UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/916 - 31

SEVENTH COLLEGE ON BIOPHYSICS:

Structure and Function of Biopolymers: Experimental and Theoretical Techniques.

4 - 29 March 1996

DNA Bending

S. PONGOR
International Centre for Genetic Engineering & Biotechnology
Area di Ricerca
Padriciano 99
34100 Trieste
ITALY.

DNA Bending[1]

A. Background

In some respects DNA is a simple molecule. A plasmid DNA of known sequence can be cut into a number of different-sized pieces by a restriction enzyme. Molecules of equal length should exhibit similar flexibility. Pieces of DNA shorter than the persistence length of 150-200 bp will behave as rather stiff rods that cannot be easily bent into a circle (see Hagerman, 1988, for discussion of persistence length). Larger pieces adopt a "random coil" shape in solution. DNA molecules of defined size behave in a very predictable way when run on agarose or acrylamide gels. Shorter molecules migrate more rapidly than larger molecules. In agarose and acrylamide gels there is a linear relationship between the log of the distance migrated and the length of the DNA in base pairs. Occasionally, a DNA band of known length does not run at its expected position on an acrylamide gel. One of the most striking examples of such anomalous migration was a 414-bp piece of kinetoplast DNA from *Crithidia fasciculata* (Marini *et al.*, 1982,1983). This DNA migrated as if it were twice as long in an acrylamide gel but migrated at its proper position in an agarose gel. The anomalous migration of the kinetoplast DNA on poly-acrylamide gels was attributed to the kinetoplast DNA being either stably bent or kinked (Figure 1). Although the migration through acrylamide gels is not completely understood in physical terms, the migration is believed to depend on the ability of DNA to "snake" through the gel matrix. A relatively straight piece of DNA that is rather flexible can easily snake through the gel. on the other hand, DNA containing a permanent bend or kink is not as flexible and can get hung up in the gel matrix. The reason is well known from gel filtration: A bent molecule has a smaller efficient diameter, so it will diffuse into many more pores of the matrix, than the straight one, which fits into much fewer pores. So while travelling through the gel, the bent molecule will make extra detours, so at the end it will stay behind.

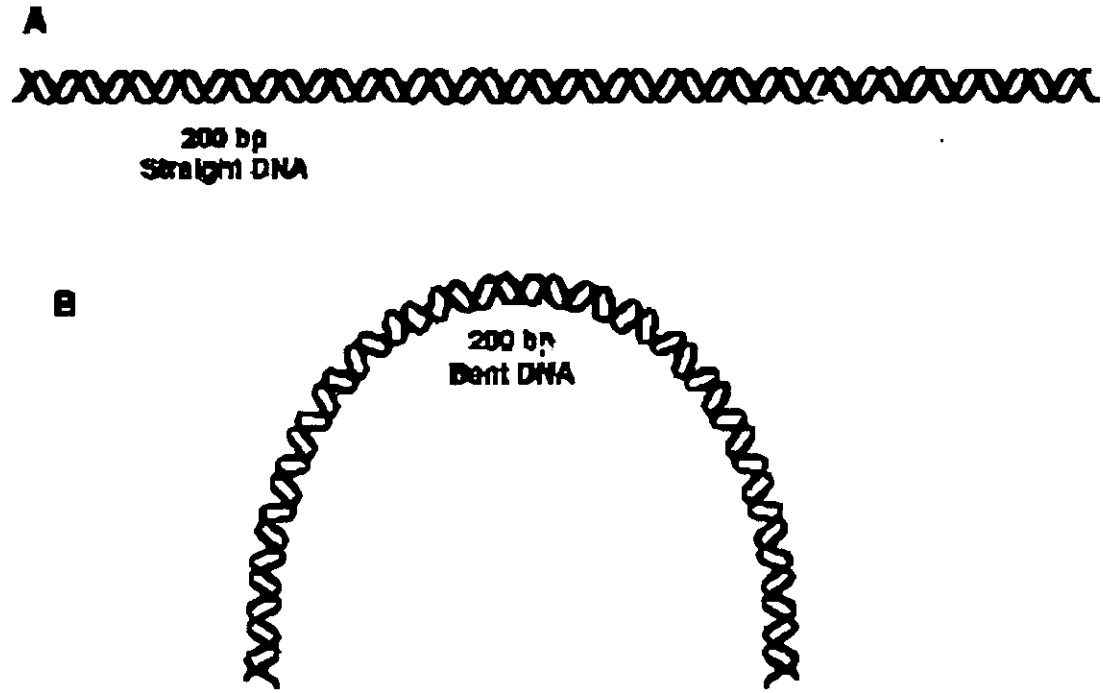


Figure 1 Straight DNA and bent DNA. A) A representation of a 200-bp straight DNA molecule. The persistence length of DNA is 150-200 base pairs. Persistence length is the length of DNA that resists easy deformation or a length that is not considered very flexible. DNA smaller than its persistence length cannot be easily bent into a circle. **B)** A 200-bp piece of bent DNA is shown.

---

[1] In these lecture notes, the description of the wedge and junction models of bending are based on Sinden's "DNA Structure and Function". The comparison of the models and the description of the vectorial flexibility model is based on the work of Ivan Brukner, Andrei Gabrielian, Michael Gromiha and Sándor Pongor at ICGEB.

The initial part of this chapter describing DNA bending experiments will be presented from a historical perspective for two reasons. First, it is a convenient way to gradually introduce a number of important concepts related to DNA bending and to introduce other important aspects of DNA structure and function. Second, it provides a small window on "the scientific process." This process involves developing hypotheses and models and then designing experiments to test or disprove the models. In the DNA bending story, the experimental design was frequently simple and the experimental results quite clear. However, the experiments designed to test the various models for the molecular basis of bending did not all support a single unified explanation. The DNA bending experiments provide an example of how the frequent paradox of scientific discovery leads to a deeper understanding of the physical realities of nature. Moreover, incongruities in the bending experiments suggest that the actual molecular mechanisms behind sequence-induced DNA bending are not yet completely understood.

B. Wedges and Junctions: Static Models for DNA Bending

1. Sequence Induced Bending of DNA

Wu and Crothers (1984) devised a clever way to map the site of bending in the kinetoplast DNA. They cloned the bent kinetoplast DNA fragment as a dimer of a 241-bp HindIII DNA fragment. By cutting the dimer with a series of different restriction enzymes that cut only once within each 241-bp fragment, different 241-bp DNA fragments were produced (Figure 2). The sequence organization of each fragment would be different from that of the original fragment. Such a set of DNA molecules is called *circularly permuted*. Although all these isomers contain the same base sequence and, for the most part, the same bend in the DNA, they do not migrate at the same rate during electrophoresis in an acrylamide gel.

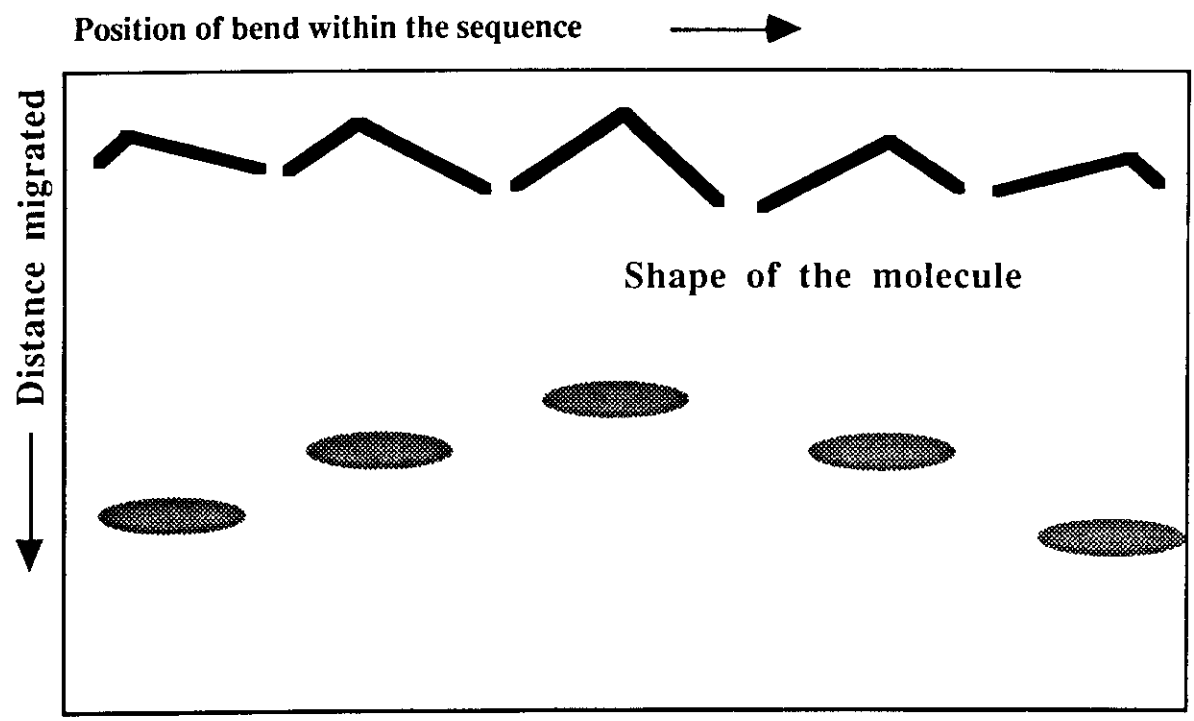


Figure 2 Mobility of molecules of the same length, with the bend situated at different positions. Gel mobility anomaly is sometimes described using the migration distances D expressed as apparent number of basepairs measured as compared with some standard. this value can be compared with the known number of basepairs to give a ratio of 1 for straight and >1 for curved molecules (remember that curved molecules migrate as if they were longer!).

Circularly permuted molecules do not migrate at the same rate because the bends are at different locations within the molecule. This creates different end-to-end distances. The end-to-end distance for a molecule at the left in Figure 2 is nearly as long as that for a straight DNA molecule. The end-to-end distance is shortest when the bend is positioned at the center of the molecule. For molecules of the same size, the end-to-end distance is important in determining gel mobility. The molecule with the shortest end-to-end distance will migrate most slowly in an acrylamide gel. This pattern was observed by Wu and Crothers (1984).

```
1         10         20         30         40
CCCAAAAATGTCAAAAATAGCAAAAATGCCAAAAATCCCAAC
```

Figure 3 The bent locus in *Crithidia Fasciculata* DNA

The DNA sequence shown in Figure 3 was found at the bend center. There are five runs of 4 or 5 As (A4 5) which, in each case, are preceded by a C and followed by a T (CA_4T). In addition, the runs of A are phased by 10 bp. The DNA bending hypothesis suggested that the runs of A and the 10-bp phasing were important in bending. Crothers also suggested that bending may occur at a *junction* resulting from the interruption of B-form DNA by an A tract that adopts a non-B-DNA helix. Trifonov proposed a second model for bending, in which a *wedge* angle was associated with the AA dinucleotide, and proposed that bending could be attributed to the summation of the wedge angles of the AA dinucleotides.

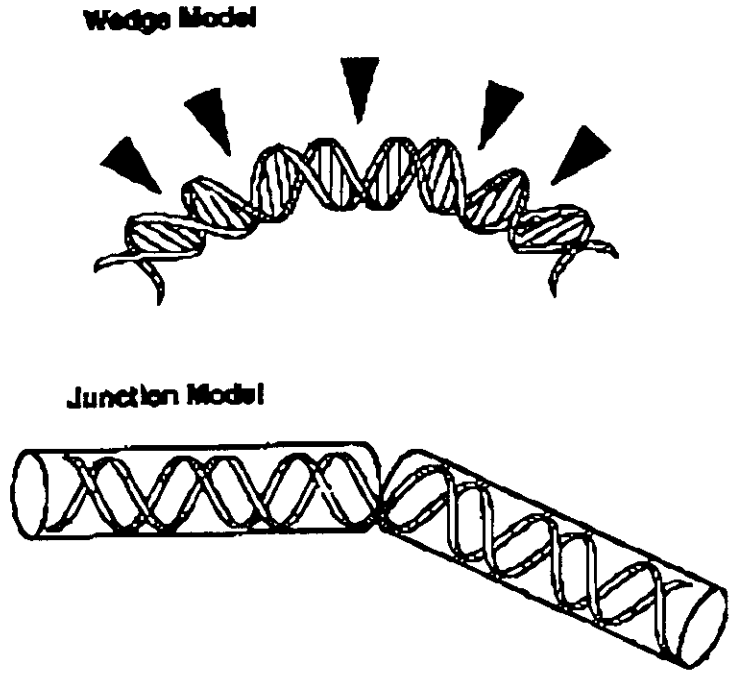


Figure 4 The wedge and junction models for DNA bending. A) The wedge model suggests that bending is the result of driving o wedge between adjacent base pairs at various positions in the DNA. The wedge model predicts a small wedge between each AA dinucleotide in all four A tracts. The total bend will be the sum of all the individual wedges. B The junction model suggests that the A tract of DNA forms a slightly different helix than the normal B-DNA helix A kink or bend is believed to exist at the junction of these two different helices The sequence shown in Figure 3 would contain four junctions one on each end of the A tract

2. The Wedge Model for DNA Bending.

The wedge model for DNA bending assumes that the AA dinucleotide contains a "wedge" angle that causes a deflection in the axis of the DNA double helix (Figure 4; Trifonov and Sussman, 198O; Ulanovsky *et al.*, 1986). The sum of wedges pointing in the same direction, a condition met by the 10-bp phasing, leads to the bending of DNA. As illustrated in Figure 5, the wedge angle can result from a wedge along the tilt axis or a wedge along the roll axis. Calladine and Drew (1992) discuss the physical basis for wedge angles for DNA.

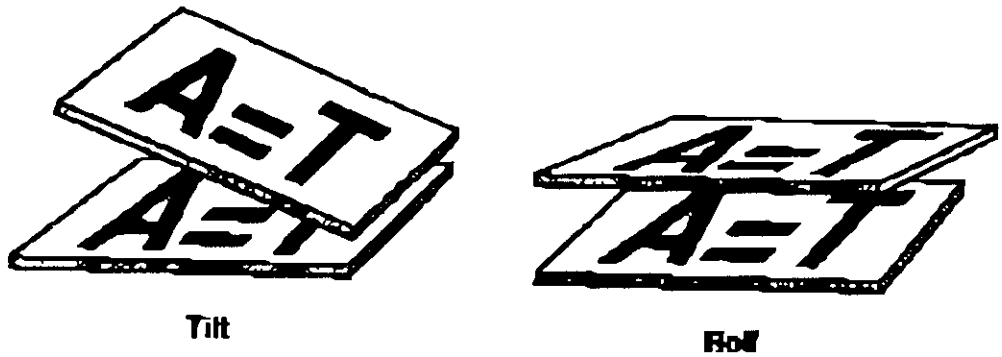


Figure 5 Tilt and roll wedges. The Tilt angle is the direction of the H-bonds and opens in the direction of the phosphate backbone. The roll is perpendicular to it rolls are positive if the bending is towards the major groove (most cases) and negative if it is towards the minor groove (isolated cases).

Ulanovsky *et al.* (1986) used measurement of the efficiency of ligation of small DNA molecules into circles to calculate the wedge angle of an AA dinucleotide. As a short piece of DNA with a defined curvature is ligated together into increasingly long polymers, at some length the total angle of curvature will result in the formation of a circle of DNA. By determining the length at which the DNA forms a circle and by knowing the number of AA dinucleotides responsible for the 36O° curvature, the individual AA wedge angle can be determined. To calculate the total wedge angle, Trifonov synthesized the following DNA sequence:

```
TCTCTAAAAAATATATAAAAA
     TTTTTTATATATTTTTAGAGA
```

This sequence contains two runs of (A),- that are l0 bp apart. Since the average helix repeat is 10.5 bp, it is necessary to make a 21-bp fragment to ensure that the A tracts will remain in phase when multiple 21-mers are ligated together. The optimal size for circle formation in polymers of this sequence was about 126 bp (Ulanovsky *et al.*, 1986). This length is well below the persistence length of DNA, which is 150-200 bp. A straight 126-bp piece of DNA will not circularize efficiently. Thus, efficient circularization is an indication of stable 36O° curvature. Within the 126-bp molecule there are 66 AA dinucleotides; the sum of their individual wedge angles is responsible for the 36O° curvature. Apparently making a few corrections, Trifonov estimated a total wedge angle (of both tilt and roll components) of 8.7o for each AA dinucleotide. This angle probably represents an upper limit of the wedge angle.

3. The Junction Model for DNA Bending

Wu and Crothers (1984) proposed the junction model for DNA bending (Figure 4) which suggests that there is a bend at the junction of B-form DNA and a non-B-DNA helix associated with A tracts. The suggestion that the A tracts might adopt a non-B-DNA helix was based, in part, on the knowledge that poly(dA) . poly(dT) can adopt a non-B-DNA helix called heteronomous DNA (Arnott et al., 1983). In addition, modelling studies had suggested that DNA would bend at the junction between A-DNA and B-DNA (Selsing *et al.*, 1979). There are actually two junctions, a 5' and a 3' junction, relative to the A-tract helix. The site of bending could be associated with the 3' end, the 5' end, or both ends of the A-tract helix. Initially Crothers felt that bending was probably localized at the

3' end of the polyA tract.(Note that an A tract contains As in one strand and Ts in the other obviously ir has two 3' ends and two 5' ends. When we talk about a 3' or 5' end of the A tract we refer ro the 3' or 5' side relative to the strand containing the As). This conclusion was supported by experiments showing greater binding associated with a molecule in which the 3' ends of A tracts were in phase at 10 bp compared with a molecule in which the 5' ends of A tracts were in ***phase*** (Koo *et al.*, 1986).

4. Tests of the Phasing Hypothesis

An important component of both the junction and the wedge model is lO-bp phasing. For DNA to contain a region of stable curvature and to produce the anomalous gel mobility, the small individual bends must be oriented in the same direction or the same plane (if bends are not all directed in the same plane. the DNA will adopt a conformation that is effectively linear). If bends were placed every 1.5 turns of the DNA helix (15-16 bp), the DNA would be maintained in a zig-zag shape. Indeed, DNA with bends every 15 bp migrates slightly faster than a molecule with out-of-phase bends (or linear DNA) upon electrophoresis in an acrylamide gel. These patterns of curvature are shown in **Figure 6.**

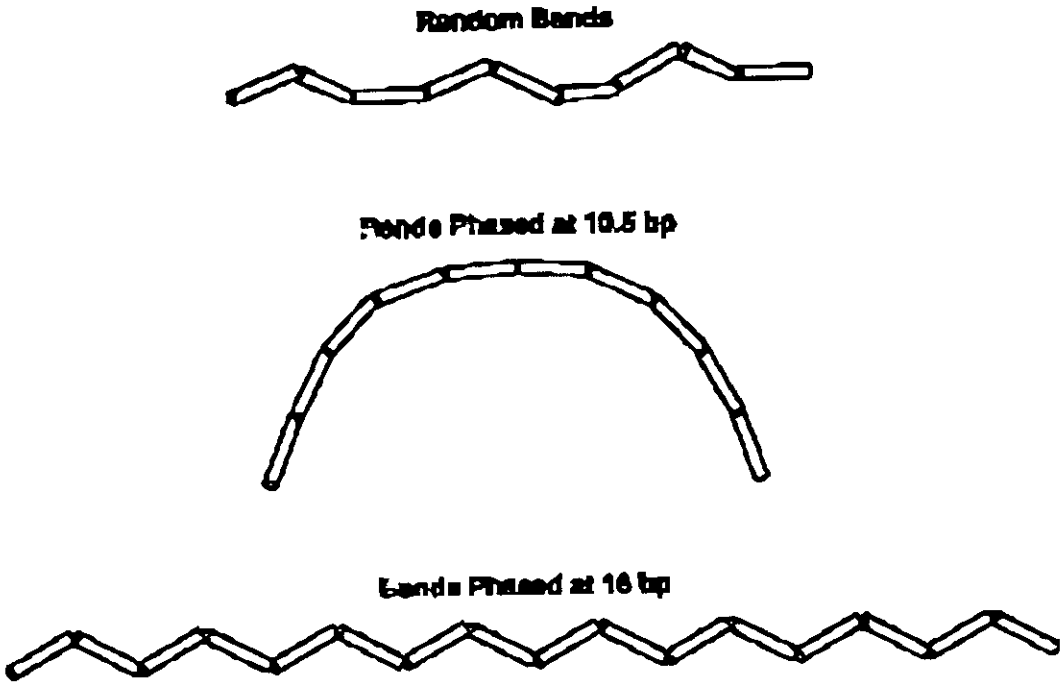


Figure 6 **The significance of phasing in DNA bending.** An A tract will introduce a small bend or deflection of the helix axis in DNA. However the phenomenon of bent "DNA" or "stably curved DNA" that exhibits anomalous migratian on polyacrylamide gel electrophoresis requires a number of small individual bends that are in phase If bends are random as shown in the top molecule the DNA will migrate true to its length in an acrylamide gel. only when bends are phased by 10.5 bp is the stable curvature shown in the middle structure observed This DNA migrates anomalously slowly on electrophoresis in an acrylamide gel. If as shown in the bottom structure, bends are phased by 16 bp successive bends will be directed alternatively up and down This creates a zigzag molecule which is unusually straight and migrates slightly more rapidly than unbent DNA (or DNA containing random bends) i in an acrylamide gel.

To test the phasing hypothesis, Hagerman (198S) synthesized oligonucleotides of sequence GA_3T_3C, $G_2A_3T_3C_2$, and G_3A3T_3C3. These oligonucleotides are self-complementary and can hybridize to form duplex DNA. These duplexes can then be ligated into longer polymers. By ligating the individual oligonucleotides, the A3 block in one DNA strand will be phased at 8-, lO-, and 12-bp intervals. If the 10-bp phasing were important, only the $G_xA_3T_3C$, polymer would exhibit anomalous migration on polyacrylamide gel electrophoresis. In **Figure 7**, mobility data are presented by dividing the apparent size in base pairs (determined by electrophoretic mobility) by the actual size of the DNA (determined by counting the ladders on the gel). The bp_{app}/bp_{seq} data points

are plotted as a function of size of the DNA. The polymers with A tracts phased at 8 and 12 bp gave a horizontal line indicative of non bent DNA. The $G_2A_3T_3C_2$ polymer with A tracts phased at l0 bp exhibited an upward slope in the plot of bp_{app}/bp_{seq} until a plateau was reached (at bp_{app}/bp_{seq} = 2.2). This pattern of electrophoretic migration is diagnostic for curved or bent DNA, and demonstrates the importance of l0-bp phasing in DNA curvature.

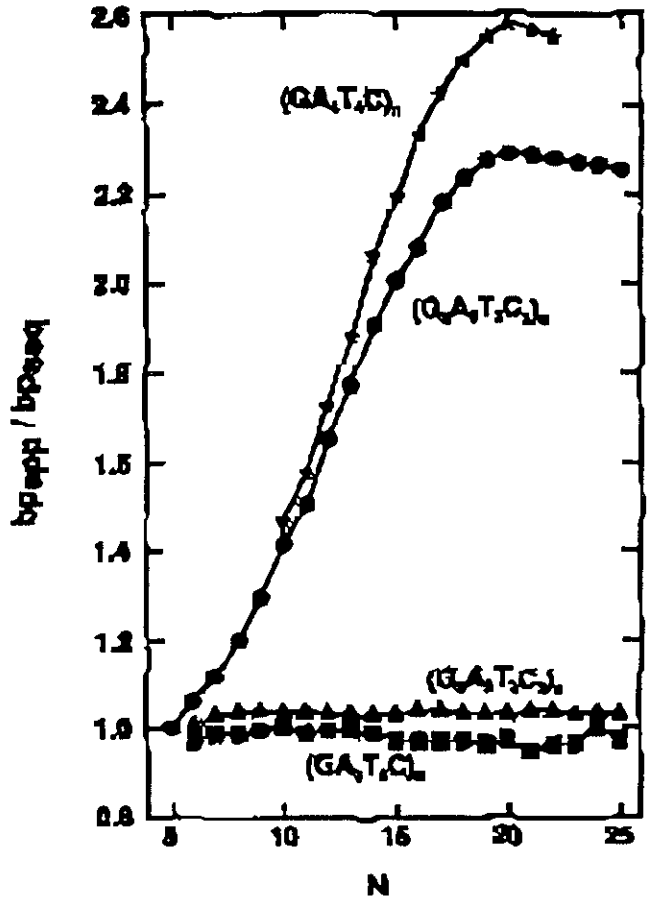


Figure 7 The analysis of DNA bending on polyacrylamide gels. Chemically synthesized l0- and 12-bp oligonucleotides were ligated into families of molecules ("ligation ladders").Their mobility on electrophoresis on a polyacrylamide gel was determined as shown in Figure 2. On the y axis is plotted the apparent length divided the actual length of the sequence on the x axis is plotted the number of individual oligonucleotides ligated (N) When DNA bends the apparent length becomes much larger than expended leading to an increase in y. The straight configuration has a ratio of 1. The $(GA_4T_4C)_n$ and (GGA3T3CC)n polymers represented by the stars and circles respectively are bent Bending is most pronounced when the polymer is $\simeq$ 2O monomers. Although the $(GA\ T_4C)_n$ sequence is not mentioned in the text it is included as another example of a bent DNA polymer In contrast the $(G_3A_3T_3C_3)$ and $(GT_3A\}C$ polymers represented by triangles and squares respectively are not bent

5. Sequence Requirements for DNA Bending

Koo, Wu, and Crothers (1986) synthesized a large number of oligonucleotides containing various lengths of A tracts that were phased at different lengths. These oligonucleotides were conceptually similar to those described by Hagerman but were not symmetrical and thus had an A tract in only one strand of the DNA. Polymers with A4_9 were bent, with bending being optimal for A6. The Koo *et al.* polymer with A3 phased at lO bp was not significantly bent. (The Hagerman bent sequence with A_3T_3contained an A3 tract in both strands that must contribute to bending.) A continuous run of As is required for bending, since replacement of the central A in A5 with C, G, or T destroys the bending. There is no sequence requirement for a particular base 5' or 3' to an A tract for bending, although flanking sequences can influence curvature. DNA sequences that do not contain runs of As can also

be bent. The bends observed in DNA lacking phased A tracts are usually not as large as A-tract bends. These sequences have not been as well studied as A-tract-induced bends.

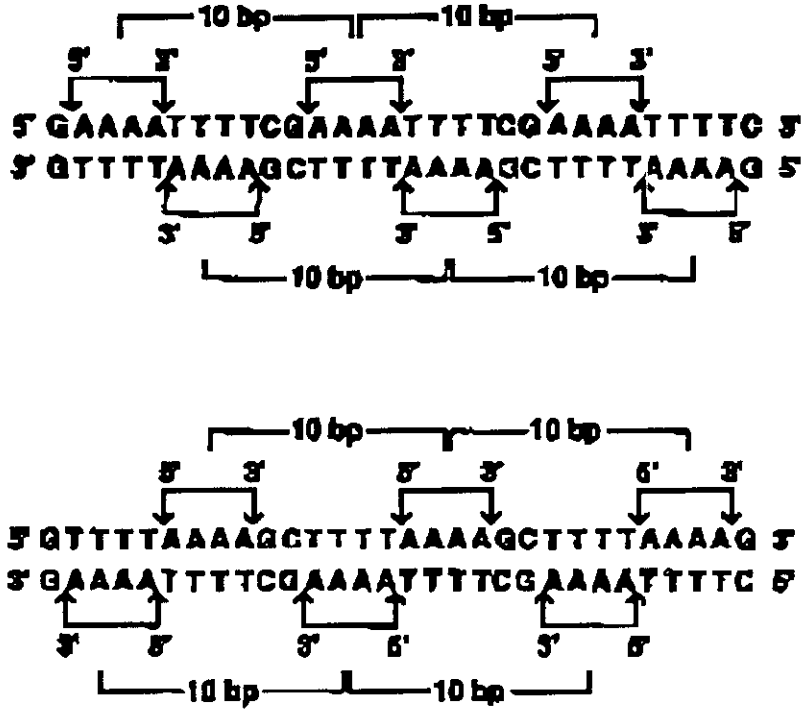


Figure 8 Asymmetry of GA_4T_4C and GT_4A_4C. At first glance the relatianship 6etween phased A trads in $(GA_4T_4C)_n$ and $(GT_4A_4C)_n$ might seem identical. In fad the A 61ocks are separated in phase by 10 bp in both strands. In opposite strands the A tracts are offset at 5 bp. If however one considers the phasing of the 5' and 3' junctions of the A tract the sequences are quite different GA_4T_4C (top) the 3' ends and the A tracts in the complementary strands are opposite cach other. In the GT_4A_4C sequence (bottom) the 5' ends of the A tracts in the complementary strands are apposite each other If one considers a simple wedge model there would be little difference between the A_4T_4 and T_4A_4 oligonucleotides. However if one considers a wedge model in which roll and tilt wedges are involved the reversal of T . A with A . T results in a different direction of tilt but not roll. In this case the AA tilt wedges reverse direction 6y 18O° with respect to the helix axis. Considering phasing introduced at a junction model the relationships 'between the 5' and 3' junctions in opposite strands are quite different in the two different molecules.

