Genetica **106:** 63–73, 1999.

Sequence-dependent modelling of local DNA bending phenomena: curvature prediction and vibrational analysis

Kristian Vlahoviček[1], Mircea Gh. Munteanu[2] & Sándor Pongor[1]
[1]*International Center for Genetic Engineering and Biotechnology, Padriciano 99, 34012 Trieste, Italy (E-mail: pongor@icgeb.trieste.it);* [2]*Universitatea Transilvania, Catedra de Rezistenta materialelor si Vibratii, B-dul Eroilor, 29 2200 Brasov, Romania*

Key words: DNA bending, DNA curvature, DNA models, 3D model, finite element method, internet server

Abstract

Bending is a local conformational micropolymorphism of DNA in which the original B-DNA structure is only distorted but not extensively modified. Bending can be predicted by simple static geometry models as well as by a recently developed elastic model that incorporate sequence dependent anisotropic bendability (SDAB). The SDAB model qualitatively explains phenomena including affinity of protein binding, kinking, as well as sequence-dependent vibrational properties of DNA. The vibrational properties of DNA segments can be studied by finite element analysis of a model subjected to an initial bending moment. The frequency spectrum is obtained by applying Fourier analysis to the displacement values in the time domain. This analysis shows that the spectrum of the bending vibrations quite sensitively depends on the sequence, for example the spectrum of a curved sequence is characteristically different from the spectrum of straight sequence motifs of identical basepair composition. Curvature distributions are genome-specific, and pronounced differences are found between protein-coding and regulatory regions, respectively, that is, sites of extreme curvature and/or bendability are less frequent in protein-coding regions. A WWW server is set up for the prediction of curvature and generation of 3D models from DNA sequences (http://www.icgeb.trieste.it/dna).

Introduction

The idea of local structural polymorphism in DNA has profoundly influenced the thinking of biologists in recent years. DNA is no longer considered as a featureless double helix but rather as a series of individual domains differing in flexibility and curvature. Local structural polymorphism is expected to contribute to the specificity of various biological events as gene regulation, packaging, for example, through regulating the affinity of protein binding [1]. In contrast to traditional structural polymorphism (e.g. B, A or Z structures), here we deal with a localised micropolymorphism in which the original B-DNA structure is only distorted but not extensively modified. The DNA segments involved in protein-induced and inherent DNA-bending are 10–50 base pairs in length [2] that is they are longer than what can be easily handled by atomic resolution molecular modelling or quantum mechanical approaches. Traditional elastic models of DNA, that represent DNA as an ideally elastic, homogeneous cylindrical rod, were used to model macroscopic behaviour of long DNA segments, such as supercoiling [3, 4]. However, local DNA conformations and recognition by DNA-binding proteins are clearly sequence-dependent, so conventional elastic rod models of DNA, which do not explicitly represent the dependence of the elasticity on the base sequence, cannot say much about them. Here, we attempt to briefly review advantages and limitations of the rod-models of DNA with particular regard to elastic modelling of local bending phenomena.

Static geometry models

The simplest form of DNA models are the so-called Rod models that picture DNA as a cylindrical rod of

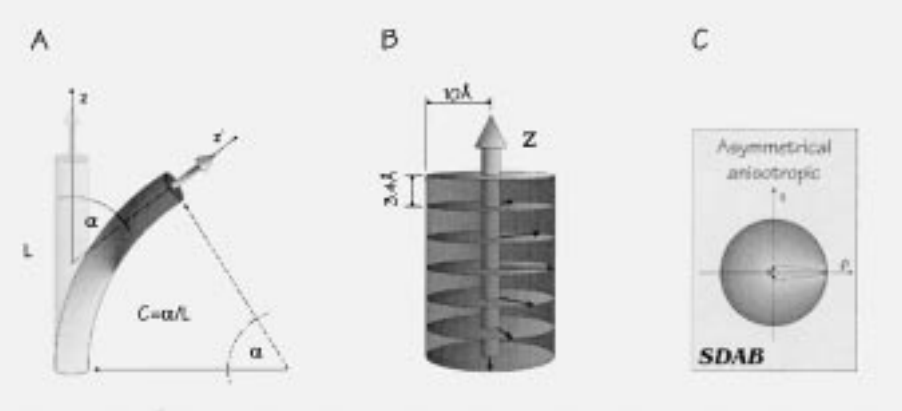

Figure 1. (A) Curvature of an elastic rod. (B) A segmented rod model. Each element corresponds to one base pair and may have different static deflection angles (roll, tilt twist) as well as different bending, torsional and stretching rigidity values. (C) The sequence-dependent anisotropic bendability model of DNA. The arrows are proportional to the flexibility in a given direction (+x, −x, +y, −y). In the SDAB model, one direction (that of the major groove) is more flexible, the other three are more rigid.

constant diameter. The shape in this case is the path or trajectory of the longitudinal axis, which can be either straight or curved (Figure 1A). The common philosophy of rod models are to divide the rod into short cylindrical segments (e.g. of the size of a base pair, Figure 1B) and then to compute some rod parameter based on the segment parameters that have to be *a priori* known. Dinucleotide models define the base pair-size unit as one between two adjacent base pairs, which results in 16 different base pairs, or 10, if we allow strand symmetry (i.e. AA = TT). Trinucleotide models define the unit around the base pair in the middle of a given trinucleotide. This amounts to 64 or 32 different units, again depending on strand symmetry.

