DNA energy landscapes deduced from well resolved structures: implications of experimental data on physical modeling

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

Sreekala Balasubramanian, Andrew Colasanti, Luke Czapla, Wei Ge, Yun Li, Fei Xu

A.R. Srinivasan
Marcia Fenley (Florida State U)
Atsushi Matsumoto (Japan Atomic Research Agency)
David Swigon (U Pittsburgh)
Michael Tolstorukov (Kharkov State U, Ukraine)
Victor Zhurkin (NIH)

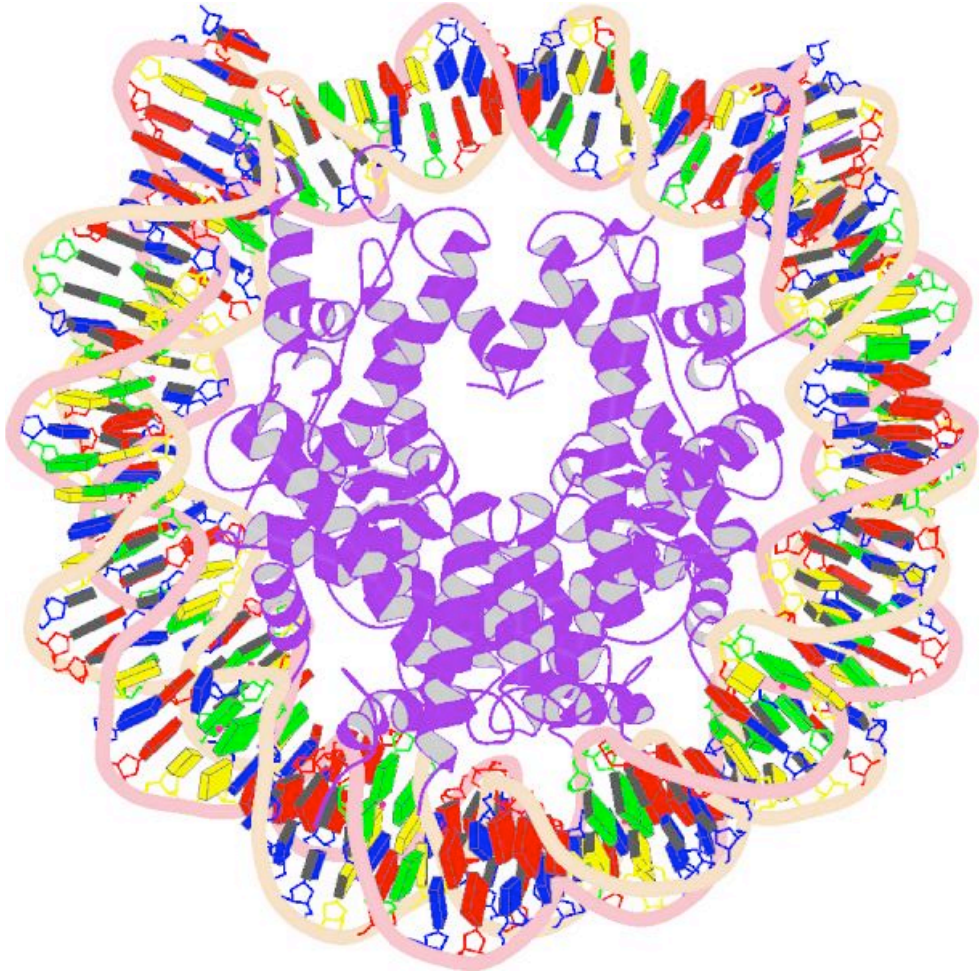
USPHS GM20861, GM34809

Outline

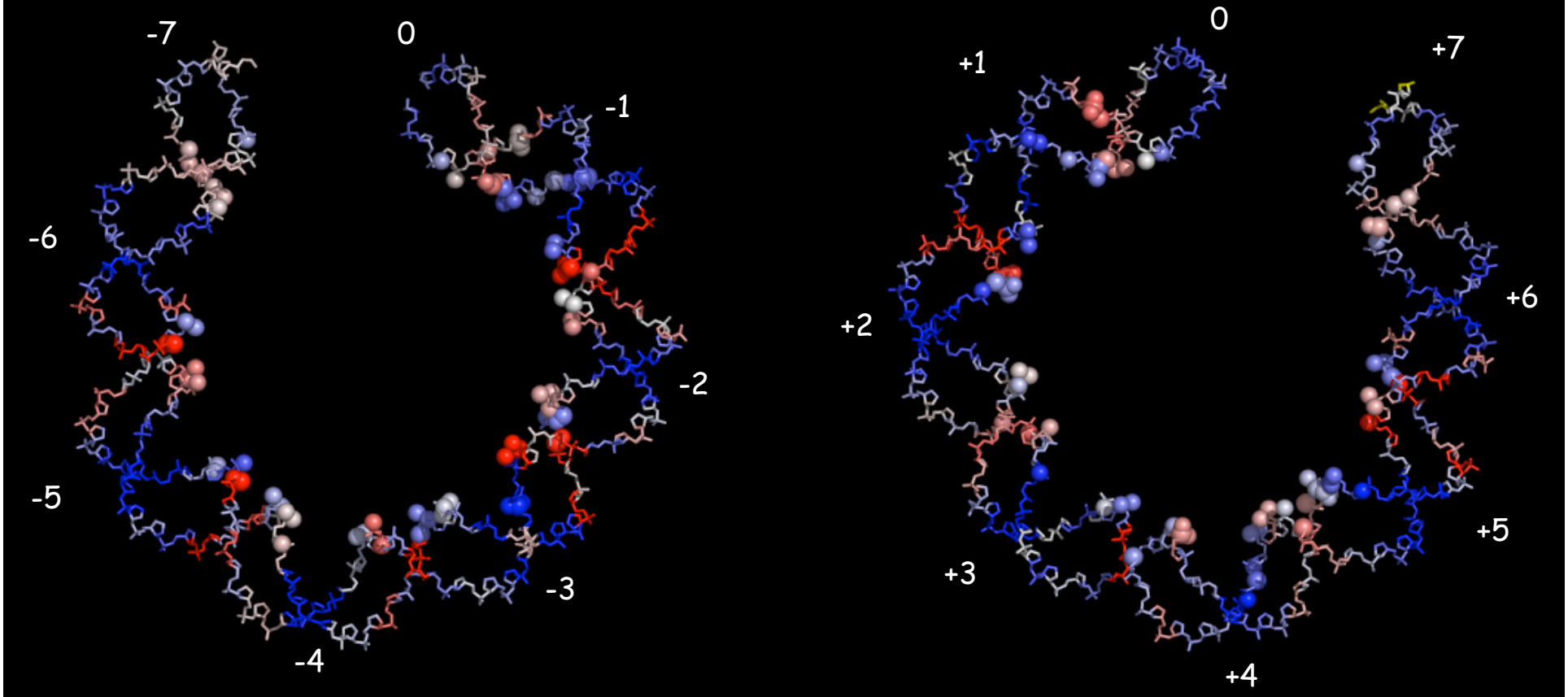
- Recognition, local deformation, and global folding of nucleosomal DNA
- DNA deformation and solvation patterns in high resolution structures, including base sequence-dependent effects
- Implications of local sequence-dependent features on binding and elastic properties of DNA at the mesoscale level

Recognition, deformation, and folding of nucleosomal DNA

The nucleosome core particle is one of the most striking examples of protein-induced DNA deformation.



Simplified, color-coded representation of a 146 base-pair DNA wrapped ~1.5 turns around a (violet) core of eight proteins in the nucleosome, the fundamental DNA packaging unit in the nucleus: NDB_ID: pd0001 (Luger *et al.*, 1997)



The tight bending of the double helix on the surface of the nucleosome is accompanied by contacts (*highlighted here by spheres*) of the positively charged histone proteins with the negatively charged DNA phosphates.

The many close interatomic contacts between charged/polar amino acids and the sugar-phosphate backbone suggest an electrostatic mechanism of nucleosomal DNA bending (Mirzabekhov *et al.*, 1978).

