The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2038-25

Conference: From DNA-Inspired Physics to Physics-Inspired Biology

1 - 5 June 2009

DNA Micromechanics and Macromolecular Organization

Wilma K. OLSON

Rutgers, the State University of New Jersey
Chemistry & Chemical Biology
Piscataway NJ 08854
USA

DNA micromechanics and macromolecular organization

Wilma K. Olson
Rutgers, the State University of New Jersey
Piscataway, NJ 08854 USA
wilma.olson@rutgers.edu

June 1, 2009

Sreekala Balasubramanian, Andrew Colasanti, Luke Czapla,
Yun Li, Fei Xu, Guohui Zheng

A.R. Srinivasan
Marcia Fenley (Florida State)
Michael Tolstorukov, Victor Zhurkin (NIH)
David Swigon (U Pittsburgh)
Vasily Studitsky (Robert Wood Johnson Medical School)

USPHS GM20861 and GM34809

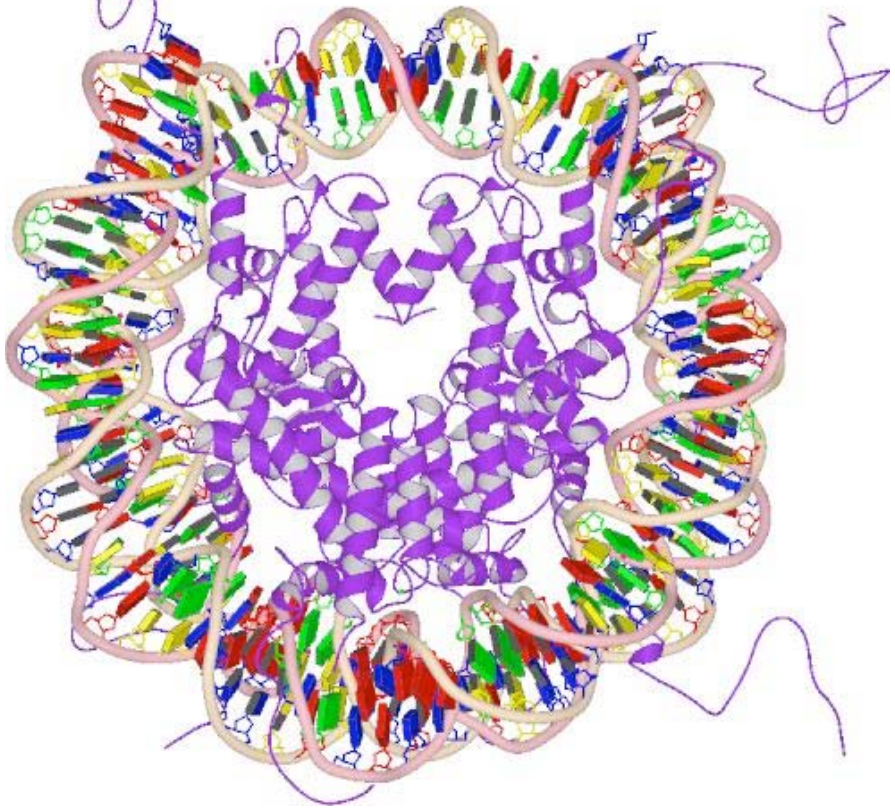
Outline

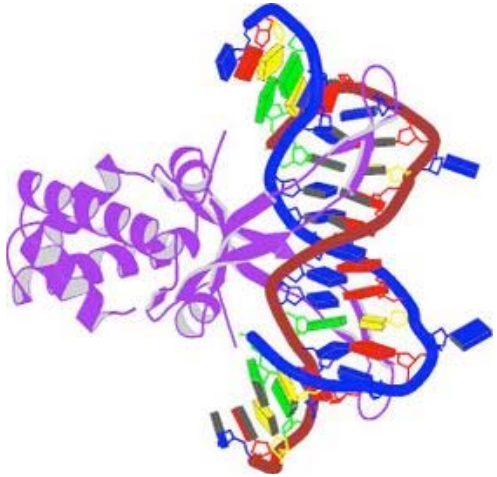
- DNA structure and base pairs
- Knowledge-based potentials for low-resolution modeling
- Sequence-dependent deformations and contacts
- Factors that enhance DNA cyclization/looping
- Structure-based predictions of nucleosome 'positioning'
- Higher-levels of DNA organization (progress report)

DNA structure and base pairs

The organization of DNA in the cell is thought to determine the parts of the genetic message that are expressed as RNA and protein.

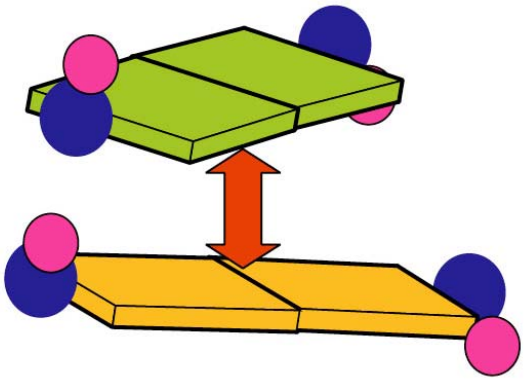
A·T
G·C
protein



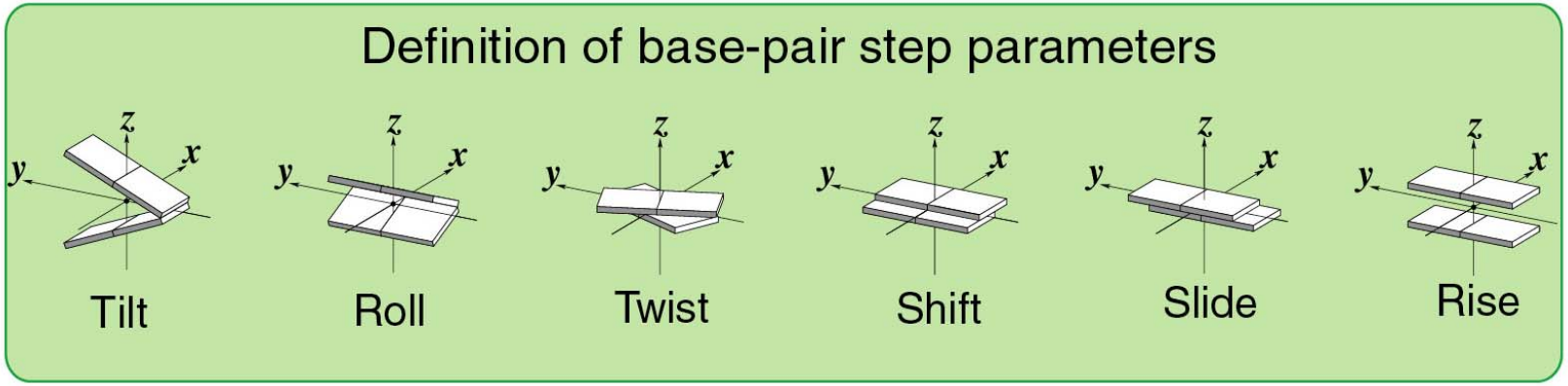


Simplified, color-coded representations of the nucleosome core particle, the fundamental DNA packaging unit in eukaryotes (NDB_ID: pd0287; Davey *et al.*, 2002), and the bacterial histone-like heat-unstable HU protein (NDB_ID pd0426; Swinger *et al* 2003).

