion  $\hat{Q}_{e_i}$  for each base in the  $i$ -th pair can be represented as:

$$\begin{cases} \hat{Q}_{e_i} = Q_{\sigma_i \pi_i} + \varepsilon \frac{1}{2} Q_{\sigma_i \pi_i} \circ Q_{S_i} \\ Q_{S_i} = [0, Sx_i, Sy_i, Sz_i] \end{cases} \quad (4)$$

Similar to dinucleotide step parameters, the local geometric parameters of base pairs can be obtained from articles or deduced from published three-dimensional DNA structures using Curves+ or 3DNA programs. The atomic coordinates for each base and sugar phosphate backbone can be obtained from relevant publications (Clowney et al., 1996; Gelbin et al., 1996; Parkinson et al., 1996; Olson et al., 2001) or structure databases.

Although displacements by *slide* and *shift* do not contribute to DNA bending, they make the main contribution to noise in further measurements. For this reason, these parameters can be ignored in subsequent *in silico* estimations of the DNA curvature. If for subsequent measurements only the DNA trajectory, excluding displacement by *shift* and *slide* is required, the use of dual quaternions is redundant and the equations above for unit quaternions can be simplified to:

$$\begin{cases} P_{(n)} = \sum_{i=2}^n [x_i, y_i, z_i] \\ x_i = \frac{Dz_i}{\|Q^i\|} (2Q_3^i Q_1^i + 2Q_2^i Q_0^i) \\ y_i = \frac{Dz_i}{\|Q^i\|} (2Q_3^i Q_2^i - 2Q_1^i Q_0^i) \\ z_i = \frac{Dz_i}{\|Q^i\|} ([Q_3^i]^2 - [Q_2^i]^2 - [Q_1^i]^2 + [Q_0^i]^2) \\ Q^i = \prod_{k=2}^i [Q_{\Omega_k \rho_k \tau_k}] \end{cases} \quad (5)$$

Calculation of the three-dimensional DNA structures was carried out in the local Cartesian coordinate system in accordance with the transformations defined by the Cambridge convention on the definitions and nomenclature of nucleic acid structure components (Dickerson, 1989). The center of the first base pair is located at the origin of the coordinate frame, at point 0.

### 3. Features of DNA molecule estimation from helical patch

#### Curvature

Overall, the curvature vector of a spatial line is defined as the derivative of the tangent unit vector along this line. Its modulus is the inverse of the curvature radius and its direction is the direction of the main normal to the curve (Landau, Lifshitz, 1970). In terms of a DNA molecule this means the angular deviation between the local helical axes of the successive base pairs. The curvature forms the basis for further estimations of the shape of the DNA molecule. For this reason, there are many ways in which it can be measured. As a rule, the curvature is estimated in DNA curvature units, where one curvature unit is defined as the average curvature of DNA in the nucleosome core particle,  $1/42.8 \text{ \AA}$  (Trifonov, Ulanovsky, 1987).

The DNA curvature of a segment can be calculated from the inverse of the radius of a circumscribed circle of a triangle

with vertices on helix axis coordinates at the center and both ends of this segment. If the sliding window has a length of  $2hw$  bp, the DNA curvature  $C$  in the  $i$ -th position is given by:

$$\begin{cases} C_i = 4\sqrt{(P(P-a)(P-b)(P-c))/abc} \\ P = (a+b+c)/2 \end{cases} \quad (6)$$

where  $a$ ,  $b$  and  $c$  are the distances between the points  $(i-hw, i)$ ,  $(i, i+hw)$  and  $(i-hw, i+hw)$ , respectively.

Curvature estimation during the least square circle (LSC) fit is based on the fitting of a circle to the coordinates of the DNA molecule curved in a plane. The radius of this circle is taken as a measure of the curvature of the DNA fragment (Kanhare, Bansal, 2003). In order to evaluate the planarity of the bend and further projection, the best fit (least squares) plane (LSP) is calculated to a set of base pair centers of the analyzed segment. The possibility of applying this method should be grounded by low values of the root-mean-square deviation (RMSD), estimated for LSC and LSP (RMSD values of distances from the reference and fitted feature (circle or plane)).

The original interpretation of DNA curvature was proposed by De Santis (De Santis et al., 1988). Since DNA curvature  $C(n)$  is a local property of the DNA axis and represents its directional change along the sequence, the distortions of the B-DNA axis along the chain gives the value of curvature per turn in modulus and phase, calculated for recurrent turns along the sequence (De Santis et al., 1988):

$$\bar{C}(n) = v(n_2 - n_1)^{-1} \sum_{s=n_1}^{n_2} [\rho_s - i\tau_s] e^{\frac{2\pi is}{v}} \quad (7)$$

where  $C(n)$  is the average curvature vector, characterizing the orientation deviation of the helical axis between  $n_1$ , and  $n_2$  sequence numbers, per turn of DNA with helical periodicity  $v$  ( $v = 10.4$ ), assigned to the position of the average sequence number of the tract ( $n = (n_2 + n_1)/2$ ), the local deviation of the  $s$ -th base pair plane from the canonical B-DNA represented as a complex vector in terms of the *roll* and *tilt* angles. This is the most popular method for estimation of DNA curvature from the nucleotide sequence directly, without modeling of the helical path in advance.