6. Analysis of $(GA_4T_4C)_n$ and $(GT_4A_4C)_n$ leads to Development of a Refined Wedge Model

The wedge model is a "nearest neighbor model" which assumes that bending can be explained by wedge properties attributed solely to an AA dinucleotide. It assumes, in its simplest form, that there will be no influence on the AA dinucleotide wedge angle from flanking bases or the sequence contexts of DNA. Considering the polymers $(GA_4T_4C)_n$ and $(GT_4A_4C)_n$, one might expect them to migrate with identical rates on electrophoresis since they each contain A_4 tracts that are phased at lO-bp intervals. However, the (GA_4T_4C) polymer is bent but the (GT_4A_4C) polymer is not bent (Hagerman, 1986). This result may seem to rule out the simple basic wedge model proposed by Trifonov.

Ulanovsky and Trifonov (1987) realized that the relationships of phased A tracts in the two strands are actually quite different for (GA_4T_4C) and (GT_4A_4C). Because GA_4T_4C and GT_4A_4C are symmetrical, A tracts in both DNA strands must be considered. Although both strands of a DNA duplex contain phased runs of As, the relationship between the A tracts in the opposite strands in GA_4T_4C compared to GT_4A_4C is different. The difference in A-tract phasing between these two molecules is illustrated in Figure 8. In GA_4T_4C, the 3' ends of the A tracts in opposite strands occur at the same position in the helix axis. In GT_4A_4C, the 5' ends of the A tracts coincide at the same point along the helix axis.

A key to understanding the difference in the electrophoretic mobility of $(GA_4T_4C)n$ and $(GT_4A_4C)n$ lies in the realization that the roll and tilt components of the wedge model are different symmetrically. For an AA dinucleotide, tilt can open either toward the As or toward the Ts. The tilt angle is believed to open toward the As, as shown in **Figures 5 and 9.** In a DNA helix containing NNNAANNN, the replacement of the As with Ts to form NNNTTNNN results in a 18O° reversal of the direction of the tilt angle. The roll angle, however, is unchanged when As are replaced by Ts because the roll angle opens toward the major groove and the positions of the grooves remain the same whether an AA or TT dinucleotide occurs in a helical region.

Assuming that bending is the result of a sum of all vectorial tilt and wedge components for AA dinucleotides in both strands of the DNA, Ulanovsky and Trifonov (1987) calculated that the AA dinucleotidc has a tilt of 2.4o opening toward the AA and a roll angle of 8.4° opening toward the major groove (see **Figure 9**).

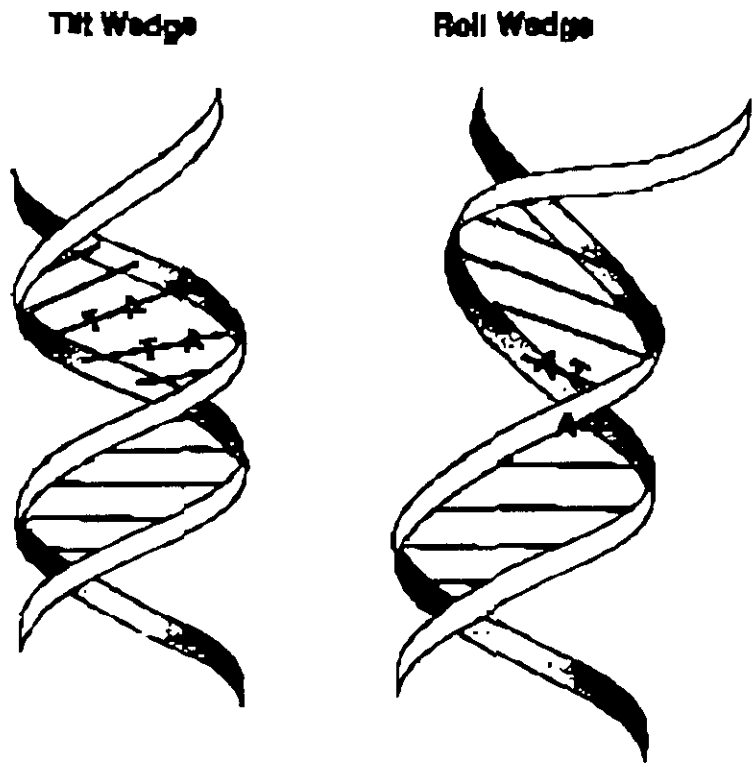


Figure 9 Tilt wedge and roll wedge. A tilt angle of a tilt wedge points toward the phosphate backbone giving a tend in the DNA that points in the direction perpendicular to the phosphate backbone. In contrast a roll wedge opens toward the major groove compressing the minor groove.

7. A Refined Junction Model

The simple junction model first proposed by Crothers, in which bending was attributed predominantly to a deflection at the 3' end of the A tract, did not adequately explain experimental results. A refined wedge model was needed to explain the normal migration behavior of DNA that seemingly should be bent. As illustrated in **Figure 10**, Koo and Crothers (1988) proposed that bending was due to a tilt deflection at both the 3' and the 5' junction. At the 5' junction the tilt was toward the T, whereas at the 3' junction the tilt was toward the A. In addition, these investigators argued that there was a roll that opened the minor groove only at the 5' junction.
Differences in bending behavior of $(GA_4T_4C)n$ and $(GT_4A_4C)n$ are also predicted by the refined junction model. Since a roll component of bending is limited to the 5' junction of the A tract, this component of bending will be at the same point in the helix axis in GT_4A_4C and at different positions in GA_4T_4C (Figure 8). Although it is not intuitively obvious why the A_4T_4 sequence is bent whereas the T_4A_4 sequence is not bent, one should be able to appreciate the symmetry differences in the relationships of the tilt and roll components at the 3' and 5' junctions from analysis of Figure 8.

8. Does Either the Junction Model or the Wedge Model Adequately Explain DNA Bending?

Neither the pure junction model nor the pure wedge model may provide a completely accurate explanation of the physical basis for DNA bending. Both models reasonably predict the bending of many DNA sequences containing runs of As. However, there are some A-tract-containing non-A-tract bent DNAs whose bending is not predicted by either model. These are related to the GGCC motif whose bending was discovered by Brukner and associates, and was also proved with X-ray crystallography by the group of R. E. Dickerson.

Burkhoff and Tullius (1988) obtained results that were inconsistent with both the wedge and the junction model. A hydroxyl radical footprinting technique was used to analyze the width of the minor groove. Hydroxyl radicals are generated by the incubation of an EDTA (ethylenediamine tetraacetic acid) iron (II) complex with hydrogen peroxide. The hydroxyl radical (. OH) removes a hydrogen from the deoxyribose sugar, resulting in a break in the phosphodiester backbone. Since the hydroxyl radical attacks deoxyribose from the minor groove, the reactivity of DNA is dependent on the width of this groove. The hydroxyl radical footprinting pattern for B-DNA showed uniform cutting at each base. Kinetoplast bent DNA showed a reduction in cutting in the A tracts. This result was interpreted as reflecting a narrowing of the minor groove along the A tract. The polymers studied by Hagerman, $(GA_4T_4C)n$ and $(GT_4A_4C)n$, showed very different results. The hydroxyl cleavage pattern in the $(GA_4T_4C)n$ bent polymer showed periodic reduction in the cleavage along the A tracts. There was no reduction in cleavage along A tracts in the $(GT_4A_4C)n$ polymer. These results suggest that in the $5'A_4T_3'$ sequence the run of As can adopt a helix different than B-form DNA, a helix in which a gradual narrowing of the minor groove occurs. This narrowing does not occur in the A tracts in the $5'T_4A_3'$ sequence.

There is evidence for cooperativity in the formation of the A-tract DNA structure. Leroy *et al.* (1988) presented imino-proton-exchange NMR (nuclear magnetic resonance) experiments to show that the structures of A_nT_n and T_nA_n sequences were quite different. Long proton exchange times for A . T base pairs were associated with longer lengths of A tracts in the A_nT_n but not T_nA_n orientation. The shorter times in T_nA_n oligonucleotides were similar to lifetimes found in B-form DNA. Long proton exchange times were found for all sequences that exhibited anomalous migration on polyacrylamide gels. Nadeau and Crothers (1989) have also confirmed that cooperative structural changes in helix structure occur in a run of As as the length of the tract is increased. Three As do not set up the "A-tract helix" responsible for bending whereas A_4 does begin to adopt a non-B-DNA helix and bend DNA. By the time a length of A6-7 is reached, the transition from a B-form to a different helix, an A-tract structure, is complete.

At present there are two static models, the wedge model and the junction model, that both explain reasonably well the bent nature of DNA containing phased A tracts. Although these models are based on a different initial premise they both predict the shape of many, but not all, bent

molecules. In both models curvature is introduced into the DNA at each A tract by a combination of deviations of the tilt angle and roll angle of bases in the A tracts.

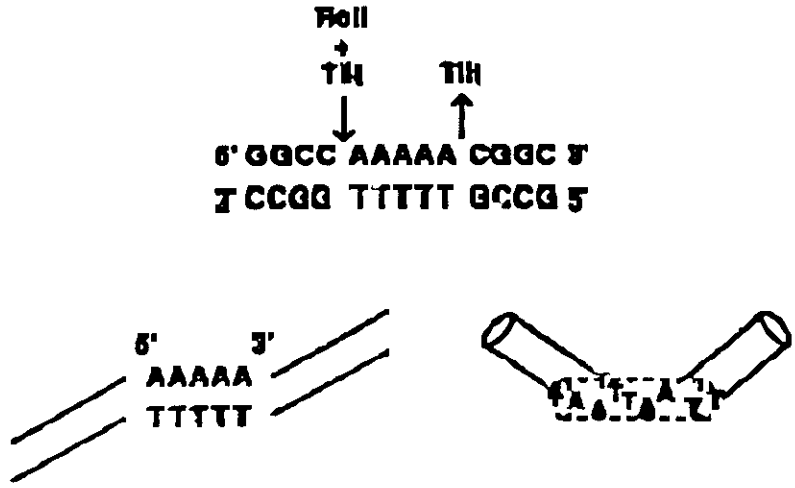


Figure 10 Refined junction model. In the refined junction model there is a tilt toward the T at the 5' junction of the A tract and a tilt toward the A at the 3~ junction. In addition there is a roll opening toward the minor groove at the 5' junction of the A tract. Therefore the bend is a composite of three short straight helices: the B-form helix 3' at the A tract the A tract helix and the B-farm helix 5' at the A tract. The bends then would occur as a mixture of roll and tilt bends at the 5~ junction of the A tracts and as a tilt and at the 3' junction of the A tracts. This relationship is shown in a linear configuration and in a helical configuration.

9. Common discrepancies of the static wedge and junction models

Until now we did not consider the actually roll and tilt values used by the wedge and junction models, nor the principle how predictions are made. The principle is in fact simple: we try to predict the trajectory i.e. the path of DNA. Simply put, the trajectory is the path of the longitudinal axis of DNA, and this will curve if DNA itself bends.

Both models calculate (predict) curvature from microscopic bending parameters (roll, tilt, twist). These parameters are small (~10°) elementary rotations, which can be approximately treated as vectors. Let's imagine a helix of an A and a non-A tract in top view, and assign the various rotations as single vectors: we will find that the two models are in fact analogous: Wedge models assign opposite vectors to the two faces while junction models assign opposite vectors to the two junctions.

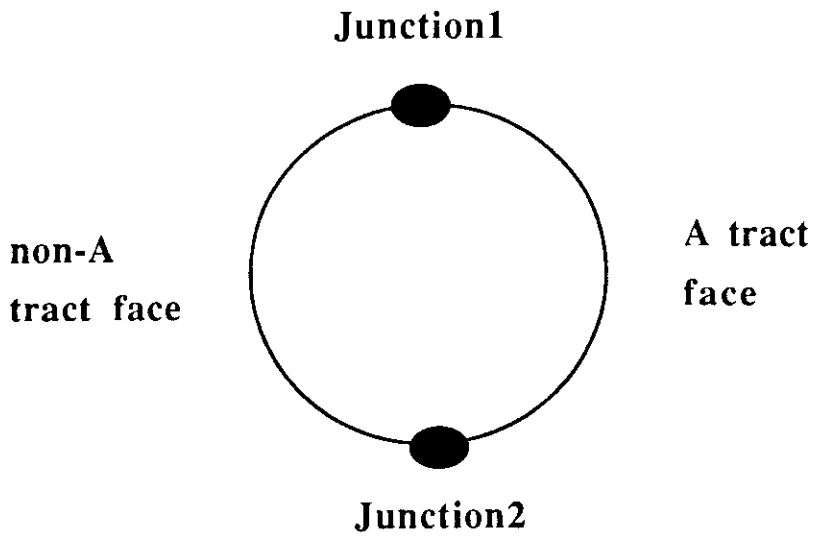


Figure 11 top-view of a curved helix "Symmetry of opposites"

How are the values assigned? Wedge models assign constant roll, tilt (and sometimes twist) values to the 10 ds dinucleotides. This is only partly correct since X-ray and NMR structures show that dinucleotides can adopt vastly different conformations. Junction models assign values to junctions. The values are shown in Table 2.

Table 1. Roll, tilt and twist parameters used in various DNA-bending models (after Goodsell and Dickerson)

1. Trinucleotide Model: Nucleosome Positioning Model of Goodsell and Dickerson, based on Satchwell, Drew & Travers (6). Tilt = 0¤: Twist = 34.3°(10.5 base pairs per turn)

Roll:									
2nd base:		G				A			
3rd base:		A	G	C	T	A	G	C	T
1st base:	A	3.3°	5.4°	7.5°	5.8°	0.0°	5.2°	3.7°	0.7°
	G	3.8°	6.0°	10.0°	5.4°	3.0°	5 4°	5 4°	5.3°
	C	8.3°	4.7°	7.5~	5.4°	3.3°	1.2°	6.5°	6.7°
	T	5.4°	5.4°	6.0°	5.2°	2.0°	2.2°	3.7°	2.8°

11. Dinucleotide Models

Base Step Type:		Purine-Purine			Purine - Pyrimidine		Pyrimid ine - Purine			
Base Step:	AA	GA	AG	GG	AT	AC	GC	TA	CA	CG
(b) Calladine, Drew & McCall Model (7)										
Roll	0°	3.3°	3.3°	3.3°	3.3°	3.3°	3.3°	6.6°	3.3°	3.3°
Tilt:	0°	0°	0°	0°	0°	0°	0°	0°	0°	0°
Twist:	35°	34°	34°	34°	34°	34°	34°	34°	34°	34°
(c) Bolshoi, McNamara, Harrington & Trifonov Model (8):										
Roll	-6.5°	-2.7°	8.4°	1.2°	2.6°	-0.9°	-5.0°	0.9°	1.6°	6.7°
Tilt:	3.2°	-4.6°	-0.3°	-1.8°	0°	-0.7°	0°	0°	3.1°	0°
Twist:	35.62°	36.9°	27.7°	33.67°	31.5°	34.4°	40.0°	36.0°	34.5°	29.8°
(d) Cacchione, De Santis, Foti. Palleschi & Savino Model (10)\										
Roll	-5.4°	-2.0°	1.0°	1.3°	-7.3°	-2.4°	-3.7°	8.0°	6.7°	4.6°
Tilt:	-0.5°	-1.7°	-1.6°	-0.6°	0°	-2.7°	0°	0°	0.4°	0°
Twist:	35.9°	35.8°	35.6°	33.0°	35.0°	34.6°	33.3°	34.6°	34.5°	33.7°
(e) Reversed Calladine, Drew & McCall Model										
Roll	3.3°	0°	0°	0°	0°	0°	0°	-3.3°	0°	0°
Tilt:	0°	0°	0°	0°	0°	0°	0°	0°	0°	0°
Twist:	35°	34°	34°	34°	34°	34°	34.0°	34.0°	34°	34°
(f) Koo & Crothers Model (12)										
Roll	0°	5.8°	0°	0°	0°	0°	0°	11.6°	5.8°	0°
Tilt:	0°	-9.6°	12.0°	0°	0°	12.0°	0°	0°	-9.6°	0°
Twist:	34.3°	34.3°	34 3°	34.3°	34.3°	34.3°	34.3°	34.3°	34.3°	34.3°

In models b-f Roll and Twist are the same for complementary pairs (i.e.—CT and AG). Tilt reverses sign for such pairs.

*** = only when associated with 3 or more successive A's. Roll and Tilt are zero otherwise.

+Not exrlicitly specirled in paper. Chosen here to correspond to 10.5 base pairs per turn.

... X-ray and NMR

In models b-f Roll and Twist are the same for complementary pairs (i.e.—CT and AG). Tilt reverses sign for such pairs.
*** = only when associated with 3 or more successive A's. Roll and Tilt are zero otherwise.
+Not exrllicitly specirled in paper. Chosen here to correspond to 10.5 base pairs per turn.

Are the assigned roll and tilt values realistic? This is a loaded question. X-ray and NMR measurements could not convincingly show that the values required for the models exist in Nature. We can calculate curvature (bending) of DNA as a result of a very large number of elementary distortions. Namely, curvature can be regarded as a fix value for which the corresponding equation system has many solutions. Wedge models and junction models are two classes of solutions we can choose from an infinite number of possible solutions. Their common principle is that they both suppose static bends in curved DNA, but the parameters that produce a realistic curvature for DNA, are not realistic at the level of atomic detail. In the following we describe an elastic model which does not require such a precondition.

A further problem results from the source of data, i.e. electrophoresis, which is used to underpin both static models. Electrophoresis is done, for historic reasons, in the absence of metal ions. Brukner et al (1992) showed that GGCC elements substantially curve in the presence of physiological metal ion concentrations and this curvature outbalances the one of A-tracts. metal ions, especially Mg, seem to be crucial for DNA conformation, so it appears that the entire body of data on which static models rely, should be reevaluated. On the other hand, non-AA curvature (the one showed by Brukner and associates) can be simply incorporated into static models, and, if we maintain the "symmetry of opposites" shown in **Figure 11**, the models will remain as valid or wrong as before. There is one trinucleotide-based static model, the one Goodsell and Dickerson developed from the nucleosome-positioning data of Satchwell et al. This model has some advantages, for example it correctly shows non-AA curvature, but as a whole it does not provide a crucial improvement on the predictive force of static methods.

C A vector description for DNA flexibility: a dynamic model

When looking for new descriptions, the starting point is to exclude some of the discrepancies of the previous models. The starting point here is that individual DNA molecules can be more or less straight and only their bendability is asymmetrical, so when a molecule is exposed to random forces (like a molecule in solution) it will bend asymmetrically. This is a **dynamic model,** as opposed to the static wedge and junction models. How can one assign sequence dependent bendability (i.e. flexibility) values to DNA? Brukner et al (1995) determined such values from deoxyribonuclease I (DNaseI)cleavage experiments. DNaseI, an enzyme with no pronounced sequence specificity, bends DNA towards the major groove. Since the cleavage rate is thought to primarily depend on the flexibility of DNA in this direction so as a first approximation we consider them as indicators of flexibility. The parameters are in fact high for those motifs that are bendable towards the major groove and low in the opposite case. It is important to note that these parameters are determined in the presence of Mg^{2+} ions.

Table 2 DNaseI-derived DNA bendability (flexibility) parameters.

Trinuc-leotide	Flexibility [a.u.]	Trinuc-leotide	Flexibility [a.u.]
AAA/TTT	0.1	CAG/CTG	9.6
AAC/GTT	1.6	CCA/TGG	0.7
AAG/CTT	4.2	CCC/GGG	5.7
AAT/ATT	0.0	CCG/CGG	3.0
ACA/TGT	5.8	CGA/TCG	5.8
ACC/GGT	5.2	CGC/GCG	4.3
ACG/CGT	5.2	CTA/TAG	7.8
ACT/AGT	2.0	CTC/GAG	6.6
AGA/TCT	6.5	GAA/TTC	5.1
AGC/GCT	6.3	GAC/GTC	5.6
AGG/CCT	4.7	GCA/TGC	7.5
ATA/TAT	9.7	GCC/GGC	8.2
ATC/GAT	3.6	GGA/TCC	6.2
ATG/CAT	8.7	GTA/TAC	6.4
CAA/TTG	6.2	TAA/TTA	7.3
CAC/GTG	6.8	TCA/TGA	10.0

It is immediately apparent that the flexibility description is based on trinucleotides, rather than dinucleotides. This improved resolution may be a itself reason for improvement. The physical meaning of the parameters is only approximate. A good visual interpretation is found if one plots these parameters as vectors pointing to the major groove, using an ideal B-DNA structural framework. This can be done in the form of helical circle diagrams. One finds that the vectors are asymmetrically distributed in curved, and symmetrically distributed in straight DNA.

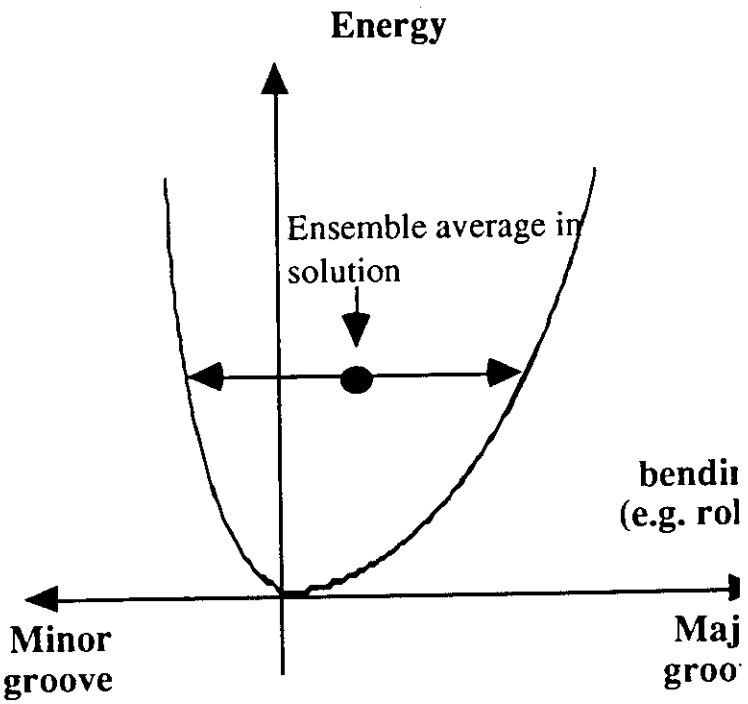


Figure 12 Schematic energy profile of the vector model. Here bending is allowed to the minor groove, but is more costly than toward the major groove. In the very simplified model, bending is not allowed to the minor groove. In a sequence-dependent model, each parameters in Table 1 correspond to one scheme like shown here, and the width of the energy well is proportional to the parameter.

This flexibility based description is in fact similar to a simplified wedge model, because it uses only one parameter at every bp. It is more refined, at the same time, because of the trinucleotide based description. And finally, it is not in contradiction with the experimental findings.

The physical picture behind the vector description is that flexibility of DNA at every point is anisotropic, i.e. bending to the major groove is allowed only . Second, it is also asymmetric, since bending is not allowed (or is more difficult) toward the minor groove. These assumptions are approximate, but in keeping with the 3D data: in fact the minor distortions in DNA are predominantly small bends toward the major groove.

The schematic energy profile in **Figure 12** shows that the **dynamic equilibrium point** (the ensemble average) will be shifted towards the positive roll angles. If one would fix these angles, one would arrive to a static wedge model built only from roll angles. In fact such a model was already proposed on a dinucleotide basis, by Calladine and Drew, but the descriptive power of the model was not sufficient due the lack of data from which the parameters were deduced.

What are the advantages of the vector model? First, it describes curvature quite realistically, for example it can differentiate between the Hagerman 10mers without having to invoke tilt values. It also allows to evaluate curvature in simple terms, since the vector average can be very simply computed. this makes it possible to demonstrate that in fact all known examples of experimentally proven curvature are correctly predicted by the model.

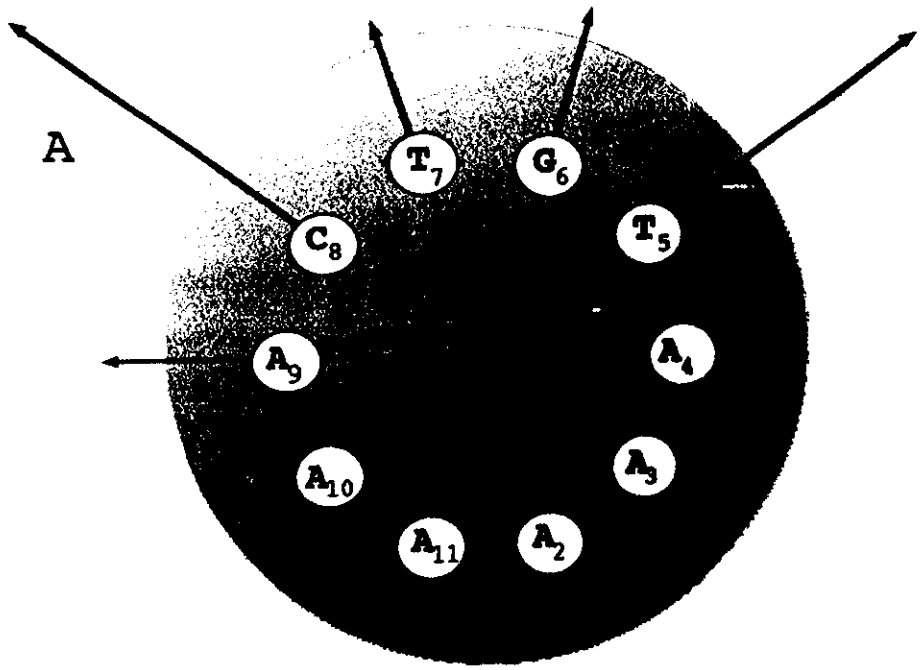
Is asymmetric distribution of flexibility a sufficient criterion for curvature? This can be proved only indirectly. One can build elastic models of DNA in which the flexibility is isotropic, anisotropic, or anisotropic and asymmetrical as shown in Figure 12. One can study the deformations of these elastic models by finite element methods. In fact, the curvature of the model will be always symmetrical, except in one case: if one uses an asymmetrical and sequence dependent model like the one outlined in Figure 12, and one builds the model of a curved DNA.

So in summary, the vector model incorporates dynamic properties of DNA and correctly describes features of bent DNA which pose problems for static models. Much of this may be simply due to the trinucleotide based description which goes beyond the nearest neighbour (dinucleotide) descriptions used in the wedge models.