Base-pair step representation of DNA

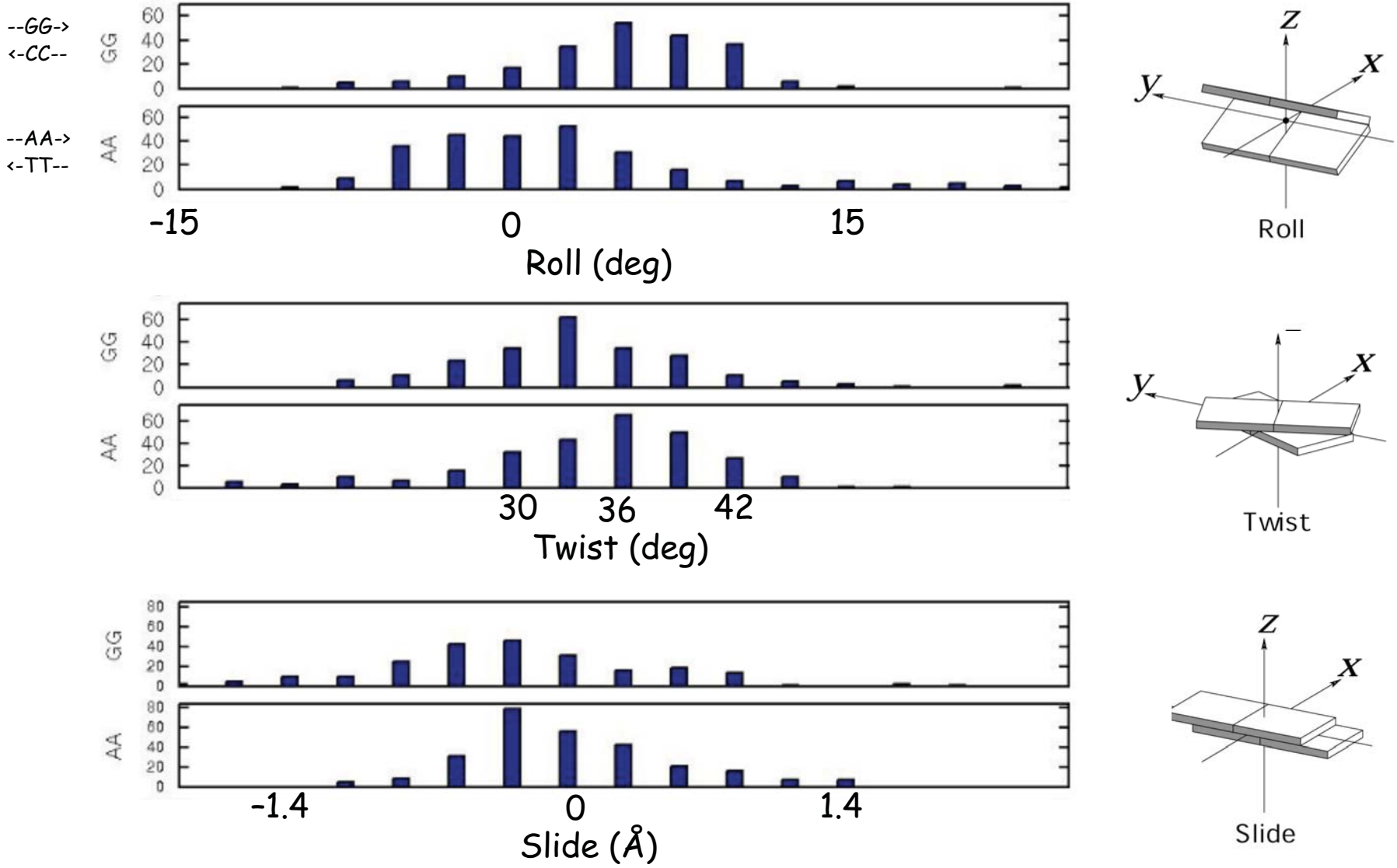


Treatment of the global organization of DNA requires models that capture key sequence-dependent features of the double helix, such as the positioning of adjacent base pairs, yet allow for easy generation of long polymer chains.



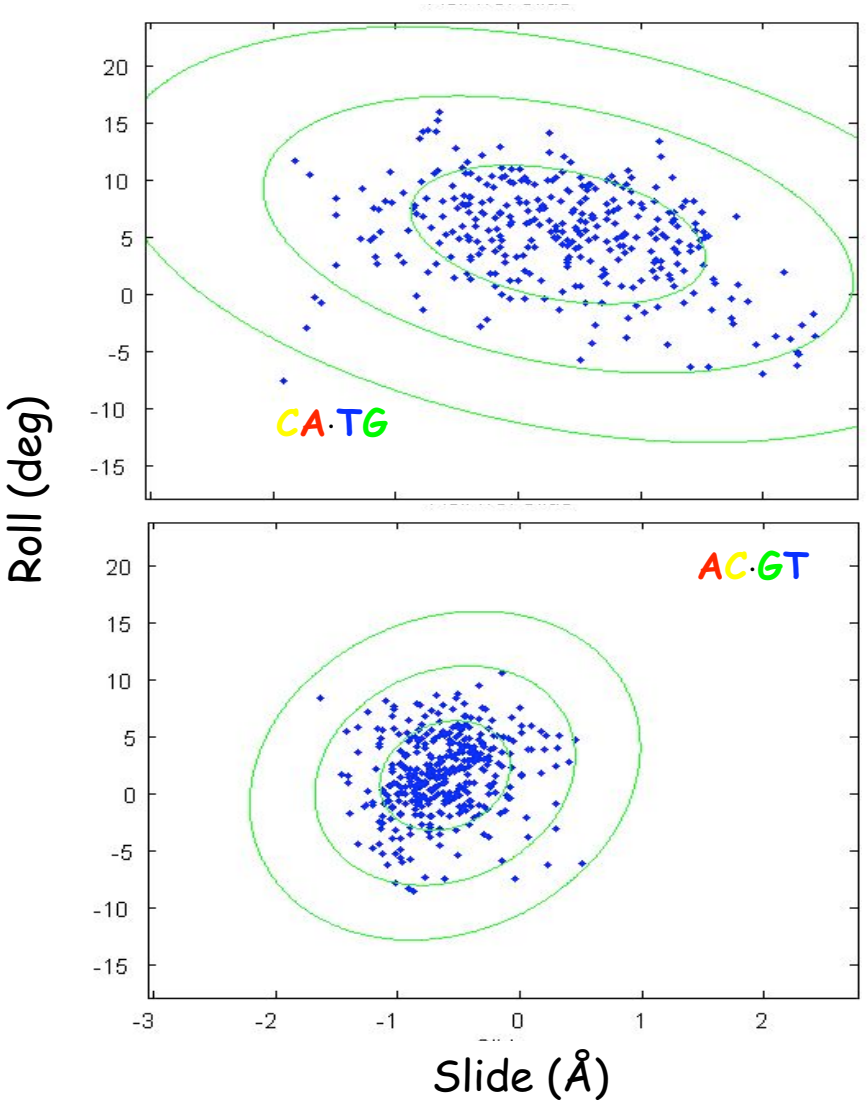
Base-pair positioning is described in terms of six rigid-body parameters.

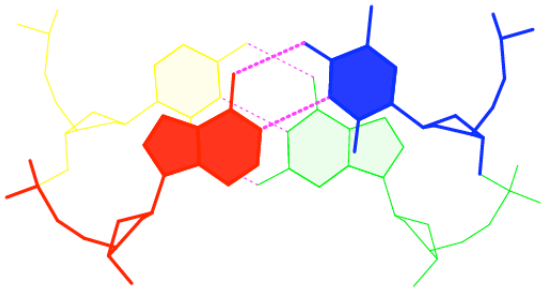
The base-pair 'step' parameters capture the local sequence-dependent structure and deformability of DNA.



Sequence-dependent variation of the three base-pair 'step' parameters, which dominate the conformational variability in high-resolution DNA structures.

