

- 32 Hiyama, H., Iavarone, A. and Reeves, S. A. (1998) *Oncogene* 16, 1513–1523
- 33 Chinery, R. *et al.* (1997) *Nat. Med.* 3, 1233–1241
- 34 Somasundaram, K. *et al.* (1997) *Nature* 389, 187–190
- 35 Zeng, Y-X., Somasundaram, K. and El-Deiry, W. S. (1997) *Nat. Genet.* 15, 78–82
- 36 Mal, A. *et al.* (1996) *Nature* 380, 262–265
- 37 Zerfass-Thome, K. *et al.* (1996) *Oncogene* 13, 2323–2330
- 38 Funk, J. O. *et al.* (1997) *Genes Dev.* 11, 2090–2100
- 39 Jones, D. L., Alani, R. M. and Münger, K. (1997) *Genes Dev.* 11, 2101–2111
- 40 Li, X-S. *et al.* (1996) *Cancer Res.* 56, 5055–5062
- 41 Timchenko, N. A. *et al.* (1997) *Mol. Cell. Biol.* 17, 7353–7361
- 42 Maki, C. G. and Howley, P. M. (1997) *Mol. Cell. Biol.* 17, 355–363
- 43 Chen, I-T., Smith, M. L., O'Connor, P. M. and Fornace, A. J., Jr (1995) *Oncogene* 11, 1931–1937
- 44 Warbrick, E., Lane, D. P., Glover, D. M. and Cox, L. S. (1997) *Oncogene* 14, 2313–2321
- 45 Chuang, L. S-H. *et al.* (1997) *Science* 277, 1996–2000
- 46 Chen, J., Chen, S., Saha, P. and Dutta, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 11597–11602
- 47 Levin, D. S. *et al.* (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 12863–12868
- 48 Watanabe, H. *et al.* (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 1392–1392
- 49 Alevizopoulos, K., Vlach, J., Hennecke, S. and Amati, B. (1997) *EMBO J.* 16, 5322–5333
- 50 Hofmann, F. and Livingston, D. M. (1996) *Genes Dev.* 10, 851–861

## Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena

Mircea G. Munteanu, Kristian Vlahoviček,  
Subbiah Parthasarathy, István Simon and  
Sándor Pongor

Local bending phenomena can be predicted by elastic models that incorporate sequence-dependent anisotropic-bendability (SDAB). SDAB models consider DNA to be an initially straight, segmented, elastic rod, in which the flexibility of each segment is greater towards the major groove than it is in other directions. While local bending can be predicted by static-geometry models as well, SDAB models, in addition, qualitatively explain such phenomena as the affinity of protein binding and kinking. A set of prediction tools is available at <http://www.icgeb.trieste.it/dna>

**LOCAL BENDING OF DNA** can contribute extensively to the specificity of biological events such as gene regulation and packaging<sup>1</sup>. In contrast to traditional structural polymorphism (e.g. the B-, A- and Z-DNA structures), bending is a localized micropolymorphism in which the original B-DNA structure is distorted but is not modified extensively. Broadly speaking, the DNA segments that are involved in the protein-induced and/or inherent DNA bending that occurs

in many promoters, enhancers and silencers are about 10–50 bp in length. DNA molecules in this size range are difficult to model because they are longer than those that can be described easily by atomic-resolution molecular modelling or quantum-mechanical approaches. Equally, they are shorter than those that can be meaningfully handled by traditional elastic models, which successfully describe macroscopic behaviour (such as supercoiling) in longer DNA segments<sup>2,3</sup>. Also, 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 tell us much about these conformations. Here, we review briefly the advantages and limitations of

rod models of DNA, particularly with regard to elastic modelling of local bending phenomena.