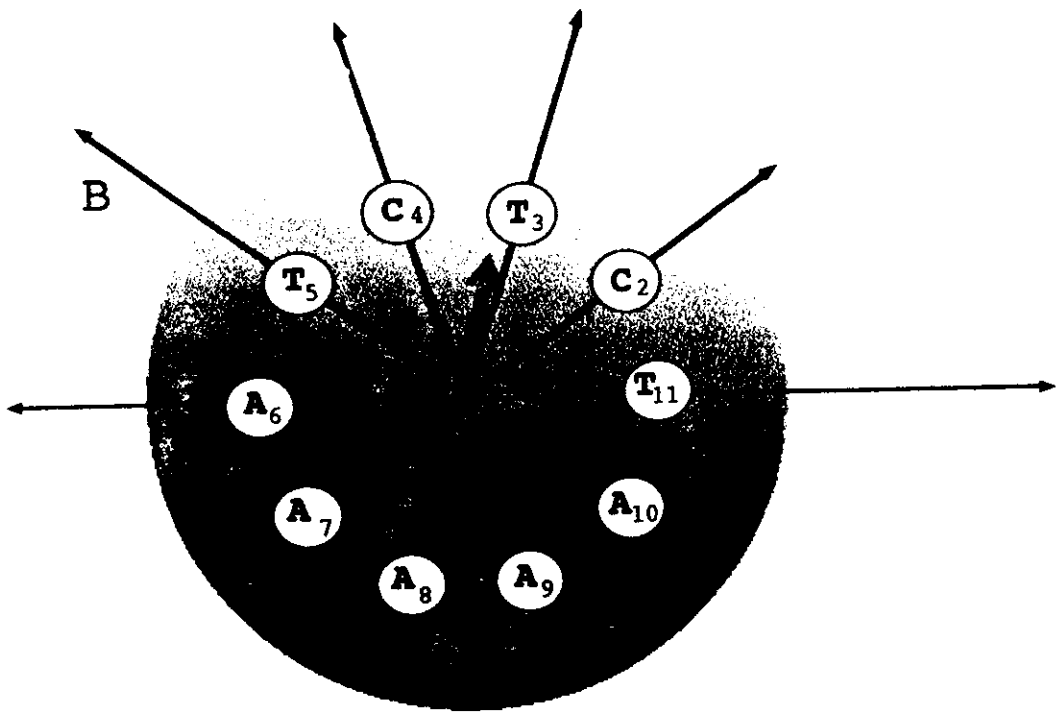
Simplicity of this model is also important since this allows evaluation of entire genomes. In this respect ,a technical trick was very useful: namely in addition to monitoring curvature, one can calculate the average flexibility and draw 2D plots. This makes it possible to distinguish curved, flexible and rigid segments in DNA.

Dynamic properties of DNA can also be studied with full-detail atomic models using molecular dynamics simulations (e.g. B. Honig, R. Lavery, V. Zhurkin, and W. Olson). The review relevant work is beyond the scope of these notes, however.

(A)AAATGTCAAA(A) from a *Leishmania tarentolae* minicircle



(A)CTCTAAAAAT(A) - a synthetic, curved oligonucleotide



Asymmetry (curvature) ~vectorial average: $\frac{1}{n}\sum \vec{f_i}$

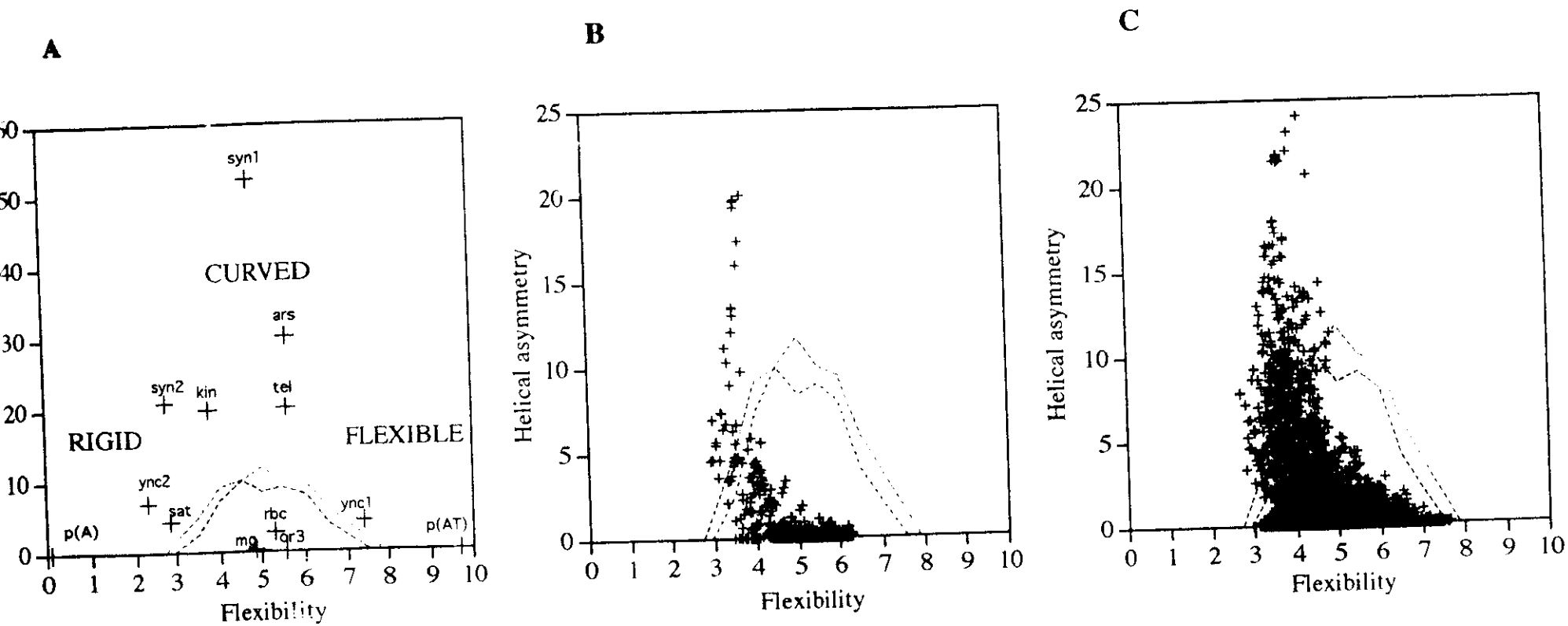


Figure 3 Helical asymmetry (curvature) vs. flexibility plots in selected sequences**A**: Curved and straight sequences from Table 1. **B**: Leishmania tarentolae class II minicircle (Genbank LEIKPMNC2). The black dots indicate a random-shuffled version of the sequence. **C**: 10 different kinetoplast sequences from various flagellatae, exluding the example in Figure B (Genbank locus name, species, length): CFU12625_S2, *C.fasciculata*, 1371bp; KILTB4MC_S2. *L.tarentolae*, 887bp; LDKMPL13_S7, *L. sp*, 821bp; LGKINET11_S2, *L. guliki*, 860bp; LIKINET10_S3, *L.infantum*, 793bp; LMKINET4_S3, *L.major*, 692bp; LTKINET3_S2, *L.tropica*, 854 bp; PBKINET6_S2, *P.brevicola*, 1477bp; TBU03908_S2, *Trypanosoma brucei*, 1014bp; TRBKINC5BA_S2, *Trypanosoma rangeli*, 1764bp). Random sequences were generated by randomly shuffling (21, 24) the *M. genitalium* genome (blue dotted line, sequence from ref. 34) and the *S. cerevisiae* chromosome III sequence (green dotted line, sequence from ref.35).

Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides

Ivan Brukner, Roberto Sánchez, Dietrich Suck[1] and Sándor Pongor[2]

International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano 99, 34012 Trieste, Italy and [1]Biological Structures and Biocomputing Division, European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69012 Heidelberg, Germany

[2]Corresponding author

Communicated by A.Falaschi

Structural parameters characterizing the bending propensity of trinucleotides were deduced from DNase I digestion data using simple probabilistic models. In contrast to dinucleotide-based models of DNA bending and/or bendability, the trinucleotide parameters are in good agreement with X-ray crystallographic data on bent DNA. This improvement may be due to the fact that the trinucleotide model incorporates more sequence context information than do dinucleotide-based descriptions.

Key words: DNA curvature/DNA flexibility/DNA–protein interaction/DNase I/trinucleotides

Introduction

Sequence-dependent bending of DNA can be inherent in the DNA molecule in a stress-free state or it can be induced by ligands (Travers and Klug, 1990). DNA binding proteins generally bend DNA toward the major groove (imposing positive roll angles), irrespective of whether they bind to the major or to the minor groove. At present there is little structural data on DNA bending and/or bendability at the level of atomic resolution, due to the fact that these can be best obtained from crystallographic analysis.

Predictive models of sequence-dependent DNA bending are based, generally speaking, on assigning parameters to motifs within DNA. The current models use parameters which are assigned to the 10 independent dinucleotide steps. The parameters themselves are determined using various experimental approaches, such as: (i) gel-mobility analysis (Bolshoy *et al.*, 1991); (ii) a combination of gel-mobility data with X-ray crystallography (Calladine *et al.*, 1988) or energy minimization (De Santis *et al.*, 1990); (iii) interactions of DNA with other agents, such as nucleosomes (Drew and Travers, 1985; Calladine and Drew, 1986; Satchwell *et al.*, 1986), catabolic activator protein (Gartenberg and Crothers, 1988) or DNase I (Brukner *et al.*, 1990).

Dinucleotide models represent the simplest form of nearest neighbour descriptions. Extension of these models to tri- or tetranucleotides is considered to be an important improvement, since this may account for the well-known fact that dinucleotides are bent and/or bendable differently in different sequential contexts (Hunter, 1993; Poncin *et al.*, 1992). On the other hand, since there are 32 independent trinucleotides and 136 tetranucleotides, such an extension would need a much larger set of experimental data than the one currently available. In this paper we show, based on a large amount of experimental data, that bovine pancreatic deoxyribonuclease I (DNase I) digestion profiles can be used to obtain realistic DNA bending propensity parameters of trinucleotides.

DNase I is considered to be a good molecular probe of DNA bending and/or bendability, since all DNase I–DNA complexes solved so far show that productive binding of DNase I requires DNA to be bent (Lahm and Suck, 1991; Weston *et al.*, 1992). DNase I interacts with a 6 bp contact surface of the minor groove and bends the DNA molecule away from the enzyme, toward the major groove (positive roll) (Lahm and Suck, 1991; Weston *et al.*, 1992; Suck, 1994). Therefore, base sequences that are flexible or inherently bent towards the major groove should be more accessible to DNase I cleavage. A qualitative connection between DNase I digestion and DNA structural properties was recognized quite early (Lomonossoff *et al.*, 1984) and correlation with minor groove width (Drew and Travers, 1984), bending and/or bendability (Drew and Travers, 1985; Hogan *et al.*, 1989; Brukner *et al.*, 1990; Lahm and Suck, 1991; Weston *et al.*, 1992; Suck, 1994) and looping (Hochschild and Ptashne, 1986; Lavigne *et al.*, 1994) have been suggested by several authors. In this paper we try to extract quantitative and sequence-specific bending and/or bendability parameters for trinucleotides, directly from the DNase I digestion data.

Results

Description of the model

The model is based on the assumptions that: (i) DNase I interacts with the window of 6 nt around the cleaved bond; (ii) one single structural parameter $p(a)$ of the trinucleotides, constituting the enzyme–DNA contact surface, will influence the cutting rate (this is an obvious simplification, since local effects, such as specific residue contacts between the enzyme and the DNA molecule, are not considered); (iii) the bending propensity $p(a)$ of each trinucleotide contributes independently to the probability of DNase I cutting, P_w. The model thus assumes that the contribution of one element (trinucleotide) does not depend on any other element being present or absent in the window around the cut. So P_w for the 6 nt window can be written as the product of the n different and assumedly independent $p(a)_i$ probabilities:

$$P_w = \prod_{i=1}^{4} p(a)_i \qquad (1)$$

Table I. DNA bending and/or bendability parameters as revealed by DNase I: parameters of trinucleotide steps[a]

Trinucleotide step	No. of occurrences in dataset	DNase I-derived trinucleotide parameter (ln p)
AAT/ATT	89	−0.280
AAA/TTT	278	−0.274
CCA/TGG	45	−0.246
AAC/GTT	81	−0.205[b]
ACT/AGT	77	−0.183[b]
CCG/CGG	73	−0.136
ATC/GAT	112	−0.110
AAG/CTT	110	−0.087
CGC/GCG	84	−0.077
AGG/CCT	101	−0.057
GAA/TTC	117	−0.037
ACG/CGT	84	−0.033
ACC/GGT	87	−0.032
GAC/GTC	81	−0.013
CCC/GGG	141	−0.012
ACA/TGT	52	−0.006[b]
CGA/TCG	84	0.003
GGA/TCC	71	0.013
CAA/TTG	74	0.015[b]
AGC/GCT	35	0.017
GTA/TAC	83	0.025
AGA/TCT	127	0.027
CTC/GAG	102	0.031[b]
CAC/GTG	55	0.040
TAA/TTA	99	0.068[b]
GCA/TGC	84	0.076
CTA/TAG	64	0.090
GCC/GGC	57	0.107
ATG/CAT	71	0.134[b]
CAG/CTG	61	0.175
ATA/TAT	80	0.182
TCA/TGA	127	0.194

[a]Parameters are given in arbitrary units. High values indicate bending towards the major groove. The parameters were determined using a system of equations given in equation (2); the linear correlation coefficient was 0.60 for a total of 709 data points. The reproducibility of the densitometric readings was characterized by, on average, a 10% error. In the parameter values, this variation results in an average error of 0.002 (SD, standard deviation of the mean). Those values having a higher SD value (between 0.003 and 0.004) are marked. The stability of the numerical model is shown by the fact that an estimated error of 20% would result in an average error of only 0.005 in the parameter values, as determined by computer simulations. The fact that there was no overall change in either the resulting parameter values or in the ranking order of the trinucleotides makes us believe that the triplet parameters presented are close to or identical with the absolute minimum of the model (Press *et al.*, 1992).
[b]Parameters which were found to be the least stable with respect to variations in the data (SD between 0.003 and 0.004).

Equating P_w with the experimentally determined frequencies of cleavage, F_w, leads to a linear system of equations of the form

$$\ln F_w = \sum_{i=1}^{4} \ln p(a)_i \qquad (2)$$

which can then be solved by the least squares method. In this work, we use a new set of 709 DNase I digestion data points for pUC18 and T7 A1 promoter DNA (determined in triplicate) in order to derive trinucleotide parameters. The densitometric readings (which are proportionate to observed DNase I cutting frequencies) were then substituted into equation (2), which was subsequently solved to give ln $p(a)$ parameters of trinucleotides (Table I).

Until now we have not mentioned the possible physical meaning of the p parameters. As DNase I can, in principle, cut at sites that are either *a priori* bent toward the major groove or flexible, the p parameters may incorporate these two properties, without necessarily distinguishing between the static and dynamic components.

Comparison with data obtained by X-ray crystallography

The trinucleotide bending and/or bendability parameters were compared with experimental X-ray crystallography data obtained on bent DNA, using parameter versus sequence plots (Figure 2). The examples include inherently curved DNA structures (Figure 2I–K) as well as protein–DNA complexes with bent DNA (Figure 2A–H). In both cases there is an obvious qualitative agreement between the published roll angles (or the direction of groove distortion) and the trinucleotide parameters; regions with major groove compression (positive roll angles) show high parameter values, while non-bent regions or motifs showing compression of the minor groves (negative roll angles) show low values. Only in one example, that of the DNA complex with the TATA box binding protein TBP (Kim and Burley, 1994; Figure 2C), is there an obvious disagreement between the predicted and experimental values. This might be due to the fact that DNA in the TBP–DNA complex is not only doubly kinked, but also substantially unwound. This is quite atypical compared with the other structures. These examples show that DNA bending propensity, extracted from DNase I digestion profiles, and distortions, seen by X-ray crystallography, are clearly correlated.

Evaluation of direction of curvature using helical phasing experiments

Oligonucleotides were designed in order to test the relative direction of curvature in GGC/GCC, CGG/CCG and CCA/TGG trinucleotide motifs. These motifs were chosen because the trinucleotide model predicts their bending propensities in a way that contradicts the current dinucleotide-based descriptions (see also Discussion). The full sequence context of the tested sequences was designed to follow the sequence-dependent bending propensity of trinucleotides presented in Table I. The principle of this technique is as follows: oligonucleotides containing curved elements which are spaced by a half-helical turn and bend DNA in the same direction will have higher mobility ($\mu-$) than those oligonucleotides in which the spacing is a full helical turn ($\mu+$) (Figure 3A). Using a reference motif, this method estimates the relative direction of the curvature of the elements which are part of the test sequence. If there is no intrinsic curvature in the test sequence, then $\mu-$ and $\mu+$ will be independent of the phasing and equal.

All designed oligonucleotides contain A tracts (as a reference motif) which are differentially spaced from the tested elements (Figure 3A and B). One should note from Figure 3 that the outcome of this experiment ranks the curvature of the tested elements in the following manner: GGCCC versus CCGGG (Figure 3B, gc) and CCAAT versus CTCAC motifs (Figure 3B, ca). The $\mu-$:$\mu+$ ratio was significantly different from unity in both examples (Figure 3D, black and white bars), clearly showing that

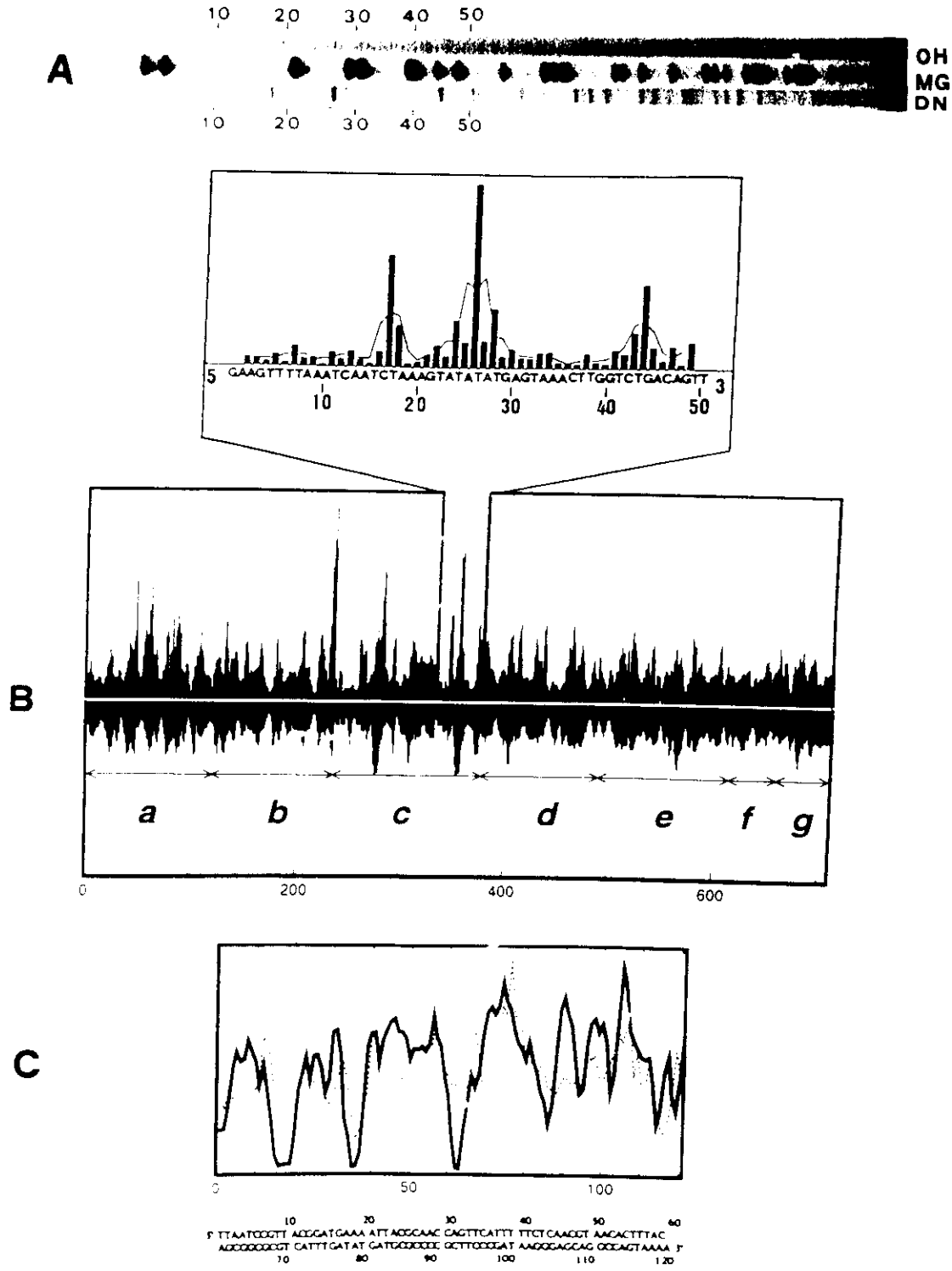


Fig. 1. (**A**) A 'typical' DNase I digestion pattern (line DN); hydroxyl radical (line OH) and Maxam–Gilbert (line MG) reactions were used to assign the cleaved bond to the sequence. Numbers refer to the sequence fragment shown in the magnification in (B). (**B**) Comparison of measured (upper) DNase I cutting frequencies with those predicted by the trinucleotide model (lower values). Origin of DNA, pUC18 restriction fragments; numbers denote site position in databank (Vector Bank 8): *a*, 2686–48; *b*, 1376–1529; *c*, 1529–1780; *d*, 2233–2534; *e*, 399–524; *f*, 245–399 and bacteriophage T7 A1 promoter DNA (EMBL databank); *g*, 80–150. Magnification: measured values (columns) and three bond average (continuous line) corresponding to the digestion profile shown in (A). The relative frequency of cleavage for each P–O3′ bond was determined, transformed to the three bond running average (full line) and presented as a function of the DNA sequence. The predicted values were calculated according to equation (2), using the parameters in Table I. (**C**) DNase I digestion data of Drew and Travers (1984), which were not included into the models (thick line), are presented together with the predictions based on the trinucleotide step (dotted line) model. The numbering corresponds to the DNA sequence written below.

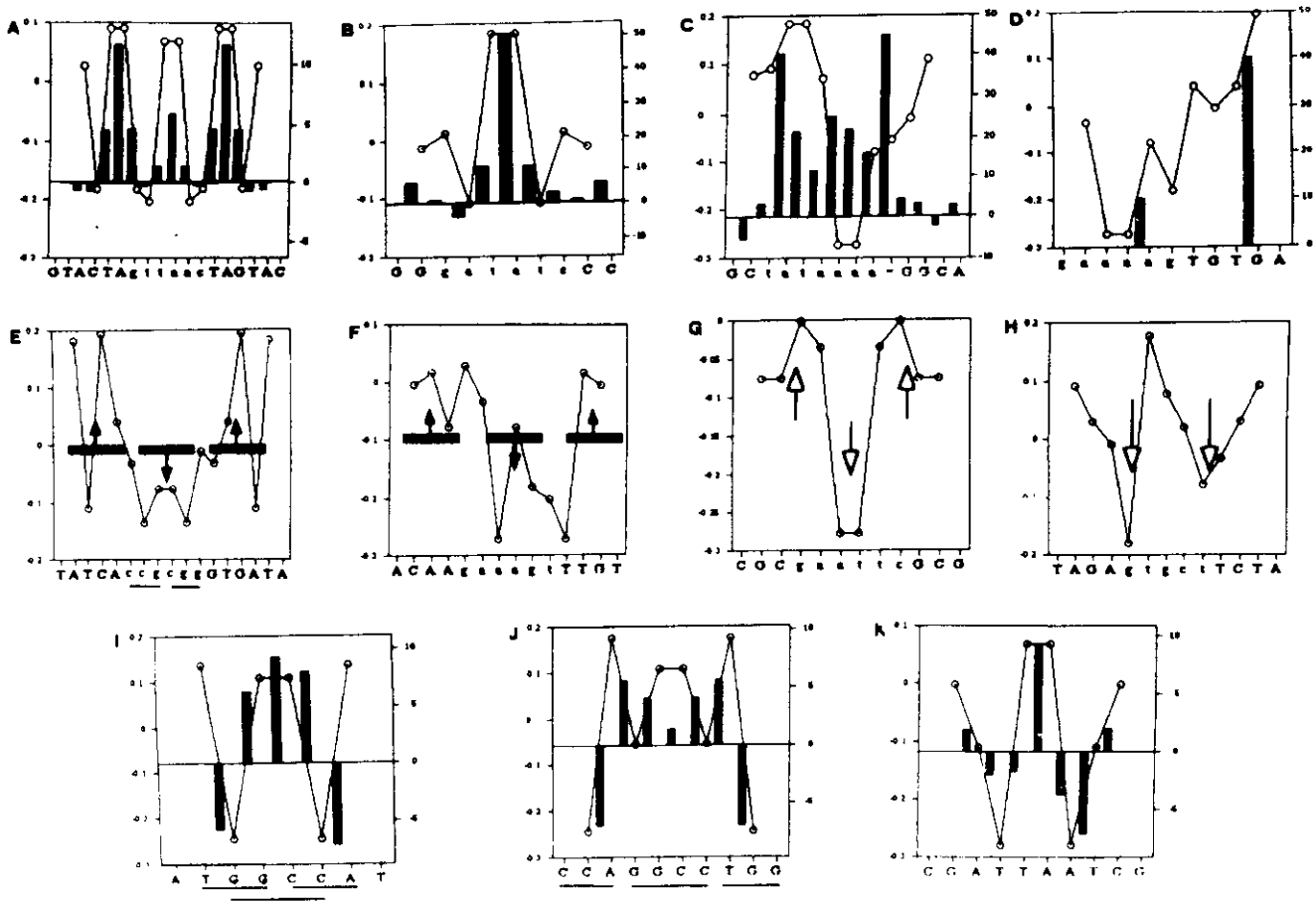


Fig. 2. Comparison of the predicted DNA bending and/or bendability with three-dimensional data. (A–H) Protein–DNA complexes with strong protein-induced bending: (**A**) *trp* repressor (Shakked *et al.*, 1994); (**B**) *Eco*RV–DNA complex (Winkler *et al.*, 1993); (**C**) TATA box binding protein TBP (Kim and Burley, 1994); (**D**) catabolic activator protein CAP–DNA complex (Schultz *et al.*, 1991); (**E**) λ Cro–operator complex (Brennan *et al.*, 1990); (**F**) the phage 434 O_R2–R1-69 complex (Shimon and Harrison, 1993); (**G**) *Eco*RI–DNA complex (McClarin *et al.*, 1986); (**H**) Arc repressor–operator complex (Raumann *et al.*, 1994). (I–K) DNA crystal structures with inherent bending: (**I**, **J**) GGCC-containing oligonucleotides (Grzeskowiak *et al.*, 1993); (**K**) a TA-containing oligonucleotide (Quintana *et al.*, 1992). *x*-axis, DNA sequence. Segments in which the minor groove faces the bound protein are written in lower case. *y*-axis, predicted bending and/or bendability values; trinucleotide values (continuous line). Roll angles, determined by X-ray crystallography, are shown by bars and are indicated on the right hand side. In cases where the roll angles were not given in the original papers, kinks (white arrows) or regions of groove compression (underlined black arrows) are indicated, as originally given by the authors.

the test sequence has intrinsically curved elements. Finally, deducing from the helical phasing experiment, we demonstrated that the major groove-directed curvature is higher for GGCCC than for CCGGG and lower for CCAAT then for CTCAC sequence elements. This exactly confirms our predictions based on bending and/or bendability revealed by DNase I, presented in Table I.

The experiments shown in Figure 3 were carried out in the presence of 10 mM Mg^{2+} ions. When Mg^{2+} was left out of the electrophoresis buffers, few or no detectable differences were found between the respective pairs (data not shown).

Discussion

What is the difference between this trinucleotide model and the current dinucleotide models? In the first place, the trinucleotide model includes more context information: instead of a single value (e.g. AA/TT) describing a particular dinucleotide, we now have several trinucleotide values in which the dinucleotide is present (i.e. AAA/TTT, CAA/TTG, GAA/TTC, TAA/TTA, AAC/GTT, AAG/CTT and AAT/ATT). Of the previous models, only the nucleosome packing data (of Calladine and Drew, 1986) incorporate some context dependence.

A second, more important, difference is that, in contrast to the dinucleotide models, the trinucleotide model is in reasonable agreement with X-ray crystallographic data (Figure 2), even though the X-ray data were not included in the model. It is worth noting that the qualitative agreement between the trinucleotide data and the experimental roll angles is reasonably good, even though the protein–DNA complexes used as examples represent a variety of different binding mechanisms. This is in keeping with the notion that bending propensity is an intrinsic property of DNA, which can be exploited by different DNA binding proteins.