The pyrimidine-purine steps stand out as deformable and the purine-pyrimidine steps as stiff. ‡

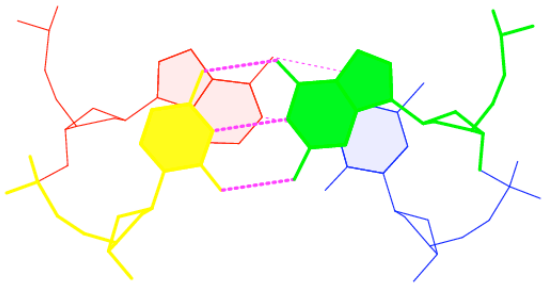




Roll Slide

$\langle \theta_2^{CA} \rangle = 5.3°$; $\langle \theta_5^{CA} \rangle = 0.33$ Å

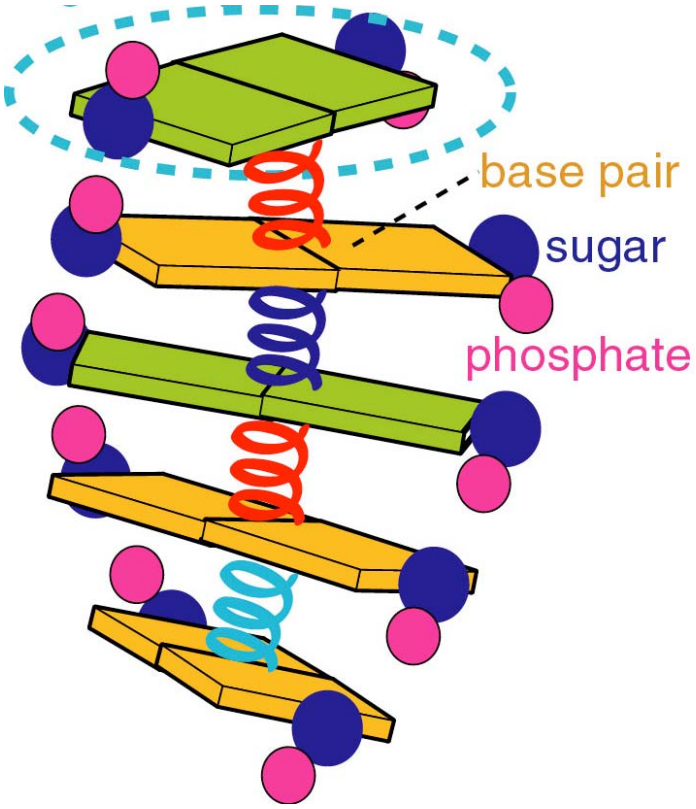
$\langle \theta_2^{AC} \rangle = 1.6°$; $\langle \theta_5^{AC} \rangle = -0.61$ Å



‡341 *CA·TG* and 418 *AC·GT* steps from 239 protein-DNA crystal complexes of 2.5 Å or better resolution

Knowledge-based potentials for low-resolution modeling

DNA sequence-dependent features can be incorporated in a coarse-grained, dimeric (base-pair step) model.



The total 'energy' Ψ of a sequence of N base pairs is the sum of the deformation scores of the N–1 base-pair steps:

$$\Psi = \sum_{n=1}^{N-1} V_n$$

$$V_n = (1/2) \sum_{i=1}^{6} \sum_{j=1}^{6} f_{ij} \left(\theta_i - \theta_i^0\right)\left(\theta_j - \theta_j^0\right)$$

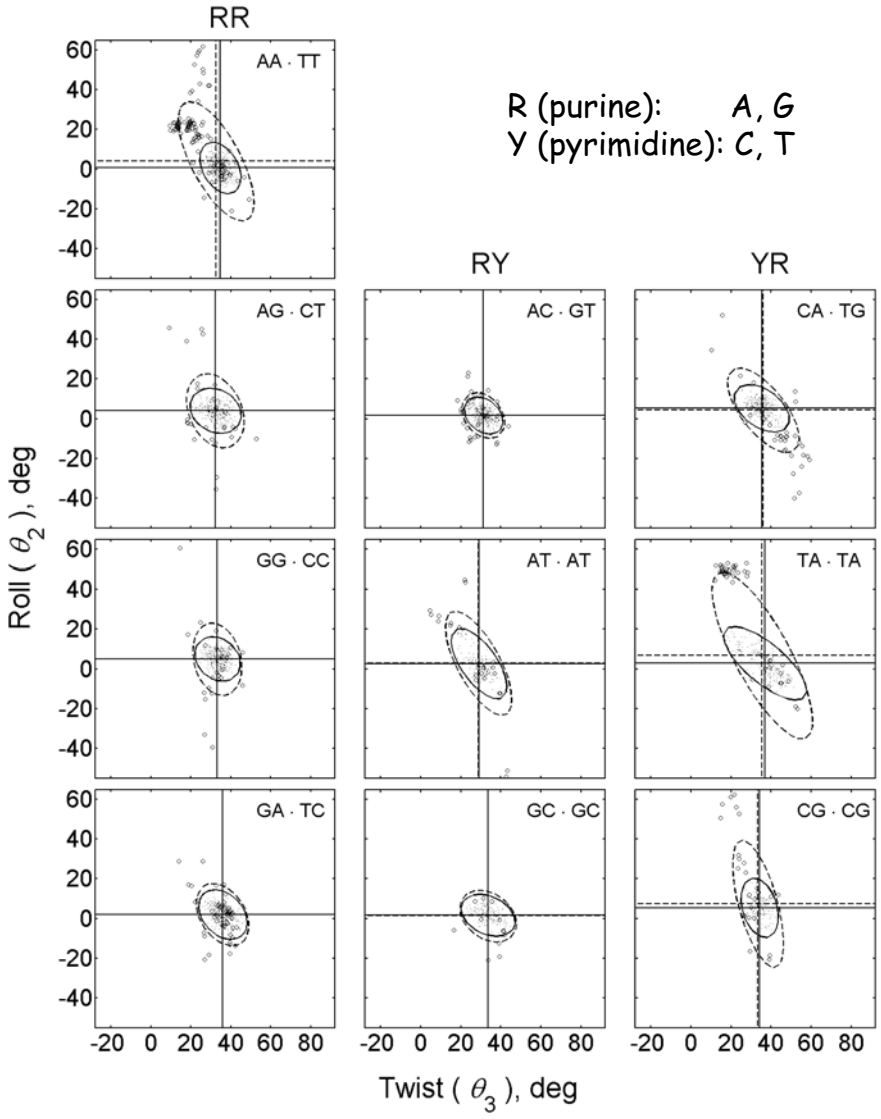
Knowledge-based dimeric potentials V_n(XZ)

The dimeric 'force constants' f_{ij} of individual XZ steps are derived from the covariance of observed parameters in protein-bound DNA structures.

Dimeric rest states θ_i^0 are equated to the mean values of observed parameters.

$$V_n = (1/2)\sum_{i=1}^{6}\sum_{j=1}^{6} f_{ij}\left(\theta_i - \theta_i^0\right)\left(\theta_j - \theta_j^0\right)$$

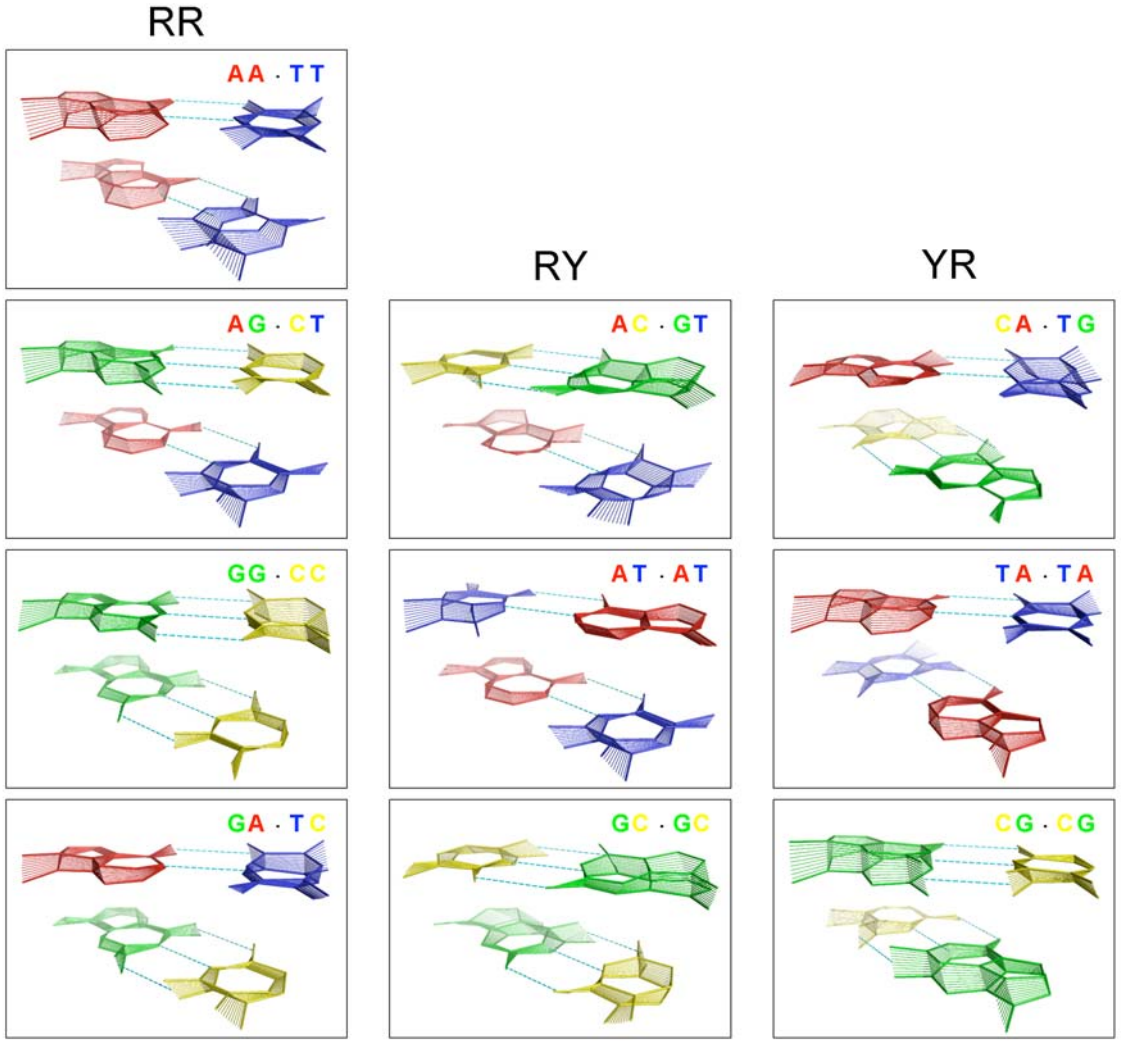
The knowledge-based potentials reveal structural and deformational codes in the DNA base-pair steps.