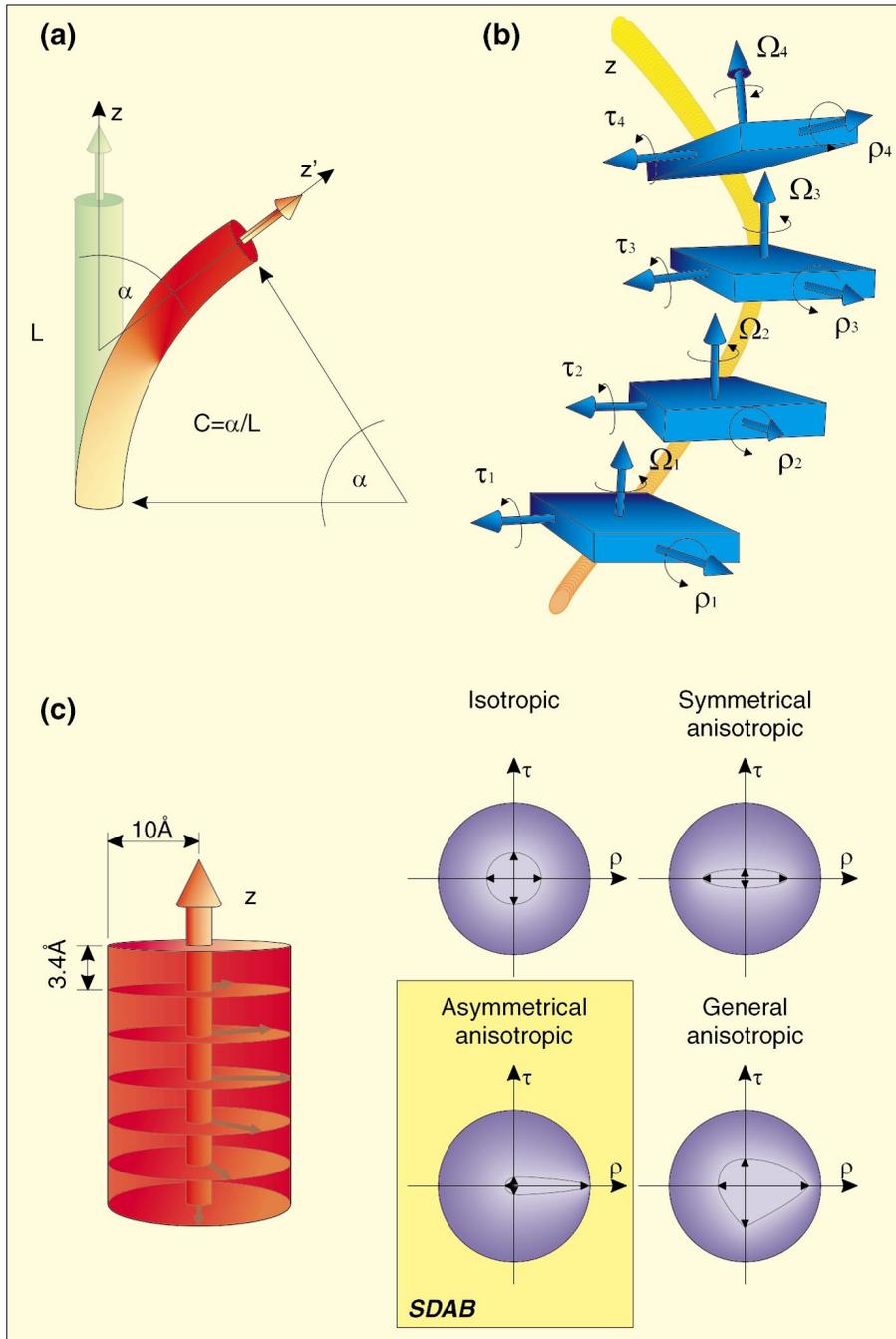
### Static-geometry models

Rod models are the simplest form of DNA models and represent DNA as a cylindrical rod of constant diameter. The shape, in this case, is the path or trajectory of the longitudinal z-axis, which can be either straight or curved (Fig. 1a). The common philosophy of rod models is to divide the rod into short cylindrical segments (e.g. the size of a base pair) and then to compute a given rod parameter on the basis of segment parameters that have to be known *a priori*. Dinucleotide models define the base-pair-size unit as two adjacent base pairs. There are therefore 16 possible units, or 10 if we allow strand symmetry (e.g. AA = TT). Trinucleotide models define the unit around the central base pair of a given trinucleotide. This yields 64 or 32 different units, again depending on whether or not strand symmetry is allowed.

Static models are rigid rod models that only consider the static geometry of a segment. Curvature in B-DNA was originally believed to be a consequence of  $A_n$  ( $n = 4-6$ ) tracts that were repeated in phase with the helical repeat in DNA. Two static models were proposed initially, to explain the phenomenon. In the nearest-neighbour model, the axial deflections of successive AA/TT dinucleotides sum to produce a curve<sup>4</sup> (Fig. 1b). In the so-called junction model, curvature is produced at the junction between the modified B-DNA structure (consisting of  $A_n$  tracts) and adjacent unmodified B-DNA structure<sup>5</sup>. More recently, it became clear that DNA curvature also involves other sequence elements<sup>6</sup>, and more sophisticated models, which included different geometries for all 16 dinucleotides, were proposed<sup>7,8</sup>.

Current static models consider dinucleotide geometries that are derived from direct measurements such as X-ray crystallography<sup>9-12</sup> or NMR<sup>13,14</sup>, from

**M. G. Munteanu, K. Vlahoviček, S. Parthasarathy** and **S. Pongor** are at the International Centre for Genetic Engineering and Biotechnology, Padriciano 99, 34012 Trieste, Italy; and **I. Simon** is at the Institute of Enzymology, BRC, Hungarian Academy of Sciences, H-1518 Budapest, PO Box 7, Hungary.  
Email: [pongor@icgeb.trieste.it](mailto:pongor@icgeb.trieste.it)



**Figure 1**

(a) An ideally elastic rod bent in a circular arc. The curvature ( $C$ ) is calculated from the deflection angle ( $\alpha$ ) and the length ( $L$ ) of the rod. (b) Schematic calculation of trajectory using a static-geometry model. The deflection can be calculated from the roll ( $\rho$ ), tilt ( $\tau$ ) and twist ( $\Omega$ ) values. (c) Sequence-dependent anisotropic-bendability (SDAB) models of DNA. Each element corresponds to one base pair. The arrow in each base pair points towards the major groove. The arrow length is proportional to the flexibility in a given direction ( $+\rho/-\rho$  and  $+\tau/-\tau$ ). In the SDAB model, the rod is more flexible in one direction (that of the major groove); the other three models are uniformly more rigid<sup>26</sup>.

statistics of nucleosome-binding data<sup>15</sup> or from gel-electrophoretic analysis of concatenated synthetic oligonucleotide repeats<sup>7,8</sup>. The parameters that are considered explicitly are roll, tilt and twist angles (see Fig. 1b); the values of all other parameters are considered to be equal to those of B-DNA. Given the geometry of the segments, one can predict the trajectory of a DNA molecule by

summing the distortions in successive overlapping dinucleotide segments. A number of programs that can calculate an approximate trajectory for a given DNA sequence are available<sup>8,10,16,17</sup>. Table I shows that current dinucleotide models give values that correlate well with the experimentally obtained values: curved and straight motifs can be distinguished, although all the models mis-

predict the curvature of some motifs. Importantly, the DNA gel-mobility anomaly that is used to validate the models is itself strongly dependent on environmental factors, such as metal ions and temperature. Our knowledge of straight and curved motifs is therefore qualitative rather than quantitative.

**Simple elastic models**

If a rod is ideally elastic (i.e. it will return to its original shape after deformation; see Box 1), one can compute the energy necessary for bending, stretching or torsional deformation<sup>18</sup>. For example, the energy ( $\Delta G$ ) that is required in order to bend a rod of length  $L$  to a given angle  $\alpha$  (see Fig. 1a) can be calculated:

$$\Delta G = \frac{1}{2} EIL\alpha^2 \quad (1)$$

In the above expression,  $E$  is the stiffness parameter (also known as Young's modulus; see Box 1);  $I$  is the moment of inertia that, for a cylindrical rod of radius  $r$ , is given by  $\pi r^4/4$ .