Static models are rigid rod models that only consider the static geometry of a constituting segment. Curvature in B-DNA was originally believed to be an attribute of $A_n (n = 4–6)$ tracts repeated in phase with DNA's helical repeat. Two static models have been proposed initially to explain the phenomenon. In the nearest neighbour model, the axial deflections of successive AA/TT dinucleotides add up to produce a curve [5]. In the so-called junction model, curvature is produced as the modified B-DNA structure of A_n tracts joins with adjacent B-DNA [6]. More recently it became clear that in addition to the A_n tracts, DNA curvature involves additional sequence elements [7, 8], and more sophisticated models were proposed which included wedge angles for all 16 dinucleotides [9, 10].

Current static models consider dinucleotide geometries derived from direct measurements such as X-ray crystallography [11, 12] and NMR [13, 14], from statistics of nucleosome binding data [7, 15] or from gel electrophoretic analysis of concatenated synthetic oligonucleotide repeats [9, 10] . The parameters explicitly considered are Roll, Tilt and Twist angles, and all other parameters are considered equal to that of B-DNA (the diameter of the rod plays no role in this calculation). Given the geometry of the segments one can calculate the trajectory of DNA by adding the distortions of the successive segments. This exercise is only seemingly simple, because addition of rotation operations like roll, tilt and twist are not commutative, so the results in principle depend on the serial order of the operations [16]. The resulting error is negligible for short DNA segments and for small angles (and generally speaking, roll and tilt angles are small). The calculation will however give large uncertainties for long segments or for large distortions. There are a number of programs available that can calculate an approximate DNA trajectory starting from a DNA sequence [10, 12, 17–19]. Table 1 shows that the current dinucleotide models give good qualitative correlations with the known experimental values, that is curved and straight motifs can be distinguished even though all models mispredict some of the motifs. It has to be mentioned that gel mobility anomaly of DNA that is used to validate the models is itself strongly dependent on environmental factors such as metal salts and temperature, so our knowledge on straight and curved motifs is qualitative rather than quantitative.

Table 1. Analysis of curved and straight sequence motifs with various methods

No.	Sequence motif	Static geometry model				Sequence dependent anisotropic bendability (SDAB) model			
		Curvature [°/helical turn]				Bendability [a.u.]		Predicted curvature [°/helical turn]	
		Electro-phoresis [9]	X-rays [41]	NMR [13]	Nucleo-some [7, 15]	DNaseI [29]	Consensus [14]	DNaseI [29]	Consensus [14]
Curved DNA									
1	$(aaaattttgc)_n$	26.2	6.9	18.3	13.7	2.8	2.4	21.1	17.4
2	$(aaaattttcg)_n$	21.0	3.8	3.8	17.7	2.2	2.3	16.6	17.2
3	$(tctcaaaaaacgcgaaaaaaccggaaaaaagc)_n$	27.1	8.2	16.7	17.1	3.2	3.2	15.4	15.9
4	$(ccgaaaaagg)_n$	14.7	6.8	13.1	23.3	3.9	4.3	17.1	20.2
5	$(tctctaaaaaatatataaaaa)_n$	27.8	3.0	7.5	10.9	4.5	3.2	27.6	18.5
6	$(ggcaaaaaac)_n$	26.8	12.0	20.1	20.4	3.2	3.3	19.2	19.5
7	ccaaaaatgtcaaaaaataggcaaaaaatgcc - *L. tarentolae* kinetoplast	26.0	6.4	15.7	19.6	3.9	3.5	21.8	20.5
8	aaaaactctctaaaaactctccctagaggggccctagagggc	19.4	7.8	10.6	13.3	5.2	4.6	13.6	12.8
9	aaaaactctctaaaaactctctagaggggccctagagggccc	17.5	7.2	13.1	16.4	5.1	4.6	13.3	13.2
10	agaattgggacaaaaattggaaatttttaaggg - *C. risortia* bent sat. DNA	18.5	8.9	8.7	12.4	3.3	3.0	13.5	12.6
11	$(aaaaactctctaaaaactctcgaggggccctagagggcccta)_n$	21.6	5.2	10.1	14.5	5.0	4.7	13.8	10.8
Straight DNA									
12	$(atctaatctaacacaacaca)_n$	0.8	0.5	2.7	1.2	5.1	4.4	0.8	0.8
13	actacgttaaatctatcaccgcaagggataaa - OR3 operator region	10.4	5.5	4.9	5.9	5.0	4.4	10.4	7.8
14	actacgttaaatctatcacc*a*caagggataaa - OR3 region, mutated	11.0	5.5	3.4	6.2	4.9	4.4	10.6	8.1
15	$(a)_n$ - poly-A[1]	0.008	0.000	0.008	0.000	0.100	0.063	0.002	0.002
16	$(ttttaaaacg)_n$	1.5	7.1	14.5	10.7	2.8	2.5	2.3	4.2
17	$(ttttaaaagc)_n$	1.7	0.8	16.0	16.3	3.6	3.3	3.0	9.6
18	$(aaaaactctctaaaaactctcgggccctagaggggccctaga)_n$	27.1	3.1	5.9	7.5	4.9	4.6	12.8	8.3

The angular deflection [°/10.5 bp hel.turn] was calculated with the bend.it server (the curvature units of Trifonov et al. correspond to 4.5°/bp or 47.25°/10.5 bp hel.turn).
[1]By definition, homopolymers should give zero curvature. The nonzero value indicates the numeric precision of the calculation.

Simple elastic models

Elastic bodies are those that regain their original shape after the stress is removed and in which the displacement (distortion) is proportional to the force applied. Single-molecule experiments show that these conditions are in fact met for small distortions of DNA molecules. If a DNA rod is ideally elastic one can compute the energy necessary for bending, stretching or torsional deformation [20]. For example, the energy required to bend the rod of length L to a given curvature can be calculated as

$$\Delta G = \tfrac{1}{2} E I L \alpha^2. \quad (1)$$

Here α is the curvature as defined in Figure 1A and I is the moment of inertia that, for a cylindrical rod of r radius is given as $\pi r^4/4$. E is the Young's modulus. Based on a variety of physico-chemical measurements and direct molecular stretching studies, the Young's modulus of average DNA is about 3.4×10^8 N/m^2, a value close to that of polypropylene or phenol resins.