Thirdly, there are significant differences in the prediction of individual motifs. (i) It is supposed by all dinucleotide models that CC and CA can be bent toward the major groove. CCA, however, frequently has negative roll angles in crystals (underlined in Figure 2I and J), which is correctly shown by the low trinucleotide parameter values. (ii) It is considered that GC can be bent towards the minor

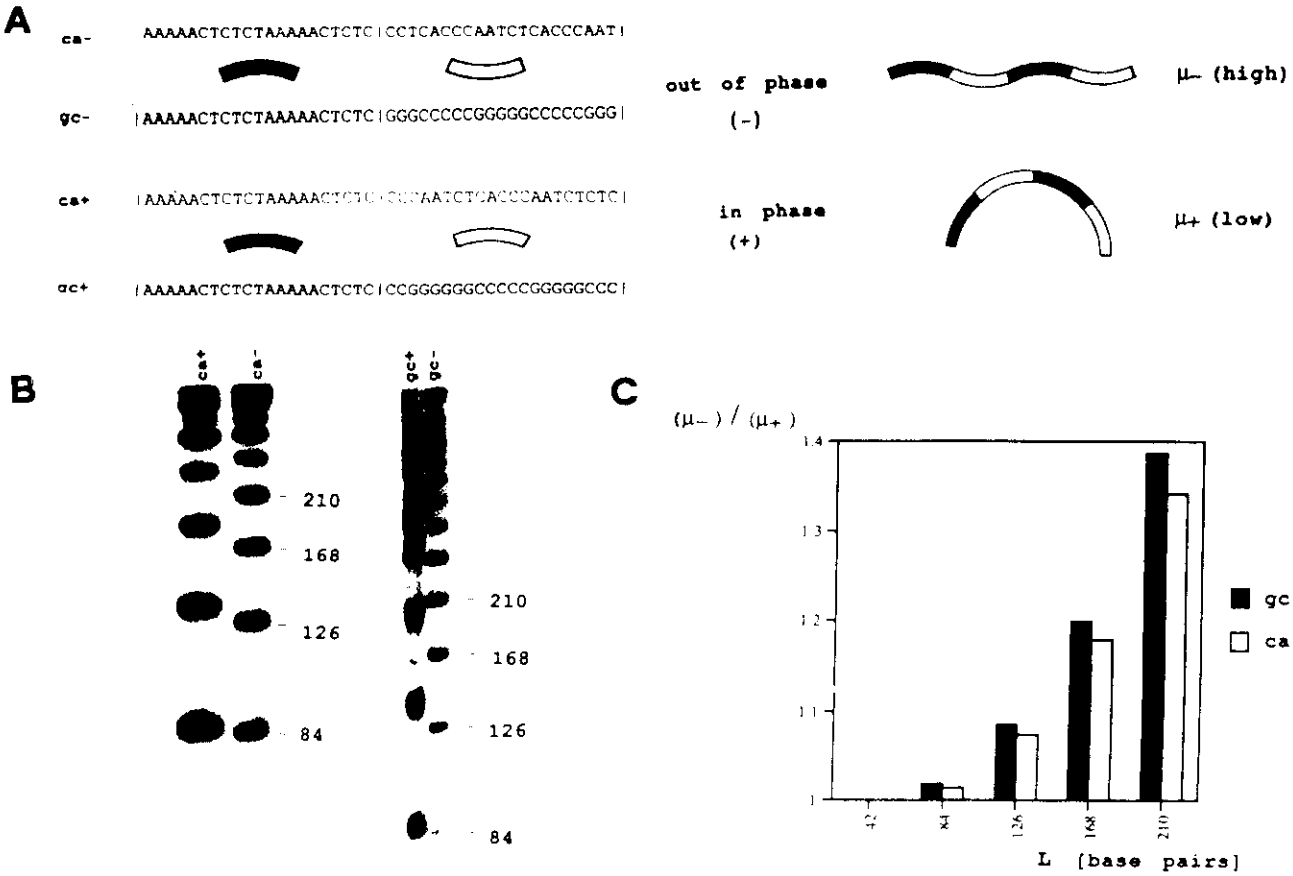


Fig. 3. Relative directions of curvature of tested sequence elements determined by helical phasing with A tracts. (**A**) Design of DNA molecules with 360° (in phase) and 180° (out of phase) angular distances between sequence elements to be tested (white arc) and A tract-containing DNA, used as reference elements (black arc). Note that the angular distance between test elements (180°) is the same in all molecules. (**B**) Dependence of gel mobility (μ) on the phasing between sequence elements of unknown direction of curvature. Lanes gc and ca are ligation multimers characterized by in phase (+) and out of phase (−) helical distance of tested elements from the A tracts. (**C**) The ratio between gel mobility μ− /μ+ of gc (black bars) and ca (white bars) multimers is plotted against linker length (L). Exerimental error in determination of μ is ±0.03. Macroscopic curvature is produced only if there is an appropriate helical phasing between curved sequence elements. This logic allows us to deduce that the major groove-directed curvature of GGGCCC is higher than CCCGGG and that CTCAC is higher than CCAAT.

groove by the dinucleotide models (except that of Calladine and Drew, 1986). In contrast, it is predicted that GGC is bent towards the major groove by the trinucleotide model, in agreement with X-ray crystallographic data (underlined in Figure 2I and J). (iii) The dinucleotide models predict that CG can be bent toward the major groove. CCG, on the other hand, has a low trinucleotide parameter, which is in agreement with the fact that CCG/CGG is found in a region of minor groove compression in the λ Cro−DNA complex (underlined in Figure 2E).

Perhaps the most intriguing of these differences are: (i) the opposite bending and/or bendability directions predicted for GCC and CGG elements respectively and (ii) the low bending propensity of the CA step in the context of the CCA motif. According to the purine/pyrimidine rule (Barber and Zhurkin, 1990) or the dinucleotide models of Bolshoy *et al.* (1991) and De Santis *et al.* (1990), bending/bendability features of CCA, GGC and CCG sequence elements show exactly the opposite trend from our trinucleotide model. We tested the curvature of GGCCC versus the CCGGG element and that of the CA step in the context of a CCA motif, using helical phasing experiments (Zinkel and Crothers, 1987) in the presence of Mg^{2+} ions (Diekmann, 1987; Shlyakhtenko *et al.*, 1990; Brukner *et al.*, 1994). DNA fragments which contain tested elements are placed at varying distances from the A tract (Figure 3). It was found that the major groove-directed curvature of GGC is higher than that of CCG and that the CCA trinucleotide has a bending propensity similar to that of the A tracts, i.e. low propensity for bending toward the major groove. Thus, both results are in firm agreement with the trinucleotide data.

The advantage of the approach used here lies in the fact that the experimental dataset (DNase I digestion data) can be expanded relatively easily. It has to be noted that the current dataset does not yet allow determination of the 136 tetranucleotide parameters. It should also be noted that the model is not meant for quantitative prediction of DNase I digestion profiles, since it does not consider local effects, such as specific enzyme−substrate interactions. In principle, local structural features recognized by DNase I can be captured by using sequence patterns. Recently, Herrera and Chaires (1994) determined such patterns, using a different and rather limited dataset. These patterns, however, show little correlation with our experimental data; only seven of the 20 preferred cleavage sites and none of the 20 most poorly cleaved sites found in our experimental dataset contain the patterns suggested by

Herrera and Chaires (1994). While it may be possible, in principle, to separate local and global effects, at present these influences are incorporated as a bias into the $p(a)$ parameters of the present model. In spite of these limitations, the trinucleotide model presented here seems to show, in a correct manner, the sequence-specific differences in DNA bending and/or bendability and offers a convenient way to predict the bending propensity of DNA using simple parameter versus sequence plots. We feel that the improvement is partly due to the trinucleotide-based representation and partly due to the larger dataset, since a previous dinucleotide model based on a smaller set of DNase I data (Brukner *et al.*, 1990) did not detect the differences which are now correctly shown by the present trinucleotide model.

The physical meaning of the trinucleotide parameters can be summarized as follows: for efficient cleavage, DNase I has to distort a particular sequence to a conformation that is optimal for cutting; the energy required for this purpose will be low for sequences which are (i) flexible (i.e. able to bend) or (ii) inherently bent to a conformation which is near to that required for cleavage. The trinucleotide parameters capture, at least in principle, these two properties at the same time, but do not in themselves enable us to distinguish between them. A discrimination between flexible and inherently curved sequence elements is possible only with additional experiments, like the helical phasing studies described here. In the future it will be possible to derive more accurate models, such as those based on tetranucleotides or on applying 'position-dependent' weighting schemes within the window. This will require, however, a larger dataset than the one currently available.

Materials and methods

DNase I digestion

pUC18 (Boehringer Mannheim, Mannheim, Germany) fragments (see Figure 1) and phage T7 A1 promoter DNA (kindly provided by Dr H.Heumann, Max-Planck Institute of Biochemistry, Martinsried, Germany) were dephosphorylated and end-labelled as described (Drew and Travers, 1984). After end-labelling, restriction cleavage was performed to remove the labelled nucleotide from one end of the DNA, which was purified on 8% polyacrylamide gels (Drew and Travers, 1984). DNA was subjected to digestion with DNase I and hydroxyl radicals, under limited digestion conditions, as described (Brukner *et al.*, 1993). These conditions produce on average 0.16 cuts/molecule, the obtained cleavage intensities therefore reflect the affinity of the cleavaging agent for different parts of the DNA. The fragments resulting from DNA cleavage and the Maxam–Gilbert G reaction (Maxam and Gilbert, 1980) were separated on 8% polyacrylamide denaturing gels containing 8 M urea. The cleavage patterns were recorded using autoradiography. Quantitative evaluation was carried out by scanning image plates in a Molecular Dynamics PhosphorImager System.

Data processing and calculations

The procedure was similar to that described before (Brukner *et al.*, 1990). Briefly, the densitometric readings (which are proportional to observed DNase I cutting frequencies) were determined in triplicate and averaged. The raw data obtained in this manner were corrected for cuts that occurred between a cut and the radioactive label and smoothed by a three bond averaging procedure (Drew and Travers, 1984; Brukner *et al.*, 1990). This was done in order to diminish local fluctuations caused by individual structural features recognized by DNase I. The value of stagger between the complementary strands was two to the 5′ direction (Brukner *et al.*, 1990).

The smoothed values were then substituted into equation (2), which was subsequently solved by the singular value decomposition least squares method (Press *et al.*, 1992) to give the ln $p(a)$ parameter of the trinucleotides (Table I).

The stability of the model was tested with computer simulation experiments. Random error terms with an expected average value corresponding to ±5, ±10, ±20, ±30 and ±40% of the measured values were generated and added to the raw data and the ln $p(a)$ parameter was calculated with the simulated dataset. This procedure was repeated 20 times and the average and standard deviation of the parameters were calculated. For the per cent error values listed, the standard deviation values (averaged for all parameters) were found to be 0.002, 0.004, 0.007, 0.012 and 0.016 respectively.

Helical phasing experiments

Oligonucleotides were synthesized using an Applied Byosystems 381A Oligonucleotide Synthesizer and purified by 8% denaturing polyacrylamide gel electrophoresis. The purified oligonucleotides (0.6 μg) were 5′-end-labelled as described (Brukner *et al.*, 1993), complementary strands were mixed, heated to 90°C and slowly cooled to form hybrids. The ligation reactions were carried out as described (Brukner *et al.*, 1993) and the ligated products were run on 8% polyacrylamide gels, in the presence of 10 mM $MgCl_2$, in both gel and gel running buffers (Diekmann, 1987). The applied voltage was 5 V/cm and the electrophoresis was carried out at 22°C.

Acknowledgements

The authors thank Professor A.Falaschi (ICGEB) for help and advice throughout this project and Ms Suzanne Kerbavcic for her help in preparing the manuscript. Drs Horace Drew, Richard Lavery, Ana Savic, Andrew Travers and Victor Zhurkin provided valuable suggestions on the work. The help of Drs Miklós Cserzö and Mensur Dlakic with a previous version of the least squares program is gratefully acknowledged. D.S. is suppoorted by a grant from Deutsche Forschungsgemeinschaft. I.B. and R.S. are recipients of an ICGEB fellowship.

References

Barber,A.M. and Zhurkin,V.B. (1990) CAP binding sites reveal pyrimidine–purine pattern characteristic of DNA bending. *J. Biomol. Struct. Dynam.*, **8**, 213–232.

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

Brennan,G.R., Roderick,S.L., Takeda,Y. and Matthews,B.W. (1990) Protein–DNA conformational changes in the crystal structure of a λ Cro–operator complex. *Proc. Natl Acad. Sci. USA*, **87**, 8165–8169.

Brukner,I., Jurukovski,V. and Savic,A. (1990) Sequence dependent structural variations of DNA revealed by DNase I. *Nucleic Acids Res.*, **18**, 891–894.

Brukner,I., Dlakic,M., Savic,A., Susic,S., Pongor,S. and Suck,D. (1993) Evidence for opposite groove-directed curvature of GGGCCC and AAAAA tracts. *Nucleic Acids Res.*, **21**, 1025–1029.

Brukner,I., Susic,S., Dlakic,M., Savic,A. and Pongor,S. (1994) Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J. Mol. Biol.*, **236**, 26–32.

Calladine,C.R. and Drew,H.R. (1986) Principles of sequence-dependent flexure of DNA. *J. Mol. Biol.*, **192**, 907–918.

Calladine,C.R., Drew,H.R. and McCall,M.J. (1988) The intrinsic curvature of DNA in solution. *J. Mol. Biol.*, **201**, 127–137.

De Santis,P., Palleschi,A. Savino,M. and Scipioni,A. (1990) Validity of the nearest-neighbour approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry*, **29**, 9269–9273.

Diekmann,S. (1987) Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. *Nucleic Acids Res.*, **15**, 247–265.

Drew,H.R. and Travers,A.A. (1984) DNA structural variations in the *E. coli tyrT* promoter. *Cell*, **37**, 491–502.

Drew,H.R. and Travers,A.A. (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.*, **186**, 773–790.

Gartenberg,M.R. and Crothers,D.M. (1988) Sequence determinants of CAP-induced bending and protein binding affinity. *Nature*, **333**, 824–829.

Grzeskowiak,K., Goodsell,D.S., Kaczor-Grezskowiak,M., Cascio,D. and Dickerson,R.E. (1993) Crystallographic analysis of CCAAGCTTGG

and its implications for bending in B-DNA. *Biochemistry*, **32**, 8923–8931.

Herrera,J.E. and Chaires,J.B. (1994) Characterization of preferred deoxyribonuclease I cleavage sites. *J. Mol. Biol.*, **236**, 405–411.

Hochschild,A. and Ptashne,M. (1986) Homologous interactions of lambda repressor and lambda Cro with the lambda operator. *Cell*, **44**, 681–687.

Hogan,M.E., Roberson,M.W. and Austin,R.H. (1989) DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl Acad. Sci. USA*, **86**, 9273–9277.

Hunter,C.A. (1993) Sequence-dependent DNA structure: the role of base stacking interactions. *J. Mol. Biol.*, **230**, 1025–1054.

Kim,J.L. and Burley,S.K. (1994) 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.*, **1**, 638–653.

Lahm,A. and Suck,D. (1991) DNase I-induced DNA conformation 2Å structure of a DNase I–octamer complex. *J. Mol. Biol.* **222**, 645–667.

Lavigne,M., Kolb,A., Yermian,E. and Buc,H. (1994) CRP fixes the rotational orientation of covalently closed DNA molecules. *EMBO J.*, **13**, 4983–4990.

Lomonossoff,G.P., Butler,P.J.G. and Klug,A. (1984) Sequence dependent variation in the conformation of DNA. *J. Mol. Biol.*, **149**, 745–760.

Maxam,A.M. and Gilbert,W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavage. *Methods Enzymol.*, **65**, 499–560.

McClarin,J.A., Frederick,C.A., Wang,B.-C., Greene,P., Boyer,H.W., Grable,J. and Rosenberg,J.M. (1986) Structure of the DNA *Eco*RI endonuclease recognition complex at 3 Å resolution. *Science*, **234**, 1526–1541.

Poncin,M., Hartmann,B. and Lavery,R. (1992) Conformation sub-states in B-DNA. *J. Mol. Biol.*, **226**, 775–794.

Press,W.H., Teukolsky,S.A., Vetterling,W.T. and Flannery,B.P. (1992) In *Numerical Recipes in C: The Art of Scientific Computing*, 2nd Edn. Cambridge University Press, Cambridge, UK, pp. 676–680.

Quintana,J.R., Grzeskowiak,K., Yanagi,K. and Dickerson,R.E. (1992) Structure of a B-DNA decamer with a central TA step: CGATTAATCG. *J. Mol. Biol.*, **225**, 379–395.

Raumann,B.E., Rould,M.A., Pabo,C.O. and Sauer,R.T. (1994) DNA recognition by β-sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.

Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.*, **191**, 639–659.

Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001–1007.

Shakked,Z., Guzikevich-Guerstein,G., Frolow,F., Rabinovich,D., Joachimiak,A. and Sigler,P.B. (1994) Determinants of repressor/operator recognition from the structure of the *trp* operator binding site. *Nature*, **368**, 469–473.

Shimon,L.J.W. and Harrison,S.C. (1993) The phage 434 OR2/R1-69 complex at 2.5 A resolution. *J. Mol. Biol.*, **232**, 820–838.

Shlyakhtenko,L.S., Lyubchenko,Y.L., Chernov,B.K. and Zhurkin,V.B. (1990) Influence of temperature and ionic strength on electrophoretic mobility of synthetic DNA fragments. *Molekulyarnaya Biologiya*, **24**, 79–95.

Suck,D. (1994) DNA recognition by DNase I. *J. Mol. Recognition*, **7**, 65–70.

Travers,A.A. and Klug,A. (1990) In Cozzarelli,N.R. and Wang,J.C. (eds), *DNA Topology and its Biological Effects*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 57–106.

Weston,S.A., Lahm,A. and Suck,D. (1992) X-ray structure of the DNase I–d(GGTATCC) complex at 2.3 Å resolution. *J. Mol. Biol.* **226**, 1237–1256.

Winkler,F.K. *et al.* (1993) The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.

Zinkel,S.S. and Crothers,D.M. (1987) DNA bend direction by phase sensitive detection. *Nature*, **328**, 178–181.

Received on November 28, 1994; revised on January 25, 1995

ELASTICITY OF DNA[1]

1.	INTRODUCTION	1
2.	THE CLASSICAL ELASTIC ROD	2
3.	DNA AS A CLASSICAL ELASTIC ROD	4
4.	PERSISTENCE-LENGTH DESCRIPTION OF DNA	6
5.	EULERIAN INSTABILITY OF DNA INDUCED BY COUNTERIONS	7
6.	BENDING BY ASYMMETRIC NEUTRALIZATION	12
7.	QUASI-ELASTIC BEHAVIOR OF KINKED DNA	13
8.	QUASI-EULERIAN INSTABILITY	15
9.	FINITE ELEMENT DESCRIPTION OF DNA	18
10.	ELASTIC PROPERTIES OF INDIVIDUAL DNA MOLECULES	22
11.	REFERENCES	23
	Appendix 1: Polyelectrolyte free energy calculation	26
	Appendix 2: Stiffness matrix calculation	28

These lecture notes review the elastic behavior of DNA. DNA is usuallymodelled as an elastic rod that smoothly bends and stretches in response to external force. DNA has two realistic properties that have to be explained: i) In a native state, the repulsion of the phosphates keeps DNA in a stretched form. If the phosphates are neutralized, DNA can collapse into compact forms necessary for packaging. If the phosphates are asymmetrically neutralized, DNA will asymmetrically bend. The effect of charges can thus be explained by elasticity theory - charge neutralization is analogous to the Eulerian instability of buckling rods. ii) In reality , DNA has kinks due to base pair opening at every 20 to 100 base pairs, but this realistic structure can be modelled by a quasi-elastic description. In summary, DNA's behaviour can be adequately modelled as an elastic rod.

1. INTRODUCTION

A long polymer molecule is flexible in solution. Its segments occupy a solvent-rich quasispherical domain, and if the radius of the domain is contracted by an applied compression, the polymer will spring open again to restore its original unconfined shape once the external constraint is removed. In a famous picture (1), a T2 phage DNA molecule has escaped from the ruptured phage head, not by spilling like liquid from a bottle, but by a dramatic elastic response, like Jack from the box.

Two general effects, acting in concert, account for the global, or bulk, elasticity of a polymer. i) Polymer segments diffuse radially outward after release from confinement (the "configurational entropy" increases). For this reason a rubber ball has its desirable properties. ii) Also, intersegmental forces influence the equilibrium shape. An excellent example is DNA itself, which bears a strong negative electrostatic charge from the ionized phosphate groups. Two aligned DNA segments close to each other exert a strong mutual repulsion, and many segments packed into a relatively tiny phage head generate an enormous ionic repulsion which dominates the diffusion component of the elastic response in the initial stages of release when the segments are still close (2).

[1] These notes were compiled, without formal editing, from reviews written by Gerald Manning, as well as from articles of W. Bauer, C. Bustamante, F. Caron and their associates. The sections on finite element methods were compiled by M. Gromiha, ICGEB.

Actually, there is a third elastic effect, of a more localized nature, on which we concentrate in this paper. On a length scale of 50 nm, or 150 base pairs (bp), a DNA polymer is stiff, like a rod. Only over longer lengths do thermal bending fluctuations gradually dominate, endowing the long polymer with global flexibility. Elastic restoring forces arise when a rodlike segment is bent, and the rod springs out to its straight form when the bending constraint is released. Resistance to local curvature is a significant impediment to packing DNA into a phage head. It is comparable in importance to the suppression of configurational entropy; both, however, are small contributions to the overall positive free energy of packing compared to ionic repulsion among segments (2).

On the other hand, there are circumstances when local bending is the only elastic response that occurs. Shore et al. (3) have showed that a 126 bp DNA fragment has an extremely small probability of end-to-end closure, attributed to severe elastic bending strain (local elastic twist also has important consequences in ring closure experiments, but does not bear on the present discussion). Their observation is illuminated by the root-mean-square bending amplitude for the thermal fluctuations of a "persistence-length" segment, which has about the same number of base pairs (see subsequent discussion). A typical bending fluctuation carries the axis of a persistence-length DNA segment onto an arc covering only about a sixth of a circle. A 126 bp DNA ring stores considerable elastic bending energy. If the constraint of ring closure is released, the circumferential axis springs open, like Jack-in-the-box, to about a sixth of a circle.

Wrapped around a nucleosome core, a full turn of the DNA axis is made over a length of 83 bp (4). If the attractive interaction between DNA and histone core were hypothetically, or by an actual cellular regulatory mechanism, to be "turned off," the DNA axis would spring out into a relatively extended form. DNA wraps twice around the core, and the adjacent turns are closely spaced. In this case elastic release involves ionic (phosphate-phosphate) repulsion between turns as well as relaxation of curvature.

In this note we review elastic bending in a DNA molecule with special regard to the role of the phosphate backbone (5-13) which seems to. Clearly, reduction of repulsive ionic interaction between DNA segments by partial or complete neutralization of the phosphate groups facilitates the formation of compact structures in the phage head, in chromatin, and in DNA toruses. But the point is that charge neutralization has a supplementary effect. Like a stiff rod, the DNA axis on a local length scale is ordinarily elastically stable when extended. If it is bent, it snaps back on release. The straight axis becomes unstable, however, if the phosphate charge is eliminated to a sufficient degree. Its stable form is then some bent configuration. If it is straightened, it snaps back to the bent form on release. This effect also facilitates compact structuring and may direct the most stable compact form. Restoration of the phosphate charge causes the compact structure to spring out to its original extended shape, an elastic response driven globally by ionic repulsion between segments and locally (as the local axis is restored to its straight form) by ionic repulsion *within* a segment.

The importance of the polyelectrolyte nature of DNA can be illustrated by the marginal stability of the double helical structure relative to two separated strands. From information contained in Chapter 6 of the book by Bloomfield et al. (14) we estimate the total free energy of stabilization against strand separation of E. *coli* DNA as -1.2 kcal (mol bp)$^{-1}$ at 25°C and 0.1M KCl. In the same conditions we calculate from the present polyelectrolyte theory (5) that phosphate-phosphate repulsion *destabilizes* the double helix relative to separated strands by the amount +1.8 kcal (mol bp)$^{-1}$. The evolutionary process barely succeeded in stabilizing the DNA double helix against internal polyelectrolyte disruption.

2. THE CLASSICAL ELASTIC ROD

Compression, bending, and twisting of a rod made of homogeneous and isotropic material are described by Hooke's Law. In its minimum free-energy (undeformed) state the rod is straight and has a circular cross-section of the same radius R along its entire length L. For the

compression-extension *(ce)*, bending, and twisting free energies ΔG we have the following relations (with $A = \pi R^2$ the cross-sectional area),

$$\Delta G_{ce} = 1/2\, EAL(\Delta L/L)^2 \quad [1]$$
$$\Delta G_{bend} = 1/2\, BL\rho^2 \quad [2]$$
$$\Delta G_{twist} = 1/2\, CL\tau^2 \quad [3]$$

These formulas are all examples of Hooke's Law: the free energy of a small elastic displacement is proportional to the square of the displacement. The *ce* displacement is $\Delta L/L$, the relative change in length of the rod, whether it be caused by compression ($\Delta L < O$) or stretching ($\Delta L > O$) along the axis of the rod. If the axis of the rod is bent to the form of an arc of a circle of radius ρ^{-1} (curvature ρ), the generalized displacement is ρ (a straight line has zero curvature), so ΔG_{bend} is proportional to ρ^2. Finally, if the rod is twisted about its axis, which remains straight, the displAcement is the rate of twist τ, that is, the angle of twist per unit length along the axis.

The Hooke's Law constants are Young's modulus, E, the bending rigidity, B, and the twisting (or torsional) rigidity, C. All three coefficients are thermodynamically intensive; they are characteristic of the material from which the rod is fashioned and in no way depend on the quantity of material in the rod. In particular, they are independent of the rod length L and of the cross-sectional area A.

The coefficients *E*, *B*, and C are related,

$$B = EI \quad [4]$$

and

$$B/C = 1 + \sigma \quad [5]$$

Here I is the "moment of inertia" of the cross-section, equal to $1/4\, \pi R^4$, and σ—an interesting and subtle quantity—is called Poisson's ratio. If a rod is stretched or compressed (axial displacement), its thickness changes (radial displacement); σ is the negative of the ratio of the radial to the axial displacement. The very general requirement of thermodynamics - that the free energy must be minimum for a stable state—dictates limits on σ,

$$-1 < \sigma < 1/2 \quad [6]$$

A negative value of σ means that the rod becomes thicker when stretched, thinner when compressed. No known homogeneous and isotropic elastic material does this. For all known materials σ is less than 1/2 but positive (16). However, the requirement of minimum free energy by itself does not rule out the existence of a homogeneous and isotropic substance with a negative Poisson ratio.

We may remark that the Hooke's Law Eqs. [1]-[3] are themselves illustrative of thermodynamic stability. An assumption in their formulation is that the stable form of the rod is unbent, untwisted, and of length L. Then all three ΔG values must be positive, so the Hooke's Law constants *E*, *B*, and C are positive. The straight, untwisted rod of length L is said to be *elastically stable.*

The straight, untwisted rod of length L can become elastically unstable in the following circumstance. Suppose an external compressive force is imposed on both ends of the rod, and that the force is constrained always to point along the original line of the rod axis. The rod is compressed at the cost of a rise in ΔG_{ce}. But if the direction of the axis of the rod becomes variable along the length (if the rod bends), then the compressive component of the fixed force becomes less than the full force, and ΔG_{ce} is smaller. The price paid for this decreased ΔG_{ce} is a positive ΔG_{bend}. It is always true that the price is too high if the force is small; the stable form of the rod is then a slightly compressed, but straight, version of the rod in the absence of the force. It is equally true that for a sufficiently large force, the gain in ΔG_{bend} is smaller than the

decrease of ΔG_{ce}. The rod then *buckles*. Its stable form, of least free energy, is bent. The straight rod has become *elastically unstable*. (More precisely, the straight form of the rod becomes an unstable equilibrium configuration, in analogy to a cone standing straight up on its vertex. The least thermal fluctuation of the axis away from a straight line will grow in amplitude instead of regress.)