By simple elastic models we mean those that consider DNA to be a straight, cylindrical rod that has a single stiffness parameter<sup>3,19</sup>. This means, on the one hand, that the model is not sequence dependent (i.e. all segments are equal) and, on the other, that the model is equally bendable (deformable) in all directions. The phenomena that can be described using such a simplified elastic model include gross shape changes in DNA, such as supercoiling, the response of plasmids to stress, etc.

The nature of the predictions that are obtained by using these models is qualitative. For example, one can show that, in response to torsional stress, a plasmid-like elastic ring adopts a shape that is reminiscent of those observed for supercoiled plasmids. This indicates that some properties of DNA are, in fact, reminiscent of those of elastic bands (i.e. they depend mostly on properties that are common to simple mechanical systems). The underlying, quite complex mathematics can be avoided by finite-element analysis – a technique that is routinely used for the analysis of deformations in engineering (see Box 1)<sup>20</sup>. This technique has been applied successfully for small DNA deformations<sup>21,22</sup>. Simple elastic models can simulate how local structures affect the elastic behaviour of a larger molecule. For example, incorporating a fixed inhomogeneity into a DNA model might influence the gross shape of plasmids<sup>23</sup> and might bring distant sites closer together<sup>23,24</sup>.

### Anisotropic, sequence-dependent elastic models

In order to model local bending phenomena, starting from a base sequence, one must incorporate sequence dependence into the elastic models. Brukner and associates<sup>25</sup> have developed trinucleotide bendability (see Box 1) parameters by using the enzyme DNaseI (Ref. 25). This enzyme bends DNA towards the major groove and binds, without any pronounced sequence specificity, to virtually all DNA sites. DNaseI cutting rates can thus be used as an estimate of DNA bendability, which in turn can be used to calculate approximate DNA rigidity<sup>26</sup>. The result is a simplified segmented-rod model (Fig. 1c).

In this sequence-dependent anisotropic-bendability (SDAB) model (see Box 1), each disk corresponds to one base pair, and the arrow (see Fig. 1c) 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. Simplifying this, one can take bendability towards the major groove as the principal parameter<sup>26</sup>. (The widely accepted concept of bending anisotropy in DNA is a theoretical postulate<sup>27,28</sup> rather than an experimental finding.)

In contrast to the isotropic elastic models, the SDAB model is non-linear in terms of displacement response, but is still amenable to finite-element analysis. The present form contains several rough approximations: (1) it does not include static-deflection components (i.e. as in isotropic elastic models, DNA is considered to be an originally straight rod); (2) torsional flexibility is not incorporated; (3) bending anisotropy is incorporated into the model in a simplified way – more sophisticated models have been suggested (e.g. in molecular dynamics calculations<sup>29</sup>); (4) the bendability measures derived using DNaseI contain both static and dynamic components – their conversion into a rigidity scale implies that these two contributions are approximately proportional, which might not of course be the case<sup>26</sup>. In summary, the SDAB model is an approximate, stripped-down model that is primarily designed to reflect one aspect of DNA – local bending phenomena.

**Bending anisotropy of curved and straight DNA<sup>30</sup>.** A simple experiment shows the macroscopic anisotropy of the SDAB model (Fig. 2b): a DNA rod model is bent in various directions, and the energy of the curved model is plotted

**Table 1. Sequence-dependent Young's modulus values calculated for various bending models<sup>26,34,40</sup>**

Trinucleotide	Relative bendability (arbitrary units)		Young's modulus ( $10^8 \text{ Nm}^{-2}$ )	
	DNaseI scale	Consensus scale	DNaseI scale	Consensus scale
AAA/TTT	0.1	0.05	2.307	2.23
AAC/GTT	1.6	2.65	2.016	1.71
AAG/CTT	4.2	4.70	1.523	1.30
AAT/ATT	0.0	0.35	2.327	2.17
ACA/TGT	5.8	5.50	1.225	1.15
ACC/GGT	5.2	5.30	1.336	1.18
ACG/CGT	5.2	5.30	1.336	1.18
ACT/AGT	2.0	7.80	1.94	1.46
AGA/TCT	6.5	4.90	1.096	1.26
AGC/GCT	6.3	6.90	1.133	0.87
AGG/CCT	4.7	5.05	1.429	1.23
ATA/TAT	9.7	6.25	0.519	0.99
ATC/GAT	3.6	4.45	1.636	1.35
ATG/CAT	8.7	7.70	0.697	0.71
CAA/TTG	6.2	4.75	1.151	1.29
CAC/GTG	6.8	6.65	1.041	0.92
CAG/CTG	9.6	6.90	0.536	0.87
CCA/TGG	0.7	3.05	2.19	1.63
CCC/GGG	5.7	5.85	1.244	1.08
CCG/CGG	3.0	3.85	1.749	1.47
CGA/TCG	5.8	7.05	1.225	0.84
CGC/GCG	4.3	5.90	1.504	1.07
CTA/TAG	7.8	5.00	0.859	1.24
CTC/GAG	6.6	6.00	1.078	1.05
GAA/TTC	5.1	4.05	1.355	1.43
GAC/GTC	5.6	5.50	1.262	1.14
GCA/TGC	7.5	6.75	0.914	0.90
GCC/GGC	8.2	9.10	0.787	0.45
GGA/TCC	6.2	5.00	1.151	1.24
GTA/TAC	6.4	5.05	1.115	1.23
TAA/TTA	7.3	4.65	0.95	1.31
TCA/TGA	10.0	7.70	0.465	0.71

against the angle that denotes the direction of bending. A model built from repeats of curved DNA motifs has a single energy minimum (blue plot in Fig. 2c). Such a rod model therefore has a preferred direction of bending and, as a result of thermal fluctuations, it will oscillate around a single conformation. In other words, the physically measurable average conformation of such a model will be curved. The straight motifs (red and green plots in Fig. 2c), by contrast, have either no minima or have two minima in opposite angular directions.