Contacts to DNA		Amino acid type			
	totals	hydrophobic	charged	polar	glycine
Base	20	0	19	1	0
Sugar	107	12	75	12	8
Phosphate	248	24	140	68	16

Residue breakdown of close protein-DNA interatomic (≤ 3.4 Å) contacts in the best resolved (1.9 Å) nucleosome crystal structure, NDB_ID: pd0287 (Davey *et. al*, 2002)

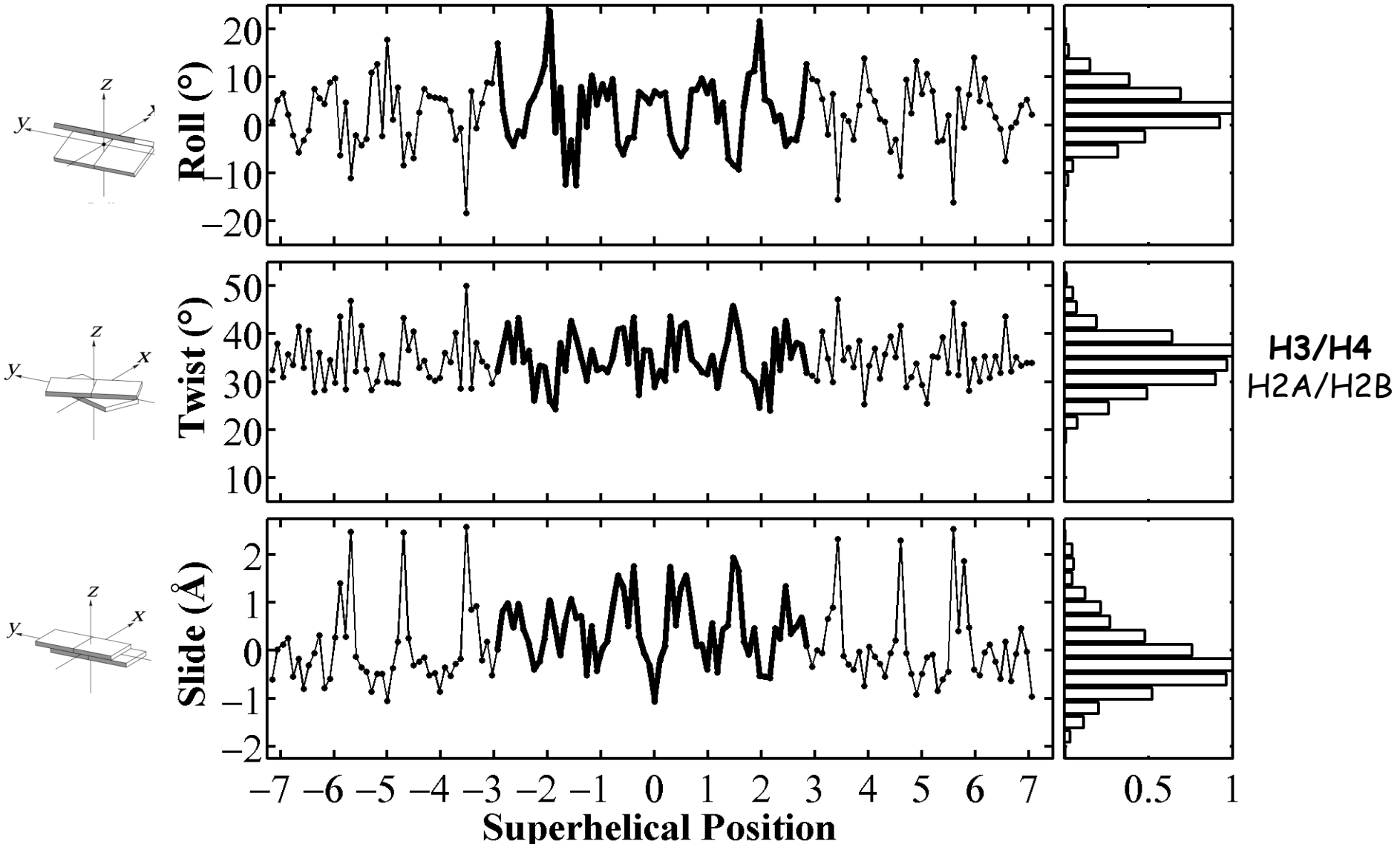
Phosphate neutralization seemingly fails to account for the nucleosome binding preferences of natural and synthetic "positioning" sequences.

GGCAAGGTCGCTGTTCAATACATGCACAGG
GTGTATGTATCCGACaCGTGCCTGGAGACT
AGGAAGTAATCCCCCTTGGCGGTTAAAATGC
GGGGGACAGCGCGTACGTGCGTTTAAGCGG
TGCTAGAGCTGTCTACGACCAATTGAGCGG
CCTCGGCACCGGGATTCTCCAGGACGGCCG
CGTATAGGGTCCATCACATAAGGGATGAAC
TCGGTGTGAAGAATCATGCTTT

Conventional structural rationales for DNA sequence recognition via the direct contact of unique base atoms by protein fail in the case of nucleosomes, where there are few such contacts.

J. Widom - 602nvp_rev_rev, 233 bp

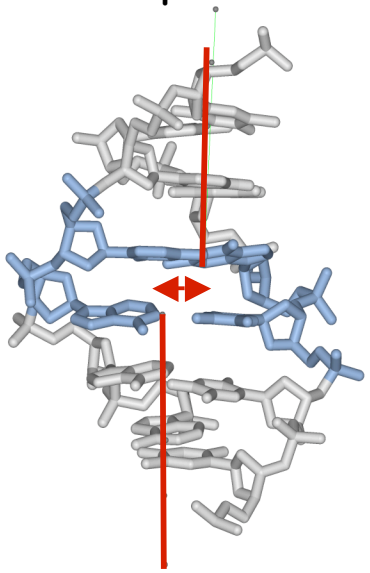
Nucleosomal DNA undergoes concerted changes of three key step parameters.



Major deformation @ contact points

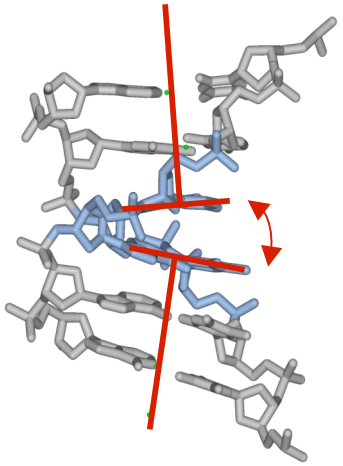
The most highly deformed dimer steps kink and displace the double helical axis.

TG:CA(38) @ SH -3.5



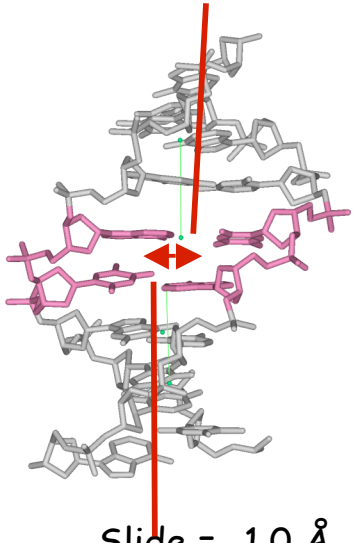
Slide = +2.7 Å
(minor-groove view)

90°

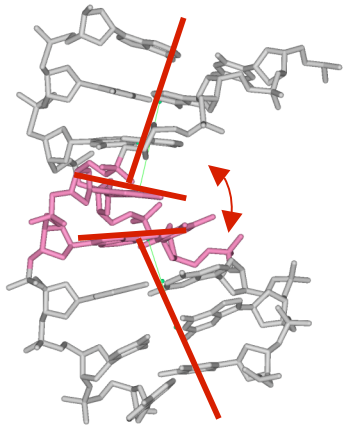


Roll = -18°
(minor-groove bend)