Euler initiated the mathematical research needed for a quantitative description of this phenomenon (26). His efforts, and Lagrange's, led to a formula for the critical value of the compressive force F_{crit}, forces larger than which buckle the rod,

$$F_{crit} = 4\pi^2 B L^{-2} \qquad [7]$$

Longer rods are easier to buckle. The less resistant is the material of the rod to bending (small B), the more easily does the rod buckle. The constant $4\pi^2$ is determined by a particular boundary condition chosen to simulate the ends of a rodlike segment internal to a polymer (10).

The rod is stabilized in its buckled, or compact, form only by the presence of the externally imposed force. If the force is removed, or, more precisely, if it is decreased below the critical value, the compacted rod will spring open to an unbent, extended form (straight rod). In this sense we may say that the bent, compacted rod is elastically unstable.

3. DNA AS A CLASSICAL ELASTIC ROD

On a length scale of about 100 bp, DNA can be expected to exhibit some of the characteristics of elastic Hookian behavior. The twisting rigidity C, defined operationally by Eq. [3], has been measured (17-19). The statistical fluctuations of the end-to-end distance of very long DNA polymers may be described in terms of the "persistence length," so (see below) the bending rigidity B of short internal segments is also an experimentally accessible quantity (9,20).

An individual DNA molecule, however, is not the homogeneous isotropic material that forms the subject for much of the classical theory of elasticity. Averaged over a turn of the double helix, its structure may be isotropic and axially homogenous, but its radial inhomogeneity is evident. Whether its radial dimension is sufficiently large to meet the requirements of macroscopic theory is also problematic.

The bending rigidity of DNA is a function of temperature and ionic conditions (9,12,20). In the range 10-2-1.0M aqueous NaCl at room temperature, it has been measured to vary from about 3×10^{-19} erg cm to about 1×10^{-19} erg cm. It decreases as the ionic strength increases, attaining a value roughly constant at the latter figure for high ionic strengths. The twisting rigidity has not been systematically studied as a function of environmental conditions, but for the purposes of illustration we may assume that a recently measured value 2.9×10^{-19} erg cm is not too sensitive to ionic conditions (also see below) (18,19). From Eq. [5] we may then calculate values of Poisson's ratio σ varying from -0.7 to 0.0 (from high to low ionic strength). Interestingly, this range is inside that allowed by Eq. [6]—the requirement of minimum free energy for stable states imposed by classical macroscopic theory—but the negative values of σ_{DNA} over much of the range stand in contrast to the positive values of all known homogeneous and isotropic macroscopic materials (16). DNA, viewed literally as a classical elastic rod, becomes radially thicker when axially stretched.

A popular model that abstracts the polyelectrolyte nature of DNA is a circular cylinder with uniform surface charge (21,22). Motivation for the model is provided by the long range of the Coulomb ionic potential. Unshielded ionic repulsions among phosphate groups axially distant from each other cause a divergent (infinitely large) positive polyelectrolyte free energy, and the minimal shielding conditions maintained by low-to-moderate ionic strengths are expected to leave the axial direction singled out as dominating the repulsive ionic free energy. A variant model

would be a uniformly charged line as the axis of a cylinder (23). or, the interior of a cylinder may be uniformly filled with charge.

In all cases the cylinder without the charge may be assumed to obey the rules of classical elasticity theory. Since the cylinder without the charge is meant to represent DNA without its phosphate charge, deviation of the models from classical elastic behavior are attributed entirely to long-range electrostatic forces. In fact such forces are present in a real DNA molecule but are assumed to be absent in the development of classical elasticity theory.

For these models we can decompose Hooke's Laws into classical elastic and polyelectrolyte parts. For example, $\Delta G_{bend} = \Delta G'_{bend} + \Delta G''_{bend}$ and $B = B' + B''$. Calculations (21,22) show that Hooke's Law, Eq. [2], is obeyed for the polyelectrolyte component ($\Delta G''_{bend}$, B") alone, so the overall bending free energies of the models follow Hooke's Law. The models predict a strong ionic-strength dependence of B through its polyelectrolyte part B''. On balance, the charges are pushed more closely together when the cylinder bends, so the electrostatic bending free energy is positive; but higher ionic strengths increasingly shield the repulsions, so the positive electrostatic bending free energy decreases. At very low ionic strengths the theoretical result is that B" diverges to $+\alpha$. The symmetry of the models makes the behavior of the polyelectrolyte component of the torsional rigidity transparent; C" is equal to zero. By plausible assumption, neither B' nor C' depend on ionic strength in the low-to-moderate range. A conclusion is that, for these models, the ratio B/C of overall bending rigidity to overall torsional rigidity becomes large at very low ionic strengths. It then follows from Eq. [5] that Poisson's ratio for the models is large and positive at low ionic strengths, in contrast to its narrow confinement by the classical elastic stability requirement Eq. [6].

In summary thus far, measured values for DNA suggest that σ_{DNA} may attain atypical negative values in the range of Eq. [6]; whereas theoretical calculations on plausible models result in large positive values of σ_{DNA} at low ionic strength, well outside the upper limit of Eq. [6]. The latter result can be attributed to a long-range polyelectrolyte effect, which may be made small (in the models) by increasing the ionic strength. But the *negative* values of σ_{DNA} are measured at the *high* end of the ionic strength range. We conclude that DNA may not behave in all respects like a classical elastic rod.

We can discover another distinctive feature of a DNA elastic rod—its variable radius. Suppose we write Young's modulus E as the sum of two terms $E' + E''$. The first term E' represents DNA without its ionic phosphate charge, whereas E" is the effect of the ionic charge. At low to moderate ionic strengths, E'' may be calculated from polyelectrolyte theory *(see* Appendix 1) and turns out to have the form $E'' = fA^{-1}$, where A is the cross-sectional area and f is independent of A.

On the other hand, E' is expected to behave like the Young's modulus of an ordinary elastic rod, that is, to be an intensive quantity independent of the cross-sectional area. From Eq. [4] we get the relation $B = E'I + fA^{-1}I$, which becomes a quadratic equation in the unknown R^2 if the relations between A, I, and the radius R are substituted ($A = \pi R^2$, $I = 1/4\ \pi\ R^4$). The result of solving the quadratic for R^2 is

$$R^2 = (f/2\pi E')\ [(1+16\pi E' f^{-2} B)^{1/2} - 1] \qquad [8]$$

In this formula E' is expected from its meaning to be independent of ionic strength. The quantity f also does not depend on ionic strength *(see Appendix 1)*. However, as an experimental result (9,20), B does depend on ionic strength for DNA, so R does too. Since B increases as the ionic strength decreases, so does R. This conclusion is not unreasonable if the effective radius R is assumed in some sense to include the ion atmosphere that surrounds DNA in solution and shields its phosphate charge, for it is known from Debye-Huckel theory that a diffuse ion atmosphere expands as ionic strength decreases.

4. PERSISTENCE-LENGTH DESCRIPTION OF DNA

Consider a collection of DNA fragments in solution with many different lengths *L*. A very short fragment bends like an elastic rod. It conforms to Hooke's Law, Eq. [2], and possesses a bending rigidity *B*. It may be that, because of end effects, B depends on L, a slightly longer fragment may have a higher value of B. Eventually, however, a fragment length *L* will be found, past which *B* will not change. This plateau value of *B*, then, is "the bending rigidity of DNA" for the temperature and solvent under consideration, and by B we always mean this value.

We note that *B* has the units erg cm, so the ratio B/kT is a length (k is Boltzmann's constant and T is the Kelvin temperature). Call this length λ the persistence length of DNA. The theory of Bresler arld Frenkel, as presented by Landau and Lifshitz (14,24), shows that a fragment of length ***L* much shorter than** λ (but longer than the minimum length L for which a fragment is free of end effects) **behaves like an elastic rod subjected to thermal fluctuations**. Its thermal bending fluctuations are of small amplitude, so its average end-to-end distance is almost equal to the fragment length *L*. A segment internal to a long DNA polymer behaves the same way if its length is short compared to λ .

The theory of Bresler and Frenkel also shows that the axis of a fragment of length *L* much longer than λ traces out a randomly fluctuating path with no correlation between the directions of distant sections of the polymer. The cumulative effect of thermal bending fluctuations overwhelms the tendency of short sections "to go straight ahead." The result is the familiar picture of a randomly coiled, long, flexible polymer in solution. The quantitative formulas of the theory relate the mean-square end-to-end distance (or the mean-square radius of gyration) of fragments of any size longer than L to the length $\lambda = B/kT$. The mean-square radius of gyration of long polymers may be accurately measured by light scattering, so λ (or equivalently, *B)* may be extracted and tabulated for different sets of conditions (9,20) as in Table 1.

It is found (see Table 1) that λ and B decrease as salt concentration increases. **A short fragment or internal segment (short relative to the range of λ values) is bent with less work at high ionic strengths than at low**. A long polymer (much longer than λ) has a more crinkled, randomly coiled, appearance at high ionic strengths than at low (25), and it occupies a smaller volume domain (in accord with the Bresler-Frenkel theory, which predicts a smaller end-to-end distance as a consequence of smaller λ).

TABLE 1
Persistence Length λ and Bending Rigidity B of
DNA in Aqueous NaCl at 25°C

c_{NaCl}, M	λ, nm	λ, bp	B, erg cm
0.001	97	290	4.0×10^{-19}
0.01	75	220	3.1×10^{-19}
0.1	53	160	2.2×10^{-19}
1	32	94	1.3×10^{-19}

We can now describe a salt effect on the elastic response of a long DNA polymer on two levels-local and global. Locally, a short section of the polymer will bend like an elastic rod in the face of an imposed bending torque. It will snap back to its extended form on release of the torque. Added salt acts like an applied torque by weakening the bending ngidity, hence increasing the root-mean-square axial curvature. Lowering the ionic strength causes the segment to snap back to a position of decreased rms curvature corresponding to the increased bending rigidity.

On the global level we look at the spherical volume domain occupied by the polymer as a whole. If the domain is confined to a smaller volume, and then the confining constraint is removed, the domain spontaneously expands, just like a rubber ball released from having been squeezed. The polymer segments diffuse away from each other in an entropically driven process. Additionally, if the confinement was severe, the DNA segments are driven apart by electrostatic repulsion. An increased ionic strength acts like a radially confining force, since its effect is to decrease the size of the polymer volume domain. Decreasing the ionic strength effectively removes this force, and the domain follows the laws of elasticity, but as applied to a sphere, not a rod.

5. EULERIAN INSTABILITY OF DNA INDUCED BY COUNTERIONS

We wish here to provide an alternate interpretation for the shape of a long randomly coiled DNA polymer. First, we recall a discussion of Eulerian instability to buckling and rewrite Eq. [7] to read,

$$L = 2\pi (B/F)^{1/2} \qquad [9]$$

For given B and F, this formula tells us the minimum length of an elastic rod that will be buckled by the compressive force F. Longer elastic rods with the same bending rigidity B will also be buckled by the same force F, but shorter rods will not.

Thus for a given force F we can predict the minimum length of a DNA segment that will be buckled by it—under the assumption that asegment of this length behaves like an elastic rod. If this assumption is correct, then longer segments will also buckle in the face of the same force F, up to a segment of maximum length that ceases to respond like an elastic rod, roughly, that is, up to a segment of length equal to the persistence length (a "persistence-length segment").

We note further that an elastic segment of length L given by Eq. [9] will just buckle; the axis will deviate from a straight line by only a small amplitude. Longer segments, if still elastic, will buckle to shapes that deviate more and more from a straight line (Figure 1).

We now indicate how increases in ionic strength (addition of NaCl) can generate compressive forces in any polyelectrolyte, including DNA. The polyelectrolyte free energy G^{polyel} is positive (*see* Appendix 1 for specific formulas), since it equals the work against phosphate-phosphate repulsion needed to assemble phosphate charges from infinite separation onto their locations in the DNA structure. Further, the dominant dimension is axial; in fact, the only length parameter that appears in the present formulas is a distance along the axis. The negative derivative of G^{polyel} with respect to axial length L is the force exerted on the DNA structure by phosphate-phosphate ionic interactions. It turns out to be positive (Appendix 1), that is, extensile, as a reflection of the obvious fact that phosphate-phosphate repulsions stretch the DNA polymer along its axis. Suppose now that the phosphate-phosphate repulsive forces are

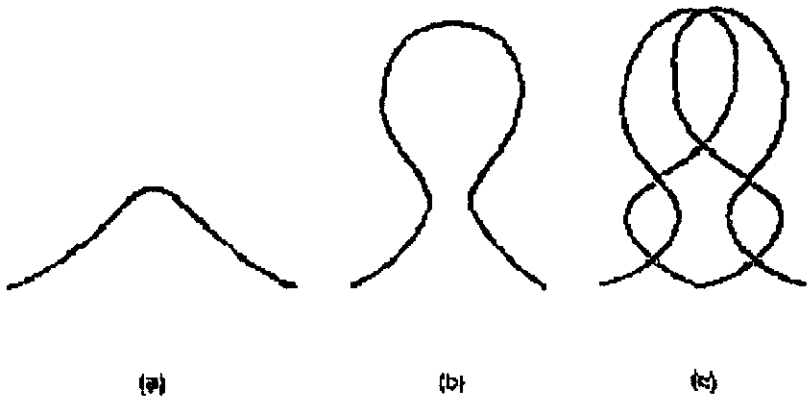


Figure 1. Buckled forms of an elastic rod in response to increasingly large (from left to right) above-critical compressive forces applied at the ends. The drawings are modified from Figure 55 of Love (16) to account for more appropriate boundary conditions *(10)*.

annulled, say, by "turning off" the charge on each phosphate group. The axial stretching force disappears. An equivalent effect is produced, if, instead of annihilating the phosphate charges, a compressive force is imposed just equal to the extensile phosphate-phosphate force. Again, *the effect of phosphate charge annihilation may be ascertained by replacing the charge annihilation with the assumption that the portions of the DNA polymer adjoining an internal segment subject the segment to a compressive force equal in magnitude to the stretching force of the phosphates.* An obvious extension of this reasoning covers the more general case of partial annihilation of phosphate charges, or of the weakening of their effect by simple ion-atmosphere screening in the sense of Debye-Huckel theory. We will implement the idea quantitatively by means of the equations for $-\delta G^{polyel}/\delta L$ given in Appendix 1.

We can now establish, in a qualitative way, that a long randomly coiled DNA polymer is a buckled form—in the Eulerian sense—of the completely extended polymer. In the same sense the completely extended polymer is elastically unstable. To this end, consider a long DNA polymer in the absence of any counterions. An array of many uncompensated charges, the phosphate groups in this case, has essentially infinite (i.e., very large) free energy because an unscreened Coulomb potential tends so slowly to zero as a function of distance. In the equilibrium configuration, then, the charges are as far apart as possible, so the DNA polymer is completely extended (26), and it has infinite bending rigidity B. The stretching polyelectrolyte force, $-\delta G^{polyel}/L$, is infinite as well.

Let Na^+ counterions be added to the solution, and excess NaCl also, so that the ionic strength (NaCl concentration, since in excess of DNA phosphate concentration) becomes 10^{-3}M. We compute (Appendix 1) a new value of the stretching force, 6.00×10^{-6} erg cm^{-1}, that, because of shielding of the phosphate charge by bound Na^+ and ion-atmosphere screening, is much lower than the infinite value obtaining before counterions were added. The equivalent compressive force F is the difference between $-\delta G^{polyel}/\delta L$ before and after addition of counterions, i.e., F is infinite. Ignoring the niceties of rigor for this extreme case, we interpret the ratio α/α in Eq. [9] as a nonzero, finite number and thus get a finite, if unspecified, value for L in Eq. [9].

How should we interpret the value of L thus obtained? Segments of longer length, which were rodlike before addition of counterions, are buckled by the infinite force *F*, the longer the length, the greater the degree of buckling (Figure 1). For a length approaching the length of the polymer, the degree of buckling is so large that a myriad of alternate configurations of almost equal energy are consistent with the degree of buckling. Thermal buffeting from one configuration to another renders the buckled polymer indistinguishable from a random coil. Segments of length shorter than *L* are not buckled by F. Their axes remain straight, subject to

slight thermal fluctuations. We expect, then, that L and the persistence length λ are approximately the same thing, $L = \lambda$, or about 290 bp at 10^{-3}M NaCl and room temperature (Table 1).

We conclude that ordinary randomly coiled DNA at 10^{-3}M NaCl is a buckled form of completely extended DNA in the absence of counterions and salt. Alternately stated, the extended form becomes elastically unstable after counterions and salt are added. The buckling force is provided by addition of counterions and salt. The bending rigidity is infinite. The persistence length after imposition of the buckling force is interpreted as the maximum segment length that is not buckled. There is nothing special about 10^{-3}M NaCl; the qualitative conclusion is the same for any salt concentration. ***We have therefore given an exnmple, at the most primitive possible level of biological organization, of a dramatic "packaging" of DNA (from complete extension to random coil) brought about by Eulerian instability of the extended form relative to the buckled form induced by ordinary*** Na^+ ***counterions.***

It is essential to lay stress on the meaning of the bending rigidity B in this, and in the following, context. B *is* the bending rigidity before, during, and after imposition of the buckling force F. B does not change in the Eulerian buckling description. B *is not* related to λ after buckling by the formula $B=kT\lambda$ discussed in the previous section. The value of B given by this relation pertains to the conventional (and perfectly correct) polymer description of the previous section. In *this* section B is the bending rigidity of a rodlike DNA segment at specified environmental conditions. A change in environment (e.g., a change of salt concentration) is conceptualized as a buckling force imposed on a segment of unchanged bending rigidity, in accord with the Eulerian theory of elastic instability of rods with characteristic invariant Hooke's Law constant.

Just as we may regard a polyelectrolyte random coil at some salt concentration as the consequence of an unstable fully extended form, we can describe a random coil at a higher salt concentration as the result of buckling rodlike segments of a random coil at a lower salt concentration. Table 2 illustrates this type of analysis. We have chosen 10^{-3} as the reference low salt concentration, but similar results are obtained for several other choices of reference, as noted below. The first column gives the NaCl concentration c. The second column lists the compressive force F at each NaCl concentration. These values were calculated in the following way. At the reference concentration $c = 10^{-3}$M, the polyelectrolyte stretching force is $-\delta G^{polyel}/\delta L = 6.00 \times 10^{-6}$ erg cm^{-1}, as calculated from the theoretical equations in Appendix 1. At a higher value of c the stretching force is computed in the same way, and, as expected (because of increased ionic shielding), it has a lower value. F is then the difference between the reference stretching force and the stretching force at the higher salt concentration. For example, we see from the table that an increase of NaCl concentration from 10^{-3} to 10^{-1}M is equivalent to application of the compressive force 1.31×10^{-6} erg cm^{-1}. Of course, the equivalent applied force is zero if the ionic strength is not increased.Now in all this, the bending rigidity B is always that at $c = 10^{-3}$M.

From Table 1 its measured value is 4.0×10^{-19} erg cm. Thus we get the third column L from Eq. [9] and convert the length to base pairs (bp) with the known structural factor 3.4×10^{-8} cm/bp. The infinite value for the reference ionic strength, which follows immediately from Eq. [9] and the lack of a compressive force, requires physical interpretation. Since Eq. [9] has meaning only for elastic rods, $L = \alpha$ means that the longest rodlike segment that cannot be buckled is the longest rodlike segment in the polymer, i.e., roughly, a persistence length. Persistence lengths λ for each salt concentration are taken from Table 1 and listed in the last column.

The reference entries for L and λ have been forced to agree, but the rough correlation between L and λ for the other salt concentrations is noteworthy and consistent with the present interpretation. For example, an increase of salt from 10^{-3} to 10^{-1}M buckles rodlike segments (rodlike at 10^{-3}M) of length greater than 100 bp, so that at 10^{-1}M the longest rodlike segment is

about 100 bp in length. The relatively long extended segments at 10^{-3}M NaCl become elastically unstable in the Eulerian sense when compressed by the force F equivalent to raising NaCl to 10^{-1} M.

TABLE 2
Interpretation of Salt Increase as an Euler Buckling Force, *F*

c_{NaCl}, M	F, erg cm^{-1}[a]	L, bp[b]	λ, bp[c]
0.001	0	α[d]	290
0.01	0.64×10^{-6}	150	220
0.1	1.31×10^{-6}	100 [130][e]	160
1	1.97×10^{-6}	83 [89,[e] 110[f]]	94

a Reference salt concentration, 10^{-3}M, as described in text.
b Longest segment unbuckled by F; reference 10^{-3}M, unless otherwise noted.
c Measured persistence lengths from Table 1.
d Interpreted in text as equaling the corresponding value of A.
e Reference salt 10^{-2}M.
f Reference salt 10^{-1}M.

We have noted that the values of L in Table 2 depend on the choice of reference ionic strength c. A different choice of reference gives the numbers in brackets for L. The variation, which gives a degree of ambiguity to the interpretation, is not large.

To this point we have provided perhaps an amusing and certainly rough description of random coiling of a polyelectrolyte molecule. Its possible real value may emerge as we analyze cases to which the usual persistence-length approach cannot be directly applied. Suppose we want to know the effect of total annihilation of the DNA phosphate charge. The persistence length cannot be measured in this state of null charge. Indefinite increase of NaCl concentration does not reduce the phosphate charge to zero (5), and besides, very high ionic strengths imply a solvent markedly different from typical aqueous conditions. The effective phosphate charge can be progressively decreased by binding of counterions of progressively higher valence [Mg (II), Eu (III), etc.] (5), but it is well known that DNA forms quasi-ordered compact structures—as distinct from a random coil with persistence length—when sufficient amounts of multivalent cations are bound. With the present method, however, we easily get a value L for the longest stable extended rodlike segment of a DNA molecule in typical aqueous conditions, but with zero phosphate charge.

Let the phosphate charge be annulled at 10^{-2}M aqueous NaCl. From Table 1 the measured value of B is 3.1×10^{-19} erg cm at 10^{-2}M NaCl. The stretching force at normal phosphate valence -1 is $-\delta G^{polyel}/\delta L = 5.36 \times 10^{-6}$ erg cm^{-1} at 10^{-2}M NaCl, as calculated from the theoretical equations in Appendix 1. The stretching force is zero when the phosphate charge is zero. The force F thus equals 5.36×10^{-6} erg cm^{-1}. From Eq. /9/, $L = 44$ bp. In this way we construct Table 3 for a range of ionic strengths. Annihilation of the phosphate charge at typical NaCl ionic strengths buckles extended rodlike segments of lengths down to 30-50 bp, depending on ionic strength. (Given the present interpretation of L as the longest rodlike segment of an uncharged DNA polymer, the variation of L with ionic stength is unexpected and may provide a measure of the roughness of the interpretation). Table 3 also lists measured persistence lengths (referring, of course, to phosphate valence -1). The small values of L compared to A mean that annihilation of the phosphate charge facilitates formation of compact DNA structures. The small values of L in Table 3 compared to Table 2 show that the effect of direct phosphate charge annihilation at any fixed NaCl concentration is much greater than the effect of an increase of NaCl from 10^{-3} to 1M.

We can bridge the gap between the relatively small effect of increased NaCl concentration (from some typical reference value) and the large contraction caused by total charge annihilation

at fixed NaCl concentration. In aqueous NaCl of any typical ionic strength (less than 1M), 76% of the DNA phosphate charge is neutralized by territorially bound

TABLE 3
The Longest Unbuckled Segment Length *L* at Zero Phosphate Valence Compared to the Persistence Length λ at Phosphate Valence -1

c_{NaCl}, M	0.001	0.01	0.1	1
L, bp	48	44	40	33
λ, bp	290	220	160	94

Na^+ ions (5). on titration with Mg^{2+} ion neutralization can be increased to an upper limit of about 88%. In general (5), titration with N-valent ion can neutralize the DNA phosphate charge by up to about 100 ($1 - 0.24N^{-1}$) percent. The cited upper limit of charge neutralization is exact in solutions of low salt.

Figure 2 shows what happens when titrations with N = 2 and 3 are performed at three different NaCl concentrations. The invariant bending rigidity B for each curve in the figure is the value measured at the NaCl concentration specified by the curve, in the absence of any di- or trivalent ions, and is obtained from Table 1. The compressive force F is, as before, the difference of two stretching forces $-\delta G^{polyel}/\delta L$. The stretching force for conditions corresponding to a point on a curve in Figure 2 is subtracted

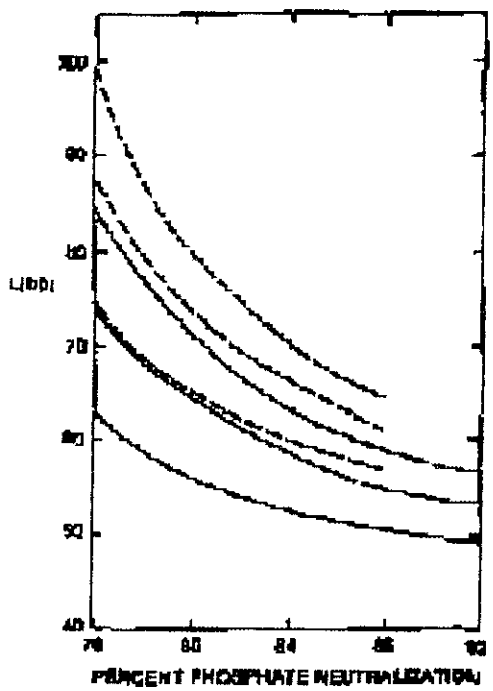


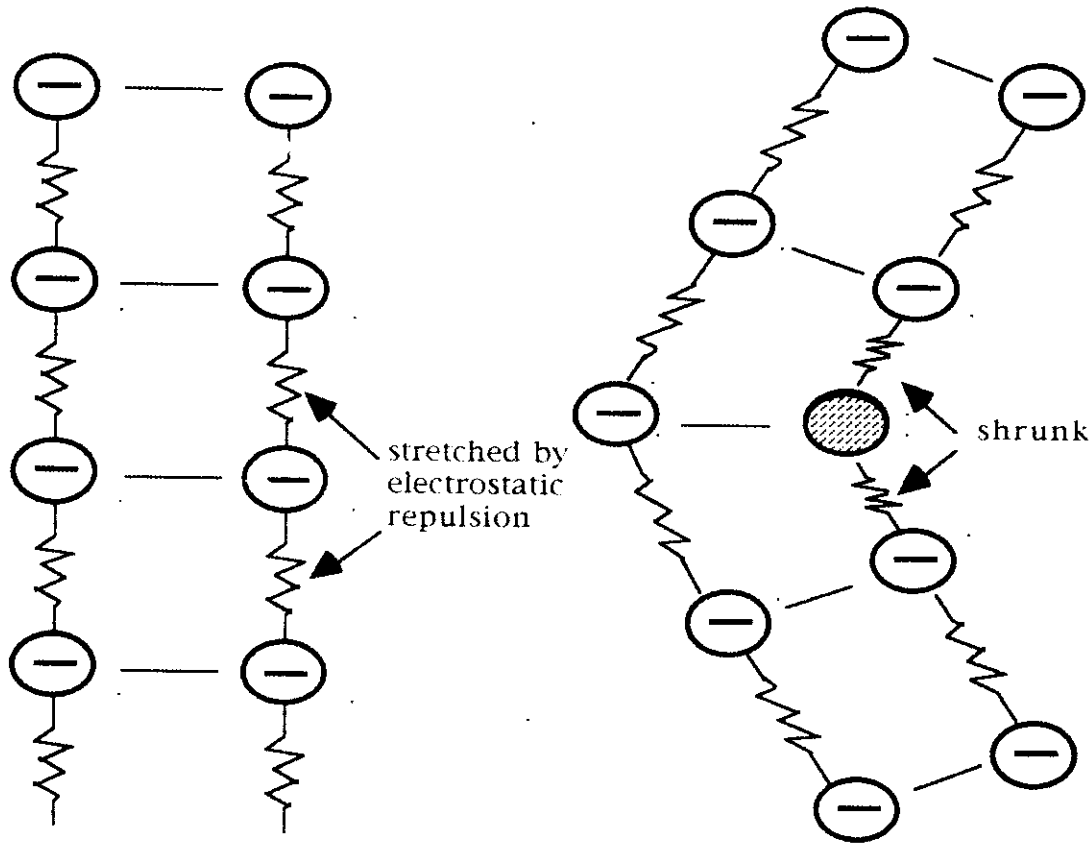
Figure 2 Minimum buckling lengths L as a function of extent of phosphate neutralization by N-valent cations in excess 1:1 salt. Dashed lines, N = 2; solid lines, *N = 3*. Within each set of three curves, the NaCl molarity varies from top to bottom: 10^{-3}M, 10^{-2}M, 10^{-1}M.

from the larger stretching force at the same NaCl concentration, but in the absence of N-valent ion. Information sufficient for this calculation is provided in the Appendix 1. The curves in Figure 2 begin at 76% neutralization, where nearly all (all, for low NaCl concentration) bound Na^+ ions have been displaced by bound N-valent ions. In accord with the discussion above, the curves for N = 2 end at 88% phosphate neutralization; for N = 3, at a higher value 92%.