R (purine): A, G
Y (pyrimidine): C, T

Roll and Twist are anticorrelated at most steps.

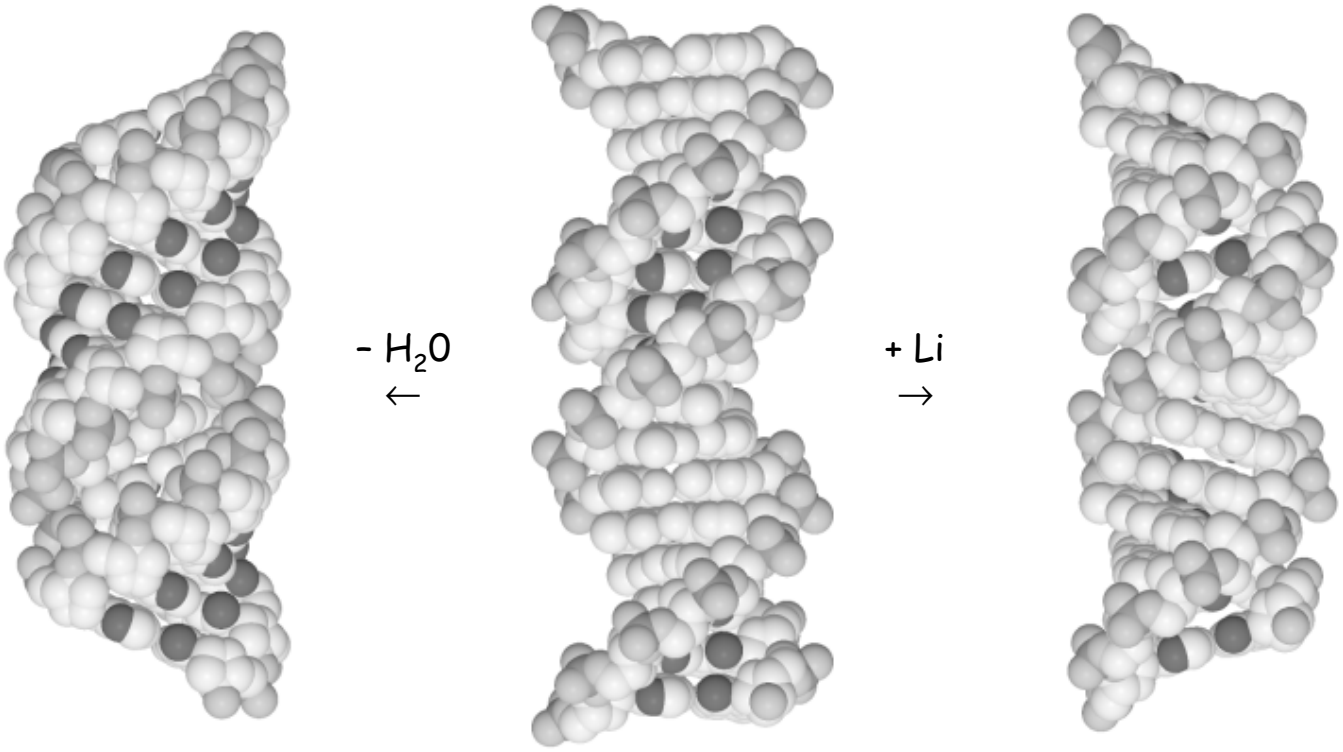
The local sequence-dependent 'rules' can be incorporated in the construction of low-resolution models of deformable polymers.



Non-equilibrium forms, corresponding to deformations along the longest principal axis, superimposed on the intrinsic (average) dimer structures. Moves typical of those sampled with Gaussian random # generator.

Sequence-dependent deformations and contacts

The likely deformations are of the type that induce well-known transitions of double-helical structure.



$- H_2O$
←

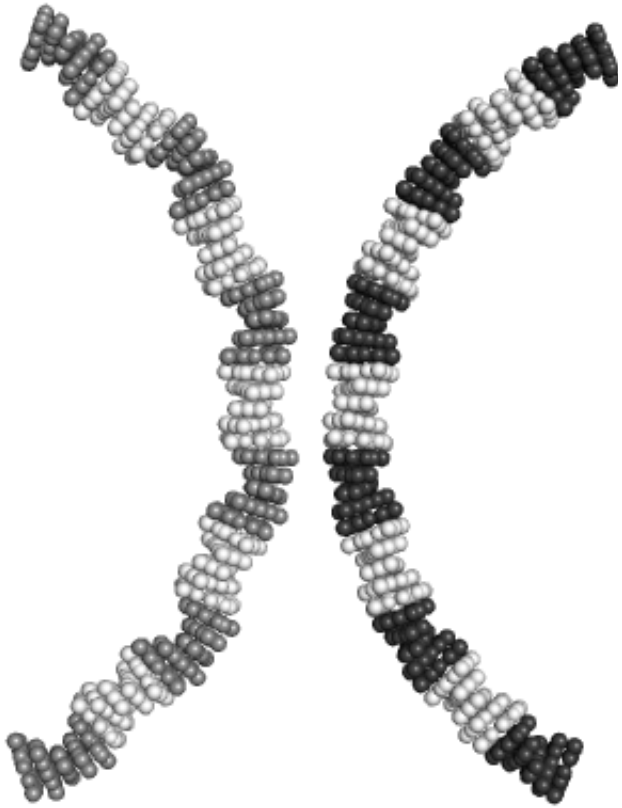
$+ Li$
→

A DNA
Roll > 0
Slide < 0
11 res/turn

B DNA
Roll ≈ 0
Slide ≈ 0
10 res/turn

C DNA
Roll < 0
Slide > 0
9 res/turn

The alternation of different helical forms introduces curvature in DNA without disruption of base stacking.