The local bend angle characterizes the deflection of the helix axis and is calculated as the angle between the tangent vectors in the direction along the contour of the DNA molecule (or as arccosine of a scalar product if these tangents are unit vectors). The tangent vectors can be represented as the unit vectors, indicating orientation of the base pair centers, or as the vectors, connecting  $n$  and  $n+s$  bp ( $s \in [3, 15]$ ), or alternatively, as the unit vectors (normals) averaged over 11–15 bp (centroids of helical turns). Some authors distinguish the bend angles estimated using of tangent vectors obtained in the latter two ways as successive and cumulative (Kanhare, Bansal, 2003). The cumulative bending angle between the averaged normal vectors  $>15$  bp apart is often used in the estimation of DNA curvature (Goodsell, Dickerson, 1994; Gabrielian, Pongor, 1996).

DNA curving can be estimated based on radius of gyration (see Supplementary materials for detail). When comparing DNA molecules of identical sizes the radius of gyration will be smaller for curved and larger for extended molecules (Dlagic, Harrington, 1998). Furthermore, information about the different shape properties such as symmetry, anisotropy,

asphericity, acylindricity and more can be extracted from the various combinations of the orthogonal components of the radius of gyration tensor (Jernigan et al., 1987; Olson et al., 1993; Dlakic, Harrington, 1998; Kanhere, Bansal, 2003; Rawat, Biswas, 2009).

The *d-max* (Tung, Burks, 1987) is the maximum orthogonal distance from the base pair center to the straight line connecting the ends of the analyzed segment, evaluated over all base pair centers of this segment. This value is estimated as the maximal perpendicular distance (*p-dist*) over all base pair centers. In its turn, the perpendicular distance is the minimal distance from the *i*-th base pair to the straight line connecting the ends of the segment.

Lastly, the curvature of the DNA molecule can be simply characterized by the ratio of the curvilinear distance (contour length) to the linear distance between the ends of the segment (*SD* value) (Eckdahl, Anderson, 1987; Tan, Harvey, 1987) and vice versa (*Qc*, (Dlakic, Harrington, 1998; Kanhere, Bansal, 2003; Matyašek et al., 2013)). Large values of *C*, *d-max* and *SD* indicate a more curved DNA fragment.

The sliding window technique is a common way for analyzing this distribution. Reducing the size of the sliding window increases the resolution and magnitude of the curvature measurements, but also increases the noise that is determined by local bends (by *roll* and *tilt*) and displacements (by *shift* and *slide*) of adjacent base pairs. In order to reduce the noise, the size of the sliding window as a multiple of a helical turn (10 or 10.4 bp), from 20 bp is recommended. Furthermore, since local bends during a helical turn, as a rule, are mutually compensated, when the curvature is estimated for a position in the center of sliding window, the size of a half window should be a multiple of 10 (for example, see the analysis of the “Hagerman paradox”, below).

### Persistence length

Persistence length is a key parameter for quantitative interpretation of the conformational properties of DNA. Theoretically, the persistence length can be formulated in terms of the magnitude of the projection vector and the tangent-tangent correlation function. In previous studies, both methods were successfully applied to estimation of the persistence length of DNA molecules from the reconstructed 3D trajectories of helical paths (Shpigelman et al., 1993; Bednar et al., 1995; Schellman, Harvey, 1995; Vologodskaja, Vologodskii, 2002).

In terms of projection vector, the persistence length (*P*) can be defined as the average projection of the end-to-end vector ( $I_N$ ) onto the unit vector of the first segment ( $I_1$ ) in the limit of infinite chain length (Flory, 1969). It is likely that for a long heterogeneous chain with average link angle the more statistically significant is the average projection of end-to-end vector on the direction (unit vector) of each segment along the chain. Thus the persistence length is averaged over position and direction is given by:

$$P = \lim_{N \rightarrow \infty} \frac{1}{N} \left\langle \sum_{j=1}^N \frac{\overline{I_N \cdot I_j}}{|I_j|} \right\rangle. \quad (8)$$

Persistence length, measured in this way characterizes the chain length through which the memory of the initial orientation persists. So that its will be strongly dependent on the direction, length, homogeneity and shape of molecule.

Nevertheless, for the cases when  $L \gg P$  this method can be effectively used for the operative estimation of the DNA persistence length.

According to the tangent-tangent correlation function (Landau, Lifshitz, 1958):

$$\langle \cos(\theta) \rangle = \left\langle \overline{I(s_i) \cdot I(s_{i+1})} \right\rangle = e^{-\frac{s_{i+1}-s_i}{P}}, \quad (9)$$

where  $\langle \cos(\theta) \rangle$  is the average cosine calculated from the complete set of local bend angles ( $\theta$ ) measured between the unit vectors ( $I$ ) tangent to the chain at the points  $s_i$  and  $s_{i+1}$  (spaced of contour length  $l$  ( $s_{i+1}-s_i$ )) collected over the entire chain. Tangent vectors can be calculated in the various ways described above keeping in mind that the method employed will dramatically affect the deflection angle and, hence, persistence length estimation. In particular the persistence length strongly depends on the scale of measurements, i. e. when unit segments of the DNA chain are bonds between base pairs, or when they are centroids of helical turns (see Supplementary materials).