Examination of Figure 2 along a horizontal line at fixed L (the longest stable rodlike segment length) shows that the higher the NaCl concentration the lower the degree of phosphate neutralization by N-valent ions required to buckle segments of length down to L. Looking along a vertical line at fixed charge neutralization, we see that trivalent ions produce greater compaction than divalent ions, even if the extent of charge neutralization is the same.

6. ASYMMETRIC NEUTRALIZATION BENDS DNA

According to the polyelectrolyte theory, repulsion of the negative charges of the phosphate backbone counteracts the elastic forces of the molecule. This model is sufficient to explain the bending phenomena caused by asymmetric charge neutralization. Strauss and Maher (60) synthesized asymmetrically neutralized ds DNA molecules that were strongly bent.



Normal DNA

Synthetic DNA with one chemically neutralized phosphate

This phenomenon is a consequence of Manning's polyelectrolyte theory. Another interesting consequence is the explanation for protein induced DNA bending: Most proteins bend DNA toward the major groove while they bind to the major groove. It appears that the positive charges of the protein neutralize the DNA phosphate charges in an asymmetric fashion, so DNA will band exactly as shown in the figure. There are proteins that bend DNA by binding to the minor groove: for those cases the polyeclectrolyte theory does not offer explanation.

7. QUASI-ELASTIC BEHAVIOR OF KINKED DNA

We have indicated that an individual DNA polymer in solution does not necessarily have the elastic response of the homogeneous, isotropic rod of classical theory. In fact, an equilibrated DNA polymer in solution has about one of every twenty to a hundred base pairs disrupted by thermal agitation *(12,13,27)*. Hydrogen atoms that form interbase hydrogen bonds in an intact base pair are free to exchange with solvent hydrogens in a disrupted pair. one or both members of the disrupted pair are therefore able to swing out into the solvent, and a relatively large amount of space is then created between the two intact pairs on either side of the open pair. We therefore expect the site of an open pair to be also a site of greatly enhanced flexibility. ***The expected dominant mode of DNA bending is kinking (abrupt change of axial direction) at the site of an open base pair. There are kinks at intervals of about 20 to 100 base pairs.***

Observation of smoothly bent DNA segments in a crystal, with all base pairs intact, does not bear on this question, since thermal disorder manifested as disrupted base pairs in solution is evidently, as expected, substantially suppressed under the relatively ordered arrangements imposed by crystalline force fields *(28)*. Modeling studies that assume all base pairs to be intact are, in this connection, also besides the point, since DNA in solution has been observed experimentally to possess a large number of open pairs.

Another expected consequence of open base pairs is a slight contraction along the helical axis of the intact segment intervening between two open pairs, relative to its unperturbed length recovered when these pairs close. Such a segment contains 20-100 base pairs. It is short, hence stiff. A force imbalance created at its ends is propagated throughout the interior, and its structure responds accordingly (consider an example of the opposite extreme: equal opposed forces acting at the ends of a relaxed piece of string). When initially intact base pairs at its ends are opened and swing out into the solvent, base pair stacking forces are partially disrupted at the ends of the segment but not in its interior. The structural response is a slight axial contraction to a new structural equilibrium, driven by imbalanced attractive stacking forces in the interior of the segment.

We make two remarks on this argument for the coupling of axial contraction to base-pair opening. The "attractive base-pair stacking forces" represent the net effect of the entire system of forces that, at structural equilibrium, counterbalance the axially repulsive phosphate-phosphate ionic forces. They include forces arising in the sugar-phosphate backbone (exclusive of the ionic forces). The statement that stacking forces are disrupted when base pairs open is an assumption, but perhaps a plausible one. Secondly, we recall that the contracted segment is internal to the polymer as a whole. The open base pairs at its ends are internal base pairs. We do not discuss structural events that may occur near the ends of the polymer as a whole.

We can form a picture of a DNA segment bent by a thermal fluctuation. On average, as indicated in Figure 3, it is a regular polygonal arc with angle (kink angle) α and side of length s. (In the notation of ref. 12, s, or more precisely s_0 below, is denoted by l'). At each vertex there is an open base pair. As this pair swings further out into the solvent, the average kink angle α increases, and the average length s of the intact segment between open pairs decreases. Note that s is measured here in ordinary length units, not as a number of base pairs as in a previous section; the number of base pairs in segment s does not change. Further, elementary geometric theorems show that the kink angle α in Figure 3 is equal to the kink angle as more commonly defined—the angle formed by a side of the polygon and the straight extension of an adjacent side.

At first glance, it may appear that this type of bending (expected to be the real bending mode of DNA) cannot be described energetically by Hooke's Law, Eq.[2]. There is no smoothly

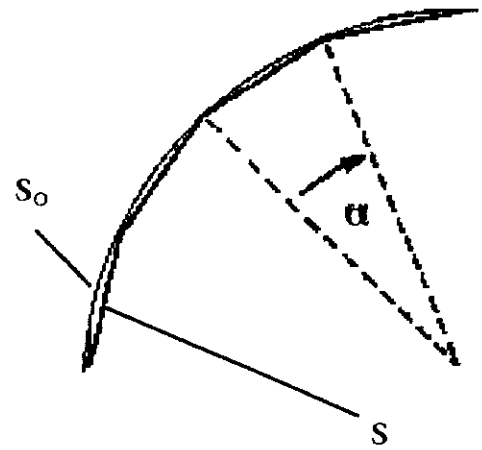


Figure 3. Fragment of average polygonal arc followed by DNA duplex axis in postulated dominant bending mode, showing kink angle α. There is an open base pair at each vertex. The circular arc has length s_o between successive vertices. Bending compresses the segment from length s_o to length s.

bent circular arc, hence no constant curvature ρ. The length L in Eq. [2] is independent of ρ. Nevertheless we can show that DNA bent in the present sense probably obeys Eq. [2].

As in Figure 3, consider the circular arc circumscribed about the polygonally bent DNA axis. Deform the helical axis so that it lies along the circular arc, while at the same time closing the open base pairs. The result is a smoothly bent, fully intact DNA segment. (This structure is within the range of those attainable by DNA, but is a high-energy, low-probability form in solution. The low-energy, high-probability structure in solution, as discussed, contains open base pairs). Since there are no open base pairs in the smoothly bent intact segment, there are no apparent structural features that would cause a change in axial length. Thus we suppose that the smoothly bent intact segment lying along the circumscribed circular arc in Figure 3 has length L_o independent of the radius ρ^{-1} of the circular arc. The length L_o is just the average length of a straight intact DNA segment. In the following we designate by s_o the corresponding length of segment s, that is, the uncontracted length. For example, $L_o/s_o = 4$ in Figure 3.

If the kink angle is small, we can write

$$\Delta G = 1/2\ \beta\ (L_o/s_o)\ \alpha^2 \qquad [10]$$

for the free energy of bringing a straight DNA segment into the kinked form of Figure 3. The energetic unit here is a side and an angle of the polygon. The free energy of forming the angle α and compressing the side from length s_o to length s is written together as $1/2\ \beta\alpha^2$ because when the angle is zero, the side is not compressed. For purposes of comparison, we note that $1/2\ \beta$ equals the quantity $(\beta_K + \beta_C l')$ discussed in ref. 12.]

By the definition of angle we have the relation $\alpha = s_o\rho$ where ρ is the curvature of the circular arc in Figure 3 (ρ^{-1} is the radius of the circle). Note that the length s_o of an uncontracted side and the length of the circular arc segment connecting the two ends of a side in Figure 3 are equal according to the above discussion. Equation [10] then becomes,

$$\Delta G = 1/2\ \beta s_o L_o \rho^2 \qquad [11]$$

that is Hooke's Law for bending a rod of length L_o, Eq. [2], with bending rigidity $B = \beta s_o$ (in agreement with Eq. [7] of ref. 12).

Because Hooke's Law for rodlike bending holds in its usual form, the Eulerian instability calculations of the previous section apply here as well. The results in the tables and figures of the previous section are in principle valid, with no modification, even if DNA bending proceeds by means of abrupt kinks rather than smoothly. The classical model loses meaning, however, when

applied to segments of length L less than $2s_0$ (the circular arc in Figure 3 is defined by three consective vertices). The numerical values for minimum buckling lengths L computed in the previous section are similar to the expected values of s_0 (about 20-100 basepairs). We can conclude that rodlike segments substantially longer than the lengths L given in the previous section do indeed buckle as described there, but that the interpretation of L as a sharp minimum length for buckling may become blurred.

8. QUASI-EULERIAN INSTABILITY OF KINKED DNA

We have seen that DNA segments of appropriate length L are unstable to bending fluctuations, in the Eulerian sense, if an axially compressive force F is imposed. We can exhibit a second sense in which segments are unstable to bending fluctuations in the presence of the same force F. The physical meaning of this situation is that two different classes of bent shapes may both be stable with respect to the straight rod; both may have negative free energies relative to the undeformed axis. A currently unexplored problem is to determine which of the two classes has the lower free energy.

An application of trigonometry to Figure 3 gives us a relation between the contraction of a side and the kink angle, $\Delta s = s_0 - s = (1/24)\alpha^2 s_0$, valid for small angles α. The contraction ΔL of a polygonal arc of undeformed length L_0 is the number of sides times the contraction of a side, $(L_0/s_0)\,\Delta s$. The work W done by a compressional force F applied to each side in the polygonal arc is $F\Delta L$. We then have the following result for the additional work ΔG needed to form the polygonal arc from the straight segment of length L_0.

$$\Delta G = 1/2\,\beta\,(L_0/s_0)\,\alpha^2 - W = 1/2\,(\beta - 1/12\,Fs_0)\,(L_0/s_0)\,\alpha^2 \qquad [12]$$

where we use Eq. [10].

According to Eq. [12] the free energy of a kinked segment of unkinked length L_0 is zero for the critical force $F = 12\beta s_0^{-1}$ and becomes negative for forces F exceeding this value. Negative values of ΔG mean that positive work $|\Delta G|$ must be done on the kinked form to straighten it. ***Kinks at open base pair sites can be stabilized by compressional forces in excess of a critical value acting along the intact segments between open base pairs.***

From the relation $B = \beta s_0$ (see Eq. [11]), we can write the critical force in the form

$$F = 12\beta s_0^{-2} \qquad [13]$$

where B *is* the Hooke's Law bending rigidity for a rod. Comparison with Eq. [7] for the Eulerian critical force reveals a strong analogy. The essential difference is that L in Eq. [7] can be assigned any numerical value (as long as it refers to a rodlike segment), while so here is a specific length of molecular-structural significance (the average distance between two open base pairs). The resemblence of Eqs. [7] and [13] prompts the designation quasi-Eulerian for the present type of instability.

Further insight may be gathered with the relation $\alpha = s_0\rho$ (see preceding section), which transforms Eq. [12],

$$\Delta G = 1/2\,(B - 1/12\,Fs_0^2)\,L_0\rho^2 \qquad [14]$$

into Hooke's Law for rodlike bending (*see* Eq. [2]), but with an effective bending rigidity that becomes negative when F exceeds the critical force. The onset of quasi-Eulerian instability is marked by a change of the bending rigidity from positive to negative values.

We give some numerical results. In Table 4 ΔG of kinking is listed for several NaCl concentrations. It is obtained from Eq. [14] with F = 0 and the value of B in Table 1 for the corresponding NaCl concentration. Figure 4 reports ΔG from Eq. [14] as a function of the percentage of phosphate charge

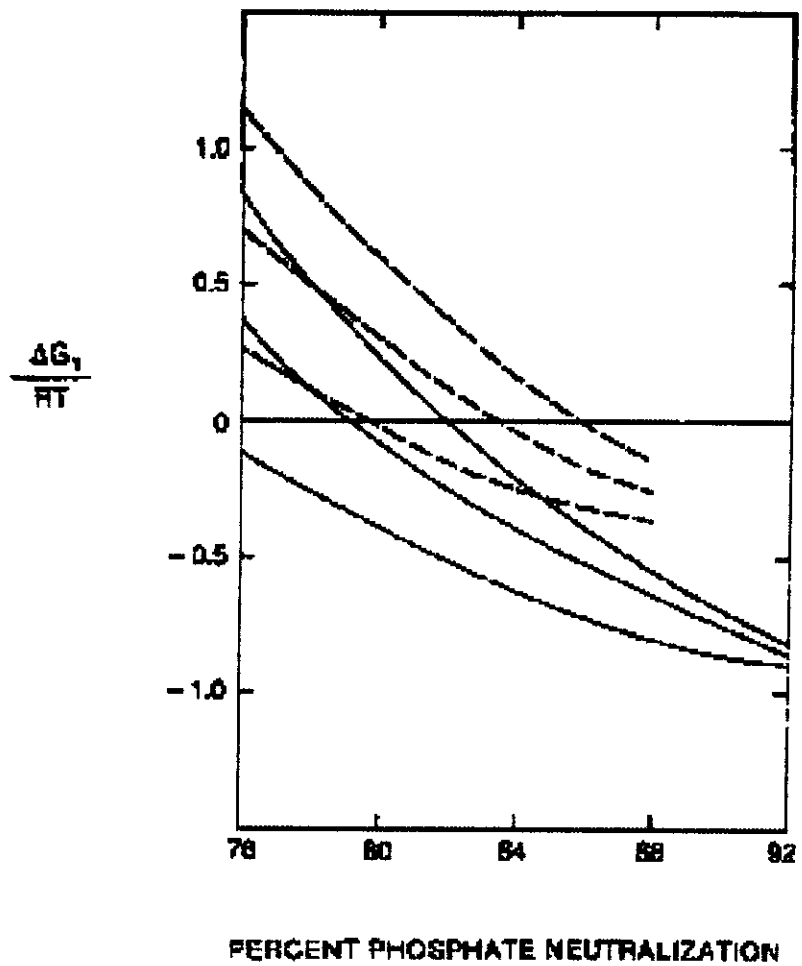


Figure 4. Unit kinking free energy as function of extent of phosphate neutralization by N-valent cations in excess 1:1 salt. Dashed lines, N = 2; solid lines, N = 3. Within each set of three curves, the NaCl molarity varies from top to bottom: 10^{-3}M, 10^{-2}M, 10^{-1}M. Horizontal line at zero free energy separates region of kink instability (top) from region of kink stability (bottom).

neutralized by territorially bound divalent or trivalent cations at various ionic strengths set by NaCl (or any other univalent buffer). Again, B is taken from Table 1 at the corresponding value of c_{NaCl}. The forces F are the same as those calculated in the construction of Figure 2; F is a reaction force equal and opposite to that part of the polyelectrolyte stretching force which is annulled when a multivalent cation neutralizes the phosphate charge from 0% to the percent indicated in the figure.

The free energies ΔG_1 in Table 4 and Figure 4 refer to a particular structural unit, a segment of length s_0 kinked at its center (*see* Figure 3); L_0 is set equal to s_0 in Eq. [14]. Longer lengths L_0 imply proportionally increased values of ΔG. For example, the "double unit" used in ref. (12) has $L_0 = 2s_0$ and $\Delta G = \Delta G2 = 2\Delta G_1$, where ΔG_1 is given in Table 4 and Figure 4.

This unit ΔG_1 depends on numerical values of s_0 and α ($\rho = \alpha s_0^{-1}$ in Eq. [14]). Our choice of s_0 is 13 nm, or 38 base pairs, based on a comparison of theory with measured persistence lengths and discussed in ref. (12). We specify 41° as the kink angle α. We determined this value on theoretical grounds in ref. (12) as a maximum value, past which Eq. [12] of Hooke's law form and thus valid only for "small" elastic displacements, is expected to break down. The matter has not been investigated, but new sets of repulsive forces (steric overlap, if nothing else) must ultimately begin to dominate the structure, increasing its energy more abruptly than the Hooke's Law quadratic dependence. Our numerical choices for s_0 and α should be understood as average values. When used in Eq. [14], α is converted to radians and s_0 to cm.

If the effective bending rigidity (parenthetical factor in Eq. [14]) is positive, ΔG increases from zero as a function of α according to Eq. [14] until α reaches a value of about 41°. The corresponding positive value of ΔG_1, is reported in Table 4 or Figure 4. If the effective bending rigidity is negative, ΔG decreases from zero, following Eq. [14] through negative values until a largest negative ΔG is attained at 41°. Then, following some other energy law (of course, the transition is broad, with Eq. [14] phased out as other forces become dominant), ΔG turns up, and a potential energy (better, free energy) well is created. The negative values of ΔG_1 in Figure 4 give the depth of the well, i.e., the stabilization energy of the kink unit.

The free energies in Table 4 and Figure 4 are in units of thermal energy *RT* (about 0.6 kcal at room temperature). For example at 10^{-1} and 1M NaCl, where (in the absence of multivalent cations) DNA is still a stiff random coil (persistence lengths 160 and 94 bp, or 53 nm and 32 nm,

TABLE 4
Free Energy of Kinking to $\alpha = 41°$ as a Function of NaCl Concentration

c_{NaCl}, M	0.001	0.01	0.1	1
$\Delta G_1/RT$	1.9	1.5	1.1	0.6

respectively—*see* Table 1). Table 4 indicates that the unit kinking free energy is positive, but rather small, barely equal to *RT* or less. In other words, a small positive ΔG_1 of order *RT is* sufficient to maintain DNA in a state of rough local extension on a length scale substantially larger than s_0 (38 bp, or 13 nm). Numerically similar values of ΔG_1, but negative, are attained in Figure 4 by the trivalent cation curves at 88-91% neutralization. We note also that if a curve had been drawn in Figure 4 for a cation of very large valence, this curve would extend nearly to 100% phosphate neutralization and there would have the value $\Delta G_1/RT = -2.1$, independently of ionic strength. (If the curves in Figure 4 are extended to the left, to 0% neutralization by multivalent cations, each curve for a given NaCl concentration converges to the corresponding entry in Table 4).

9 DNA MODELLING WITH THE FINITE ELEMENT METHOD

The previous sections proved that DNA's behaviour can be well aproximated with elasticity theory. The Finite Element Method (FEM) is a general purpose numeric approach widely used to calculate elastic deformations of complex structures in architecture and mechanic engineering (59) Recently it has been applied to study the fundamental properties of DNA. The great advantage of FEM is that it gives analytical solutions to complex non-linear problems. Such problems in molecular structure studies are usually tackled with energy minimization or with molecular dynamics approaches on full-scale atomic modells. FEM uses the much simpler models based on elastic rodes, but offers analytical solutions.

In FEM, a given physical or mathematical problem is modelled by dividing it into small or fundamental parts called "elements". Next, an analysis of the physics or mathematics of the problem is made on these elements. Simple, very often linear equations are used for the individual elements. Finally, the elements are "reassembled" and the system of equations (consisting of all equations for all elements) is solved. There are sophisticated and widely used computer programs for this kind of calculations. Given the very large number of simple elements, quite complex problems can be modelled.

Typical aplications are the changes of shape of objects under pressure or load. For example, one can modell the behavior of a simple beam (or of an entire bridge) under load. In such cases, the object (e.g. a beam) is subdivided into small elements with defined elastic behavior. Often one describes the element as being ideally elastic (This is not an incorrect approximation since the elements as well as their distortions, are very small). Then he force is applied at the given point(s) and the equations of elastic deformations are solved for the entire object.

The procedure used in FEM analysis can be explained by the following steps:

1. The boundary and the interior of the continuum are subdivided by points, lines or surfaces into a finite number of discrete sized sub-regions or finite elements.
2. The nodal points and the elements themselves are given numeric identifiers, the so called node numbers and element numbers. Using these (and other) identifiers one can define the entire topology of the problem (e.g. the object one is trying to modell).
3. The nodal parameters like e.g. pressure, velocity, displacement are defined for each node.
4. The material parameters (stiffness matrix) and external parameters (forces, temperature etc) are established; this process usually includes physical intuition. Some problems require the definition of properties at the nodal points. eg. stress analysis, require the thickness of material at each node point.
5. The element equations are established using the information from stiffness matrix and then the contribution of each element. The system of equations resulting from a finite element analysis will always be symmetric.
8. Apply the boundary conditions by (i) defining explicit values of the unknowns on the boundary and by (ii) defining constraint equations that are linear combination of the nodal quantities.
9. The system equations is solved using th appropariate and programmed.

In general we need the following basic inputs:
i. co-ordinates of the nodes (nodal co-ordinates) and the linkage of nodes to describe the geometry.
ii. material properties (Young's modulus, Poison ratio etc.)

iii. load and boundary conditions.

For modelling with finite element methods, DNA is usually considered to be a linearly elastic rod having the appropriate cross section and material constants ("moduli"), including the modulus of elasticity and the shear modulus. The cross section is usaually taken to be circular, with a diameter of 20 or 30 Å. If we use the radius of 10Å , the moment of inertia, I, will be equal to $\pi/4$ nm^4. The modulus of elasticity, *E*, is usually taken as

$$E = 3.4 \times 10^{-19} \text{ dyne cm}^{-2}.$$

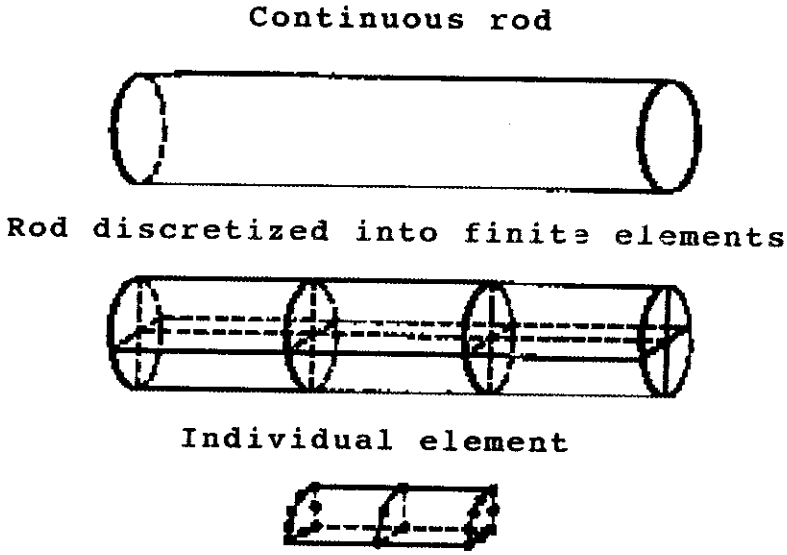
This can be calculated from the persistence length λ of DNA, the temperature and the Boltzmann constant as $E = \lambda k_B T/I$.

The shear modulus, G, is obtained from the torsional rigidity, C, with the expression $C = C/2I$. For the value of $C = 2.5 \times 10^{-19}$ erg.cm, the value of G is

$$G = 1.59 \times 10^{-19} \text{ dyne.cm}^{-2}.$$

The next step is to model the initial DNA configurations in the FEM context, as three-dimensional continua. The DNA rod is divided along its axis into a discrete number of appropriate segments

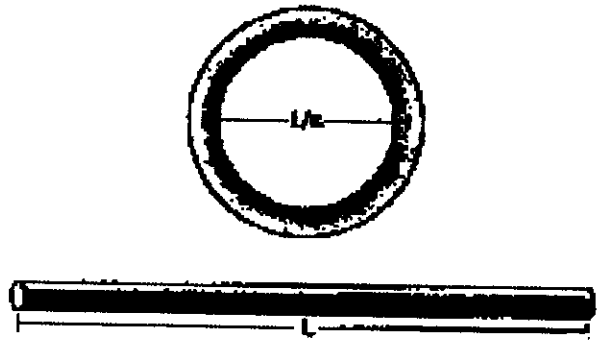
William Bauer and his associates at U. of NY Stony Brook and UCLA use the following division in which each element is about 160 Å or 46bp long.



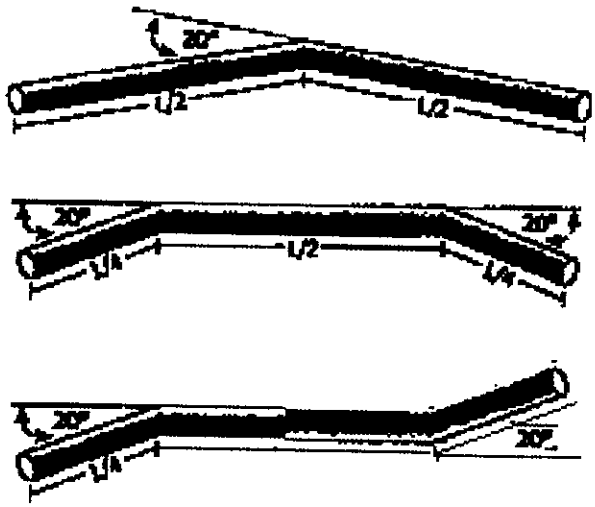
Wilma Olson and her associates ar Rutgers University (New Jersey), use cilindrical elements of 21 bp.

The questions that can be asked at this level of approximation are mostly concerned with **supercoiling**. Simply put, the DNA molecule modelled in this particular way, is twisted and the potential shapes are compared with those that can be observed experimentally. In lab experiments, stable supercoiled DNA can be prepared in a circular form (by exposing a piece of DNA to chemical agents, then circularizing it with enzymes) so also FEM studies concentrate on circular DNA. According to electron microscopic observations, supercoiled circular DN will

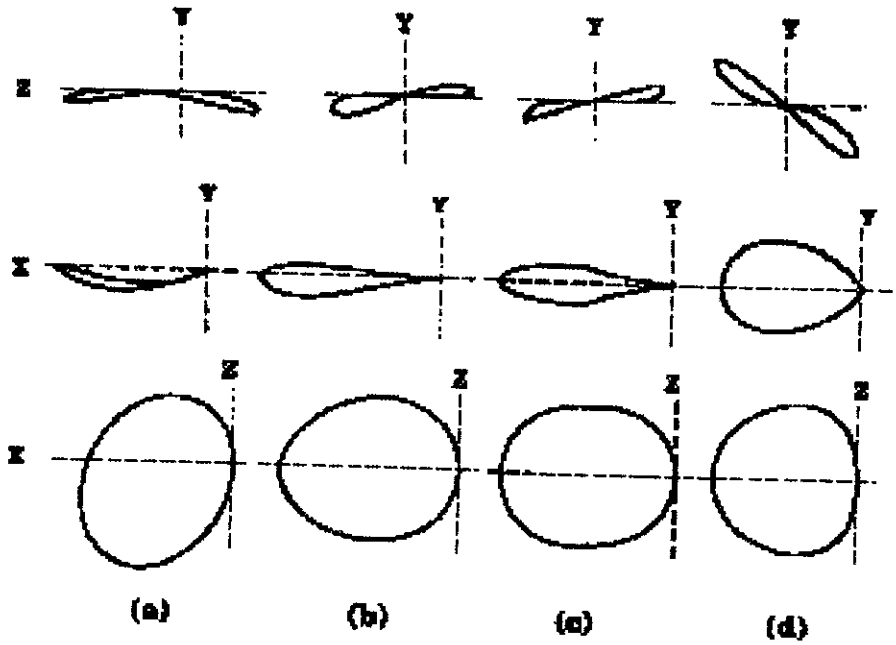
take avariety of shapes at any given degree of supercoiling. These shapes are (sometime) obtained as low energy solutions provided by the FEM programs.



In addition to supercoiling, one can also introduce stable bends into DNA, by adding elements of fixed angular distortions, e.g.:



The top two elements represent curves caused by one and two bends, respectively, the third one is a combination of two opposite bends. Including these into the circle and applying various levels of "overtwisting" i.e. supercoiling, one gets shapes that clearly resemble the experimentaly observable ones. The following are modelling results on a circular piece of 630 nm long DNA piece (approx. 2100 basepairs in length).



a) Initially straight structures, 2 turns of rotation before circularization;
b) Single bend, 1.6t urns of rotation before circularization;
c) Double bend, 1.5 turns of rotation before circularization;
d) O-ring (originally closed, 0.375 rotation applied;
(W. Bauer et al.)