In the minimum-energy conformation of a curved motif, the high-positive roll values occur at one face of the arc, while the negative rolls occur at the opposite face (Fig. 2b) – as predicted for DNA that is wrapped around a nucleosome. Clearly, the lowest-energy conformation has slight kinks; this is reminiscent of the mini-kink model postulated by Zhurkin and co-workers<sup>28</sup> and also of the experimentally observed polygonal shape of DNA that is bent into minicircles<sup>31</sup>. The high-energy conformations, by contrast, have much smaller

#### Box 1. Glossary

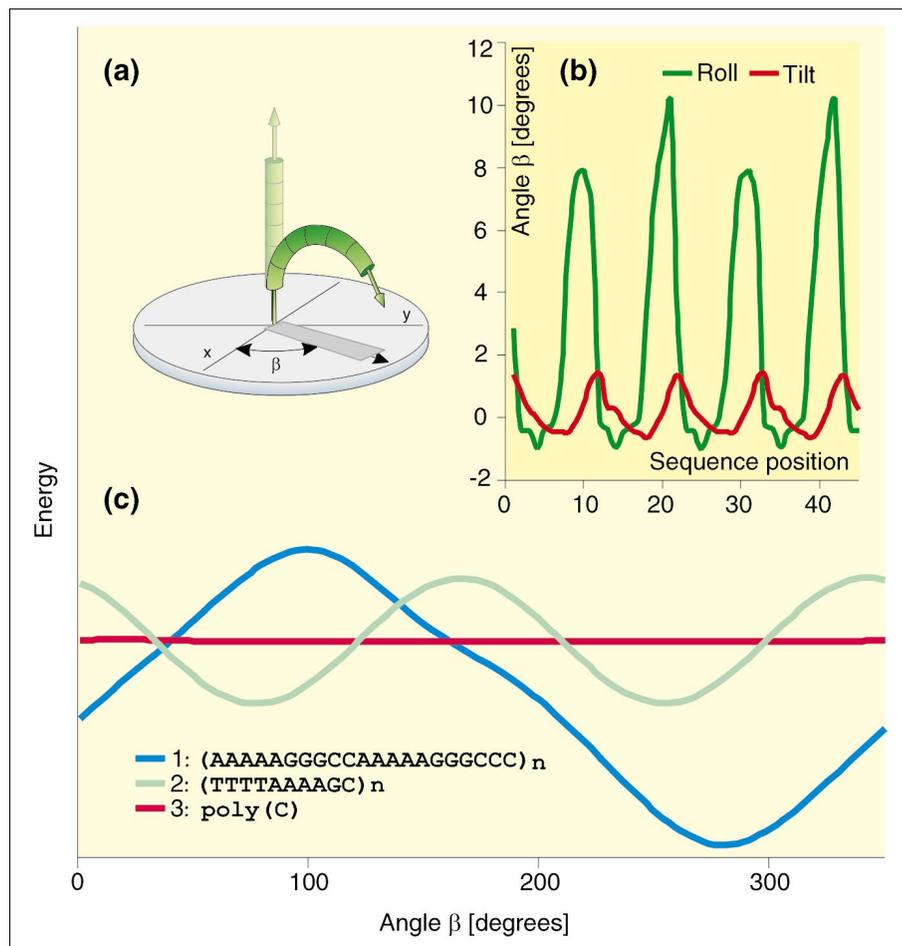
**Bendability.** The relative ability of DNA to bend, usually expressed in arbitrary units. Bendability can be estimated from nucleosome-binding data or from the DNaseI-cutting frequency. It is a form of mechanical flexibility, which is a property that is dependent on the shape of the body.

**Elastic body.** A body that returns to its original shape when forces that have deformed it are no longer present.

**Finite-element method.** A numerical method for solving differential equations that is widely used for calculating the deformation of elastic bodies.

**Sequence-dependent anisotropic-bendability (SDAB) model.** An elastic rod model of DNA, which represents DNA as an initially straight, segmented rod, in which the flexibility (rigidity) of DNA is sequence dependent and anisotropic.

**Young's modulus.** A shape-independent material-stiffness parameter (expressed in  $\text{Nm}^{-2}$ ) that can be used to indicate the stiffness of DNA. In mechanics, Young's modulus is determined by force-against-deformation experiments. The average stiffness of the DNA can be deduced by a variety of direct and indirect methods, and is similar to that of plastic materials such as polypropylene or phenol resins.



**Figure 2**

Testing bending anisotropy in DNA by finite-element methods<sup>30</sup>. Three DNA models were bent to a given curvature in various directions [angle  $\beta$  in (a)], and the bending energy of the model was calculated for each direction by finite-element methods. (b) Distribution of roll and tilt angles along the model corresponding to sequence 1 in the minimum-energy conformation. Curved motifs (e.g. motif 1, blue plot<sup>31</sup>) exhibit a single energy minimum (c), which corresponds to a bending preference in one direction. Straight motifs have either no minimum (motif 2; green plot) or minima in two opposite directions (motif 3; red plot). The three sequences used are shown in (c). 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 normalized to one base-pair unit, for better comparison<sup>30</sup>.

kinks (M. G. Munteanu, K. Vlahoviček, S. Parthasarathy, I. Simon and S. Pongor, unpublished).

**Prediction of bendable and curved segments.** The single energy minimum shown in Fig. 2 is a consequence of two facts: the base pairs are themselves anisotropically bendable and, moreover, their sequence dictates that the direction of principal bendability is along one face of the DNA helix. This can be illustrated best by a vector representation in helical-circle diagrams – a technique originally developed for amphipathic  $\alpha$ -helices in proteins<sup>32</sup> (Fig. 3a). In these diagrams, the bendability value for each base pair is plotted as a vector that points towards the major groove. In fact, inherently curved motifs show an asymmetrical bendability distribution: large vectors are on one side (Fig. 3a).

The vectors for base pairs in straight sequences, by contrast, have a rather symmetrical distribution. This amounts to saying that, in randomly chosen DNA, thermal fluctuations in various directions will cancel out, while inherent bending is dependent upon a specific sequence.