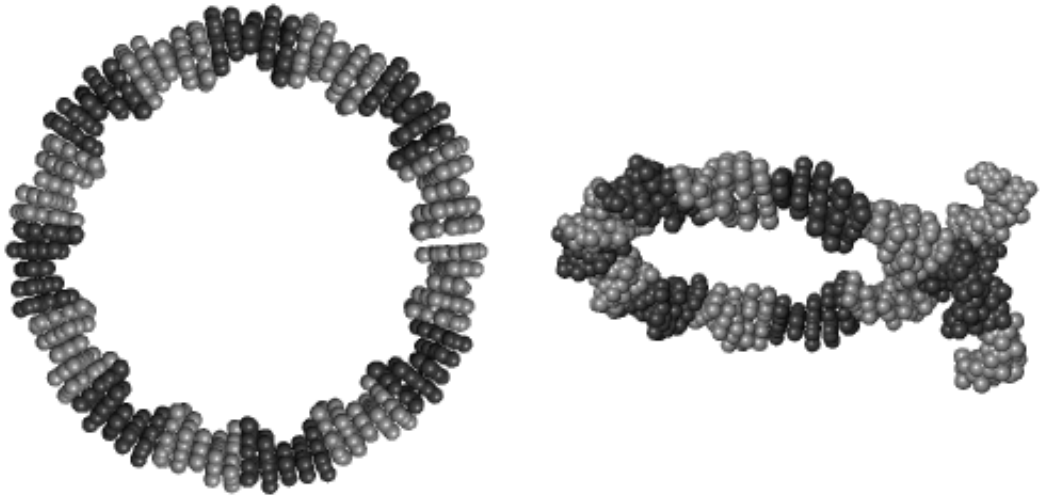
A_5B_5 repeat
150 bp/turn

C_5B_5 repeat
150 bp/turn

A DNA	B DNA	C DNA
Roll = 7.4°	Roll ≈ 0	Roll = –7.4°
Slide = –1.2 Å	Slide ≈ 0	Slide = 1.2 Å

(Tilt, Twist, Shift, Slide, Rise) = (0°, 36°, 0 Å, 0 Å, 3.4 Å)

The alternation of different helical forms introduces curvature in DNA without disruption of base stacking.

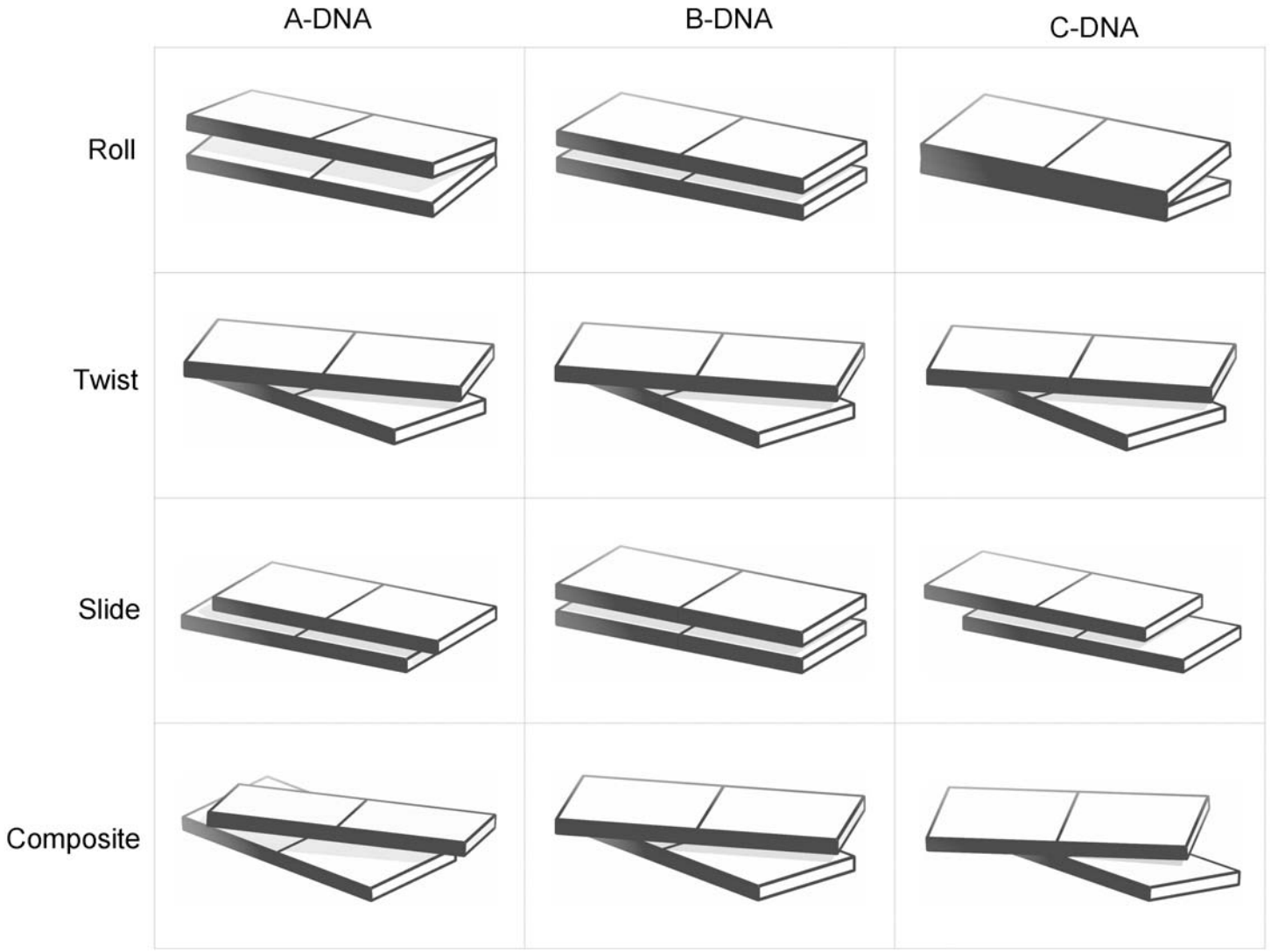


A_5C_5 repeat
75 bp/turn

A DNA	C DNA
Roll = 7.4°	Roll = –7.4°
Slide = –1.2 Å	Slide = 1.2 Å

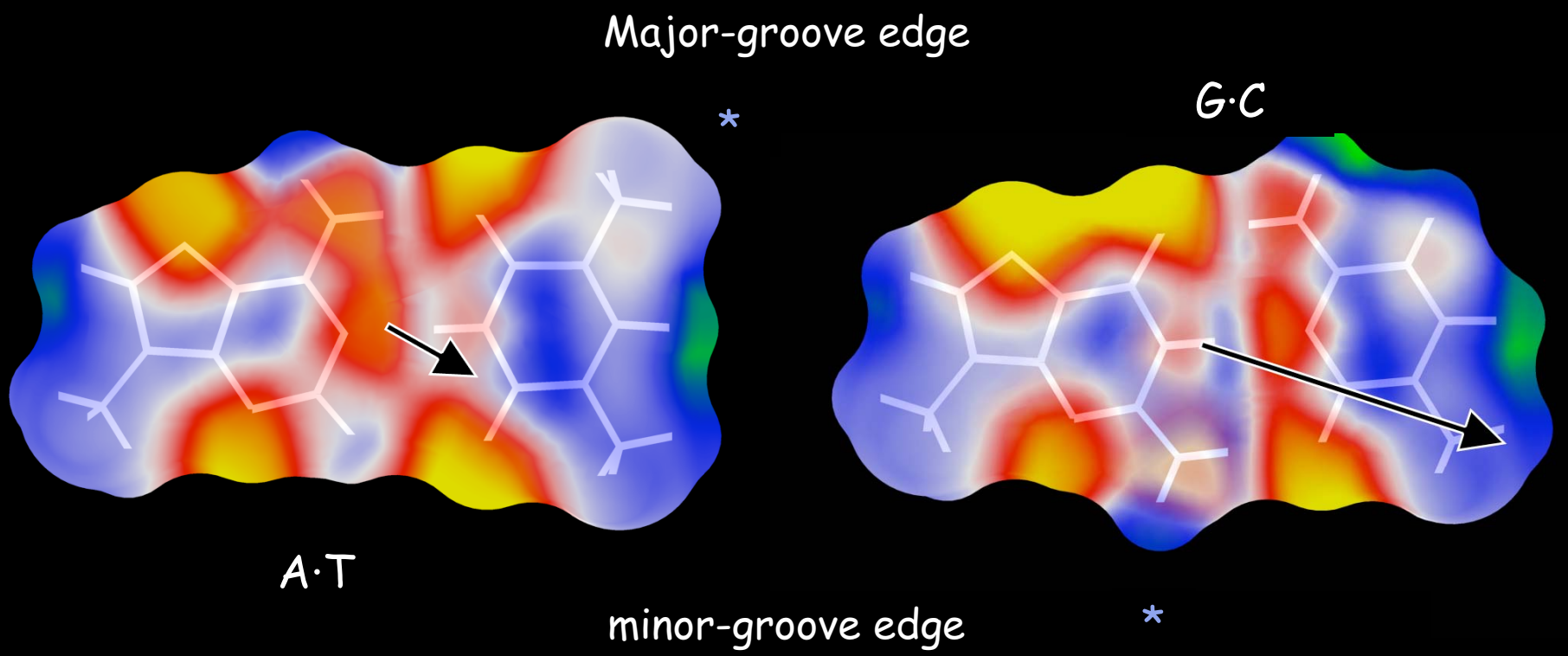
(Tilt, Twist, Shift, Slide, Rise) = (0°, 36°, 0 Å, 0 Å, 3.4 Å)

The subtle distinctions in the three key rigid-body parameters of A-, B-, and C-like base-pair steps are linked to sequence.