TA:TA(23) @ SH -5

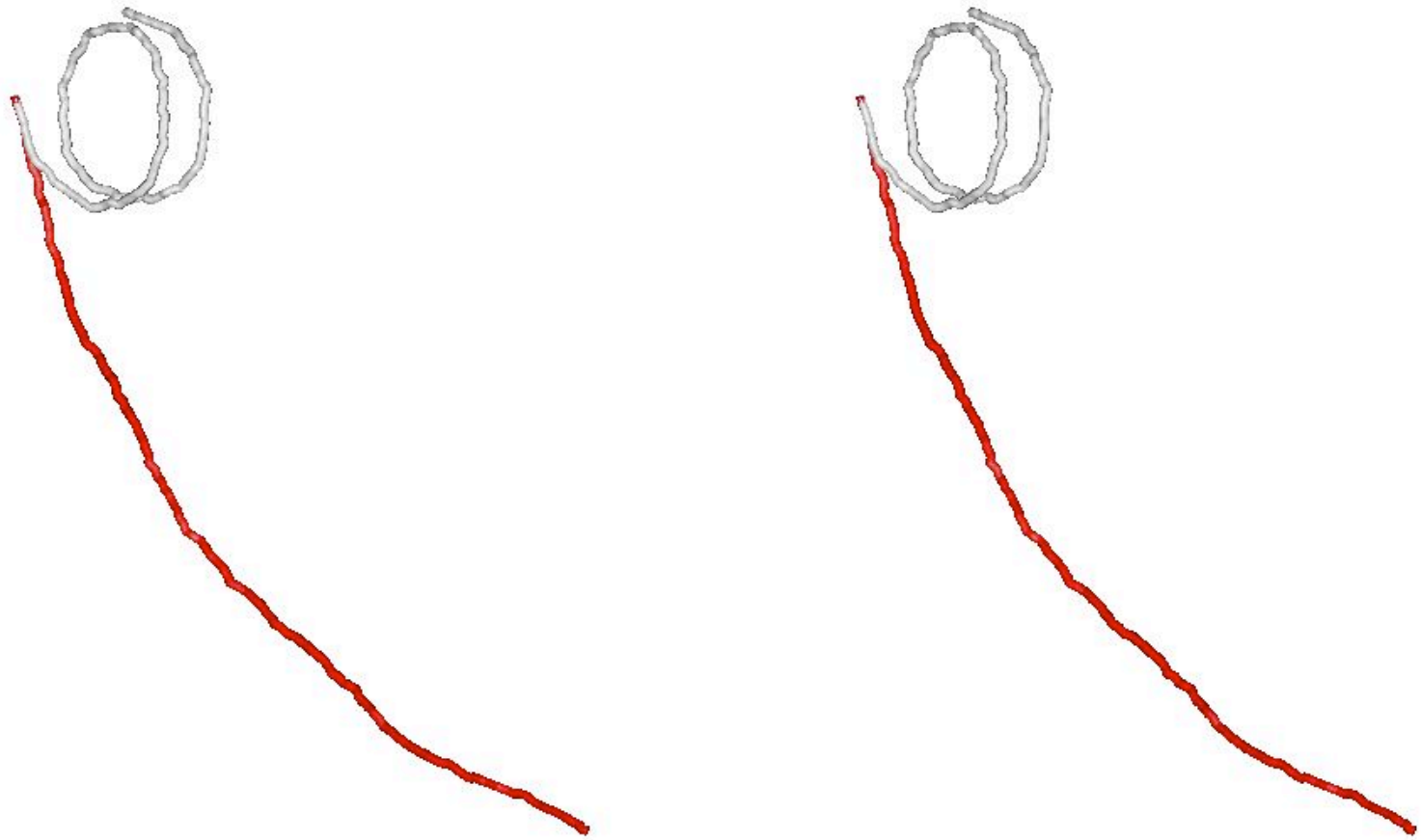


Slide = -1.0 Å
(major-groove view)



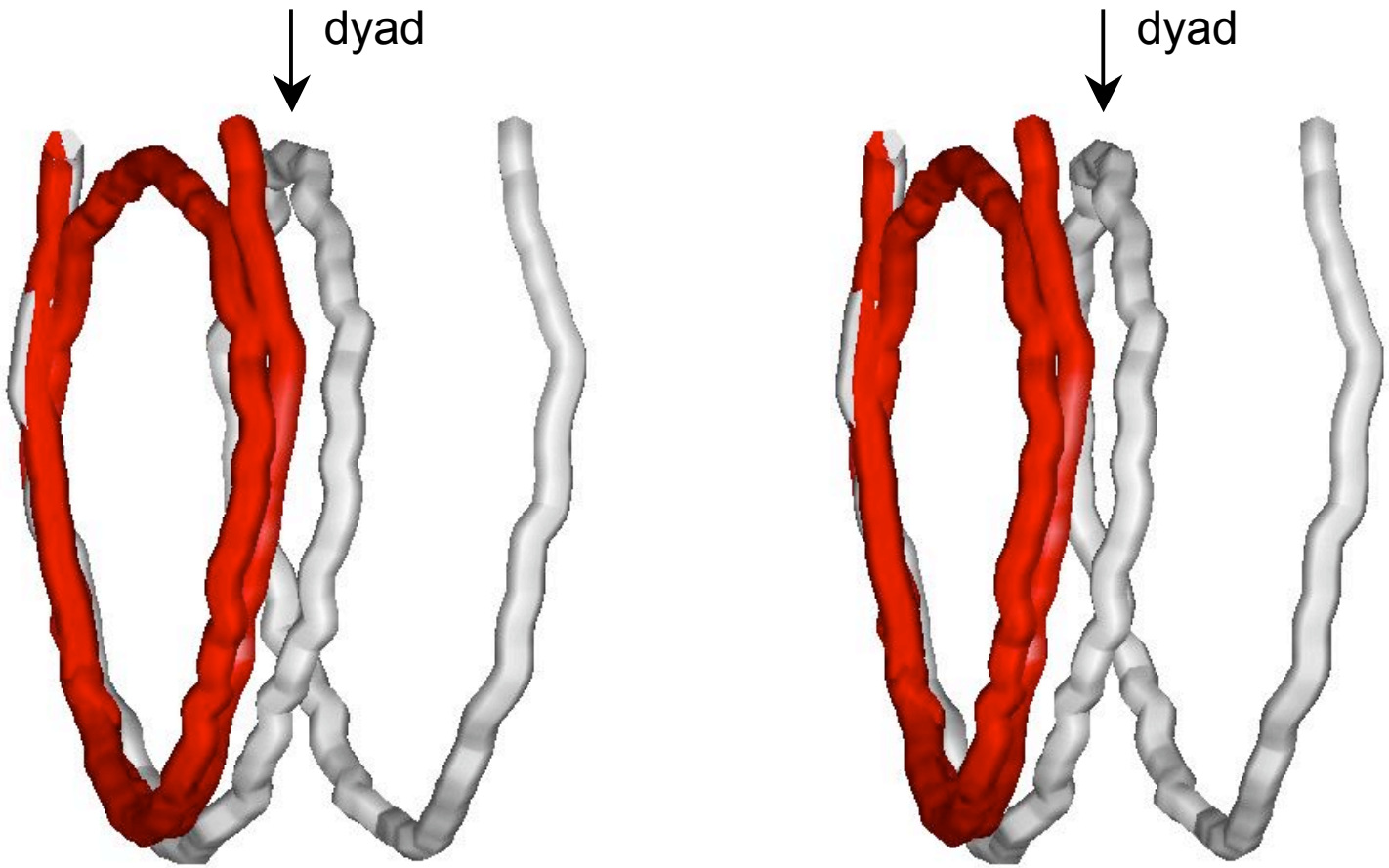
Roll = +18°
(major-groove bend)

Bending via Roll underlies the tight bending of nucleosomal DNA.