The average cosine  $\theta$  is a multiplicative function of segment length  $l$ , assuming that the average directional correlation between two segments decays exponentially along the chain. The distribution of  $-\ln\langle \cos(\theta) \rangle$  is evaluated for the segments of different length  $l$ , where  $l$  is the average curvilinear distance between base pair centers, separated by the  $n$  bp window, moving along the sequence with a step of  $\geq 1$  bp. Since correlation in the orientation of segments decays exponentially to their distance, the  $-\ln\langle \cos(\theta) \rangle$  will be linear. The persistent length can be calculated from the inverse of the regression slope of the  $-\ln\langle \cos(\theta) \rangle$  distribution plot. Application of this method should be justified by a good linear fit of the  $-\ln\langle \cos(\theta) \rangle$  distribution (coefficient of determination,  $R^2 \geq 0.9$ ).

The equations above are suitable only for intrinsically straight, homogeneous polymers. However, in most cases the analyzed DNA is not straight, but contains numerous planar and coplanar static bends as well as segments with different flexibilities, located at any position along the chain. For such heterogeneous polymers, the persistence length strongly depends on the starting position, the direction, and the length of the fragment. One of possible solutions is the division of the DNA path into a set of large fragments between the bends and a set of small fragments overlapping these bends, then estimation of  $P$  can be conducted for each of these fragments separately (Supplementary material contains examples for different cases). Following that, the average persistence length of a chain can be derived from summation of  $P$  over all fragments ( $N_j$ ) with length  $l$  (Rivetti et al., 1998):

$$\frac{L}{P} = \sum_{i=1}^{N_j} \frac{l_i}{P_i}. \quad (10)$$

Thus, the resultant persistence length (averaged over all fragments of molecule) will be mainly determined by the minimal  $P$  among all fragments. Furthermore, it seems that the contribution of intrinsic bends to persistence length calculated in this way will be decreased.

The apparent persistence length ( $P_a$ ) of the DNA molecule includes contributions from both static ( $P_s$ ) and dynamic ( $P_d$ ) persistence lengths, which are related as follows (Trifonov et al., 1988; Schellman, Harvey, 1995):

$$\frac{1}{P_a} = \frac{1}{P_s} + \frac{1}{P_d}. \quad (11)$$

Static persistence length is determined by the intrinsic DNA curvature (shape) at the minimum energy conformation, without accounting for thermal fluctuations. The equations above, applied to the predicted DNA trajectories, estimate the static persistence length. Conversely, the dynamic persistence length characterizes the rigidity of the DNA molecule in thermal fluctuations of the angles between adjacent base pairs. Assuming crystal packing, interactions with proteins and thermal fluctuations as external perturbing force fields, eliminating correlations in the bend directions, the dynamic persistence length can be estimated from the DNA sequence using a set of dispersions of *roll* ( $\rho$ ) and *tilt* ( $\tau$ ) angles (which can be obtained from the same sources as for the base pair parameters or during averaging of base pair parameters, see Section 1) for all dinucleotide steps (Vologodskaya, Vologodskii, 2002):

$$\frac{1}{P_d} = \frac{1}{2l} \sum_{i=1}^{16} p_i \theta_i^2, \quad (12)$$

where  $p_i$  is the probability of dinucleotide  $i$ , the length of segment  $l \ll P$  is an average *rise* ( $\sim 3.38 \text{ \AA}$ ) and  $\theta_i^2$  is the variance of the local bend angle (in radians) of  $i$ -th dinucleotide ( $\theta_i^2 = \Delta\rho_i^2 + \Delta\tau_i^2$ ,  $\rho$  and  $\tau$  in radians,  $\theta_i \ll \pi/2$ ).

However it should be noted that whereas flexibility of base pair steps in overall correlates well with local bends, as determined by *roll* and *tilt* (Olson et al., 1998; Packer et al., 2000; McConnell, Beveridge, 2001), the bends produced by A-tracts are highly stable with a midpoint of structural transition near 30 °C (Chan et al., 1993) and a melting temperature above 37 °C (Chan et al., 1990; Jerkovic, Bolton, 2000). Thus, although the static persistence length of A-tracts-produced bends is very small, the dynamic persistence length of these bends will be very large. Since  $P_a$  is always less than either  $P_s$  or  $P_d$  the correlation between apparent persistence length, calculated from the summation of  $P_s^{-1}$  and  $P_d^{-1}$ , and the stiffness of DNA containing A-tract induced bends will be lost. The extremely low static ( $P_s$ ) and exceptionally high dynamic ( $P_d$ ) persistence lengths at relatively large apparent persistence length ( $P_a$ ) were noted in a recent study of DNA molecules containing A-tract related large intrinsic bends (Mitchell et al., 2017).