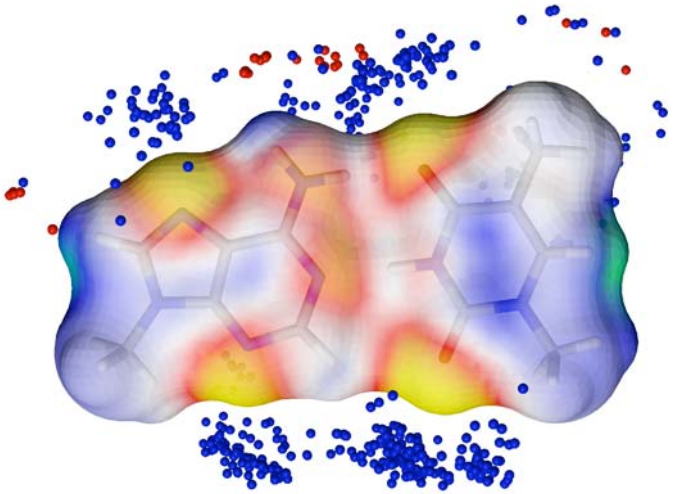


Base-pair 'bricks' drawn such that minor-groove edges are shaded.

The physical basis of these sequence-dependent deformational propensities is also relevant to molecular recognition.

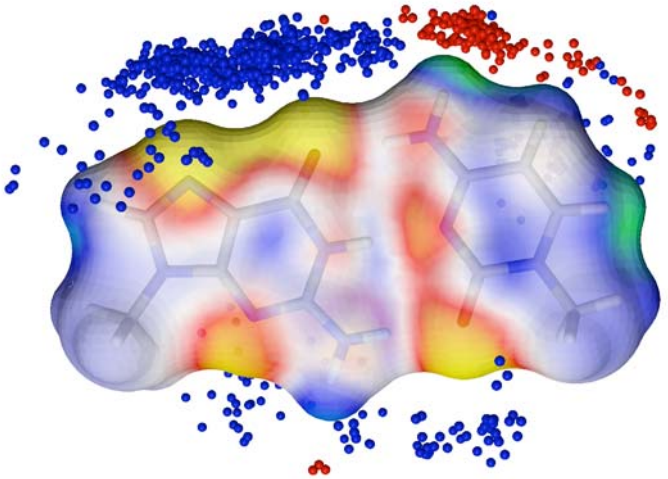


The sequence-dependent, asymmetric build-up of 'ions' on the surfaces of base pairs in well-resolved protein-DNA structures mirror the observed deformational propensities.



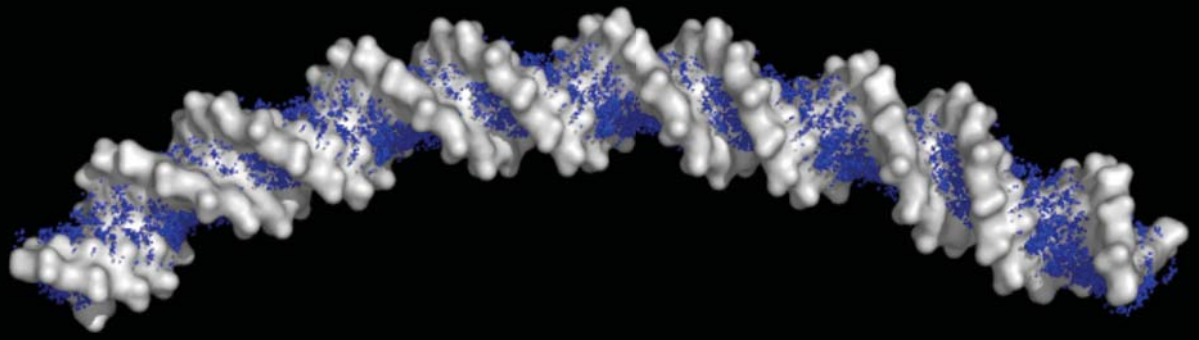
A·T minor groove edge

G·C major groove edge



Arg/Lys_N+
Asp/Glu_O–

The asymmetric build-up of 'cationic' charge on one face of DNA in concert with perturbation of double-helical state offers a structural rationale for the intrinsic curvature of specific sequences.



Factors that enhance DNA cyclization/looping

The unexpected, spontaneous cyclization of short DNA molecules has renewed interest in the intrinsic structure and deformability of double-helical DNA.

$$J = K_1 / K_2$$

$DNA_n \Leftrightarrow cDNA_n$ $\qquad$ $K_1 = [cDNA_n]/[DNA_n]$

$2DNA_n \Leftrightarrow DNA_{2n}$ $\qquad$ $K_2 = [DNA_{2n}]/[DNA_n]^2$

The ease of polymer cyclization is described in terms of the Jacobson-Stockmayer *J* factor, or cyclization constant, which is defined as the ratio of the equilibrium constants for unimolecular cyclization vs. bimolecular ligation of a linear molecule.

Conventional models fit the observed chain-length dependent cyclization of DNA to the properties of an ideal naturally straight, inextensible elastic rod.

Structural and deformational features within the DNA are subsumed in the parameters of the model, *i.e.*, the *uniform* bending and twisting moduli and the double-helical repeat.

Ideal DNA:

$$\psi = \frac{A}{2}\Delta\theta^2 + \frac{C}{2}\left(\tau - \tau_0\right)^2$$

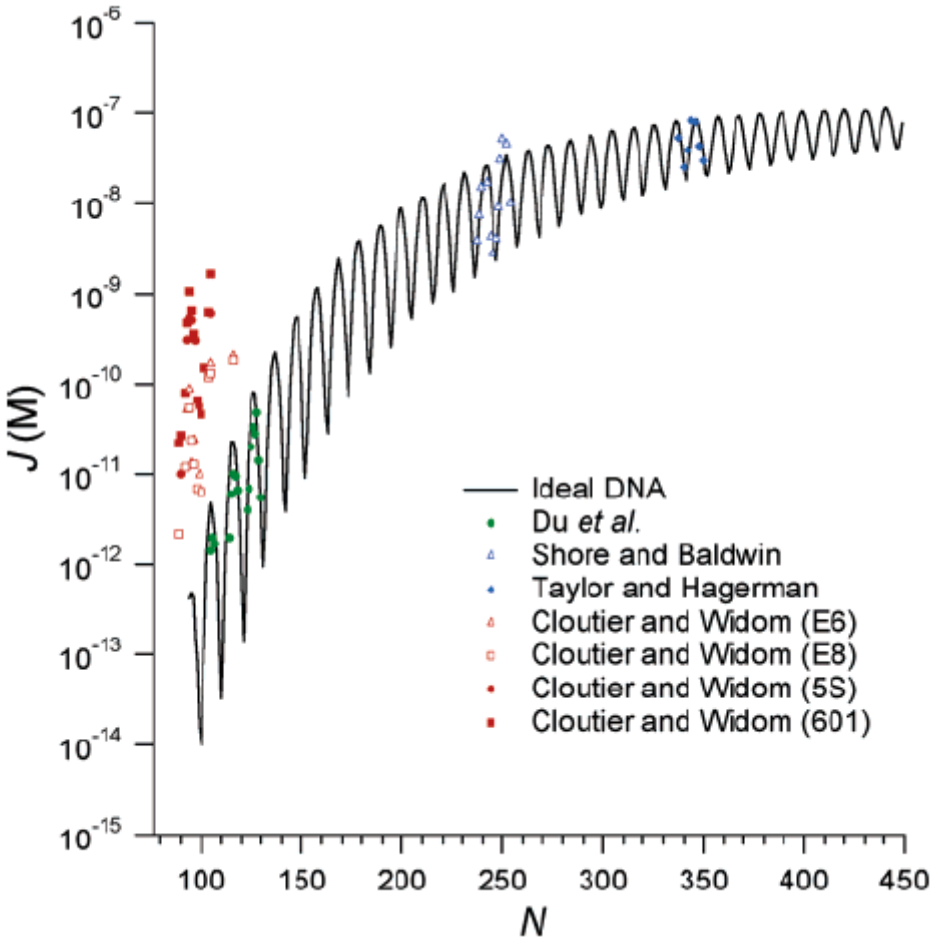
A: bending modulus

$$A = ak_BT$$

(a: persistence length)