By 'simple elastic models' we mean those that consider DNA as an originally straight cylindrical rod with one *homogeneous isotropic elasticity* parameter (for extensive reviews see [2, 4, 21]). This means on the one hand that the model is not sequence-dependent, that is all segments are equal, and on the other hand, the model is equally bendable (deformable) in all directions. The phenomena that can be described with such a simplified elastic model include the gross shape changes of DNA like supercoiling, response of plasmids to stress, etc. The nature of the answer is qualitative, for example one can state that the shape of a plasmid-like elastic ring model adopts, in response to torsional stress, shapes that are reminiscent of the shapes observed for supercoiled plasmids. This fact shows that some properties of DNA are in fact reminiscent of those of rubber strings, that is, depend mostly on properties that are common to simple mechanical systems. The underlying, very complex mathematics can be performed by *finite element analysis*, a technique routinely used for the analysis of stresses and displacements in engineering [22]. This technique has been applied with success for small DNA displacements [23, 24]. Even though simple elastic models do not contain local (i.e. sequence-dependent) information, they allow on the other hand to demonstrate how local structures might act on the elastic behaviour of the entire molecule. For example, incorporating a fixed curve into a DNA model will influence the gross shape of plasmids [25] and will bring distant sites into each other's vicinity [26]. One can subjectively conclude that the isotropic elastic models can, efficiently model an impressive range of phenomena connected with the propagation of stress along the molecule.

SDAB: A sequence dependent anisotropic model of DNA bendability

For the modelling of local phenomena sequence dependence has to be incorporated into the elastic models. It is also known that bending is anisotropic [27, 28], that is DNA bends much more easily towards the major groove than in any other direction, moreover, bending in the tilt direction is not favourable in energetic terms (for a review see [2]). We have developed anisotropic bendability parameters using the enzyme DNaseI [29]. This enzyme bends DNA towards the major groove and binds virtually to all DNA sites without pronounced sequence specificity. DNaseI cutting rates can thus be used as an estimate of DNA bendability, which in turn can be scaled to an approximate DNA rigidity scale [30]. The result is a simplified segmented rod model, shown in Figure 1B. In this sequence dependent anisotropic bendability (SDAB), each disk corresponds to one base pair and the arrow indicates the direction of facilitated flexibility (i.e. that of the major groove). In principle, such an anisotropic bendability model can have different bending flexibility in all directions. As a simplification, we take bendability towards the major groove as the principal parameter, and consider DNA 10-times stiffer in all other directions (see sketch in Figure 1C). The model does not include static deflection components, that is similar to isotropic models, it considers DNA as an originally straight rod. Neither does the model incorporate torsional flexibility. In other terms, it is a stripped down model that is designed to reflect one aspect: local bending phenomena. In contrast to the isotropic elastic models, the SDAB model is non-linear in terms of displacement response. The non-linearity of the SDAB model is due to two factors: (i) Material non-linearity is the non-linear bending characteristics of each base pair, that we term bending anisotropy; (ii) Geometric non-linearity occurs in the case of non-negligible displacements. Finite element analysis is very appropriate to study such non-linear problems.

The SDAB model pictures DNA as an anisotropically bendable rod whose bendability is deduced from

DNaseI experiments. However, the concept bendability was used also previously, by Travers and co-workers who determined the frequency of trinucleotides on the inside and outside faces of DNA bent around the nucleosome particle. These frequency data were then considered as a measure of bendability (without an explicit model similar to the one outlined here). The ranking of trinucleotides is however slightly different according to the DNaseI and nucleosome scales, respectively.

The first attempt to incorporate the nucleosome frequency data into a physical model was that of Goodsell and Dickerson who devised a static rod model in which the roll angles were derived from rescaled nucleosome frequency data. This model was quite efficient in predicting, for example GC-type curvature, and Goodsell and Dickerson designed a simple and fast algorithm, BEND, for the prediction of curvature. This gave us the idea to use the DNaseI data in an analogous manner, using the BEND algorithm. In the following sections we apply the SDAB model first to three problems: (i) Estimation of bending energies and their correlation with observable quantities (sequence-dependent curvature, protein–DNA binding free energy). (ii) Modelling of vibrational phenomena and finally (iii) Prediction of curvature from sequence.

SDAB: Calculation of bending energy

Modelling of minicircles of curved and straight DNA [31]

A simple experiment can show the macroscopic anisotropy of the SDAB model (Figure 2). A rod model is bent to the same extent into various directions as shown in the inset, and the bending energy is determined by finite element methods and plotted against the direction of bending (angle β). Sequences that are repeats of curved DNA motifs, in fact, will have a rotational preference, that is there will be one stable energy minimum. This means that such a rod-model has a preferred direction of bending, so as a result of thermal fluctuations it will preferentially bend into one direction. In other terms, the physically measurable average conformation of such a model will be curved. The straight motifs, on the other hand, have either no minima or have two minima in opposite angular directions.

Another consequence of the minimum energy found in the circles is that, in the minimum energy conformation of helically phased repeats, certain motifs will face inward and the others outward. For example, in repeats of AAAAGGGCCC, the major groove of the GGGCCC motif faces inwards and the major groove of AAAAA are on the outer side of the circle, as found, for example in nucleosomes. The roll values at the central GC are the highest while there are slight negative rolls in the AAAAA tract Figure 2B. As the roll/twist profile of the lowest energy conformation illustrates, the shape of the circle is reminiscent of a polygon – in fact quite similar to those recently obtained for the same molecule using AFM microscopy [32]. This means that the lowest energy conformation has slight kinks, which is reminiscent of the mini-kink model postulated by Zhurkin and co-workers [28]. The high-energy conformers on the contrary, have smaller kinks (Figure 2C).