### Helical phasing. DFT analysis

Discrete Fourier transform is extremely useful for macroscopic DNA curvature analysis. This is due to its ability to reveal periodicity in analyzed DNA properties, such as local curvature in sliding window, as well as the relative strength of any periodic components. Subsequent DFT frequency spectrum analysis allows us to detect and quality evaluate the regular alternation of DNA bends along the molecule. The discrete Fourier transform is defined as:

$$F_k = \sum_{j=0}^{N-1} x_j e^{2\pi ijk/N}, \quad k \in [0, N-1], \quad (13)$$

where  $x_j$  is a property value in the  $j$ -th position, represented as a complex number with a zero-valued imaginary part, and  $k$  is the frequency domain (for positive frequencies  $k \in [0, N/2-1]$ ).

In the case of DNA analysis, the sampling frequency ( $f_s$ , sampling rate) is the ratio of the data set size ( $N$ ) to the step of the sliding window. The minimal frequency (frequency resolution) is given by  $\Delta f = f_s/N$  and maximal frequency (Nyquist frequency) is  $f_s/2$ . Since this DFT spectrum is  $N$ -periodic, each

frequency domain characterizes the distribution of values of the analyzed features with frequencies of  $f$  per  $N$  bp and a maximum  $1/2 \text{ bp}^{-1}$  ( $f = \Delta f x = x$ ,  $x \in [0, N/2]$ ).

The magnitude of the frequency domain for the real input data is calculated as:

$$|F_k| = \sqrt{\left[ \sum_{j=0}^{N/2-1} x_{2j} e^{2\pi ijk/(N/2)} \right]^2 + \left[ e^{2\pi ik/N} \sum_{j=0}^{N/2-1} x_{2j+1} e^{2\pi ijk/(N/2)} \right]^2}. \quad (14)$$

The phasing of systematic bends can be evaluated from the data of distribution of the local curvature in the course of frequency analysis of the DFT power spectrum. The frequency of helix turns ( $fh$ ) in the DFT spectrum is  $N/10$  (where 10 is the floor of average *twist*, 10.4 bp) and the initial phase is 0°. The local bends repeated with a frequency multiple of  $fh$  or with variations around this are phased with the helix screw: they have the same direction and progressively increase the macroscopic curvature of the DNA molecule. The fractional part of the ratio of  $fh$  to the frequency of local bends repeated in antiphase is near to 0.5: these bends are oriented oppositely (phase shift is 180°,  $2\pi \cdot 0.5$ ) and mutually compensate each other.

In order to demonstrate how phasing of DNA bends affects the shape of molecules, the trajectories of the  $(A_5N_5)_n$  (curved DNA, local bends in phase) and  $(A_5N_{10})_n$  (straight DNA, local bends out of phase) sequences were modeled and the DFT power spectrum of curvature distribution was analyzed (Fig. 2). If sampling is 128, the  $fh = 13$  (128/10). The  $(A_5N_{10})_n$  DNA fragment contains systematic bends, repeated with a frequency between 8 and 9 that corresponds to  $\{fh/8.5\} \approx 0.5$ . Hence, the phases of neighboring bends will be opposite and the molecule is almost straight (see Fig. 2). The systematic bends of the  $(A_5N_5)_n$  DNA fragment repeated with a frequency of 13 corresponds to  $fh$ , hence, the local bends alternate in phase with the helix periodicity, and the resulting global curvature of DNA molecule will be large (see Fig. 2).

For operative estimation of the phasing of DNA bends, the DFT power spectrum can be evaluated only for frequencies near to the frequency of the helix screw ( $\omega$ , average twist angle, 34.6°) (Gabrielian, Pongor, 1996). In this case, the period is  $2\pi$  (360°, one helical turn), hence frequency ( $k/N$ ) is 34.6°/360° and the average twist corresponds to angular frequency  $\omega = 2\pi 34.6^\circ \pi / 2\pi 180^\circ$  (in radians). The discrete Hartley transform (DHT) is an alternative for the real DFT. DHT for angular frequency  $\omega$  is given by:

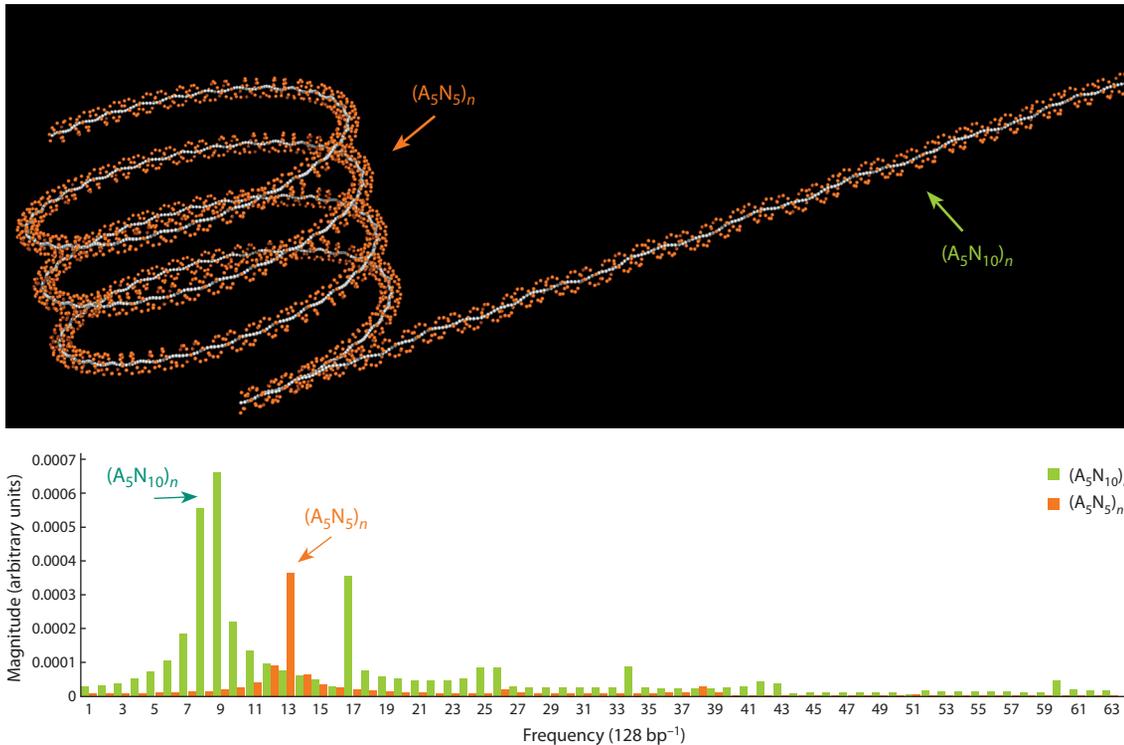
$$H_\omega = \frac{1}{N} \sum_{j=0}^{N-1} x_j \text{cas}(j\omega), \quad (15)$$

where  $\text{cas}(j\omega) = \cos(j\omega) + \sin(j\omega)$ .

The magnitude of this frequency bin in terms of DHT can be expressed as:

$$|H_\omega| = \sqrt{\left[ \frac{1}{N} \sum_{j=0}^{N-1} x_j \cos(j\omega) \right]^2 + \left[ \frac{1}{N} \sum_{j=0}^{N-1} x_j \sin(j\omega) \right]^2}. \quad (16)$$

The DFT power spectrum should be calculated for frequencies in range of minimal and maximal values of *twist* angle (as a rule from 27° to 42° for B-DNA). The supplementary material contains some recommendations on the use of DFT analysis in studying of the DNA curvature distribution.



**Fig. 2.** Reconstructed DNA paths of the experimental molecules, containing repeated A-tract in phase  $((A_5N_5)_n)$  and out of phase  $((A_5N_{10})_n)$  with the helix screw. DFT power spectral for both molecules were calculated using values of DNA curvature as input data, estimated in a sliding window of 40 bp and step of 1 bp, using equation (6). The dominant frequencies for each spectrum are indicated.

**Phasing of coplanar and non-periodic bends**

It seems obvious that non-periodically bends in phase as well as bends in opposite directions in antiphase will increase the macroscopic DNA curvature. Furthermore, the orientation of bends is not distinguished by the DFT method, so that coplanar bends will be taken as planar, leading to distortion of the results. For this reason, the definition of phase relative to the direction of the DNA trajectory will be useful. In order to define the relative phase  $(\varphi_{vb})$  of the vector position of the center of bend (“bend vector”, for example this can be the center of an analyzed segment:  $vb_{n+s} = v_{n+s} - v_n$ ) the “directional vector”  $vd_n = v_{n+2s} - v_n$  was aligned with the  $X$  axis and then the coordinate frame was rotated around the  $Y$  axis by  $90^\circ$  to align  $vd$  with the  $Z$  axis (the direction of this rotation is determined by the sign of  $x$  coordinate of  $vd$ ). The quaternion for realizing these rotations is given by  $Q^\varphi$ :

$$\begin{cases} Q^\varphi = \left[ 0.7, 0, -0.7 \frac{x_{vd}}{|x_{vd}|}, 0 \right] \cdot Q^{vx} \\ Q^{vx} = \left[ \sqrt{\frac{1+qR_0}{2}}, \frac{qR_1}{\langle qR \rangle} \sqrt{\frac{1-qR_0}{2}}, \frac{qR_2}{\langle qR \rangle} \sqrt{\frac{1-qR_0}{2}}, \frac{qR_3}{\langle qR \rangle} \sqrt{\frac{1-qR_0}{2}} \right], \\ qR = [0, x_{vd}, 0, 0] \cdot \frac{\overline{v_d}}{|v_d|} \end{cases} \quad (17)$$

where  $\sqrt{\frac{1+qR_0}{2}}$  equivalent to the  $\cos(\text{acos}(qR_0)/2)$ ,  $\sqrt{\frac{1-qR_0}{2}}$  is the  $\sin(\text{acos}(qR_0)/2)$ , and  $\{qR_1, qR_2, qR_3\}/\langle qR \rangle$  are the direction cosines at  $X, Y$  and  $Z$  axes respectively.