Apparently, the closed circle (d) is sensitive to overtwisting. The linear-then-closed molecules sustain a larger amount of twisting to give the same shape.

It is important to realize that FEM-based calculations yield (a few) minimum-energy structures that correspond to structures of minimum elastic energy at 0 Kelvin. Work in this field differs in the extent of distortions - highly distorted modells need sophisticated numeric approaches and may be less reliable. Sophisticated modells give a large number of solutions that need to be analyzed by first clustering them into groups - similar to how structures obtained by molecular dynamics simulations are evaluated.

10. ELASTIC PROPERTIES OF INDIVIDUAL DNA MOLECULES

Recently, several groups (A. Bensimmon, C. Bustamante, F. Caron) determined the elastic properties of individual DNA molecules. The principle is the following: polymer beads are attached to the end ov individual DNA molecules, and the beads are stretched with a constant force so that the displacement is accurately measured with a microscope.

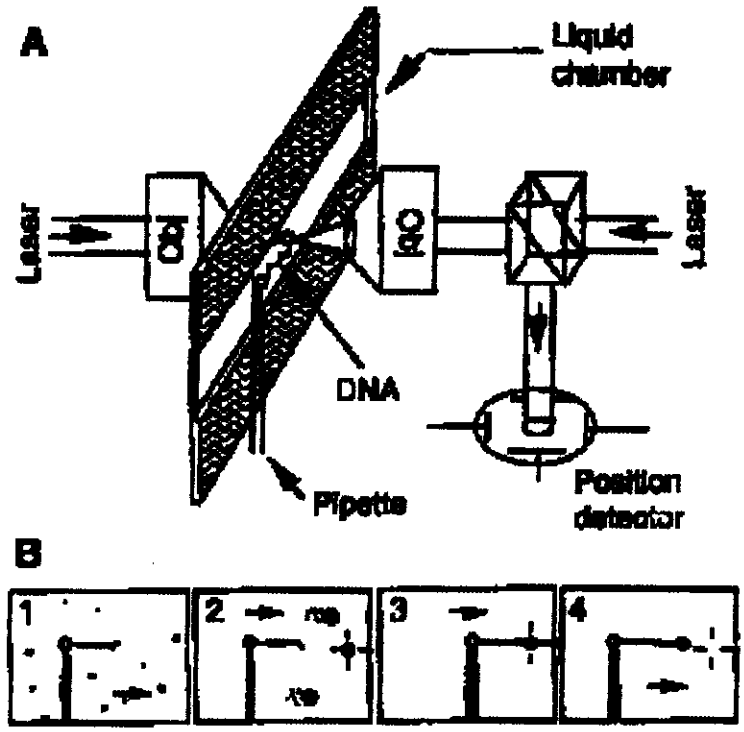
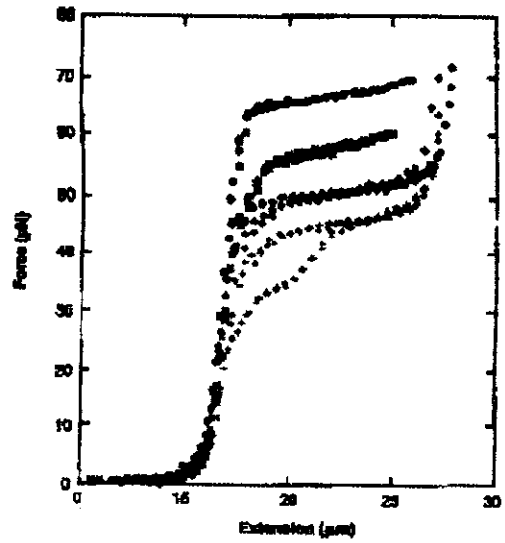


Fig. 1. (A) Two diode lasers with 100-mW power and 800-nm wavelength (SDL-5311-G1) are oriented with orthogonally polarized beams. These beams converge to a common focus by means of two objective lenses (Nikon CF Plan Achromat 60x with correction, NA = 0.85), thus forming an optical trap. Only the rays exiting the trap are directed to position detectors (UDT SC- 1 OD) by polarizing cube beamsplitters (only right cube and detector are shown). White-light illumination and video camera (not shown) are coupled through the objectives with dichroic beamsplitters, thus allowing measurement of the distance between bead centers. The molecular length is inferred from a set of such distances (27). (B) Bead-DNA assembly from solution. Arrows indicate buffer flow inside the fluid chamber. (1) DNA molecules are compact random coils in solution except when one attaches by an end to the bead on the pipette tip. (2) Extra beads are carried by buffer and one is caught by the laser trap. (3) The pipette is moved closer to the trap so DNA can span the gap. (4) The presence of the invisible DNA is evident because it pulls the bead upstream when the pipette is moved away from the trap (60).

The results proove that DNA behaves like a truely elastic molecule. There are however three phases of elastic behaviour: i) Coiled DNA can be stretched with small forces ranging from 0.01 to 10 pN. In this phase the force that counteracts the pulling is a result of entropy. ii) upt ot forces of 50 pN, DNA reacts as an elastic body with a Young's modulus of 3.4 ±0.3 Nm^{-2}. This value is equivalent to a persistence length of 193 bp aassuming a rod radius of 10Å. iii) Beyond forces of 65 pN the polymer suddenly yields and lengthens to about 1.7 times its original contour length. This stretching is ***reversible*** and increases helical rise to 5.8Å from a value of 3.4Å in B-DNA. R. Lavery has modelled the structure of this peculiar conformation (S-DNA) - it corresponds to a highly elongated ds helix. More importantly, molecular mechanics principles seem to explain the stretchnig process.



Stretching of dsDNA molecules in 150 mM NaCl, 10 mM tris, 1 mM EDTA, pH 8.0 buffer (black curve), Na2 sEDTA buffer (pH 8) with Na+ concentrations of 5 mM (red curve), 2.5 mM (green curve), and 0.625 mM (blue curve). For labeling dsDNA at both ends but on opposite strands, bio11-dCTP (Sigma), dATP, dGTP, and dUTP were polymerized opposite A's 12-bp sticky ends using Klenow enzyme. For cross-linking studies, DNA was exposed to a saturated solution of 4,5',8-trimethyl psoralen and irradiated with ultraviolet light according to the method of Ussery et al. (29). This method should produce 5% intercalation of psoralen (30), and indeed the contour length was increased by ~5% over the molecule in Fig. 2.

11. REFERENCES

1. Kleinschmidt, A. K., Lang, D., Jacherts, D., and Zahn, R. K. (1962), *Biochim. Biophys. Acta* 61, 857.
2. Riemer, S. C., and Bloomfield, V. A. (1978), *Biopolymers 17*, 785.
3. Shore, D., Langowski, I., and Baldwin, R. L. (1981), *Proc. Natl. Acad. Sci. USA 78*, 4833.
4. McGhee, J.D., and Felsenfeld, G. (1980) Ann. Rev. Biochem. 49, 1115
5. Manning, G. S. (1978), *Quart. Revs. Biophys.* 11, 179.
6. Manning, G. S. (1979), *Biopolymers 18*, 2929.
7. Manning, G. S. (1980), *Biopolymers 19*, 37.
8. Manning, G. S. (1981), *Biopolymers* 20, 1261.
9. Manning, G. S. (1981), *Biopolymers* 20, 1751.
10. Manning, G. S. (1981), *Biopolymers* 20, 2337.
11. Manning, G. S. (1982), *Comments Mol. Cell. Biophys. 1*, 311.
12. Manning, G.S. (1983), *Biopolymers* 22, 689
13. Manning, G.S., in *Structure and Dynamics: Nucleic Acids and Proteins*, Clementi, E., and S
14. Bloomfield, V. A., Crothers, D. M., and Tinoco, Jr., I., *Physical Chemistry of Nucleic Acids*, Harper & Row, New York, 1974.
15. Landau, L. D., and Lifshitz, E. M., *Theory of Elasticity*, Pergamon Press, oxford, 1970.
16. Love, A. E. H., A *Treatise on the Mathematical Theory of Elasticity*, Dover, New York, 1944.
17. Shore, D., and Baldwin, R. L. (1983). J. *Mol. Biol* 170, 957.
18. Shore, D., and Baldwin, R. L. (1983), J. *Mol. Biol* 170, 983.
19. Horowitz, D. S., and Wang, J. C. (1984), J. *Mol. Biol 173*, 75.
20. Post, C. B. (1983), *Biopolymers 22,1087.*

21. Le Bret, M. (1982), J. *Chem. Phys.* 76, 6243.
22. Fixman, M. (1982), J. *Chem. Phys.* 76, 6346.
23. Manning, G. S. (1969), J. *Chem. Phys. 51,* 3249.
24. Landau, L. D., and Lifshitz, E. M., *Statistical Physics,* Pergamon Press, London, 1958.
25. Frontali, C., Dore, E., Ferrauto, A., Gratton, E., Bettini, A., Pozzan, M. R., and Valdevit, E. (1979), *Biopolymers* 18, 1353.
26. Pfeuty, P., Velasco, R. M., and de Gennes, P. G. (1977), J. *Phys. (Paris) Lett.* 38, 5.
27. Englander, S. W., and Kallenbach, N. R. (1984), *Quart, Revs. Biophys.* 17, 1.
28. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980), *Nature* 287, 755.
29. Gosule, L. C., and Schellman, J. A. (1978), J. *Mol. Biol.* 121, 311.
30. Thomas, T. J., and Bloomfield, V. A. (1983), *Biopolymers* 22, 1097.
31. Yen, W. S., Rhee, K. W., and Ware, B. R. (1983) J. *Phys. Chem.* 87, 2148.
32. Widom, J., and Baldwin, R. L. (1980), J. *Mol. Biol.* 144, 431.
33. Widom, J., and Baldwin, R. L. (1983), *Biopolymers* 22, 1595.
34. Benbasat, J. A., *Biochemistry.*
35. Chattoraj, D. K., Gosule, L. C., and Schellman, J. A. (1978), J. *Mol. Biol.* 121, 327.
36. Ruben, G. C., Marx, K. A., and Reynolds, T. C., *39th Ann. Proc. Electron Microscopy Soc. Amer., Atlanta, Georgia, 1981,* G. W. Bailey, ed.
37. Marx, K. A., and Ruben, G. C. (1983), *Nucleic Acids Res.* 11, 1839.
38. Marx, K. A., and Ruben, G. C., in *The Molecular Basis of Cancer* (Rein, R. R., ed.), Liss, New York, in press.
39. Castleman, H., and Erlanger, B. F. (1983), *Cold Spring Harbor Symp. Quant. Biol.* 47, 133.
40. Marx, K. A., and Reynolds, T. C. (1982), *Proc. Natl. Acad. Sci.* USA 79, 6484.
41. Marx, K. A., and Reynolds, T. C. (1983), *Biochim. Biophys. Acta* 741, 279.
42. Grosberg, A. Yu. (1979), *Biofizica 24, 32.*
43. Bina, M., Sturtevant, J. M., and Stein, A. (1980), *Proc. Natl. Acad. Sci. USA* 77, 4044.
44. McGhee, J. D., and Felsenfeld, G. (1979), *Proc. Natl. Acad. Sci.* USA 76, 2133
45. McGhee, J. D., and Felsenfeld, G. (1982), J. *Mol. Biol.* **158, 685.**
46. McGhee, J. D., and Felsenfeld, G. (1980), *Nucleic Acids Res.* 8, 2751.
47. Simpson, R. T. (1978), *Biochemistry* 17, 5524.
48. Suau, P., Bradbury, E. M., and Baldwin, J. P. (1979), *Eur. J. Biochem. 97,* 593.
49. Renz, M., Nehls, P., and Hozier, J. (1977), *Cold Spring Harbor Symp. Quant. Biol.* 42, 245.
50. Renz, M., and Day, L. A. (1976), *Biochemistry* 15, 3220.
51. Thomas, J. D., and Butler, P. J. G. (1980), J. *Mol. Biol. 144,* 89.
52. Butler, P. J. G., and Thomas, J. D. (1980), J. *Mol. Biol. 100,* 505.
53. Belmont, A., and Nicolini, C. (1981), J. *Theoret. Biol. 90,169.*
54. Thoma, F., Losa, R., and Koller, T. (1983), J. *Mol. Biol. 167,* 619.
55. Cole, R. D., Lawson, G. M., and Hsiang, M. W. (1977), *Cold Spring Harbor Symp. Quant. Biol.* 42, 253.
56. Finch, J. T., and Klug, A. (1976), *Proc. Natl. Acad. Sci. USA 73,1897.*
57. Sperling, L., and Klug, A. (1977), J. *Mol. Biol.* 112, 253.
58. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984), *Nature* 311, 532.
59. Huston, R.L. and Passerello, C.E. (1984) Finite Element Methods,Marcel Dekker Inc. New York and Basel, and Ed Akin, J. (1986) Finite Element Analysis for Undergraduates, Academic Press Inc, London

60. Steven B. Smith, Yujia Cui, Carlos Bustamante (1996) *Science, 271, 795-799*
61. Philippe Cluzel, Anne Lebrun, Christoph Heller, Richard Lavery, Jean-Louis Viovy, Didier Chatenay and Francois Caron (1996) *Science, 271, 792-795*

Appendix 1: CALCULATION OF POLYELECTROLYTE FREE ENERGY

The basic tool used in the quantitative analyses of this paper is an equation (5) for the polyelectrolyte free energy G^{polyel} of a polymer segment bearing Np monovalent charged groups, phosphates in the case of DNA, immersed in NaCl (or some other 1:1 salt) solution of ionic strength c (NaCl molarity),

$$G_{polyel}/N_pkT = (1-\xi^{-1})\ln[10^3\,(1-\xi^{-1})/vc] - \xi^{-1}\ln(kb) \qquad [a]$$

The equation is not an exact result. Its accuracy is expected to be enhanced at lower ionic strengths. It also presupposes the negligibility of specific counterion binding, that is, Na^+ ions bound to the polyion are hydrated to the same extent as free Na^+ ions and also possess complete translational freedom of motion near the polyion, a condition expected to hold for DNA as a good, but not exact, approximation. The first term represents the counterion diffusion potential (5); the local concentration (molarity) of bound Na^+, $10^3(1 - \xi^{-1})/v$, is greater than the average solution concentration c. The second term accounts for Debye-Huckel screened repulsions among the phosphate groups. The linear charge spacing b is the ratio of polymer length to number of charged groups on the polymer, 0.17 nm for B-form DNA. The dimensionless reduced charge density ξ equals 4.2 for DNA in aqueous solution. The Debye screening parameter κ is proportional to $c^{1/2}$ and equals 1.04 nm^{-1} when c is 10^{-1}M aqueous. Bound Na^+ ions are located in a volume v near the polyion; v equals 646 cm^3/mol P for DNA [see Table 1 in ref. (5), where V_p is used for v]. Finally, *kT is* Boltzmann's constant times Kelvin temperature (ξ, v, and κ are also slightly temperature dependent).

In Eq. [a] the quantities ξ, v, and b depend on the length L of the polymer segment. If N_p is held fixed,

$$-\delta G^{polyel}/\delta L = kT(\xi b)^{-1}\,[2(\xi - 1) - \ln(\kappa b)] \qquad [b]$$

a polyelectrolyte stretching force that becomes infinite as c tends to zero.

The polyelectrolyte contribution to Young's modulus, E" in the text, can be calculated from the second derivative with respect to L with N_p fixed. The result is,

$$AE^{polyel} = kT(\xi b)^{-1}\,(2\xi + 1) \qquad [c]$$

where A is the apparent cross-sectional area πR^2 of the effective circular cylinder of radius R assumed to underlie the elastic response. Equations [a]-[c] apply to a polyion immersed in 1:1 salt solution. To extend Eq. [b] to the model discussed in connection with Figure 5, in which a "protein" neutralizes a fraction x of the phosphate charges, replace b by $b/(1 - x)$ with $b = 0.17$ nm, and replace ξ by $\xi(1-x)$ with $\xi = 4.2$. Phosphate charge neutralization by multivalent small cations like Mg^{2+} and $Co(NH_3)_6^{3+}$, which bind like Na^+ and hence compete with it in subtle ways, is a more complicated situation. The following expression for the polyelectrolyte stretching force is applicable when (i), the bathing solution is of molarity c in NaCl; (ii), the concentration of free N-valent ion, $N \geq 2$, is negligible, so that κ has the same value as in Eq. [b]; (iii), all bound Na^+ is displaced by bound N-valent ion:

$$-\delta G^{polyel}/\delta L = kTb^{-1}\,\{\xi(1-x)^2\,[1-\ln(\kappa b)] + N^{-1}x[(2\xi - 3N^{-1})/(\xi - N^{-1}]\} \qquad [d]$$

In this equation, b and ξ are, as before, based on the phosphate groups, hence equal to 0.17 nm and 4.2, respectively, for DNA, while x is the fraction of the phosphate charge neutralized by bound N-valent ion. Condition (iii) requires that $x \geq 0.76$ (5).

We would like to put to rest a criticism that has been leveled at the application of the Euler instability theory to polymers. In Euler's theory the force F is ordinarily externally imposed. In contrast, it is said, weakening of interatomic repulsive forces within a polymer can only act to contract the polymer along its length and, in particular, cannot cause a stiff polymer to buckle. But consider a counterexample. Suppose a certain polymer is flexible, so that its axial trajectory follows a more or less random path. Modify this polymer by attaching along its length a set of mutually repulsive chemical groups. If the repulsions are sufficiently strong, the modified polymer is stiff and fully stretched. Now weaken the repulsive interactions within the modified polymer by some environmental change that "turns off" the repulsive forces among the added groups. The result is obviously a buckled polymer.

Appendix 2 DEVELOPMENT OF A STIFFNES MATRIX FOR FEM ANALYSIS

We demonstrate the development of the stiffness matrix by means of a simple example , segment of a truss element (Figure1) as follows.

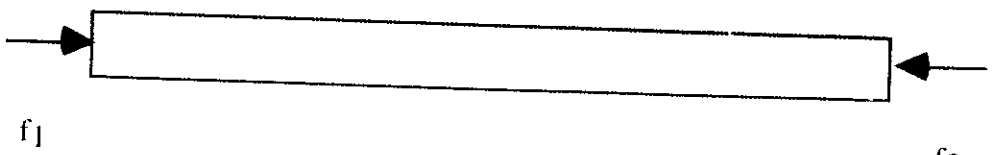


Figure 5

Hooke's law states that if the tensile load in the segment is f, then the elongation u is given by,

$$u = fl/AE \qquad [a]$$

where, l is the length of the segment; A is the area of cross section; E is the Young's modulus of elasticity.

This can be rewritten as:

$$f = k\,u, \text{ where } k = AE/l \qquad [b]$$

In FEM we use one of two basic notations, the **force notation** or the **displacement notation** (Figure 2 and 3)

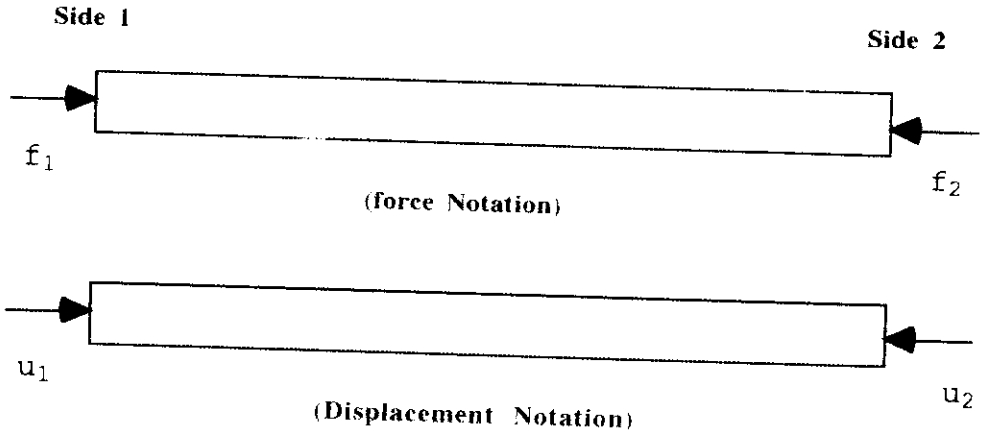


Figure 6

Considering the object to be in equilibrium one gets

$$f_1 + f_2 = 0; \quad f_1 = -f_2 = -f \qquad [c]$$

$$u = u_2 - u_1 \qquad [d]$$

then,

$$f_1 = ku_1 - ku_2 \text{ and } f_2 = -ku_1 + ku_2 \qquad [e]$$

Or, in a matrix form:

$$[f] = [k]\,[u] \qquad [f]$$

where

$$k = \begin{bmatrix} k & -k \\ -k & k \end{bmatrix} = k \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The square matrix [k] is called the stiffness matrix. Eqn. 6 can be rewritten as:

$$\begin{bmatrix} f1 \\ f2 \end{bmatrix} = k \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u1 \\ u2 \end{bmatrix}$$

We have two unknowns (u_1 and u_2) which can be determined from the two equations.

In a similar manner, we can solve the equations for all the segments in a large system.. Combining all the segments leads to a system of equations. Applying the suitable boundary conditions one can solve the entire problem.

Rationally designed helix–turn–helix proteins and their conformational changes upon DNA binding

Piergiorgio Percipalle, András Simoncsits, Sotir Zakhariev, Corrado Guarnaccia, Roberto Sánchez and Sándor Pongor[1]

International Centre for Genetic Engineering and Biotechnology (ICGEB), Area Science Park, Padriciano 99, 34012 Trieste, Italy

[1]Corresponding author

Circular dichroism and electrophoretic mobility shift studies were performed to confirm that dimerized N-terminal domains of bacterial repressors containing helix–turn–helix motifs are capable of high-affinity and specific DNA recognition as opposed to the monomeric N-terminal domains. Specific, high-affinity DNA binding proteins were designed and produced in which two copies of the N-terminal 1–62 domain of the bacteriophage 434 repressor are connected either in a dyad-symmetric fashion, with a synthetic linker attached to the C-termini, or as direct sequence repeats. Both molecules bound to their presumptive cognate nearly as tightly as does the natural (full-length and non-covalently dimerized) 434 repressor, showing that covalent dimerization can be used to greatly enhance the binding activity of individual protein segments. Circular dichroism spectroscopy showed a pronounced increase in the α-helix content when these new proteins interacted with their cognate DNA and a similar, although 30% lower, increase was also seen upon their interaction with non-cognate DNA. These results imply that a gradual conformational change may occur when helix–turn–helix motifs bind to DNA, and that a scanning mechanism is just as plausible for this motif class as that which is proposed for the more flexible basic-leucine zipper and basic-helix–loop–helix motifs.

Key words: bacterial repressors/circular dichroism/conformational changes/DNA-binding proteins/helix–turn–helix proteins

Introduction

DNA-binding proteins undergo various conformational changes when binding to their target sites in DNA (for a review, see Spolar and Record, 1994). It is reasonable to suppose that, before reaching a specific recognition site, the protein would first interact with other, non-specific, DNA segments. However, pronounced conformational changes due to non-specific interactions could not be shown convincingly by either spectroscopic (O'Neil *et al.*, 1990) or thermodynamic methods (cf. Spolar and Record, 1994). As experimental evidence for this phenomenon is extrapolated from peptide models devoid of specific DNA-binding activity (Johnson *et al.*, 1994), the existence of a conformational change upon non-specific DNA binding is considered an open question (von Hippel, 1994).

DNA-binding protein domains frequently contain helical segments that interact specifically with DNA. Eventual changes in the conformation of the DNA-binding helices can thus be conveniently monitored by circular dichroism difference spectroscopy, provided that the binding is detectable at sufficiently low DNA–protein ratios. This is in fact the case for segments of the basic-leucine zipper [bZIP] (Vinson *et al.*, 1989; O'Neil *et al.*, 1990, 1991; Talanian *et al.*, 1990; Weiss *et al.*, 1990; Saudek *et al.*, 1991; Anthony-Cahill *et al.*, 1992; Ellenberger *et al.*, 1992; König and Richmond, 1993) and for the basic-helix–loop–helix (bHLH) domains (Anthony-Cahill *et al.*, 1992; Ferré-D'Amaré *et al.*, 1993). The helix–turn–helix (HTH) motif (Harrison and Aggarwal, 1990) seems to be a special case in this respect since it is believed to dock without major conformational change, even to specific target sites (von Hippel and Berg, 1986; Berg and von Hippel, 1988; Spolar and Record, 1994). For example, the N-terminal domain of the phage 434 repressor has almost the same conformation in solution (Neri *et al.*, 1992) as in the DNA complex (Aggarwal *et al.*, 1988). In contrast, NMR spectroscopy of the Trp repressor showed that the recognition helix of the HTH motif undergoes a conformational stabilization upon binding to cognate DNA (Arrowsmith *et al.*, 1990; Youderian and Arvidson, 1994; Zhang *et al.*, 1994). A similar change—an elongation of the recognition helix—was found to take place upon the binding of the NK-2 homeodomain HTH motif to cognate DNA (Tsao *et al.*, 1994). However, conformational changes upon binding to non-cognate DNA have not been reported.

In this paper, we ask whether a helix–turn–helix based DNA-binding molecule which is capable of sequence-specific DNA recognition *in vitro*, changes its conformation when interacting with cognate or non-cognate DNA. As a criterion of specific DNA binding, we used a positive electrophoretic mobility shift assay in the presence of a large (>1000-fold) excess of competitor DNA. Specific DNA binding is an important criterion, in our view, since helical transitions in short, amphiphilic peptides can be relatively easily induced by a variety of agents, and not only by DNA (von Hippel, 1994). Our working hypothesis is that an induced-fit like conformational fluctuation may be a prerequisite for a protein to scan a large number of DNA sites, i.e. it is a necessary step in explaining the speed by which protein ligands reach their target sites. With the present work we sought to prove, on the one hand, that the helix–turn–helix motif is not an exception to this rule, and on the other that the protein conformation appreciably changes upon contact with non-specific DNA.