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.*<sup>33</sup> used the length of the vectorial average (as shown in Fig. 3a) as a measure of predicted curvature; however, more-rigorous geometry calculations can also be used: bendability values (see Table I) 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-geometry model in which the dynamic contribution of thermal fluctuations is included. When these assumptions are made, the SDAB model gives accurate predictions for DNA curvature that compare favourably to those of the static-geometry models (see Table II). Table II also includes figures that were calculated using the so-called consensus-bendability scale<sup>14</sup>, which was developed in order to increase the sensitivity of the prediction towards GC-based curved motifs – these are often mispredicted by the static-dinucleotide models<sup>15</sup> and the original DNaseI scale.

Predicted curvature can be plotted along a sequence, which allows one to identify segments of potential curvature as peaks (Fig. 3b). Because the SDAB model also allows the calculation of average bendability, in parallel with predicted curvature, one can plot these two values against each other for every segment of a long sequence. The resulting plot is two dimensional and allows one to distinguish rigid, flexible and curved segments as outliers (Fig. 3c).

**Protein–DNA interactions: DNA rigidity versus stability in Cro–cognate–DNA complexes.** Repressor proteins, such as Cro (Ref. 34), exhibit high-specificity binding to short DNA motifs, and the DNA is often bent in the resulting protein–DNA complex – bending is induced by the binding of the protein. The rigidity of the operator DNA is therefore likely to play a role in the binding. By plotting the experimental free energy values against the rigidity of the oligonucleotides (Fig. 4a), we find that cognate (operator) and non-cognate (non-operator) DNA follow adverse, quasi-linear relationships. In the operator sequences (red plot in Fig. 4),  $\Delta G$  is higher for stiffer molecules (i.e. the stiffer the molecule, the weaker the binding). This is, in fact, expected, because Cro must bend the molecule, and the energy required is directly proportional to the stiffness of the DNA (see Eqn 1). Such a relationship indicates that Cro bends all cognate DNA to the same degree. Cognate DNAs that exhibit different bendability will require different bending energies; the stability of the resulting complexes will therefore be different.

In non-operator sequences, by contrast,  $\Delta G$  is lower for stiffer sequences (i.e. the stiffer the sequence, the stronger the binding). To explain this phenomenon, we present a simple model (Fig. 4b): first, Cro binds to the oligonucleotide in a non-specific

Table II. Analysis<sup>a</sup> of curved and straight sequence motifs using various methods

No.	Sequence motif <sup>b</sup>	Parameter Method Ref.	Static geometry models				Sequence-dependent anisotropic bendability (SDAB) model			
			Curvature (degree/helical-turn)				Bendability (arbitrary units)		Predicted curvature (degree/helical-turn)	
			Gel-electrophoresis	X-rays	NMR	Nucleosome	DNaseI	Consensus	DNaseI	Consensus
			7	11	13	15	25	14	25	14
<b>Curved DNA</b>										
1	(aaaatttgc) <sub>n</sub>		26.2	6.9	18.3	13.7	2.8	2.4	21.1	17.4
2	(aaaatttgc) <sub>n</sub>		21.0	3.8	3.8	17.7	2.2	2.3	16.6	17.2
3	(tctcaaaaaacgcgaaaaaacccgaaaaaacg) <sub>n</sub>		27.1	8.2	16.7	17.1	3.2	3.2	15.4	15.9
4	(ccgaaaaagg) <sub>n</sub>		14.7	6.8	13.1	23.3	3.9	4.3	17.1	20.2
5	(tctctaaaaaatatataaaaa) <sub>n</sub>		27.8	3.0	7.5	10.9	4.5	3.2	27.6	18.5
6	(ggcaaaaaac) <sub>n</sub>		26.8	12.0	20.1	20.4	3.2	3.3	19.2	19.5
7	ccaaaaatgtcaaaaaataggcaaaaaatgcc – <i>Leishmania tarentolae</i> kinetoplast		26.0	6.4	15.7	19.6	3.9	3.5	21.8	20.5
8	aaaaactctctaaaaactctccctagagggccctagagggc		19.4	7.8	10.6	13.3	5.2	4.6	13.6	12.8
9	aaaaactctctaaaaactctctagagggccctagagggccc		17.5	7.2	13.1	16.4	5.1	4.6	13.3	13.2
10	agaatgggacaaaaattggaatttttaagg – <i>Columba risoria</i> bent satellite DNA		18.5	8.9	8.7	12.4	3.3	3.0	13.5	12.6
11	(aaaaactctctaaaaactctcagggccctagagggcccta) <sub>n</sub>		21.6	5.2	10.1	14.5	5.0	4.7	13.8	10.8
<b>Straight DNA</b>										
12	(atctaactcaacacacaca) <sub>n</sub>		0.8	0.5	2.7	1.2	5.1	4.4	0.8	0.8
13	actacgttaaatctatcaccgcaaggataaa – OR3 operator region		10.4	5.5	4.9	5.9	5.0	4.4	10.4	7.8
14	actacgttaaatctatcaccgcaaggataaa – OR3 region, mutated		11.0	5.5	3.4	6.2	4.9	4.4	10.6	8.1
15	(a) <sub>n</sub> – poly-A <sup>c</sup>		0.008	0.000	0.008	0.000	0.100	0.063	0.002	0.002
16	(ttttaaaccg) <sub>n</sub>		1.5	7.1	14.5	10.7	2.8	2.5	2.3	4.2
17	(ttttaaaccg) <sub>n</sub>		1.7	0.8	16.0	16.3	3.6	3.3	3.0	9.6
18	(aaaaactctctaaaaactctcggccctagagggccctaga) <sub>n</sub>		27.1	3.1	5.9	7.5	4.9	4.6	12.8	8.3

<sup>a</sup>The angular deflection was calculated by the bend.it server, using the BEND algorithm<sup>15</sup> and expressed as degree per helical turn of 10.5 bp (note that the curvature units of Trifonov *et al.* correspond to 4.5°/bp or 47.25° per 10.5-bp helical turn). The nucleosome, DNaseI and Consensus models are calculated using tilt = 0 and twist = 36 (ideal B-DNA). It is noted that the dimer angles used to compute curvature are not all defined on the same basis, and other algorithms might give slightly different curvature values.