Protein/DNA binding: DNA rigidity versus complex stability [33]

Repressor proteins bind to short DNA motifs with high specificity and the DNA is often bent in the resulting protein/DNA complex since bending is induced by the binding of the protein. We can consider a simple model (Figure 3A) in which Cro first binds to the oligonucleotide in a non-specific manner and reduces the free movement (thermal fluctuations) of DNA, which results in an entropy loss. Since the elastic entropy can be calculated from the $\langle\theta^2\rangle^{1/2}$ root mean square fluctuations of the model, the entropy change can be calculated as

$$\Delta S = nR \ln\left[\frac{\langle\theta^2_{\text{bound}}\rangle^{1/2}}{\langle\theta^2_{\text{free}}\rangle^{1/2}}\right] = nR \ln\left[\frac{E_{\text{free}}}{E_{\text{bound}}}\right], \quad (2)$$

where E_{free} is the average Young's modulus of the segment, n is the number of degrees of freedom and E_{bound} is the Young's modulus of the bound (quasi immobilised) DNA. Since E_{free} is smaller than E_{bound}, this equation shows an adverse relationship. In the second step, the cognate will bend to a curvature of 78.5 degrees found in the crystal structure of the Cro/DNA complex. This second step will take place only for cognate DNA. Takeda and co-workers have measured a set of free energy data for cognate and non-cognate DNA that showed a puzzling diversity of values. If we now calculate the free energy values according to the simple model outlined here (Figure 3A) the cognate and non-cognate molecules will clearly separate into two groups, that in turn clearly follow the tendencies predicted by Equation 1 and 2. Namely,

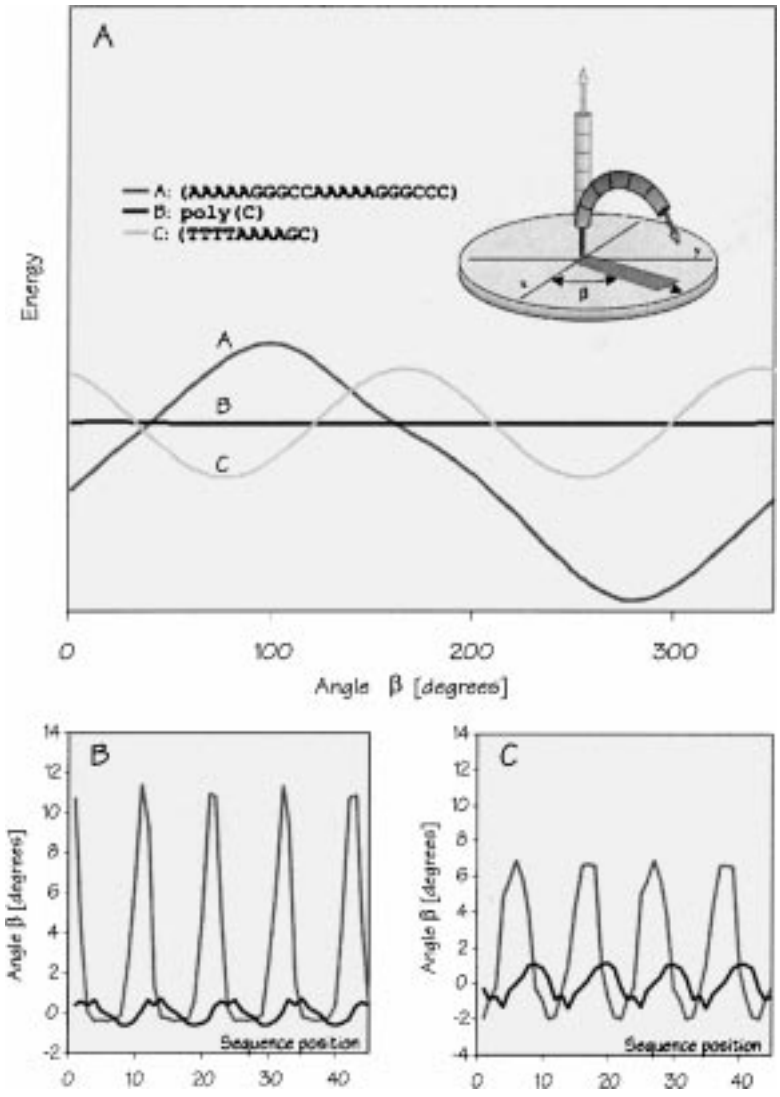

Figure 2. Testing bending anisotropy in DNA by finite element methods30. A DNA model was built (as outlined in Figure 1C) from the repeat motifs shown. The models were bent to a given curvature in various directions [angle β in A] and the bending energy of the model was calculated for each direction by finite-element methods. Curved motifs (e.g. motif 1) exhibit a single energy minimum, which corresponds to a bending preference in one direction. Straight motifs have either no minimum (motif 2) or minima in two opposite directions (motif 3). The calculations were carried out using the Cosmos/M program (release 1.75A); the curvature value was uniformly 3.43° per base pair, and the energy was normalised to one base-pair unit, for better comparison [31].

cognate (operator) and non-cognate (non-operator) DNA follow two adverse, quasi-linear relationships. In the operator sequences, DG is higher for stiffer molecules, that is the stiffer the molecule, the weaker the binding (energy is wasted on bending a stiff oligonucleotide). This in fact can be expected since Cro has to curve the molecule, and the energy required is linearly proportional to the stiffness (Equation 1).