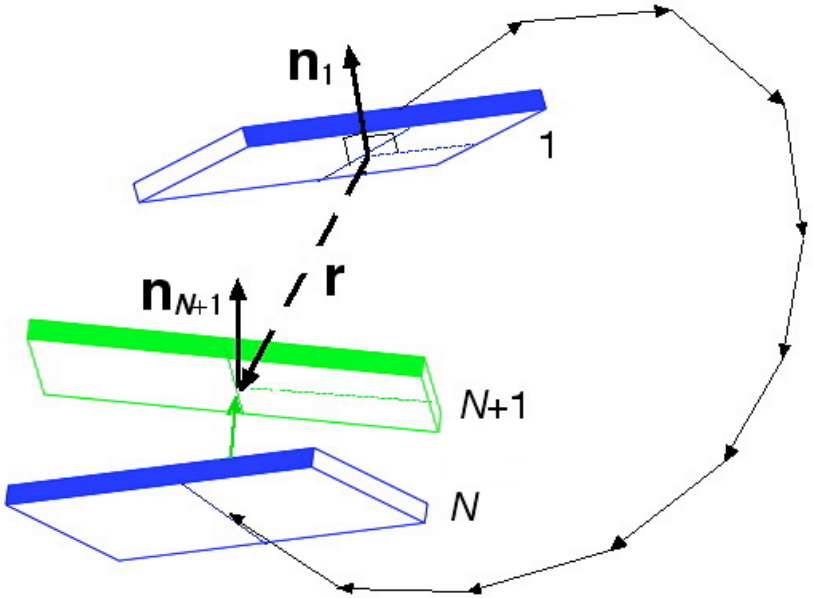
C: torsional modulus

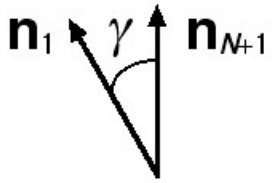
τ_0: double-helical twist

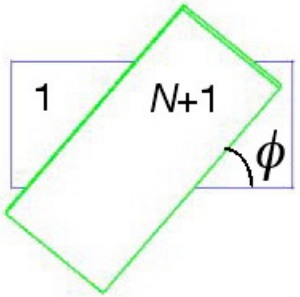


Monte-Carlo estimates of the dependence on chain length of the J factor (cyclization propensity) of an ideal, inextensible DNA with persistence length ~500 Å, 10.5 base pair/turn helical repeat, and torsional modulus C = 1.4 A (Czapla *et al* 2006).

The likelihood of DNA cyclization, measured by the *J* factor, is estimated from the number of simulated configurations of a linear molecule with terminal residues positioned so as to insure successful ligation.







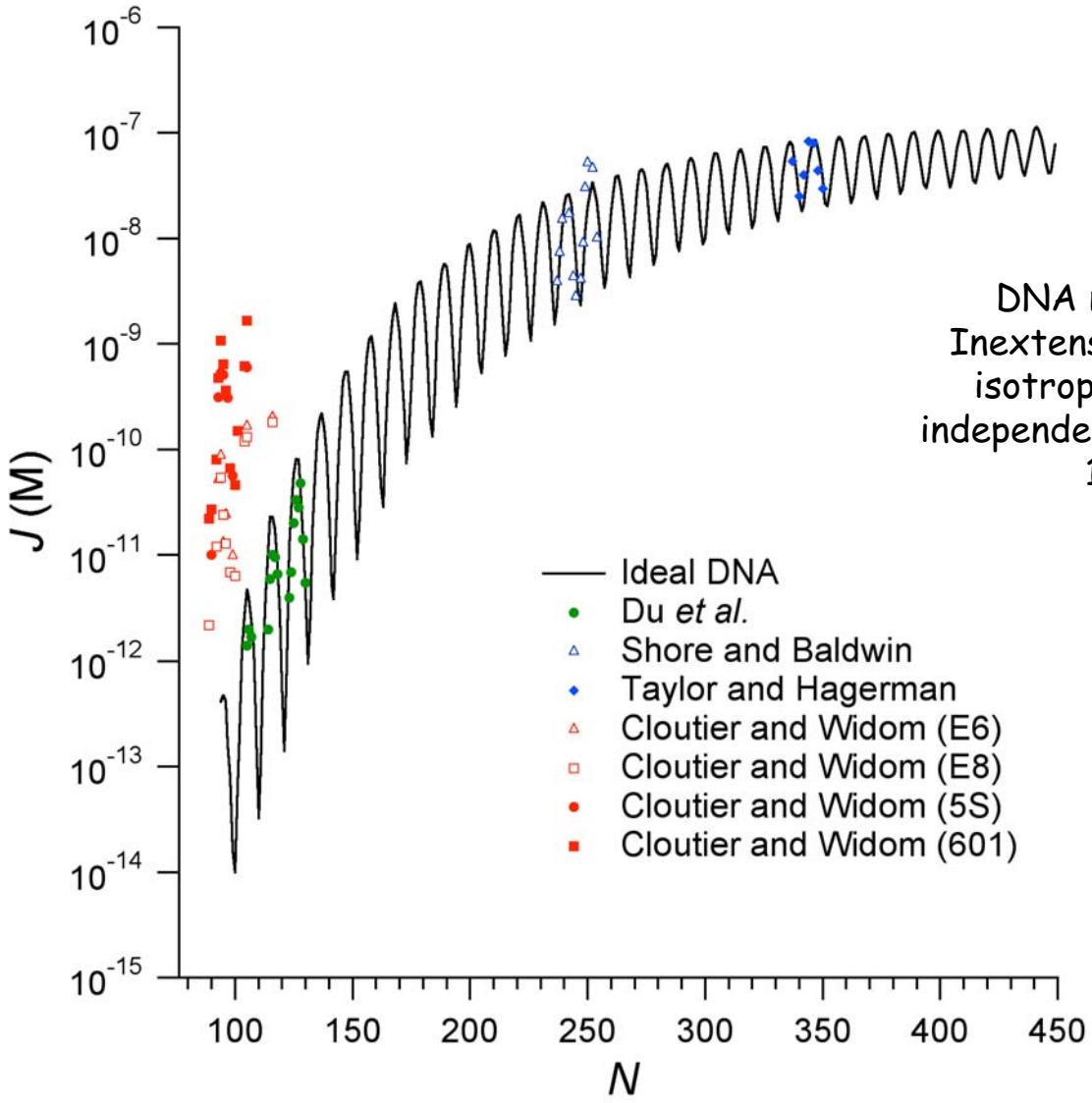
The hypothetical base pair N+1 must overlap the first base pair in a perfectly closed DNA of N base pairs.

$$\left| \mathbf{r} \right| < r_0$$

$$\cos\gamma > 1 - \varepsilon$$

$$\phi < \phi_0$$

The computed dependence of J on chain length N of ideal DNA underestimates the measured ease of cyclization for very short (89-105 bp) chains.



DNA naturally straight
Inextensible ($f_{44}=f_{55}=f_{66}>>0$)
isotropic bending ($f_{11}=f_{22}$)
independent twisting ($f_{13}=f_{23}=0$)
10.5 bp/turn

The DNA molecules found to close most easily into tight minicircles contain a well-known 'nucleosome-positioning' sequence,

'601TA-94'

with regularly spaced TA steps in phase with the double helical repeat
and AT-containing dimers that alternate at half helical turns with GC-containing steps.