<sup>b</sup>See Refs 30,33 for the experimental methods used to measure curvature in the sequence motifs.

<sup>c</sup>By definition, homopolymers should give zero curvature. The non-zero value indicates the numeric precision of the calculation.

manner and reduces the free movement (thermal fluctuations) of the DNA, which results in an entropy loss. Because 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:

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

In the above expression  $E_{free}$  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-immobilized) DNA. Because  $E_{free}$  is smaller than  $E_{bound}$ , this equation gives an inverse relationship between stiffness and  $\Delta G$ . This relationship is similar to that shown by the blue plot in Fig. 4, which is in fact very close to being linear in the range of the experimental data. Clearly, the relationship shown in Fig. 4 can be explained quantitatively by the entropy loss that is involved, which is in accordance with the intuitive expectation that immobilization

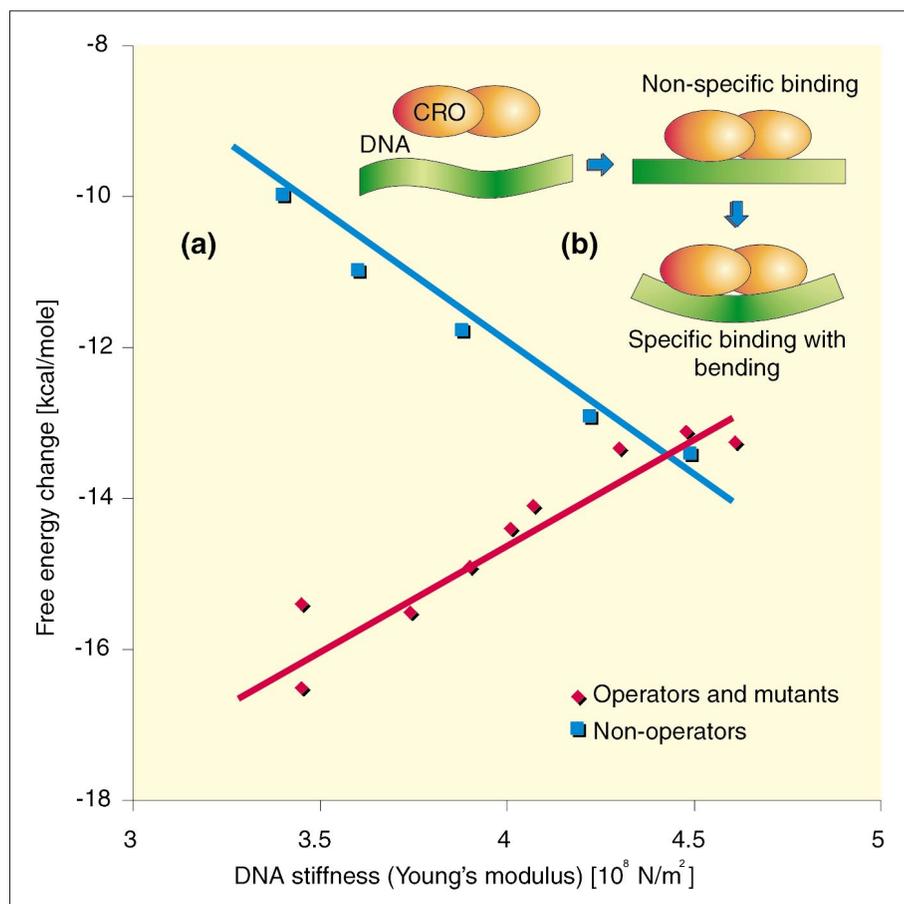
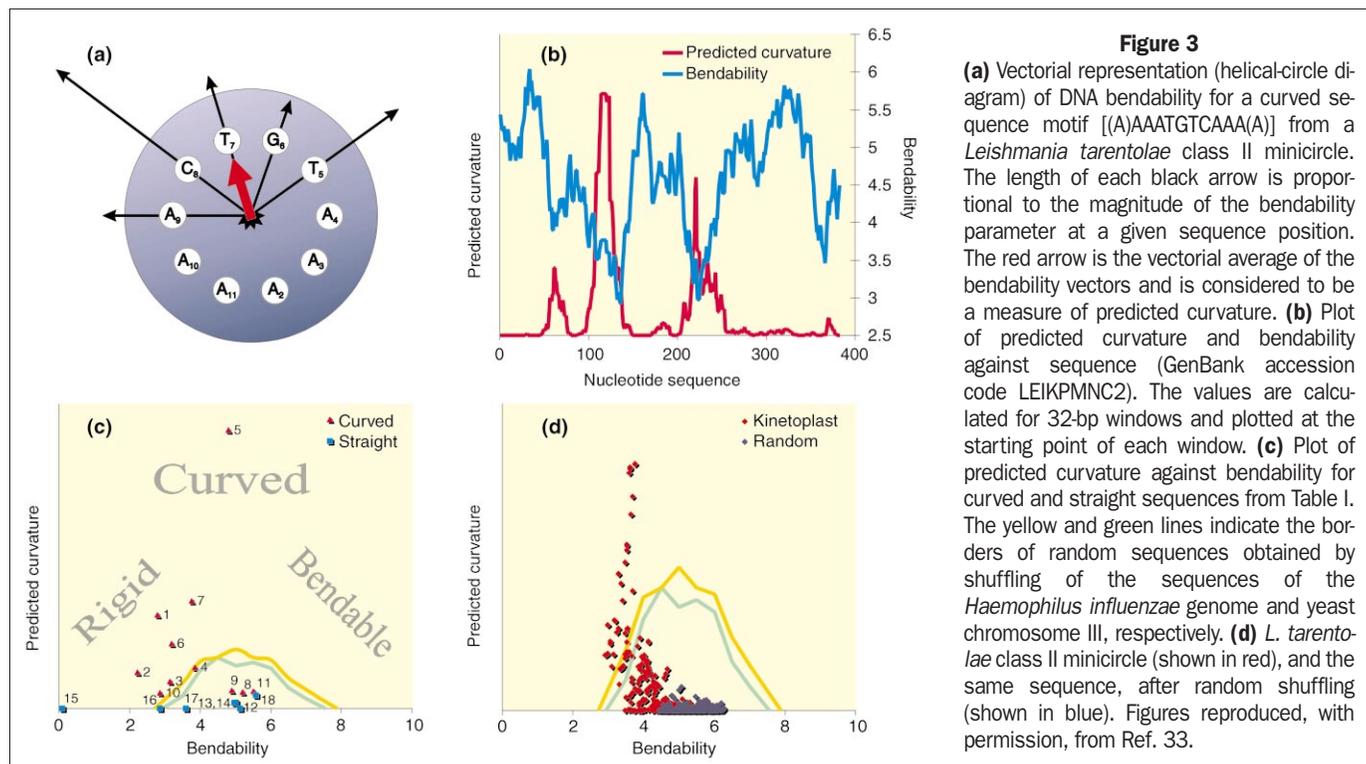
of a stiff DNA cognate will require less energy during binding. In the second step, the free-energy change will have two components: that of the specific interaction (energy gain) and that of bending. Bending energy can be calculated using Eqn 1 and is directly proportional to the rigidity of the DNA. Figure 4, in fact, shows that complexes that involve more-rigid DNA molecules are less stable. Moreover, the slope of the curve corresponds to a degree of curvature that is found in the crystal structure of Cro complexed with cognate DNA<sup>34</sup>. The Cro protein, however, is only one example; whether or not similar relationships exist for other DNA-binding proteins remains to be shown.