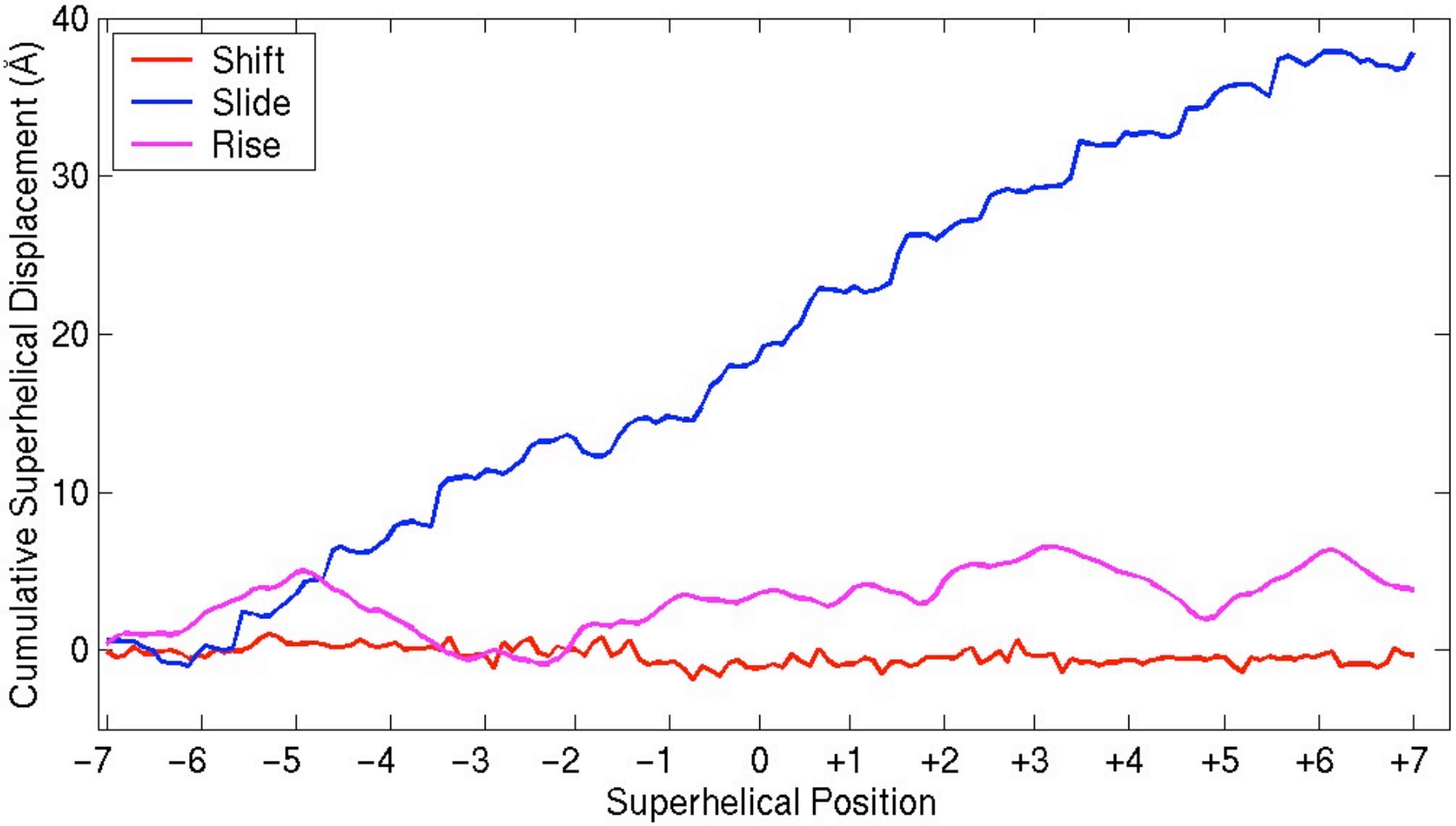
Roll = 0° at all nsm steps

Displacement via Slide governs the pitch of nucleosomal DNA.

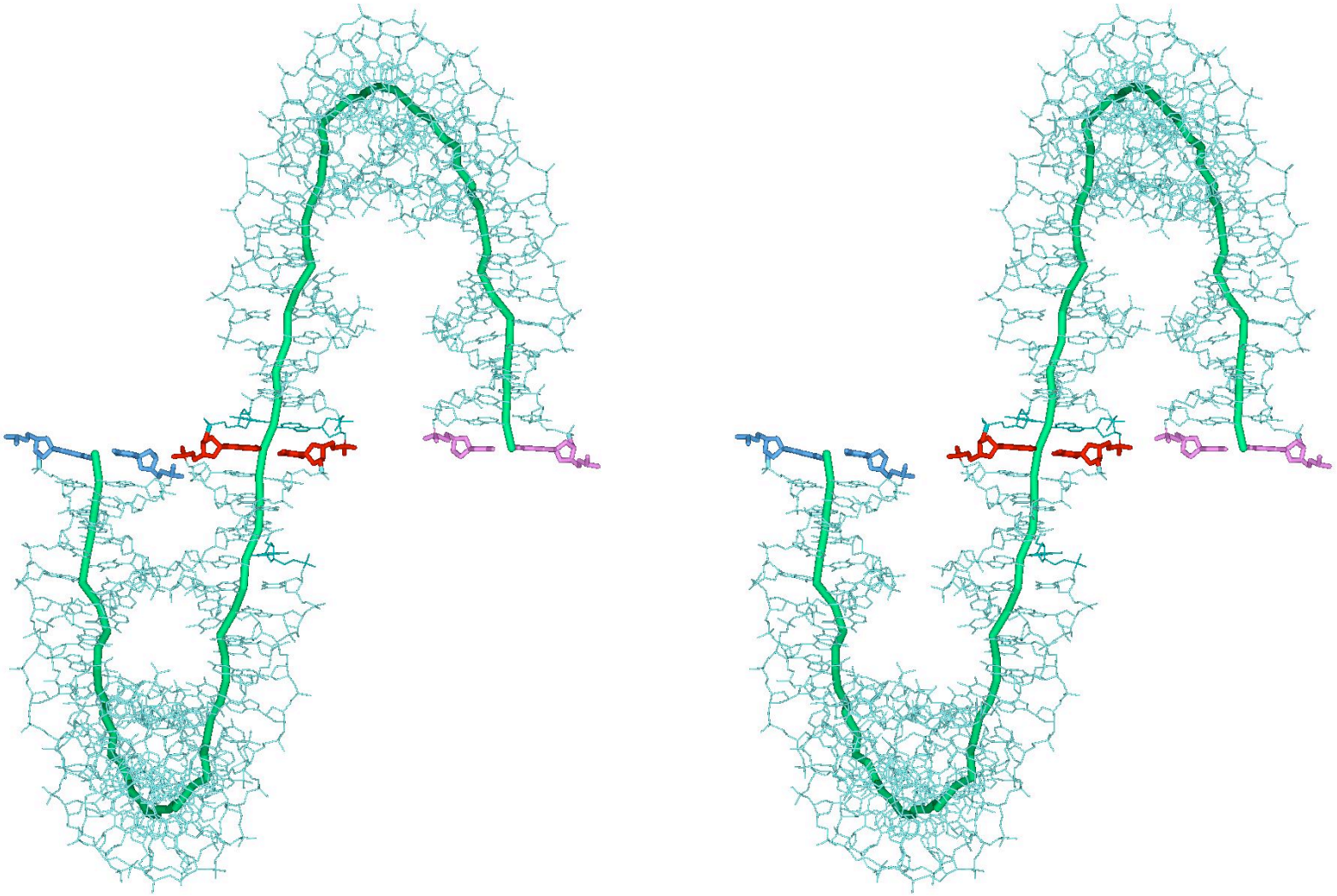


Slide = 0 Å at all nsm steps

Large jumps in Slide account for ~90% of the superhelical pitch.



The largest deformations in Slide occur at base-pair steps where the long axis runs parallel to the superhelical axis.

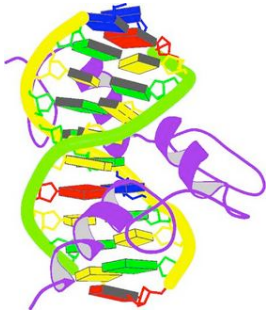


stereoview of nucleosomal DNA bps #33-115

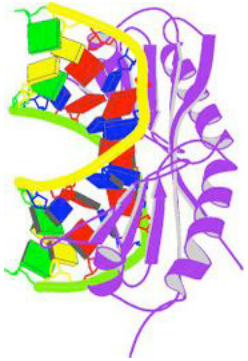
θ_5(33-34) = –0.86 Å; θ_5(74) = –1.06 Å; θ_5(114-115) = –0.74 Å

Recognition and deformational properties of non-nucleosomal DNA

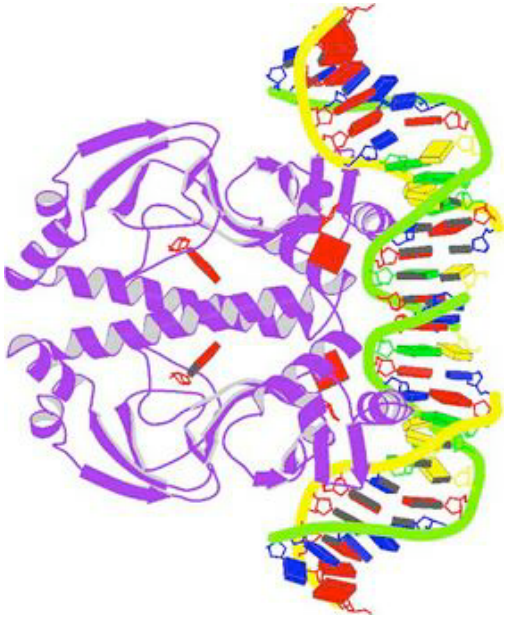
The growing library of protein-DNA crystal complexes reveals implicit sequence-dependent structural information embedded in the DNA base pairs.