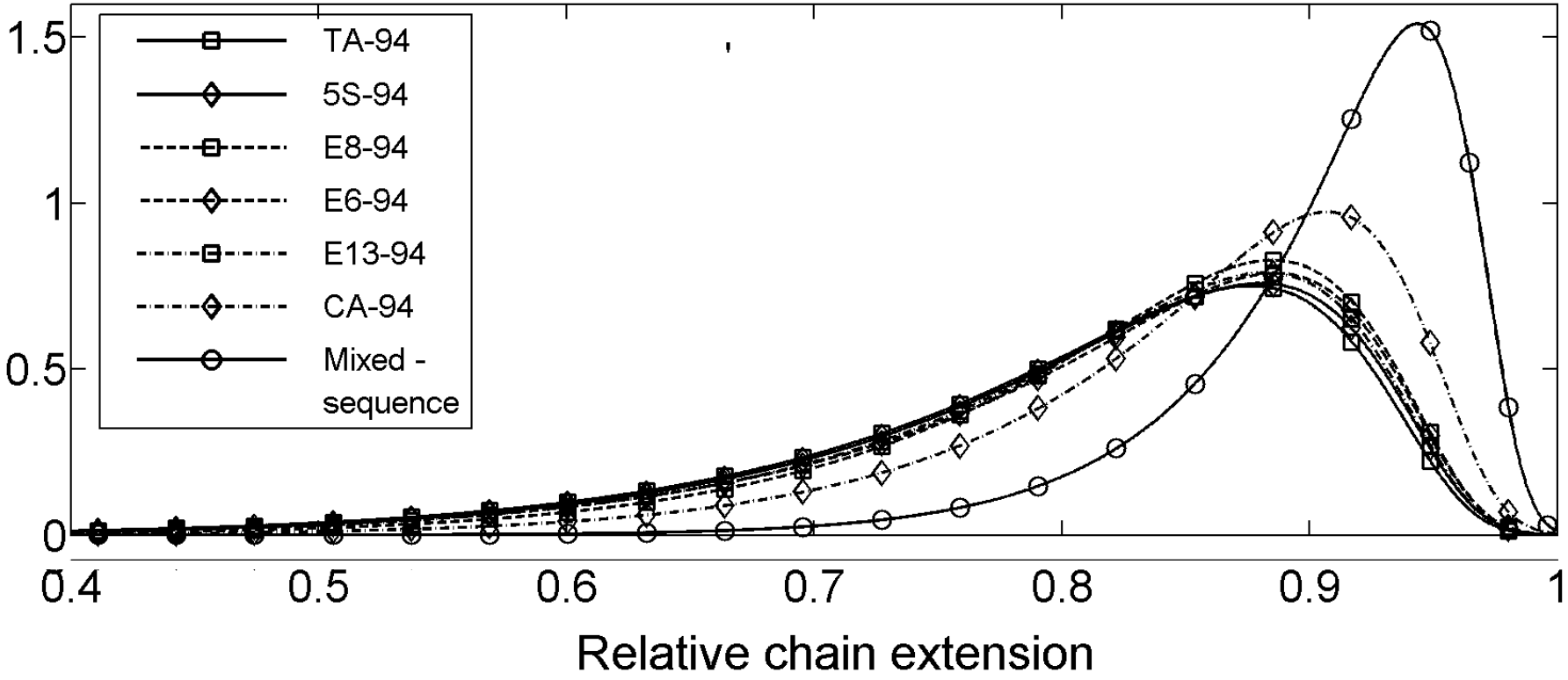
ggccgggtcg TAgCAagctc TAgCAccgct TAaacgCAcg
TAcgcgcTGt cTAccgcgtt tTAaccgcCA aTAggatTAc
tTAcTAgtct cTAc

$$J_{obs} = 1.07\times10^{-9} \pm 2.3\times10^{-11}$$

$$J_{ideal\_DNA} = 2.7\times10^{-12}$$

Cloutier & Widom

The knowledge-based dimeric potentials account approximately for the observed cyclization propensities of DNA.



The simulated distributions of the end-to-end distances of 94-bp sequences that are more easily cyclized are shifted to smaller values than the corresponding distribution for mixed-sequence DNA.

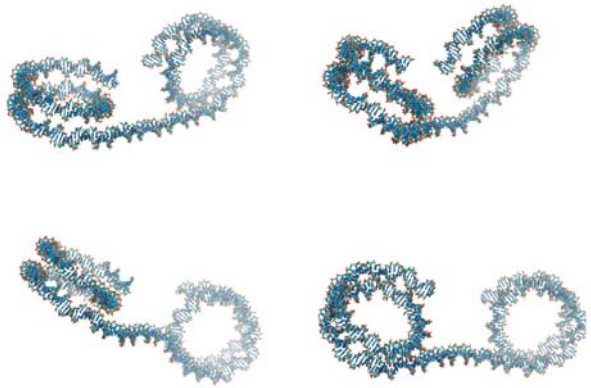
$$f_{ij}^{\dagger}\left(\text{mixed-sequence}\right)=\sum_{XZ}\zeta f_{ij}^{\dagger}\left(XZ\right);\ \zeta=0.85,\ a=500\ \text{Å}$$

The boundary delimiting the 10% shortest configurations, $r_{0.1}$, is smaller for the positioning sequences than the control sequences , and both limits are substantially smaller than the $r_{0.1}$ boundaries for mixed-sequence DNA.

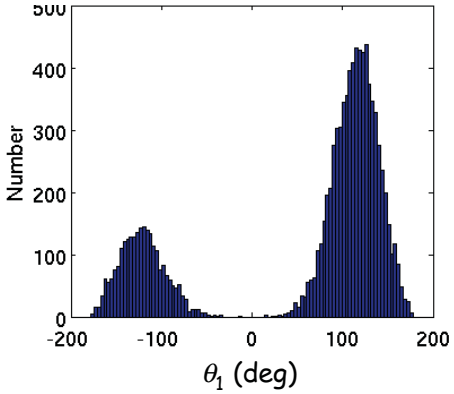
DNA	Base-pair sequence	r_{01}	log J
TA-94	ggccgggtcg TAgcaagctc TAgcaccgct TAaacgcacg TAcgcgctgt cTAccgcgtt tTAaccgcca aTAggatTAc tTAcTAgtct cTAc	206	-9.0
5S-94	ggccgacatc cctgaccctt TAaaTAgctT Aactttcatc aagcaagagc cTAcgaccaT Accatgctga aTATAccggt tctcgtccga tcac	208	-9.3
E6-94	ggccgtgcgc acgaaatgcT Atgccgaaga ttggatggac atgctTATAa aaggaatccc cagaggTAat ccttgatctg atgatgatcc gccc	211	-10.1
E8-94	ggccgtgcgT AgaacTActt tTAttTAtcg cctccacggt gctgatcccc tgtgctgttg gccgtgtTAt ctcgagtTAg TAcgacgtcc gccc	218	-10.3
E13-94	ggccgtgcgt tcggTAaggt gcgatggcct catcaaggcg ccaTATAaga tcactcgTAg tgaaaaccTA cccttcattT Aatgttgatc gccc	211	-10.2
CA-94	ggccgtccca gcaagctcca ggtgcgccca aacggctgca gacgccctgc acggcagccc aagcgcaccc agagcccccc ctccggaatt cacc	231	-10.2
Mixed-sequence		259	-11.6

Higher-levels of DNA organization (progress report)

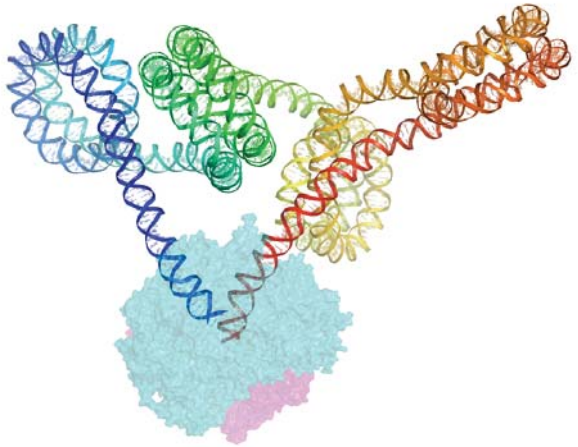
The arrangements of realistically simulated nucleosomes in minichromosomes can similarly be used to model the next level of DNA organization.



Simulated dinucleosomal configurations



Dinucleosome parameters



Model system: relative enhancement of promoter-enhancer communication on nucleosome-positioned vs. free DNA