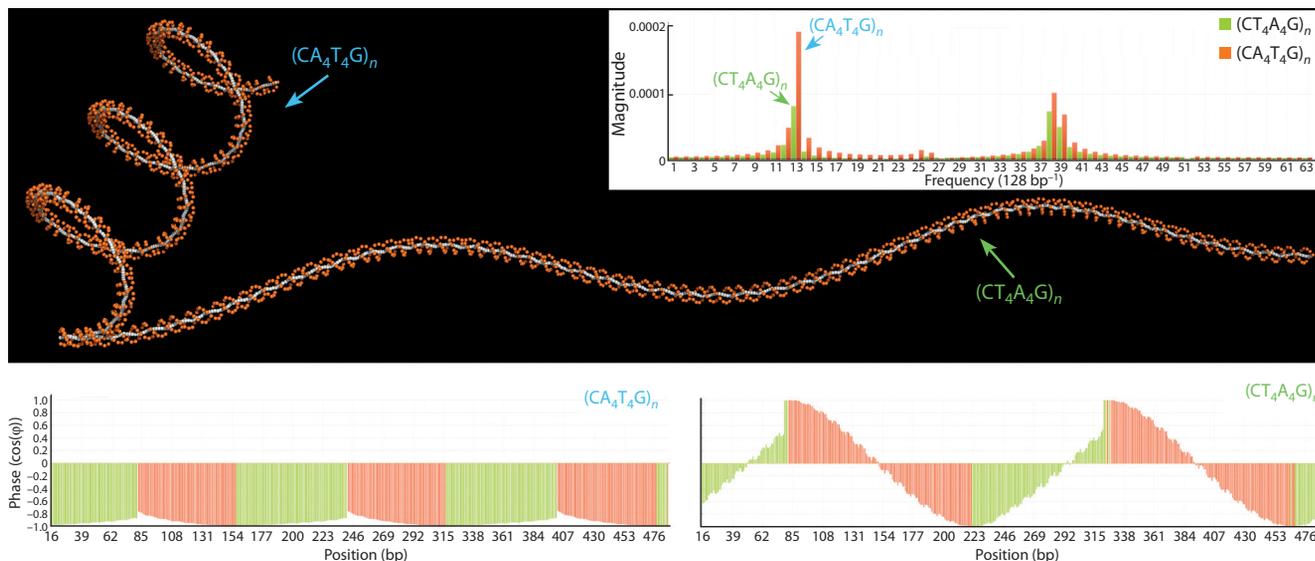
To align the unit vector with the  $X$  axis we can simplify the calculation of  $qR$  to:

$$qR = \left[ |x_{vd}|, 0, z_{vd} \frac{x_{vd}}{|x_{vd}|}, -y_{vd} \frac{x_{vd}}{|x_{vd}|} \right], \quad (18)$$

where the  $\frac{x_{vd}}{|x_{vd}|}$  ratio is necessary to get a sign. The phase is estimated by all four quadrants as the cosine of the angle between the projection of curvature vector on  $XY$  plane and the direction of the  $X$  axis  $(1, 0, 0)$ :

$$\begin{cases} vb = Q^\varphi \cdot vb \cdot (Q^\varphi)^{-1} \\ \varphi_{vb} = [1, 0, 0] \cdot \frac{vb}{|vb|} \end{cases} \quad (19)$$

For example, Hagerman has shown that DNA sequences containing repeating runs of  $A_4T_4$  in phase were significantly bent, whereas those with  $T_4A_4$  (A-tract in opposite polarity) were almost straight (Hagerman, 1986). Thus, there is a fundamental difference in the structure of  $A_nT_n$  and  $T_nA_n$  DNA segments, despite their identical nucleotide composition and phased bends. In fact, the distribution of curvature in these DNA molecules have the same DFT frequency spectrum with a peak of magnitude for 10.04 bp period, which is multiplied to the period of helix axis  $(\{10.4/10.04\} = 0.036 \ll 5)$  that is also consistent with experimental data (Price, Tullius, 1993). Thus, at first glance in both cases the phased bends should increase the macroscopic curvature of DNA molecules significantly. However analysis of the phase spectrum makes it clear (Fig. 3) why in the first case  $((CA_4T_4G)_n)$  DNA fragments



**Fig. 3.** The “Hagerman paradox” analysis using discrete Fourier transform and the phase spectrum calculated by equation (17). Curved (( $CA_4T_4G$ )<sub>n</sub>) and “straight” (( $CT_4A_4G$ )<sub>n</sub>) DNA molecules have similar frequency spectral, but differ in the phase spectrum of curvature, calculated relative to direction of the DNA path. The macroscopic bends of the “straight” fragment orient in opposite directions and compensate each other. Curvature was analyzed in a sliding window of 30 bp and 1 bp step.

are exclusively curved, but in the other case (( $CT_4A_4G$ )<sub>n</sub>) the fragments are almost straight (“zig-zag” structure). In particular it shows that the bends produced by (( $CT_4A_4G$ )<sub>n</sub>) are oriented in opposite directions and mutually compensate each other (see Fig. 3).