Zif268

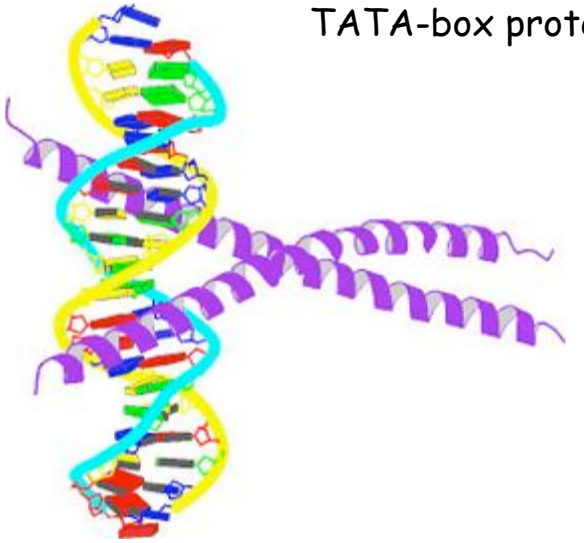


TATA-box protein

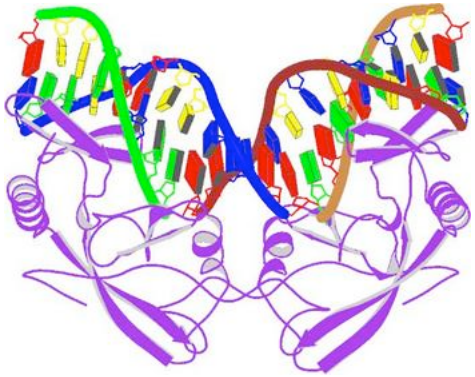


Catabolite activator protein

Tc3 transposase



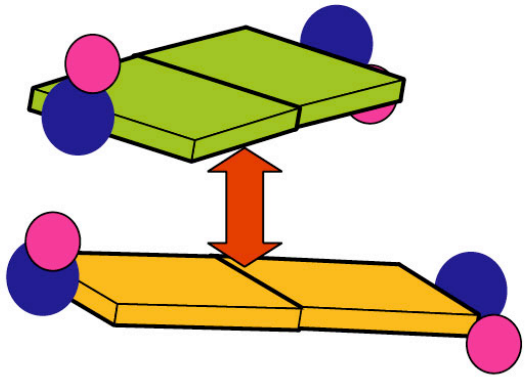
α enhancer protein



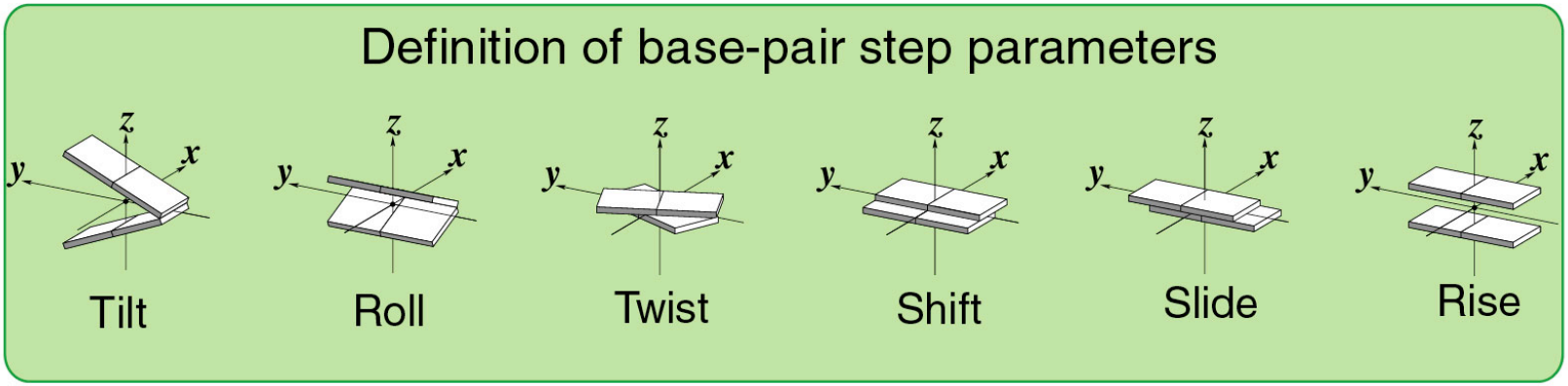
Homing endonuclease

Serum reponse factor

Base-pair step representation of DNA

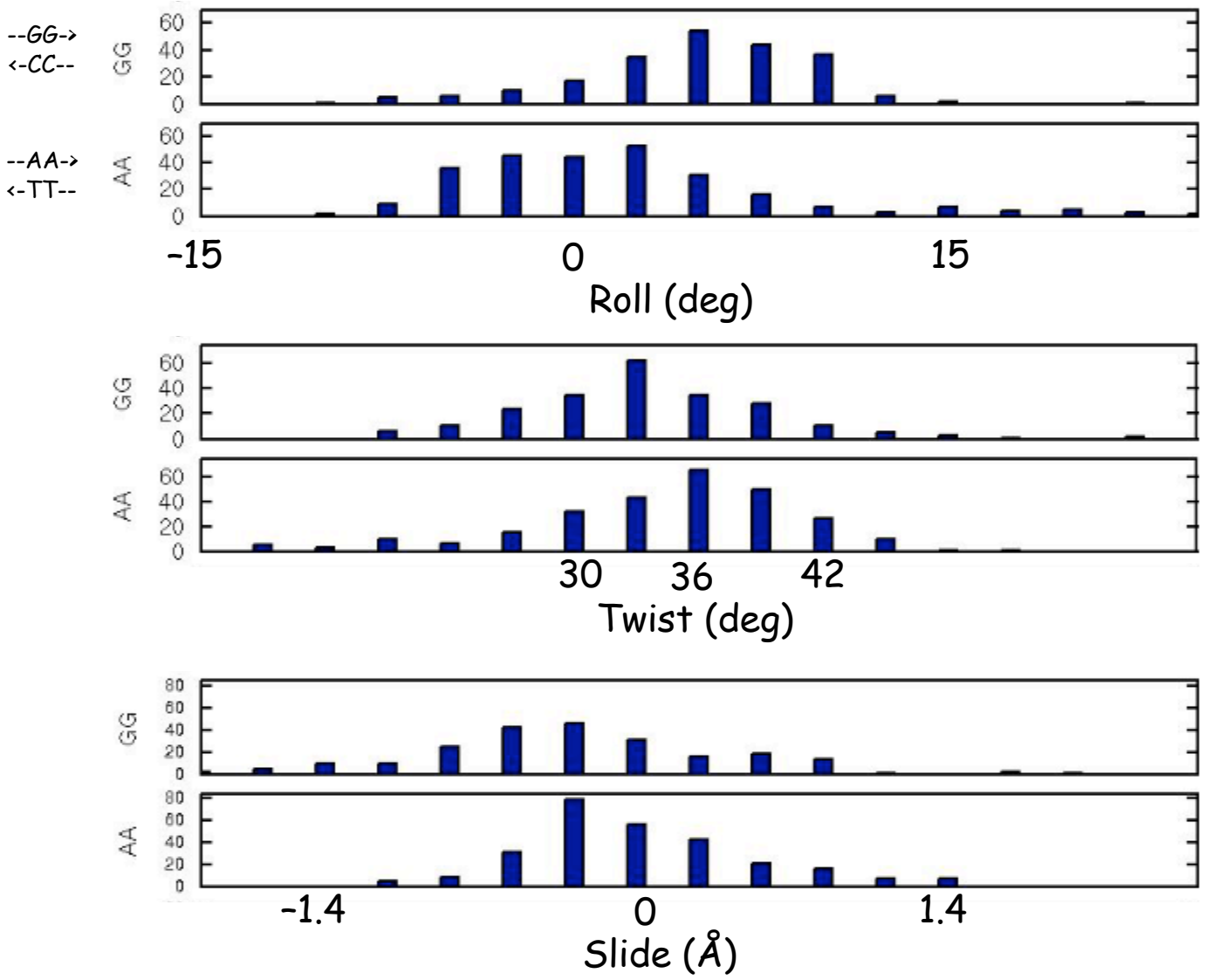


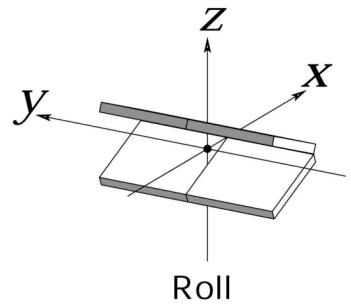
The relative position and orientation of adjacent base pairs, expressed in terms of six rigid-body "step" parameters, show characteristic, sequence-dependent patterns in well-resolved structures of DNA.