### Conclusions and future directions

Static and dynamic rod models describe different aspects of the DNA molecule by using as few parameters as possible. If they succeed – and, surprisingly, they do – then the parameters used by the model are sufficient to explain a given aspect of the molecule's behaviour. Both types of model can predict curvature

in short DNA segments – in this respect the models can be considered to be equivalent<sup>35</sup>. The differences in predictive accuracy could be due to the 'parameterization' (e.g. trinucleotide as opposed to dinucleotide representation, or electrophoresis as opposed to nucleosome data) and not to the models themselves. Considering the significance of these structural features, it is worth mentioning that bendability/curvature characteristics are conserved in evolutionarily<sup>30,33</sup> and functionally related sequences<sup>36</sup>, and correlate well with the positions of known regulatory sites<sup>37</sup>.

Refinements of the rod models, especially the incorporation of tetranucleotide-based description, will probably increase predictive accuracy and the scope for the models' application. Future models might combine both static and dynamic features and, in particular, take into account the principles of non-linear DNA dynamics. Molecular mechanic models should also improve the extent to which we can model a wider range of local phenomena – as illustrated recently by Lavery and



**Figure 4**

Binding of the Cro repressor to oligonucleotides of different stiffness<sup>34</sup>. (a) The plot shows the relationship between the average stiffness (Young's modulus) of DNA and free-energy changes ( $\Delta G$ ) for both operator (shown in blue) and non-operator (shown in red) DNA sequences. (b) A model for the binding of Cro to the operator sequences is also shown. Figure reproduced, with permission, from Ref. 34.

colleagues<sup>38</sup>. In the foreseeable future, however, the simple rod models will continue to play an important role in the large-scale analysis of genomic data<sup>39</sup>.

**Acknowledgements**

We thank J. Langowski and A. Bolshoy for their advice. M. M. G. also thanks the Universitatea Transilvania, Brasov, Romania. I. S. also thanks the International Centre for Theoretical Physics, Trieste, Italy and the Hungarian National Research Foundation (OTKA) project T017652 for financial support.

**References**

- 1 Travers, A. A. and Klug, A. (1990) in *DNA Topology and its Biological Effects* (Cozzarelli, N. R. and Wang, J. C., eds), pp. 57–106, Cold Spring Harbor Press
- 2 Olson, W. K. (1996) *Curr. Opin. Struct. Biol.* 6, 242–256
- 3 Langowski, J. et al. (1996) *Trends Biochem. Sci.* 21, 50
- 4 Trifonov, E. N. and Sussman, J. L. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 3816–3820
- 5 Wu, H. M. and Crothers, D. M. (1984) *Nature* 308, 509–513
- 6 Brukner, I. et al. (1993) *Nucleic Acids Res.* 21, 1025–1029
- 7 Bolshoy, A., McNamara, P., Harrington, R. E. and Trifonov, E. N. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 2312–2316
- 8 De Santis, P., Palleschi, A., Savino, M. and Scipioni, A. (1990) *Biochemistry* 29, 9269–9273
- 9 Olson, W. K. et al. (1995) *Biophys. Chem.* 55, 7–29
- 10 Bansal, M., Bhattacharyya, D. and Ravi, B. (1995) *Comput. Appl. Biosci.* 11, 281–287