However, it should be noted that magnitudes of DFT power spectrum of the (( $CT_4A_4G$ )<sub>n</sub>) straight fragment will be relative high only if the size of the sliding window wherein DNA curvature is estimated is in multiples of an odd number of turns. If the size of the half window is a multiple of 10 (one helical turn) the amplitude of curvature of the (( $CT_4A_4G$ )<sub>n</sub>) fragment is near to 0. This is consistent with the fact that oppositely oriented bends, in this motif, are mutually compensated in one turn (Steffl et al., 2004). The present example shows how a change in the size of the sliding window can be adopted for specific research tasks.

### Conclusion

More than 30 years ago it was shown that computer modeling of DNA sequences is a viable approach to the study of the biological implications of DNA structure. A lot of research has been devoted to this problem. The estimation of the intrinsic curvature of relatively short DNA fragments (< 10 kb) can be useful in investigating various features, such as analysis of promoter and regulatory regions of specific genes, or the parts of the genome associated with recombination suppression and heterochromatin packing, nucleosome positioning, design of shape-related DNA markers and many others. Furthermore, the actual aim is the monitoring of DNA curvature and bendability over hundreds and thousands base pairs. Similar studies, previously performed for many prokaryotic and eukaryotic genomes, provided important information about their spatial organization and its influence on various biological processes. The algorithms discussed in the present guide formed the basis for such investigations and will be useful in analysis

of relatively short DNA fragments as well as for topological mapping of whole genomes.

### Acknowledgments

This work was supported by the Siberian Branch of the Russian Academy of Sciences, integrated program 0324-2016-0025, project II.2P/VI.53-17.

### Conflict of interest

The author declares no conflict of interest.

### References

Babcock M.S., Pednault E.P., Olson W.K. Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* 1994; 237(1):125-156.

Bednar J., Furrer P., Katritch V., Stasiak A.Z., Dubochet J., Stasiak A. Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.* 1995;254(4): 579-594.

Berman H., Henrick K., Nakamura H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* 2003;10(12):980.

Bolshoy A., McNamara P., Harrington R.E., Trifonov E.N. Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. USA.* 1991;88(6):2312-2316.

Chan S.S., Breslauer K.J., Austin R.H., Hogan M.E. Thermodynamics and premelting conformational changes of phased (dA)<sub>5</sub> tracts. *Biochemistry.* 1993;32(44):11776-11784.

Chan S.S., Breslauer K.J., Hogan M.E., Kessler D.J., Austin R.H., Ojemann J., Passner J.M., Wiles N.C. Physical studies of DNA premelting equilibria in duplexes with and without homo dA.dT tracts: correlations with DNA bending. *Biochemistry.* 1990;29(26):6161-6171.

Clowney L., Jain S.C., Srinivasan A.R., Westbrook J., Olson W.K., Berman H.M. Geometric parameters in nucleic acids: Nitrogenous bases. *J. Am. Chem. Soc.* 1996;118(3):509-518.