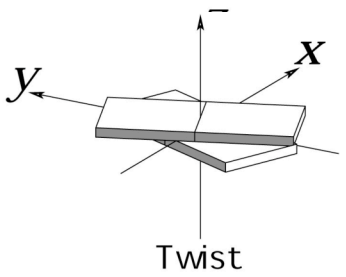


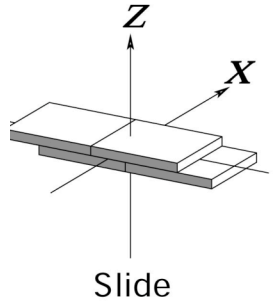
60 atoms/bp

Sequence effects appear in the observed base-pair "step" parameters.



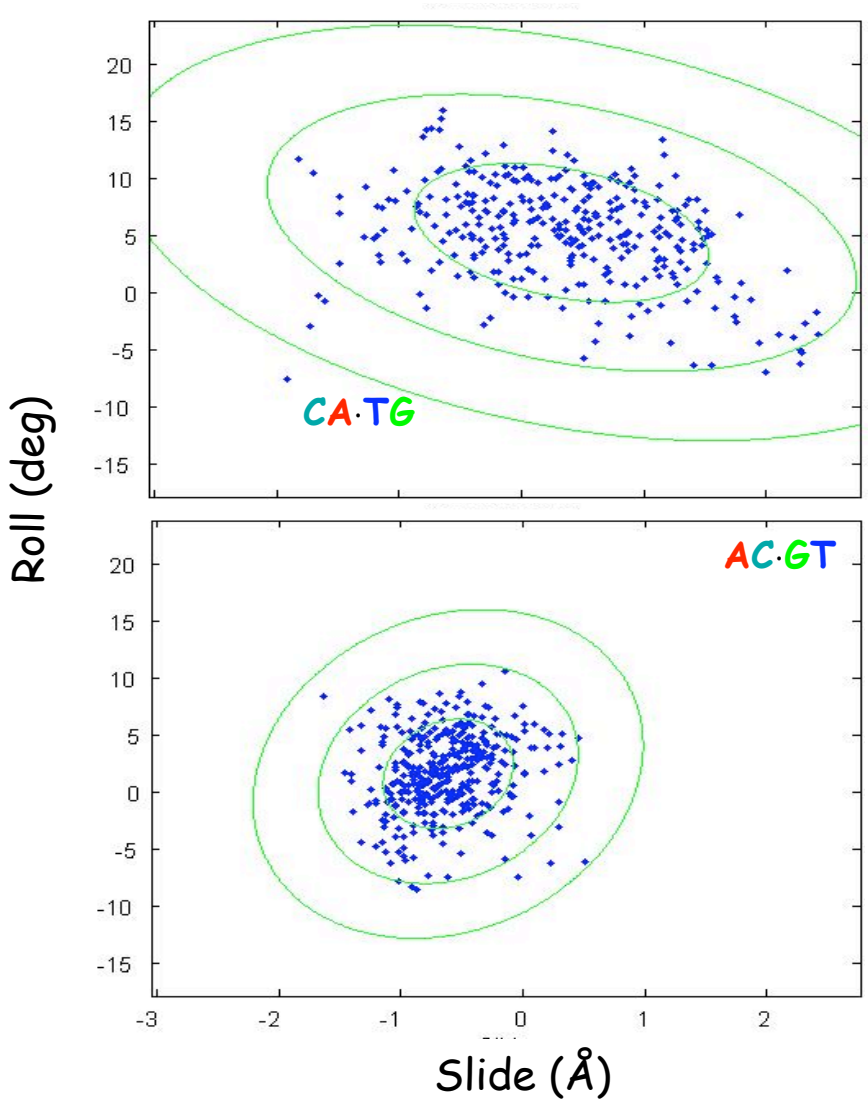






Sequence-dependent variation of the three base-pair "step" parameters, which dominate the conformational variability of DNA.

Pyrimidine-purines stand out as deformable steps and purine-pyrimidines as stiff steps in the knowledge-based potentials. ‡



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 dimer steps taken from 239 protein-DNA crystal complexes of 2.5 Å or better resolution filtered to exclude over-represented structures

Knowledge-based potentials

Dimeric "force constants" of individual steps are derived from the covariance of observed parameters in non-nucleosomal structures.

$$F^{-1} = \begin{bmatrix} \langle \Delta\theta_1 \Delta\theta_1 \rangle & \langle \Delta\theta_1 \Delta\theta_2 \rangle & \cdots & \langle \Delta\theta_1 \Delta\theta_6 \rangle \\ \langle \Delta\theta_1 \Delta\theta_2 \rangle & \langle \Delta\theta_2 \Delta\theta_2 \rangle & \cdots & \langle \Delta\theta_2 \Delta\theta_6 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \Delta\theta_1 \Delta\theta_6 \rangle & \langle \Delta\theta_2 \Delta\theta_6 \rangle & \cdots & \langle \Delta\theta_6 \Delta\theta_6 \rangle \end{bmatrix}$$

where

$$\langle \Delta\theta_i \Delta\theta_j \rangle = \langle \theta_i^{XZ} \theta_j^{XZ} \rangle - \langle \theta_i^{XZ} \rangle \langle \theta_j^{XZ} \rangle$$

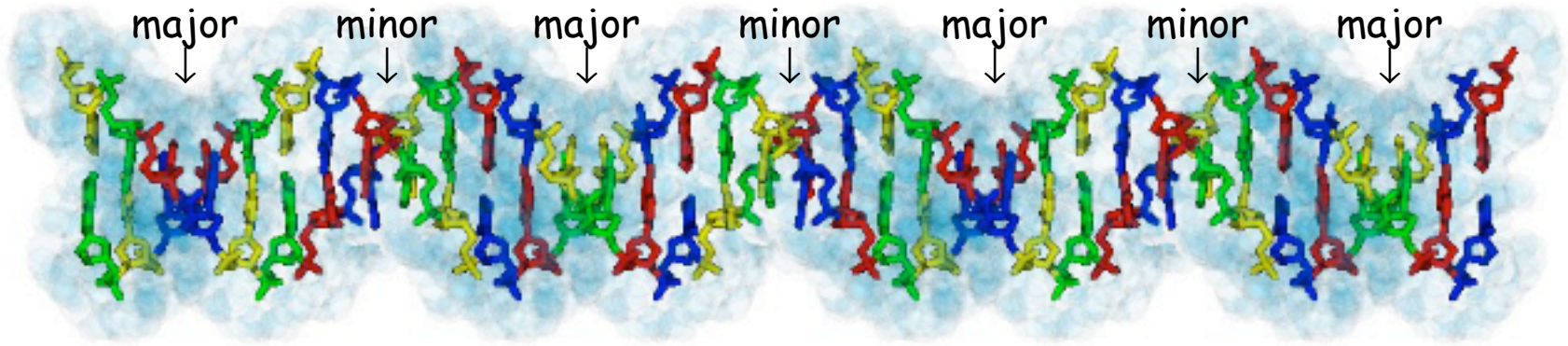
The covariance matrix is equated to the inverse of the force-constant matrix.

M. Go & N. Go (1976) "Fluctuations of an α-helix." *Biopolymers* **15**, 1119-1127.

Amino acid-nucleotide interatomic contacts in well resolved (≤ 2.5 Å) protein-DNA structures offer insight into sequence recognition.