The slope of the curve corresponds to a curvature actually found in the crystal structure [33]. In non-operator sequences, on the other hand, ΔG is lower for stiffer sequences, that is the stiffer the sequence, the stronger the binding. Namely, relatively less energy is needed for restricting a rigid molecules movement.

SDAB: Modelling vibrational phenomena in DNA

The interest in vibrational properties of DNA is understandable, as it has been speculated for a long time that energy can propagate along the chain not only in terms

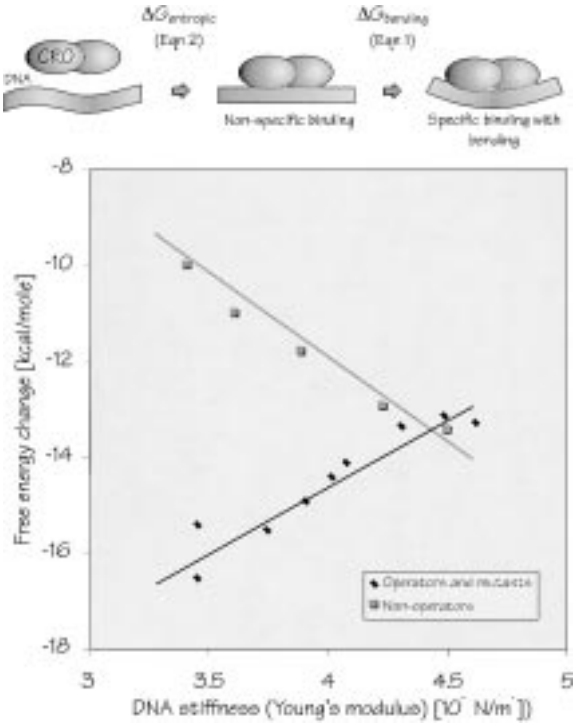

Figure 3. (A) A simple model of Cro/DNA interactions (B) Binding of Cro operator to oligonucleotides of different stiffness. Relationship between the average stiffness (Young's modulus) of DNA and the free energy change (ΔG) of operator (blue) and non-operator (red) DNA sequences.[33].

of stress (like in supercoiling or bending) but also as waves in a vibrating string. Consequently sequence-dependent vibrational differences may be responsible in promoting or blocking energy flow along the DNA molecule.

Even though to our knowledge, there is no convincing experimental evidence to this supposition, there is a vast body of theoretical modelling work on the vibrational behaviour of DNA. The non-linear dynamic models used are quite different from the rod-models and usually do not consider sequence-dependence. Moreover, DNA is usually considered as vibrating *in vacuo* and effects like solvation and binding of molecular ligands can not easily be taken into consideration. In this section, we show preliminary results on the vibrational properties of the SDAB model of DNA.

Finite element methods make it possible to calculate natural vibration frequencies and/or dynamic response of a mechanical model. However, the SDAB model is a non-linear case since the bending flexibility is different in various directions. There are possibilities to derive approximate analytical solutions for the calculation of normal modes in this particular ('bi-linear') case, however numerical solutions are also straightforward and are widely used in the vibrational analysis of mechanical structures. In these calculations, the mechanical model is first exposed to an initial bending impact or bending, and then it is left to vibrate by itself. The frequencies are then determined by Fourier analysis of the displacements in the time domain. These vibrations are calculated for a model *in vacuo*. The surrounding environment can be modelled as a viscous medium that damps the vibrations – in principle damping is not expected to influence very much the frequencies as the damping ratio has small values. Figure 4 shows a simple modelling experiment. A model of 32 residues was fixed on one end as shown in the inset and the free end was deflected to a moderate curvature of 7 degrees per helical turn and then left to vibrate. The movement of the free end was analysed by Fourier analysis to give the frequency spectrum shown here. The frequency values in such an experiment depend on the size of the molecule so they are not expected to correspond to physically measurable frequencies. However, the frequencies as well as the amplitudes (intensity) of the vibrations do depend on the sequence. Whether or not the differences are meaningful in the physical sense, will have to be determined by experimental work.

SDAB model: Prediction of bendable and curved segments

The single energy minimum in Figure 2 is a consequence of two facts: the base pairs are themselves anisotropically bendable and, moreover, they are in a sequence that dictates that the direction of principal bendability is along one face of the DNA helix. This can be best illustrated by a vector representation in helical circle diagrams – technique originally developed for amphipathic α-helices in proteins32 (Figure 5A). In these diagrams, the bendability value for each base pair is plotted as a vector pointing towards the major groove. In fact, inherently curved motifs show asymmetrical distribution of bendability; large vectors are on one side (Figure 5A). The vectors for base pairs in straight sequences, by contrast, have a rather symmetrical distribution. This amounts to saying that in randomly chosen DNA the thermal fluctuations in various directions will cancel out, while inherent bending is dependent upon a specific sequence. It is interesting to note that there is a mathematical