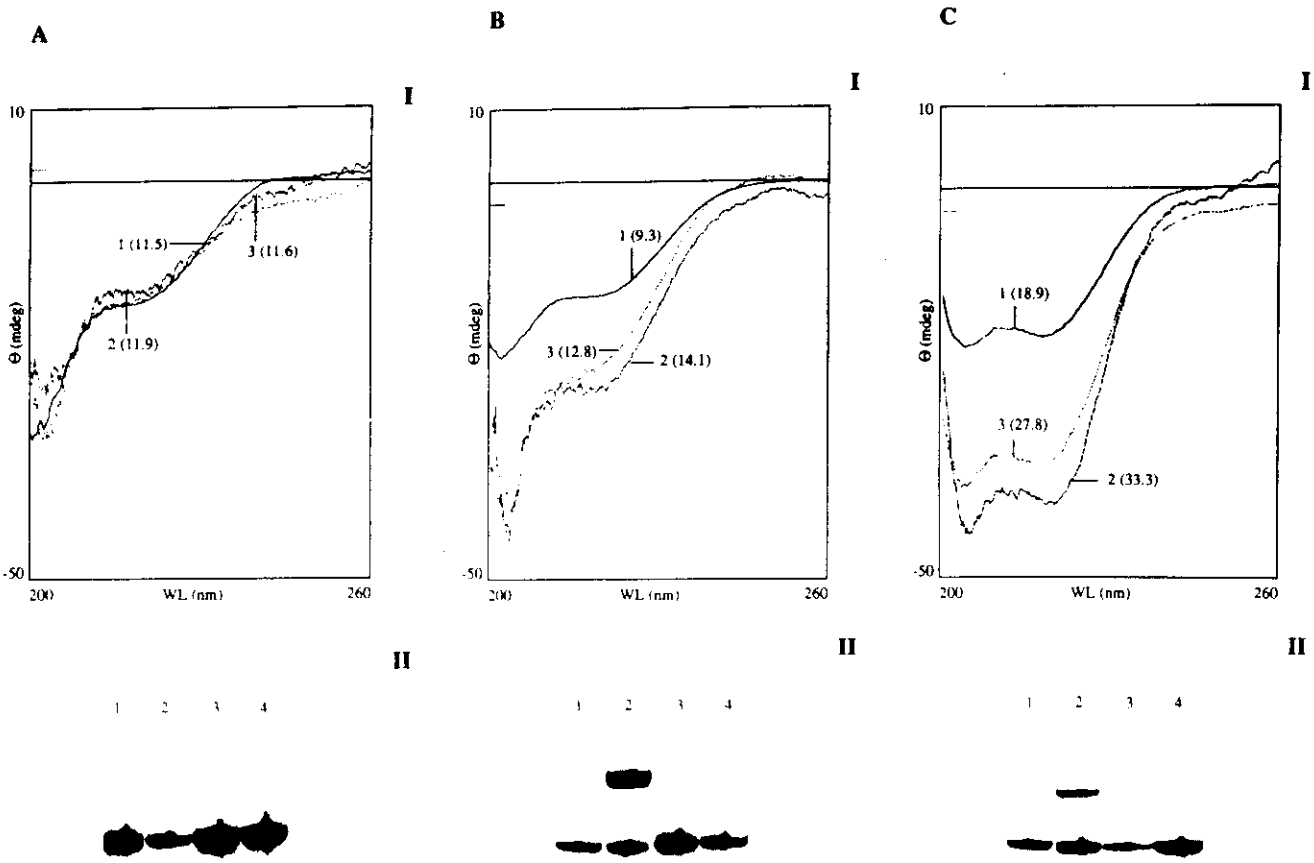


Fig. 1. Induction of the α-helical structure and the specific DNA-binding activity of the peptides studied (sequences shown in Figure 3). (**A**) Nter (the N-terminal 1–63 domain of the phage 434 repressor) (**B**) ChD, a dimeric single-chain repressor with palindromic symmetry (schematic structure Figure 2, sequence and synthesis: Figure 3) (**C**) ReD, a dimeric single-chain repressor molecule with direct repeat symmetry (schematic structure Figure 2, sequence and synthesis Figure 3). I: induction of α-helical structure by cognate and non-cognate DNA as determined by circular dichroism difference spectroscopy (O'Neil *et al.*, 1990; Talanian *et al.*, 1990). Top curves (1) are the spectra of the peptides in the absence of DNA. The bottom curves (2) are the spectra induced by cognate DNA (OR). The centre curves (3) are the spectra induced by non-specific DNA (NS, see sequences in Figure 3). The numbers in parentheses indicate the nominal α-helical content of the respective molecule, calculated according to Wu *et al.* (1984). Spectra were recorded at 24 μM DNA duplex and 24 μM protein dimer (48 μM Nter) concentrations in 10 mM Na_2HPO_4, 100 mM NaCl, pH 7.5, at 25°C. II: specific DNA binding as detected by electrophoretic mobility shift assay. Lanes 1, cognate DNA (OR) alone; lanes 2, cognate DNA (OR) + protein (Nter for section A, ChD for section B and ReD for section C); lanes 3, non-cognate DNA (NS) alone; lanes 4, non-cognate DNA (NS) + proteins as listed at lanes 2.

Results

As an experimental model we chose the N-terminal (Nter) domain of the phage 434 repressor (Aggarwal *et al.*, 1988; Neri *et al.*, 1992). This domain can be co-crystallized with cognate DNA, but its DNA-binding affinity is low (Aggarwal *et al.*, 1988). We found that Nter does not bind to the cognate DNA in the presence of competitor DNA (Figure 1AI) and it shows no appreciable increase in its helical content upon addition of *ds* oligonucleotides that contain either cognate or non-cognate sequences (Figure 1AII). This could mean, in principle, that the Nter domain is essentially rigid, as is sometimes supposed (Johnson *et al.*, 1994; Spolar and Record, 1994). On the other hand, the native 434 repressor, which has a high and specific DNA-binding activity, binds to cognate DNA as a dimer, non-covalently connected through a large C-terminal dimerizaton domain (Anderson *et al.*, 1984, Figure 2A). Since structural transitions may not be monitored accurately in the presence of a large dimerization domain, we decided to design molecular probes in which the specificity of the dimeric repressor is mimicked by a covalent linkage. As the resulting molecules consist of a single chain, we call these covalent dimers single-chain repressors. The first covalent dimer, ChD (Figure 2B, sequence shown in Figure 3) has a palindromic (dyadic) symmetry since the two Nter domains are connected through their C-termini with a flexible and symmetric synthetic linker of ~36 Å maximum length. The second single-chain repressor, ReD (Figure 2C, sequence in Figure 3), contains two direct sequence repeats of the Nter sequence connected with a longer linker (~96 Å). This molecule was produced by recombinant DNA methods (A.Simoncsits, F.Chen, S.Wang, P.Percipalle and S.Pongor, manuscript in preparation). Both molecules are capable of specifically binding cognate DNA in the presence of competitor DNA and do not show gel-mobility shift with non-cognate DNA under the same conditions (Figure 1BII and 1CII, respectively). The dissociation constants (K_ds) were estimated to be $5–6\times10^{-9}$ for ReD and $2–3\times10^{-8}$ for ChD, based on the protein concentrations at which 50% binding was detected in electrophoretic mobility shift assays. Under the same conditions, the native repressor had a K_d of $2–4\times10^{-9}$, while the binding of Nter was not measurable (data not shown).

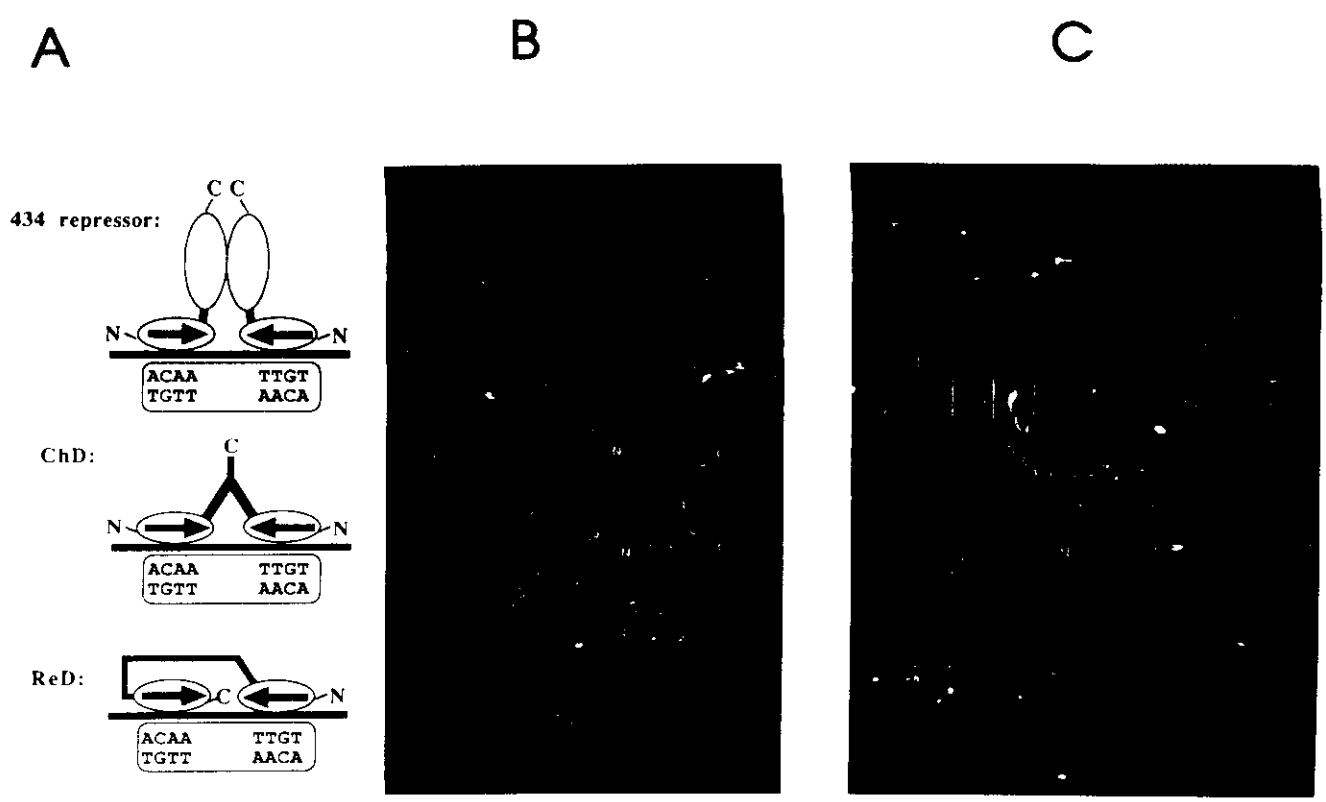


Fig. 2. Schematic structure of the 434 repressor and of the single-chain analogues, ChD and ReD. (**A**) Scheme of binding of the 434 repressor, ChD and ReD to cognate DNA. The arrows denote the N-terminal domain (N to C direction). (**B**) Predicted 3-D structure of the dimeric single-chain repressor ChD. (**C**) Predicted 3-D structure of the dimeric single-chain repressor ReD. The linker in ReD is deemed to be flexible, its position is only symbolically shown in (C).

```
                1                   20                  40                  60
                :                   |                   |                   |
Nter:           SISSRVKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKTKRPRFLPELASALGVSVDWLLNGT
                |-------H1-----|--H2--|  T|----H3----|---L---|-----H4----|--H5-->

                          cI repressor of 434 phage (1-63)

ChD  [Nter(1-62)-Ahx2-]Lys[Nter(1-62)-Ahx2]-Gly-OH

ReD  Nter-SDSNVRFVGHVEPKGKYPLISMVRAG-Nter-SDSNVR
               |------cI 434 64-89-----|

OR      5'-TACAAGAAAGTTTGTTA
              ATGTTCTTTCAAACAAT-5'

NS      5'-AGAGGAGCGCCTGTCATGCA
                   CTCGCGGACAGTACGTCGCG-5'
```

Fig. 3. Amino acid and nucleotide sequences. Nter, sequence of the 434 phage repressor protein N-terminal domain Nter (Kuziel and Tucker, 1987; Aggarwal *et al.*, 1988); H, helix; T, turn; L, loop regions; ChD, sequence of ChD, the dimeric single-chain repressor produced by chemical synthesis; Ahx, ε-aminocaproic acid; ReD, sequence of ReD, the dimeric single-chain repressor produced by recombinant DNA in *E.coli*. The sequence corresponds to (1–89)(1–69) of the cI repressor of 434 phage. OR, cognate DNA, the O_R1 operator site of the 434 phage; NS, oligonucleotides used as non-specific DNA.

In contrast to Nter, both ChD and ReD show a marked increase in their helical content (signal at 222 nm) upon the addition of cognate DNA (curves 2 in Figure 1BI and 1CI, respectively). A qualitatively very similar conformational change is observed in the presence of non-cognate DNA (curves 3 in Figure 1BI and 1CI); the change induced by non-cognate ds DNA is ~70% (69% for ChD and 73% for ReD) compared with that induced by cognate DNA. Single-stranded DNA molecules (either cognate or non-cognate sequences) failed to induce helicity in either of the peptides (not shown). Cognate and non-cognate DNA seems to increase ordered structure both in ChD and in ReD, even though the original conformations of these two molecules are slightly different; ChD apparently contains more 'random-coil' (200 nm signal) than does ReD. This difference (as well as the higher K_d value of ChD) may be attributed to the fact that ChD has a much shorter linker than ReD: moreover, the relatively harsh synthesis conditions may result in partial denaturation. If we suppose that: (i) all molecules in solution have a

conformation corresponding to the X-ray structure; and (ii) none of the residues in the linkers has helical conformation, then the expected maximal α-helix content of ChD would be 57% while that of ReD would be 51%. The nominal helicity calculated from the CD data is lower than these values (9.3–14.1 for ChD and 18.9–33.3 for ReD). The differences may originate either from the approximate nature of the CD-based helicity calculation (Wu *et al.*, 1981) or, more probably, from the fact that in the solution environment used for CD studies, not all molecules have ordered helical conformations. In fact, adding 10% trifluoroethanol to the medium approximately doubles the α-helix signal both in ChD and in ReD indicating that the peptides can adopt more helical conformations under favourable conditions (Percipalle, 1995).

Discussion

Our results show that: (i) high-affinity DNA-binding molecules can be obtained by combining two HTH-containing DNA-binding domains; and (ii) the single-chain repressors described here change their conformation upon interaction with non-specific as well as with specific DNA sites.

The combination of two HTH-containing domains to give artificial DNA-binding proteins is not new in itself. Several DNA-binding domains were fused with dimerization domains to give molecules that yield DNA-binding dimers. Examples include the lambda repressor head–GCN4 leucine zipper fusions (Hu *et al.*, 1990, 1993), the Lex A repressor head–jun leucine zipper fusions (Schmidt-Dörr *et al.*, 1991), the LexA core–gal4 fusions (Brent and Ptashne, 1985), the 434 repressor N-terminal domain/leucine zipper fusions (Pu and Struhl, 1993) and the lambda repressor N-terminal domain/*rop* fusion (Castagnoli *et al.*, 1994). While all of these fusion proteins show specific DNA binding, they are not especially useful as probes to monitor conformational changes in the recognition helices, since they contain either large, or highly helical dimerization domains that would make CD difference spectroscopy impractical. A further distinctive feature of our constructs, in comparison with earlier fusion proteins, is the fact that we based our design on covalent dimerization ('single-chain architecture') which in itself proved sufficient to reach a DNA binding affinity that is comparable with that of the corresponding full-length repressor molecule.

Fundamentally, the results confirm the principle on which the design of the molecule was based: the tightness and the specificity of the binding critically depend on a dimeric architecture. In fact, both very different covalent dimerization strategies adopted here successfully mimic the non-covalent dimerization of the full-length 434 repressor since they have K_ds that are comparable with that of the natural repressor. The dramatic enhancement of binding strength as compared with that of a monomeric Nter domain may be explained partly by an increase of the binding surface and partly by a decrease in the dissociation rate of the bound protein. A possible, structural explanation of the strong DNA-binding affinity could be that the Nter domains have a structured dimer interface sufficient to bring about dimerization once the two Nter domains are brought together by a suitable linker.

The data indicate that the Nter domain is not rigid, but seems rather to adopt its binding conformation when it comes into contact with DNA. More interestingly, a conformational change is already detectable when the protein comes into contact with non-specific DNA. The difference between 'rigid' (HTH) and 'flexible' (bZIP, bHLH) motifs may, thus, be less dramatic as previously thought (von Hippel and Berg, 1986; Berg and von Hippel, 1987, 1988; Harrison and Aggarwal, 1990; Johnson *et al.*, 1994; Spolar and Record, 1994; von Hippel, 1994) and an induced-fit like model of recognition (Koshland, 1958) also seems to be valid for the HTH-domain. In fact, one may speculate that the flexibility of the Nter domain may be essential for the repressor to scan DNA in search of a specific target site, in the same manner as is proposed for more flexible DNA-binding domains (O Neil *et al.*, 1990; Johnson *et al.*, 1994). The conformational change induced by non-cognate DNA is qualitatively similar and only somewhat weaker (70%) compared with that caused by cognate DNA. It is tempting to argue that the change is in fact gradual: the protein fluctuates in a larger conformational space when in contact with non-cognate DNA and a final, smaller, set of helical conformations is stabilized only at the target site. It must be emphasized that the present results show the conformational changes taking place within the N-terminal domain but do not prove unequivocally that these take place solely within the DNA-recognition helix. The fact that the Nter domain in solution (Neri *et al.*, 1992) has a conformation almost identical to the Nter domain–O_R1 complex (Aggarwal *et al.*, 1988) does not contradict our findings. CD spectroscopy gives a signal proportional to the number of molecules in the ordered (α-helical) conformation. This signal will be higher if the DNA drives the conformational equilibrium of the protein toward the helical state (which seems to be the case), irrespective of the fact that the helical conformations of the free and complex-bound states are identical or different.

Finally, we mention that a series of dimeric peptides incorporating shorter segments of the Nter domain had been designed and synthesized (Percipalle *et al.*, 1994; Percipalle, 1995), but they failed to show specific DNA binding under the experimental conditions used in this study.

Materials and methods

Computer modelling

The structure of both ChD and ReD were modelled using the Insight/Discover package (Biosym), by manually inserting the linker sequences into the DNA–N-terminal domain complex built using the published 3D co-ordinates of the 434 repressor N-terminal domain–DNA complex (Aggarwal *et al.*, 1988), and then subjecting the linker structures first to a 100 ps Molecular Dynamics at 300 K with a distance-dependent dielectric constant and then to 200 cycles steepest descent energy minimization. The known part of the structure was kept fixed during the entire simulation.

Peptide and oligonucleotide synthesis

Nter and ChD (sequences shown in Figure 3) were prepared via automated continuous-flow solid-phase peptide synthesis, using the Fmoc chemistry, as described (Percipalle *et al.*, 1995). ChD is a branched synthetic peptide in which the C-terminal residue of the Nter sequence is joined to both amino groups of a central lysine residue linked to the resin via a glycine. The peptides were purified by reverse-phase HPLC to electrophoretic homogeneity; the yields were 25% for Nter and 1% for ChD.

The oligonucleotides OR and NS (sequences shown in Figure 3) were prepared by the ICGEB oligonucleotide synthesis service.

Recombinant DNA methods

ReD was produced by recombinant DNA techniques and was expressed in *Escherichia coli* cells. The plasmid was constructed from λgt10 template (Huynham *et al.*, 1984; Kuziel and Tucker, 1987), by PCR amplification of DNA fragments corresponding to amino acids 1–89 (as a *RcaI–Bam*HI fragment) and to amino acids 1–69 (a *Bam*HI–*Hin*dIII fragment). Restriction sites to obtain these fragments were included into the PCR primers. The fragments were cloned consecutively between *Nco*I and *Bam*HI sites and between *Bam*HI and *Hin*dIII sites of *E.coli* expression vectors pRIZ′ (Simoncsits *et al.*, 1994), and a pRSET5d (Schoepfer, 1993) derivative lacking the second *Nco*I site in the cloning region of pRSET5d (A.Simoncsits, unpublished results), resulting in pRIZ′RR69 and pRSETkRR9, respectively (where RR69 represents the region coding for ReD). High-level expression of ReD was achieved using the pRSET5dRR69 in BL21(DE3)[pLysS] *E.coli* (Novagen). Purification to >95% homogeneity was achieved using cell lysis by the freeze–thaw method (Studier *et al.*, 1990) followed by batch adsorption to SP–Sepharose (Pharmacia) and HPLC on a Shodex SP 825 cation exchange column.

Circular dichroism spectroscopy

Spectra were recorded at 24 μM DNA duplex and 24 μM protein dimer (48 μM Nter) concentrations in 10 mM Na_2HPO_4, 100 mM NaCl, pH 7.5, at 25°C. Data were collected on a Jasco J-600 spectrometer (Jasco, Inc., Easton, MD, USA) and were smoothed using software provided by Jasco. The nominal α-helix content was calculated according to Wu *et al.* (1981). Difference spectra were calculated as described by O'Neil *et al.* (1990) and by Talanian *et al.* (1990). The spectrum of the peptide–DNA complexes did not show any change in the 250–400 nm range as compared with free DNA that would indicate a major conformational change in DNA (O'Neil *et al.*, 1990).

Electrophoretic mobility shift assays

Binding experiments were performed in 20 μl buffer [50 mM NaCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 5% glycerol, 20 mM HEPES, pH 7.9, and 1 μg poly(dI–dC)] containing 20 nM [^{32}P]phosphate DNA-probe and protein (19 μM Nter or 1.6 μM ChD or 1.2 μM ReD), at 4°C for 40 min. Electrophoresis was performed on 8% polyacrylamide gel using Tris-borate buffer at 4°C.

Acknowledgements

The authors thank Professors A.Falaschi and E.E.Baralle (ICGEB) for help and advice throughout this project. The help of Professor Sergio Paoletti and Dr Antonella Flaibani (Polybios Research Centre, Trieste) with circular dichroism measurements, and Professor Salvatore Foti (Universita di Catania) with mass spectrometry, are gratefully acknowledged. Thanks are due to our former colleagues, Drs Alessandro Tossi (presently at the Universita degli Studi di Trieste) and Miklós Cserzö (presently at Université de Nancy) for their valuable advice during the earlier stages of this project. Ms Suzanne Kerbavcic provided excellent assistance with the manuscript. P.P. was supported by a doctoral fellowship of the SISSA School of Advanced Studies, Trieste, Italy.

References

Aggarwal,A.K., Rodger,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) Recognition of DNA operator by the repressor of phage 434: a view at high resolution. *Science*, **242**, 899–907.

Anderson,J.E., Ptashne,M. and Harrison,S.C. (1984) Co-crystals of the DNA-binding domain of the phage 434 repressor and a synthetic 434 operator. *Proc. Natl Acad. Sci. USA*, **81**, 1307–1311.

Anthony-Cahill,S.J., Benfield,P.A., Fairman,R., Wasserman,Z.R., Brenner, S.L., Stafford,W.F.III, Altenbach,C., Hubbell,W.L. and DeGrado,W.F. (1992) Molecular characterization of helix–loop–helix peptides. *Science*, **255**, 979–983.

Arrowsmith,C.H., Pachter,R., Altman,R.B., Iyer,S.B. and Jardetzky,O. (1990) Sequence-specific 1H NMR assignment and secondary structure in solution of *Escherichia coli trp* repressor. *Biochemistry*, **29**, 6332–6341.

Berg,O.G. and von Hippel,P.H. (1987) Selection of DNA binding sites by regulatory proteins. Statistical mechanical theory and application to operators and promoters. *J. Mol. Biol.*, **193**, 723–750.

Berg,O.G. and von Hippel,P.H. (1988) Selection of DNA binding sites by regulatory proteins. II. The binding specificity of cyclic AMP receptor protein to recognition sites. *J. Mol. Biol.*, **200**, 709–723.

Brent,R. and Ptashne,M. (1985) A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell*, **43**, 729–736.

Castagnoli,L., Vitriani,C. and Cesareni,G. (1994) Linking an easily detectable phenotype to the folding of a common structural motif. *J. Mol. Biol.*, **237**, 378–387.

Ellenberger,T.E., Brandl,C.J., Struhl,K. and Harrison,S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted helices: crystal structure of the protein–DNA complex. *Cell*, **71**, 1223–1237.

Ferré-D'Amaré,A.R., Prendergast,G.C., Ziff,E.B. and Burley,S.K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **363**, 38–45.

Harrison,S.C. and Aggarwal,A.K. (1990) DNA recognition by proteins with the helix-turn–helix motif. *Annu. Rev. Biochem.*, **59**, 933–969.

Hu,J.C., O'Shea,E.K., Kim,P.S. and Sauer,R.T. (1990) Sequence requirements for coiled coils: analysis with λ-repressor-GCN4 leucine zipper fusions. *Science*, **250**, 1400–1403.

Hu,J.C., Newell,N.E., Tidor,B. and Sauer,R.J. (1993) Probing the role of residues at the e and g positions of the GCN4 leucine zipper by combinatorial mutagenesis. *Protein Sci.*, **2**, 1072–1084.

Huynham,T.V., Young,R.A. and Davis,R.W. (1984) In Glover,D. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, Vol. I, pp. 49–78.

Johnson,N.P., Lindstrom,J., Baase,W.A., von Hippel,P.H. (1994) Double-stranded DNA templates can induce α-helical conformation in peptides containing lysine and alanine: functional implications for leucine zipper and helix loop helix transcription factors. *Proc. Natl Acad. Sci. USA*, **91**, 4840–4844.

König,P. and Richmond,T.J. (1993) The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.*, **233**, 139–154.

Koshland,D.E.Jr (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA*, **44**, 98–102.

Kuziel,W.A. and Tucker,P.W. (1987) Determination of vector: insert junctions in λgt10 cDNAs that do not recut with *Eco*RI. Nucleotide sequence of the λ*imm*434 *Hin*dIII–*Eco*RI DNA fragments encoding part of the cI protein. *Nucleic Acids Res.*, **15**, 3181.

Neri,D., Billeter,M. and Wüthrich,K. (1992). Determination of the nuclear magnetic resonance solution structure of the DNA-binding domain (residues 1 to 69) of the 434 repressor and comparison with the X-ray crystal structure. *J. Mol. Biol.*, **223**, 743–767.

O'Neil,K.T., Hoess,R.H. and DeGrado,W.F. (1990) Design of DNA binding peptides based on the leucine zipper motif. *Science*, **249**, 774–778.

O'Neil,K.T., Shuman,J.D., Ampe,C. and DeGrado,W.F. (1991) DNA-induced increase in the α-helical content of C/EBP and GCN4. *Biochemistry*, **30**, 9030–9034.

Percipalle,P. (1995) *Engineering DNA-Binding Proteins Based on the Helix-Turn-Helix Motif*. Ph.D. thesis, SISSA School of Advanced Studies, Trieste, Italy.

Percipalle,P., Saletti,S., Pongor,S., Foti,S., Tossi,A. and Fisichella,S. (1994b) Structural characterization of synthetic model peptides of DNA-binding cI434 repressor by electrospray ionization and fast atom bombardment mass spectrometry. *Biol. Mass Spectrom.*, **23**, 727–733.

Percipalle,P., Zakhariev,S., Tossi,A., Guarnaccia,C., Cserzö,M., Simoncsits,A., and Pongor,S. (1995) Synthetic dimeric peptides based on the phage cI434 repressor. In Maia,H.L.S. (ed.), *Peptides '94*. ESCOM, Netherlands, pp. 391–392.

Pu,W.T. and Struhl,K. (1993) Dimerization of leucine zippers analyzed by random selection. *Nucleic Acids Res.*, **21**, 4348–4355.

Saudek,V., Pasley,H.S., Gibson,T., Gausepohl,H., Frank,R. and Pastore,A. (1991) Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry*, **30**, 1310–1317.

Schmidt-Dörr,T., Oertel-Buchheit,P., Pernelle,C., Bracco,L., Schnarr,M. and Granger-Schnarr,M. (1991) Construction, purification, and characterization of a hybrid protein comprising the DNA-binding domain of the LexA repressor and the Jun leucine zipper: a circular dichroism and mutagenesis study. *Biochemistry*, **30**, 9657–9664.

Schoepfer,R. (1993) The pRSET family of T7 promoter expression vectors for *Escherichia coli*. *Gene*, **124**, 83–85.

Simoncsits,A., Bristulf,J., Tjörnhammar,M.L., Cserzö,M., Pongor,S., Rybakina,E., Gatti,S. and Bartfai,T. (1994) Deletion mutants of human interleukin 1 with significantly reduced agonist properties: search for

the agonist/antagonist switch in ligands to the interleukin 1 receptors. *Cytokine*, **6**, 206–214.
Spolar,R.S. and Record,M.T.,Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
Studier,W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
Talanian,R.V., McKnight,C.J. and Kim,P.S. (1990) Sequence-specific DNA binding by a short peptide dimer. *Science*, **249**, 769–771.
Tsao,D.H., Gruschus,J.M., Wang,L.H., Nirenberg,M. and Ferretti,J.A. (1994) Elongation of helix III of the NK-2 homeodomain upon binding to DNA: a secondary structure study by NMR. *Biochemistry*, **33**, 15053–15060.
Vinson,C.R., Sigler,P.B. and McKnight,S.L. (1989) Scissors-grip model for DNA recognition by a family of leucine zipper protein. *Science*, **246**, 911–916.
von Hippel,P.H. (1994) Protein DNA recognition: new perspectives and underlying themes. *Science*, **263**, 769–770.
von Hippel,P.H. and Berg,O.G. (1986) On the specificity of DNA–protein interactions. *Proc. Natl Acad. Sci. USA*, **83**, 1608–1612.
Weiss,M.A., Ellenberger,T., Wobbe,C.R., Lee,J.P., Harrison,S.C. and Struhl,K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, **347**, 575–578.
Wu,C.-S.C., Ikeda,K. and Yang,J.T. (1981) Ordered conformation of polypeptides and proteins in acidic dodecylsulfate solution. *Biochemistry*, **20**, 566–570.
Youderian,P. and Arvidson,D.N. (1994) Direct recognition of the *trp* operator by the *trp* holorepressor – a review. *Gene*, **150**, 1–8.
Zhang,H., Zhao,D., Revington,M., Lee,W., Jia,X., Arrowsmith,C. and Jardetzky,O. (1994) The solution structures of the trp repressor–operator DNA complex. *J. Mol. Biol.*, **238**, 592–614.

Received on February 16, 1995; revised on April 19, 1995