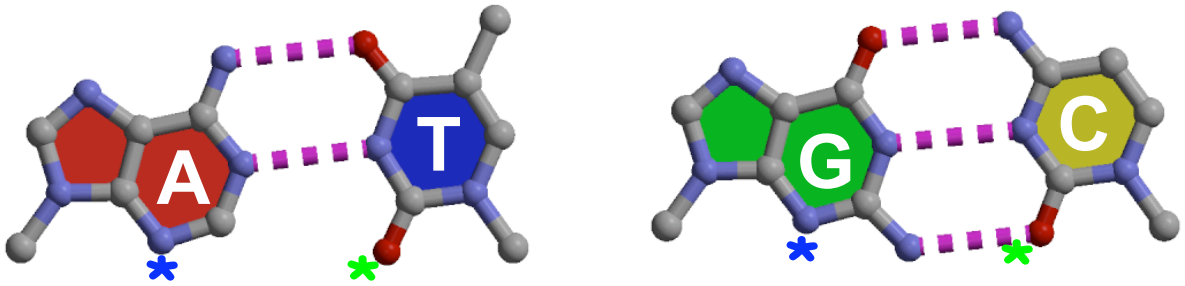
Type	# proteins	# major-groove contacts	# minor-groove contacts
Enzymes	101	1154	320
Regulatory proteins	121	1710	560
Structural proteins	16	20	113
Multi-functional proteins	1	0	3
Total	239	2884	996

Number of close (≤ 3.4 Å) contacts between protein and DNA atoms in 239 protein-DNA complexes



Whereas the bases present unique hydrogen-bond recognition patterns in the major groove, there is a common hydrogen-bond acceptor pattern of Watson-Crick base pairs in the minor groove (Seeman *et al.*, 1976).

Major-groove edge

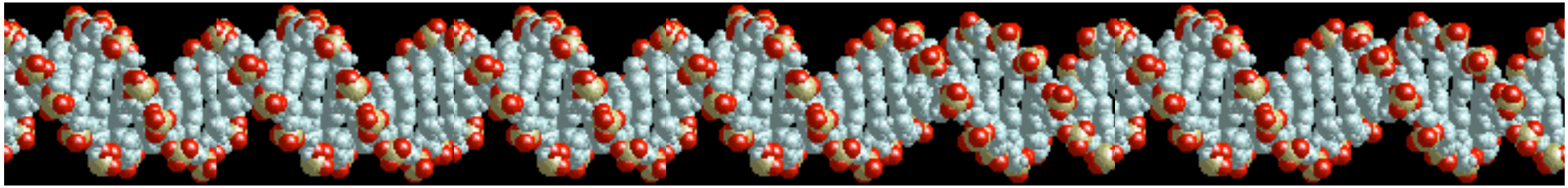


minor-groove edge

pseudo-symmetrically positioned (blue, green) hydrogen-bond acceptor atoms

Proteins contact DNA preferentially through both base and phosphate atoms using atoms on cationic or polar side groups.

		Protein residues			
DNA	Sample size	Arg, Lys +	Asp, Glu −	Ala, Ile, Leu, Met, Phe, Pro, Val	Asn, Cys, Gln, His, Ser, Thr, Trp, Tyr
Base	8240	**1958**	231	274	**1417**
Phosphate	7352	**2411**	175	289	**2237**
Sugar	8240	986	87	267	1023



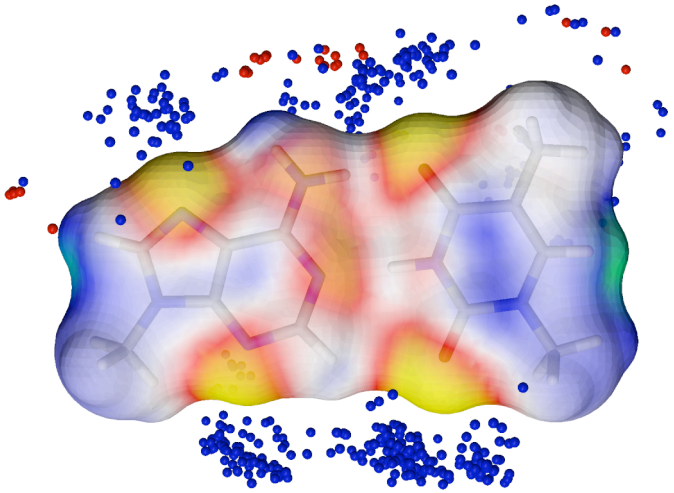
Number of close (≤ 3.4 Å) contacts between protein and DNA atoms in 239 protein-DNA complexes

Close interatomic contacts in protein-DNA complexes show base sequence-dependent binding preferences.

DNA fragment	Sample size	Amino acid atom type			
		Cationic	Anionic	Nonpolar	Polar
Base	8240	1958	231	274	1417
A	**2234**	**275**	32	90	**510**
T	**2234**	**408**	20	**141**	**376**
G	**1886**	**1074**	4	26	**353**
C	**1886**	201	**175**	17	178
Phosphate	7352	2411	175	289	2237
Sugar	8240	986	87	267	1023
Total	23832	5355	493	830	4677

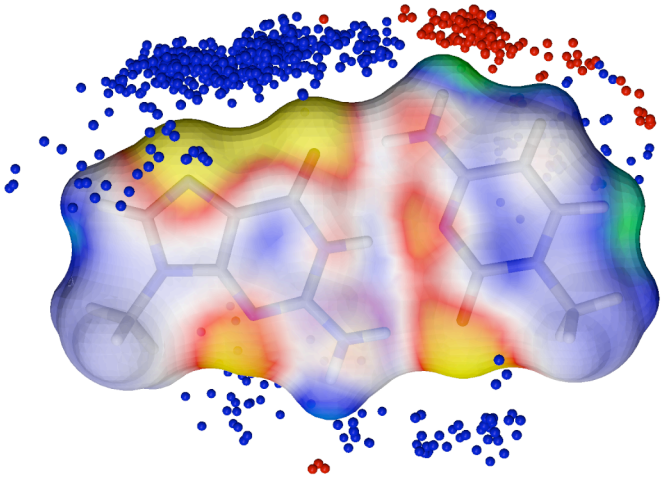
Number of close (≤ 3.4 Å) contacts between protein and DNA atoms in 239 protein-DNA complexes

"Ions" found in well-resolved protein-DNA structures exhibit a sequence-dependent, asymmetric build up on the surface of the base pairs.



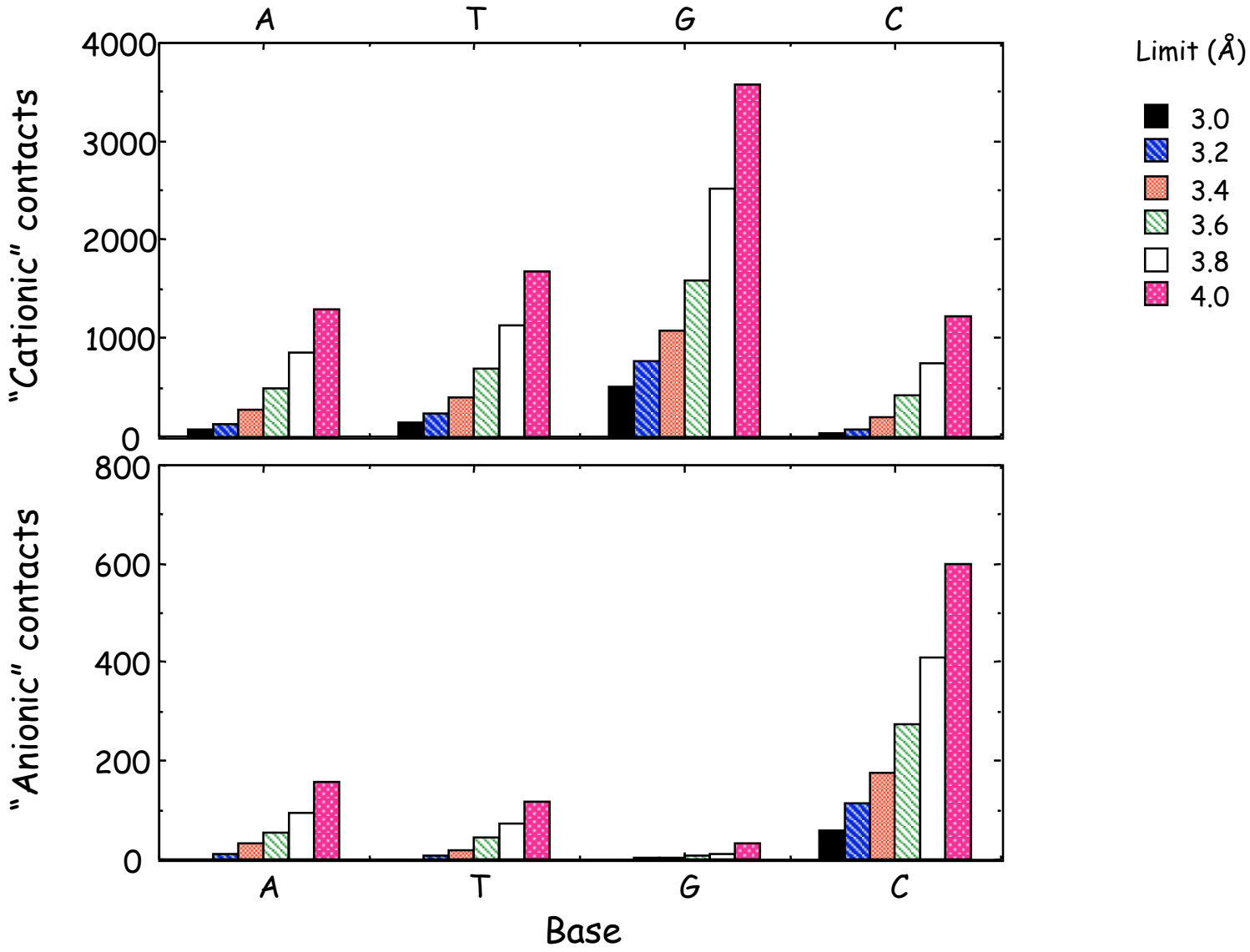
A·T minor groove edge

G·C major groove edge



Arg/Lys_N+
Asp/Glu_O–

Sequence-dependent accumulation of "ions" on the edges of the DNA bases is independent of assumed contact limit.