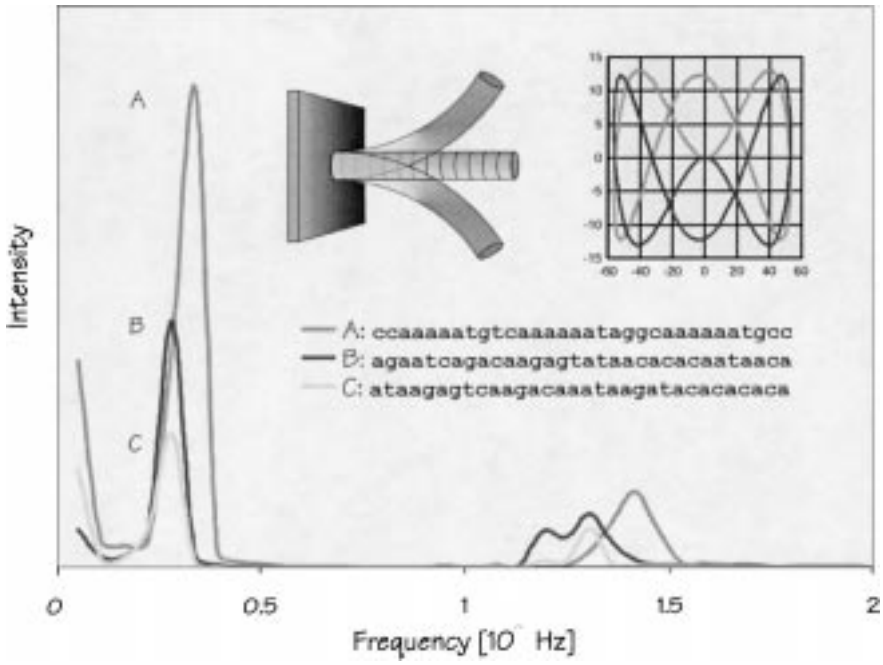

Figure 4. Sequence-dependent vibrations of DNA-models. A is an inherently curved sequence from *Leishmania tarentolae*, B and C are random shuffled variants of the same sequence, with no predicted curvature.

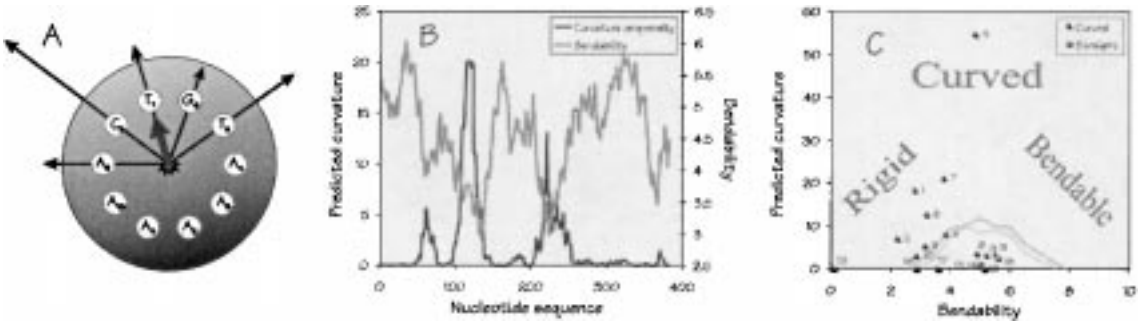

Figure 5. (A)Vectorial representation (helical circle diagram) of DNA bendability in a curved sequence motif [(A)AAATGTCAAA(A)] from a *Leishmania tarentolae* class II minicircle. The length of each black arrow is proportional to the magnitude of the bendability parameter at a given sequence position. The gray arrow is the vectorial average of the bendability vectors and is considered to be a measure of predicted curvature. (B) Predicted curvature and bendability against sequence (accession code Genbank LEIKPMNC2) plot. The values are calculated for 32-bp-sequence windows and plotted at the starting point of each window. (C) Predicted curvature against bendability plots of curved and straight sequences from Table 1. The lines indicate the border of random sequences obtained by random shuffling of the sequences of *Haemophilus influenzae* genome and yeast chromosome III.

analogy between bent DNA and amphipathic α-helices in proteins. Helical circle diagrams were originally developed for amphipathic α-helices in proteins. Furthermore, the vectorial average shown in the figures is analogous to the hydrophobic moment of Eisenberg [34].

Given the fact that the bendability of the SDAB model is asymmetrical towards the major groove, thermal fluctuations will result in bending of the model. How can one calculate the curvature from these values? Gabrielian et al. [35] used the length of the vectorial average (as shown in Figure 3A) as a measure of predicted curvature; however, more rigorous geometry calculations can also be used: bendability values (see Table 1) can be considered to be proportional to (but not necessarily identical with) static trinucleotide roll values. In other words, the bendability parameters can be considered to be analogous to a static geometric

model in which the dynamic contribution of thermal fluctuations is included.

With this assumptions, the SDAB model gives accurate predictions of DNA curvature that compare quite favourably to those of the static geometry models (see Table 1). Table 1 also includes figures calculated using the so-called consensus bendability scale14, which was developed in order to increase the sensitivity of the prediction towards GC-based curved motifs – these are often mispredicted using both the static dinucleotide models15 and the original DNaseI scale.

One application of this prediction is to analyse long DNA sequences, segment by segment, and then to visualise the distribution of predicted curvature (Figure 5B). We have developed simple visualisation methods in which several sequence-dependent properties can be combined, either as 1D-plots (along the sequence). In a 2D plot two properties of a segment (like average bendability and predicted curvature) are plotted against each other so each segment of a genome will produce a point in the plane (Figure 5C). The distribution of these points is genome specific, and also it allows one to visualise 'outliers' that is segments and regions that are different from the rest of genome. For example, protein coding and non protein coding regions show quite different distributions, extremely curved, rigid or flexible segments are much more frequent in the latter than in the former (Figure 6). Several of these calculation methods are now included into an Internet-based server *http://www.icgeb.trieste.it/dna*

A second, related, application of the method is to visualise curved DNA molecules. Once such a potential motif is identified in the sequence, the experimenter may wish to see what it looks like in three dimensions. One possibility – now included in the www server – is to reconstruct atomic co-ordinates from the idealised trajectory calculated by a simplified rod model. Standard modelling programs can be used to generate the atomic co-ordinates which then have to be refined with a combination of energy minimisation by restricted simulated annealing (molecular dynamics) and geometry optimisation. The latter steps are necessary because introduction of bends into a DNA molecule leads to atomic clashes (especially in the backbone). This method accurately reproduces the trajectory predicted by the rod model and the phosphate atoms are reproducibly re-positioned by the simulated annealing procedure. However, the accuracy of these predicted atomic models is of course limited and in our view they can be used mainly for the visualisation