- 11 Gorin, A. A., Zhurkin, V. B. and Olson, W. K. (1995) *J. Mol. Biol.* 247, 34–48
- 12 el Hassan, M. A. and Calladine, C. R. (1995) *J. Mol. Biol.* 251, 648–664
- 13 Ulyanov, N. B. and James, T. L. (1995) *Methods Enzymol.* 261, 90–120
- 14 Gabrielian, A. and Pongor, S. (1996) *FEBS Lett.* 393, 65–68
- 15 Goodsell, D. S. and Dickerson, R. E. (1994) *Nucleic Acids Res.* 22, 5497–5503
- 16 Shpigelman, E. S., Trifonov, E. N. and Bolshoy, A. (1993) *Comput. Appl. Biosci.* 9, 435–440
- 17 Lu, X. J., El Hassan, M. A. and Hunter, C. A. (1997) *J. Mol. Biol.* 273, 681–691
- 18 Barkley, M. and Zimm, B. (1979) *J. Chem. Phys.* 70, 2991–2997
- 19 Vologodskii, A. V. and Frank-Kamenetskii, M. D. (1992) *Methods Enzymol.* 211, 467–480
- 20 Zienkiewicz, O. C. and Taylor, R. L. (1991) *The Finite Element Methods*, McGraw-Hill
- 21 Bauer, W. R., Lund, R. A. and White, J. H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 833–837
- 22 Yang, Y., Tobias, I. and Olson, W. K. (1993) *J. Chem. Phys.* 98, 1673–1686
- 23 Zhang, P., Tobias, I. and Olson, W. K. (1994) *J. Mol. Biol.* 242, 271–290
- 24 Rippe, K., von Hippel, P. and Langowski, J. (1995) *Trends Biochem. Sci.* 20, 500–506
- 25 Brukner, I., Sanchez, R., Suck, D. and Pongor, S. (1995) *EMBO J.* 14, 1812–1818
- 26 Gromiha, M. M., Munteanu, M. G., Gabrielian, A. and Pongor, S. (1996) *J. Biol. Phys.* 22, 227–243
- 27 Schellman, J. A. (1974) *Biopolymers* 13, 217–226
- 28 Zhurkin, V. B., Lysov, Y. P. and Ivanov, V. I. (1979) *Nucleic Acids Res.* 6, 1081–1096
- 29 Olson, W. K., Marky, N. L., Jernigan, R. L. and Zhurkin, V. B. (1993) *J. Mol. Biol.* 232, 530–554
- 30 Gabrielian, A. *et al.* (1998) in *Structure, Motion, Interaction and Expression* (Vol. 1) (Sarma, R. H. and Sarma, M. H., eds), pp. 117–132, Adenine Press
- 31 Han, W., Lindsay, S. M., Dlakic, M. and Harrington, R. E. (1997) *Nature* 386, 563
- 32 Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984) *J. Mol. Biol.* 179, 125–142
- 33 Gabrielian, A., Simoncsits, A. and Pongor, S. (1996) *FEBS Lett.* 393, 124–30
- 34 Gromiha, M. M., Munteanu, M. G., Simon, I. and Pongor, S. (1997) *Biophys. Chem.* 69, 153–160
- 35 Calladine, C. R. and Drew, H. R. (1996) *J. Mol. Biol.* 257, 479–485
- 36 Schatz, T. and Langowski, J. (1997) *J. Biomol. Struct. Dyn.* 15, 265–275
- 37 Langst, G., Schatz, T., Langowski, J. and Grummt, I. (1997) *Nucleic Acids Res.* 25, 511–517
- 38 Sanghani, S. R., Zakrzewska, K., Harvey, S. C. and Lavery, R. (1996) *Nucleic Acids Res.* 24, 1632–1637
- 39 Gabrielian, A., Vlahovicek, K. and Pongor, S. (1997) *FEBS Lett.* 406, 69–74
- 40 Brukner, I., Sanchez, R., Suck, D. and Pongor, S. (1995) *J. Biomol. Struct. Dyn.* 13, 309–317

## Catalytic triads and their relatives

Guy Dodson and Alexander Wlodawer

Interactions among the residues in the serine protease Asp-His-Ser catalytic triad, in the special environment of the enzyme-substrate complex, activate the nucleophilic potential of the seryl O<sub>γ</sub>. In the subtilisin and trypsin families, the composition and arrangement of the catalytic triad do not vary significantly. However, the mechanisms of action of many other hydrolytic enzymes, which target a wide range of substrates, involve nucleophilic attack by a serine (or threonine) residue. Review of these enzymes shows that the acid-base-ser/thr pattern of catalytic residues is generally conserved, although the individual acids and bases can vary. The variations in sequence and organization illustrate the adaptability shown by proteins in generating catalytic stereochemistry on different main-chain frameworks.

Figure 1 shows the well-known series of chemical events that occurs during catalysis by a serine protease. It also shows the different chemical branches that can exist on each side of the amide or ester bond, although we discuss here only reactions involving the amide bond. In the scheme shown, the unifying chemistry of nucleophilic attack at the carbonyl carbon can be appreciated. In reviewing the variations in the catalytic structures among different enzyme families, we consider the acid, the base and the nucleophile separately.

### The architecture of the classical triad in serine proteases

The first protease catalytic site, revealed by David Blow and colleagues<sup>4</sup> about 30 years ago using X-ray crystallography, was that of α-chymotrypsin. Their analysis showed initially that two residues were directly involved in catalysis: Ser195 and His57. Chemical evidence for the involvement of Ser195 and His57 in the catalytic reaction already existed, and the crystal structure revealed that the catalytic serine residue was indeed near enough to His57 to form a hydrogen bond. However, the nearby residue 102 was not an asparagine residue, as originally thought, but rather an aspartate residue that was hydrogen-bonded to His57 and could potentially form a salt bridge with this histidine residue. The invariance of this aspartate residue in all the related sequences showed that the residue was an important component of the catalytic structure, which the authors identified as a Ser-His-Asp triad; because of the triad's polarization, they referred to it as a charge-relay system. An account of the

**CHEMISTRY IS THE** engine that drives biology, and many enzyme families are responsible for making this chemistry possible. The catalytic ability of enzymes rests on the spatial organization of the active atoms, through which the chemical and structural steps of the reaction are orchestrated. Structural and sequence comparisons of a variety of proteins have been carried out by a number of groups (see Ref. 1, for example). Such comparisons of enzyme families

show that catalytic groups have distinctive patterns of variation<sup>2</sup> and that the proteins themselves sometimes have unexpected, even surprising, evolutionary and structural relationships with other protein families that have utterly different functions.

In this article, we examine how the catalytic groups of enzymes that cleave amide or ester bonds by nucleophilic attack, a particularly important reaction in biology, can vary among different enzyme families. This kind of analysis has been made possible by the remarkable activity in the field of X-ray crystallography – activity that is producing an avalanche of new protein structures and accurate details of their functional surfaces<sup>3</sup>.

**G. Dodson** is at the Dept of Chemistry, University of York, York, UK YO10 5DD and at the NIMR, London UK NW7 1AA; and **A. Wlodawer** is at the Macromolecular Structure Laboratory, NCI-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, MD 21702, USA.