- De Santis P., Palleschi A., Savino M., Scipioni A. A theoretical model of DNA curvature. *Biophys. Chem.* 1988;32(2-3):305-317.
- De Santis P., Palleschi A., Savino M., Scipioni A. Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry.* 1990;29(39):9269-9273.
- Dickerson R.E. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* 1989;17(5):1797-1803.
- Dlatic M., Harrington R.E. Unconventional helical phasing of repetitive DNA motifs reveals their relative bending contributions. *Nucleic Acids Res.* 1998;26(18):4274-4279.
- Eckdahl T.T., Anderson J.N. Computer modelling of DNA structures involved in chromosome maintenance. *Nucleic Acids Res.* 1987;15(20):8531-8545.
- Flory P. *Statistical Mechanics of Chain Molecules.* N.Y.: Inter-science, 1969.
- Gabrielian A., Pongor S. Correlation of intrinsic DNA curvature with DNA property periodicity. *FEBS Lett.* 1996;393(1):65-68.
- Gelbin A., Schneider B., Clowney L., Hsieh S.H., Olson W.K., Berman H.M. Geometric parameters in nucleic acids: Sugar and phosphate constituents. *J. Am. Chem. Soc.* 1996;118(3):519-529.
- Goodsell D.S., Dickerson R.E. Bending and curvature calculations in B-DNA. *Nucleic Acids Res.* 1994;22(24):5497-5503.
- Hagerman P.J. Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. *Biochemistry.* 1985;24(25):7033-7037.
- Hagerman P.J. Sequence-directed curvature of DNA. *Nature.* 1986;321(6068):449-450.
- Jerkovic B., Bolton P.H. The curvature of dA tracts is temperature dependent. *Biochemistry.* 2000;39(40):12121-12127.
- Jernigan R.L., Sarai A., Shapiro B., Nussinov R. Relationship between curved DNA conformations and slow gel migration. *J. Biomol. Struct. Dyn.* 1987;4(4):561-567.
- Kanhere A., Bansal M. An assessment of three dinucleotide parameters to predict DNA curvature by quantitative comparison with experimental data. *Nucleic Acids Res.* 2003;31(10):2647-2658.
- Koo H.S., Wu H.M., Crothers D.M. DNA bending at adenine thymine tracts. *Nature.* 1986;320(6062):501-506.
- Kratky O., Porod G. Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas.* 1949;68:1106-1123.
- Landau L.D., Lifshitz E.M. *Statistical Physics, Course of Theoretical Physics.* London, UK: Pergamon, 1958.
- Landau L.D., Lifshitz E.M. *Theory of Elasticity.* Oxford, NY: Pergamon Press, 1970.
- Lavery R., Moakher M., Maddocks J.H., Petkeviciute D., Zakrzewska K. Conformational analysis of nucleic acids revisited: curves+. *Nucleic Acids Res.* 2009;37(17):5917-5929.
- Liu Y., Beveridge D.L. A refined prediction method for gel retardation of DNA oligonucleotides from dinucleotide step parameters: reconciliation of DNA bending models with crystal structure data. *J. Biomol. Struct. Dyn.* 2001;18(4):505-526.
- Lu X.J., Olson W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 2003;31(17):5108-5121.
- Masoudi-Nejad A., Movahedi S., Jáuregui R. Genome-scale computational analysis of DNA curvature and repeats in Arabidopsis and rice uncovers plant-specific genomic properties. *BMC Genomics.* 2011;12:214.
- Matyášek R., Fulneček J., Kovařík A. Evaluation of DNA bending models in their capacity to predict electrophoretic migration anomalies of satellite DNA sequences. *Electrophoresis.* 2013;34(17):2511-2521.
- McConnell K.J., Beveridge D.L. Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol.* 2001;314:23-40.
- Mitchell J.S., Glowacki J., Grandchamp A.E., Manning R.S., Maddocks J.H. Sequence-dependent persistence lengths of DNA. *J. Chem. Theory Comput.* 2017;13(4):1539-1555.
- Olson W.K., Bansal M., Burley S.K., Dickerson R.E., Gerstein M., Harvey S.C., Heinemann U., Lu X.J., Neidle S., Shakked Z., Sklenar H., Suzuki M., Tung C.S., Westhof E., Wolberger C., Berman H.M. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 2001;313(1):229-237.
- Olson W.K., Gorin A.A., Lu X.J., Hock L.M., Zhurkin V.B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA.* 1998;95(19):11163-11168.
- Olson W.K., Marky N.L., Jernigan R.L., Zhurkin V.B. Influence of fluctuations on DNA curvature. A comparison of flexible and static wedge models of intrinsically bent DNA. *J. Mol. Biol.* 1993;232(2):530-554.
- Packer M.J., Dauncey M.P., Hunter C.A. Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.* 2000;295(1):85-103.
- Parkinson G., Vojtechovsky J., Clowney L., Brünger A.T., Berman H.M. New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. Sect. D.* 1996;52(1):57-64.
- Price M.A., Tullius T.D. How the structure of an adenine tract depends on sequence context: a new model for the structure of TnAn DNA sequences. *Biochemistry.* 1993;32(1):127-136.
- Rawat N., Biswas P. Size, shape, and flexibility of proteins and DNA. *J. Chem. Phys.* 2009;131(16):165104.
- Rivetti C., Walker C., Bustamante C. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.* 1998;280(1):41-59.
- Schellman J.A., Harvey S.C. Static contributions to the persistence length of DNA and dynamic contributions to DNA curvature. *Biophys. Chem.* 1995;55(1-2):95-114.
- Shpigelman E.S., Trifonov E.N., Bolshoy A. CURVATURE: software for the analysis of curved DNA. *Comput. Appl. Biosci.* 1993;9(4):435-440.
- Steff R., Wu H., Ravindranathan S., Sklenar V., Feigon J. DNA A-tract bending in three dimensions: solving the dA4T4 vs dT4A4 conundrum. *Proc. Natl. Acad. Sci. USA.* 2004;101(5):1177-1182.
- Tan R.K., Harvey S.C. A comparison of six DNA bending models. *J. Biomol. Struct. Dyn.* 1987;5(3):497-512.
- Trifonov E.N., Tan R.K.Z., Harvey S.C. Static persistence length of DNA. In: W.K. Olson, M.H. Sarma, R.H. Sarma, M. Sundaralingam (Eds.). *DNA Bending and Curvature.* N.Y.: Adenine Press, 1988; 243-253.
- Trifonov E.N., Ulanovsky L.E. Inherently curved DNA and its structural elements. In: R.D. Wells, S.C. Harvey (Eds.). *Unusual DNA Structures.* Berlin, Germany: Springer-Verlag, 1987;173-187.
- Tung C.S., Burks C. A quantitative measure of DNA curvature enabling the comparison of predicted structures. *J. Biomol. Struct. Dyn.* 1987;4(4):553-559.
- Vermeulen A., Zhou H., Pardi A. Determination DNA global structure and DNA bending by application of NMR residual dipolar coupling. *J. Am. Chem. Soc.* 2000;122:9638-9647.
- Vologodskaya M., Vologodskii A. Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.* 2002;317(2):205-213.
- Wu H.M., Crothers D.M. The locus of sequence-directed and protein-induced DNA bending. *Nature.* 1984;308(5959):509-513.