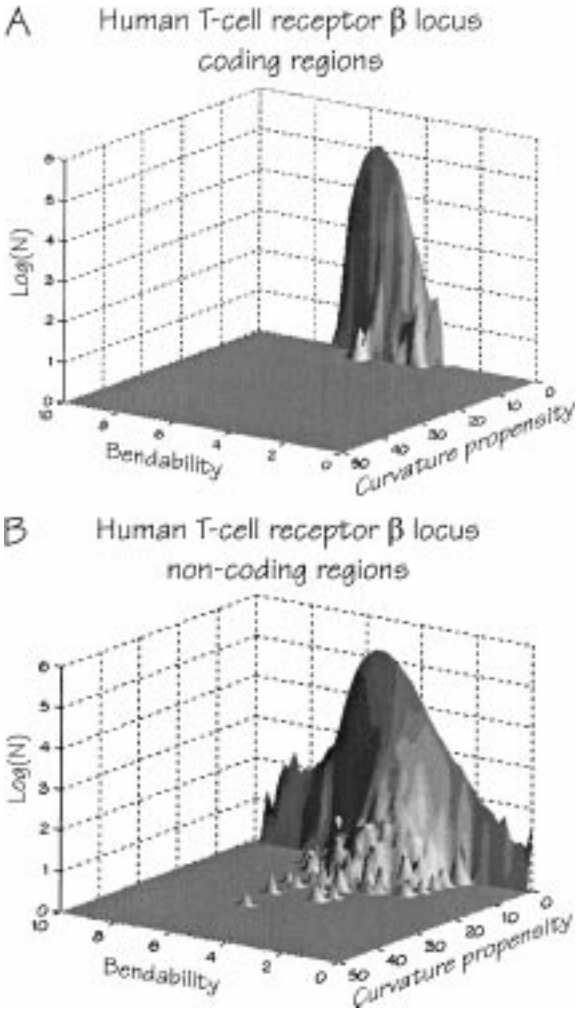

Figure 6. 3D Curvature propensity vs. bendability histogram of the Human T-cell receptor locus (Genbank: humtcrb). The protein coding regions A have a tight distribution around average bendability and low curvature. The non-protein-coding regions B have a higher number of stiff (low bendability), flexible and curved segments.

of the sidedness of the curve, that is localisation of the major and minor grooves as well as sequence motifs with respect to the curvature.

Conclusions and future directions

We can conclude that static and dynamic rod models describe quite different aspects of the DNA molecule with as few parameters as possible. If they succeed – and they do, quite surprisingly – it means that the parameters used by the model sufficiently explain a given aspect of the molecule's behaviour. Both type of models can predict curvature in short DNA segments –

in this respect they can be considered equivalent [36]. The differences in predictive accuracy may be due to the parameterisation (e.g. trinucleotide versus, dinucleotide representation, electrophoresis versus nucleosome data, etc.) and not to the models themselves. Considering the significance of these structural features, it is worth to mention that bendability/curvature characteristics are conserved in evolutionarily [31, 35] and functionally related sequences [37], and they correlate well with regulatory sites known from experiment [38]. So one can predict that these methods will have a continued role in locating functional sites in genomes and will hopefully contribute to our understanding how the putative conformational signals may operate at the genomic level.

Refinements of rod models, especially the incorporation of tetranucleotide-based description will probably increase the predictive accuracy and increase the scope of applications. Future models may combine both static and dynamic features with particular reference to the principles of non-linear DNA dynamics [39] . At the same time molecular mechanic models are also expected to expand to a wider range of local phenomena as illustrated recently by Lavery and associates [40]. In the foreseeable future however, the simple rod models will continue to play a key role in the large-scale analysis of genomic data.

References

1. Travers A.A. and Klug A.: Cold Spring Harbor laboratory, Cold Spring Harbor, Bending of DNA. In: Cozzarelli N.R. and Wang J.C. (eds) *DNA Topology and its Biological Effects.* (1990) pp. 57–106.
2. Olson W.K. and Zhurkin V.B.: Twenty years of DNA bending. In: Sarma R.H. and Sarma M.H. (eds), *Biological Structure and Dynamics*. Adenine Press, Schenectady, Vol. 2, (1996), pp. 341–370.
3. Olson W.K.: Simulating DNA at low resolution, *Curr. Opin. Struct. Biol.* **6** (1996): 242–256.
4. Langowski J., Olson W.K., Pedersen S.C., Tobias I., Westcott T.P. and Yang Y.: DNA supercoiling, localized bending and thermal fluctuations (letter) *Trends. Biochem. Sci.* **21** (1996): 50.
5. Trifonov E.N. and Sussman J.L.: The pitch of chromatin DNA is reflected in its nucleotide sequence, *Proc. Natl. Acad. Sci. USA* **77** (1980): 3816–3820.
6. Wu H.M. and Crothers D.M.: The locus of sequence-directed and protein-induced DNA bending, *Nature* **308** (1984): 509–513.
7. Satchwell S.C., Drew H.R. and Travers A.A.: Sequence periodicities in chicken nucleosome core DNA, *J. Mol. Biol.* **191** (1986): 659–675.
8. Brukner I., Dlakic M., Savic A., Susic S., Pongor S. and Suck D.: Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements, *Nucleic Acids Res.* **21** (1993): 1025–1029.
9. Bolshoy A., McNamara P., Harrington R.E. and Trifonov E.N.: Curved DNA without A–A: Experimental estimation of all 16 DNA wedge angles, *Proc. Natl. Acad. Sci. USA* **88** (1991): 2312–2316.
10. De Santis P., Palleschi A., Savino M. and Scipioni A.: Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature, *Biochemistry* **29** (1990): 9269–9273.
11. Olson W.K. et al.: Flexing and folding double helical DNA, *Biophys. Chem.* **55** (1995): 7–29.
12. Bansal M., Bhattacharyya D. and Ravi B.: NUPARM and NUCGEN: Software for analysis and generation of sequence dependent nucleic acid structures, *Comput. Appl. Biosci.* **11** (1995): 281–287.
13. Ulyanov N.B. and James T.L.: Statistical analysis of DNA duplex structural features, *Methods Enzymol.* **261** (1995): 90–120.
14. Gabrielian A. and Pongor S.: Correlation of intrinsic DNA curvature with DNA property periodicity, *Febs. Lett.* **393** (1996): 65–68.
15. Goodsell D.S. and Dickerson R.E.: Bending and curvature calculations in B-DNA, *Nucleic Acids Res.* **22** (1994): 5497–5503.
16. el Hassan M.A. and Calladine C.R.: The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme *J. Mol. Biol.* **251** (1995): 648–664.
17. Wheeler D.: A gel-concentration-independent retardation detected in two fragments of the rrnB P1 promoter of E. coli using transverse polyacrylamide pore gradient gel electrophoresis, *Biochem. Biophys. Res. Commun.* **193** (1993): 413–419.
18. Shpigelman E.S., Trifonov E.N. and Bolshoy A.: CURVATURE: Software for the analysis of curved DNA, *Comput. Appl. Biosci.* **9** (1993): 435–440.
19. Dlakic M.: DIAMOD: Display and modeling of DNA bending, *Bioinformatics* **14** (1998): 326–331.
20. Barkley M. and Zimm B.: Models of DNA, *J. Chem. Phys.* **70** (1979): 2991–2997.
21. Vologodskii A.V. and Frank-Kamenetskii M.D.: Modeling supercoiled DNA, *Methods Enzymol.* **211** (1992): 467–480.
22. Zienkiewicz O.C. and Taylor R.L.: *The Finite Element Method*. McGraw-Hill, (1991).
23. Bauer W.R., Lund R.A. and White J.H.: Twist and writhe of a DNA loop containing intrinsic bends, *Proc. Natl. Acad. Sci. USA* **90** (1993): 833–837.
24. Yang Y., Tobias I. and Olson W.K.: Finite element analysis of DNA supercoiling, *J. Chem. Phy.* **98** (1993): 1673–1686.
25. Zhang P., Tobias I. and Olson W.K.: Computer simulation of protein-induced structural changes in closed circular, *J. Mol. Biol.* **242** (1994): 271–290.
26. Rippe K., von Hippel P. and Langowski J.: Action at a distance: DNA-looping and initiation of transcription, *Trends Biochem. Sci.* **20** (1995): 500–506.
27. Schellman J.A.: Flexibility of DNA, *Biopolymers* **13** (1974): 217–226.
28. Zhurkin V.B., Lysov Y.P. and Ivanov V.I.: Anisotropic flexibility of DNA and the nucleosomal structure, *Nucleic Acids Res.* **6** (1979): 1081–1096.
29. Brukner I., Sanchez R., Suck D. and Pongor S.: Sequence-dependent bending propensity of DNA as revealed by DNase I: Parameters for trinucleotides, *Embo J.* **14** (1995): 1812–1818.

30. Gromiha M.M., Munteanu M.G., Gabrielian A. and Pongor S.: Anisotropic elastic bending models of DNA, *J. Biol. Phys.* **22** (1996): 227–243.
31. Gabrielian A., Vlahovicek K., Munteanu M.M., Gromiha M.M., Brukner I., Sanchez R. and Pongor S.: Prediction of bendability and curvature in genomic DNA. In: Sarma R.H. and Sarma M.H. (eds), *Structure, Motion, Interaction and Expression*, Adenine Press, Cininnati, NY, Vol. 1, 1998, pp. 117–132.
32. Han W., Lindsay S.M., Dlakic M. and Harrington R.E.: Kinked DNA (letter), *Nature* **386** (1997): 563.
33. Gromiha M.M., Munteanu M.G., Simon I. and Pongor S.: The role of DNA bending in Cro protein–DNA interactions, *Biophys. Chem.* **69** (1997): 153–160.
34. Eisenberg D., Schwarz E., Komaromy M. and Wall R.: Analysis of membrane and surface protein sequences with the hydrophobic moment plot *J. Mol. Biol.* **179** (1984): 125–142.
35. Calladine C.R. and Drew H.R.: A useful role for 'static' models in elucidating the behaviour of DNA in solution, *J. Mol. Biol.* **257** (1996): 479–485.
36. Gabrielian A., Simoncsits A. and Pongor S.: Distribution of bending propensity in DNA sequences, *Febs. Lett.* **393** (1996): 124–130.
37. Schatz T. and Langowski J.: *J. Biomol. Struct. Dynam.* **15** (1997): 265–275.
38. Langst G., Schatz T., Langowski J. and Grummt I.: Structural analysis of mouse rDNA: Coincidence between nuclease hypersensitive sites, DNA curvature and regulatory elements in the intergenic spacer, *Nucleic Acids Res.* **25** (1997): 511–517.
39. Yakushevich L.V.: Nonlinear DNA dynamics: Hierarchy of the models *Physica D.* **79** (1994): 77–86.
40. Sanghani S.R., Zakrzewska K., Harvey S.C. and Lavery R.: Molecular modelling of (A4T4NN)n and (T4A4NN)n: sequence elements responsible for curvature, *Nucleic Acids Res.* **24** (1996): 1632–1637.
41. Gorin A.A., Zhurkin V.B. and Olson W.K.: B-DNA twisting correlates with base-pair morphology, *J. Mol. Biol.* **247** (1995): 34–48.