Close "anionic" contacts to cytosine occur in diverse protein-DNA complexes.

Enzymes (100)

DNA endonucleases	78
Recombinases	10
DNA methyltransferases	6
DNA lyases	3
DNA polymerases	2
Excisionases	1

Structural proteins (4)

Centromere proteins	3
HMG-box domains	1

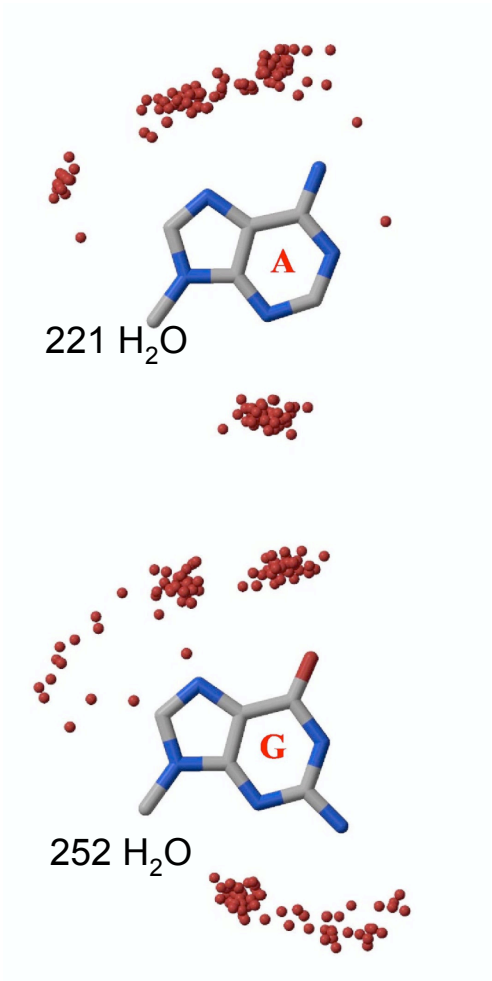
helix	180
turn	59
coil	31
sheet	20
other	3

Regulatory proteins (189)

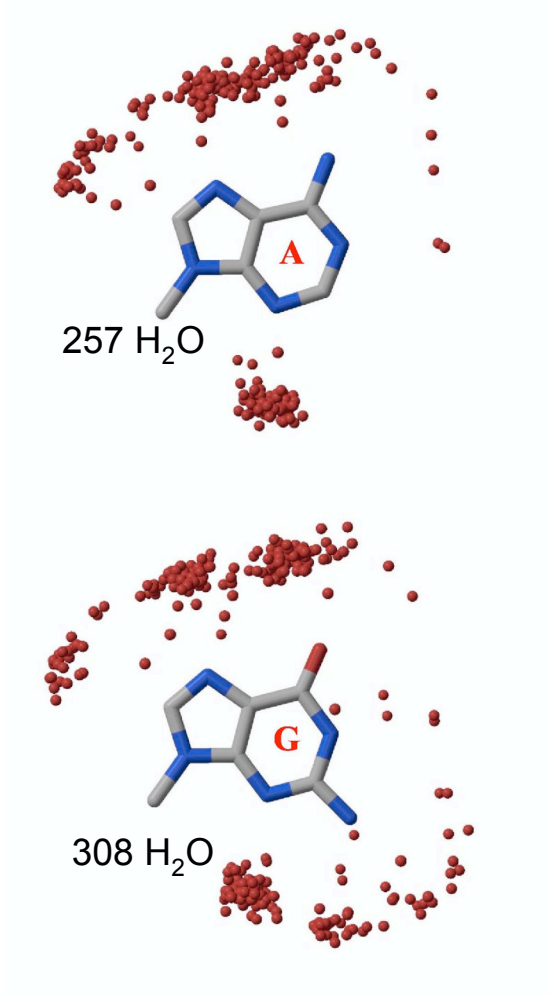
Zinc fingers	48
Helix-loop-helix domains	26
Glucocorticoid receptor-like domains	25
Nucleic acid-binding three-helical bundles	20
λ repressor-like domains	12
Immunoglobulin-like β-sandwich	9
TraR-like domains	8
Leucine zipper domains	8
Trp repressor-like domains	6
Spo0A-like domains	3
Zn_2Cys_6 domains	2
σ_4-like domains	2
Ferredoxin-like domains	2
Zinc finger + leucine zipper	11
Three-helical bundle + leucine zipper	6
σ_4 + λ repressor	1

Direct contacts of amino-acid atoms with the DNA bases mimic the distributions of bound H_2O molecules in well resolved structures

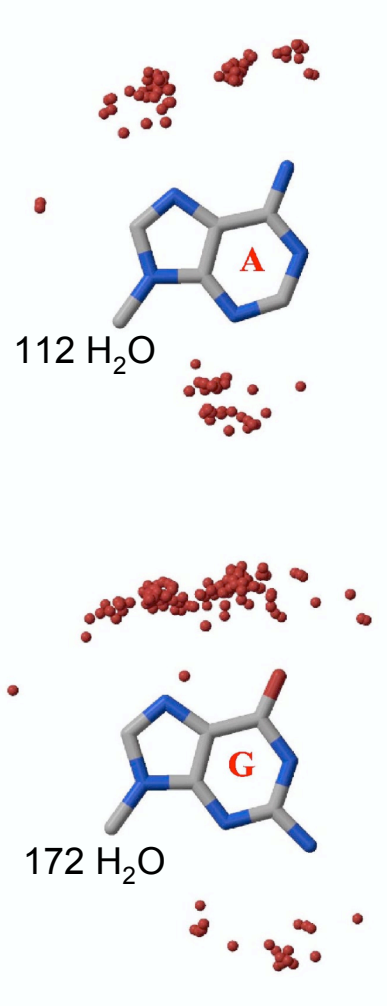
(H_2O from 27 B DNA; H_2O and hydrogen-bond donor/acceptor atoms from 27 P+DNA structures)



H_2O·B DNA

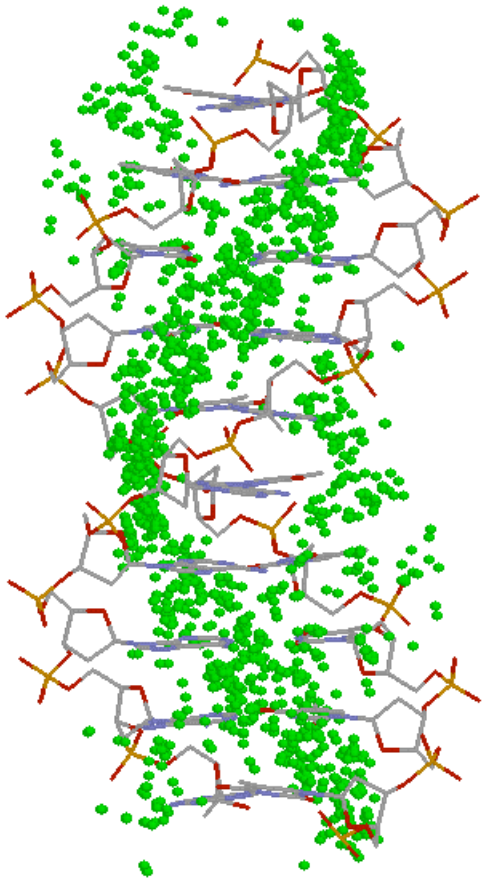


Protein·H_2O·DNA

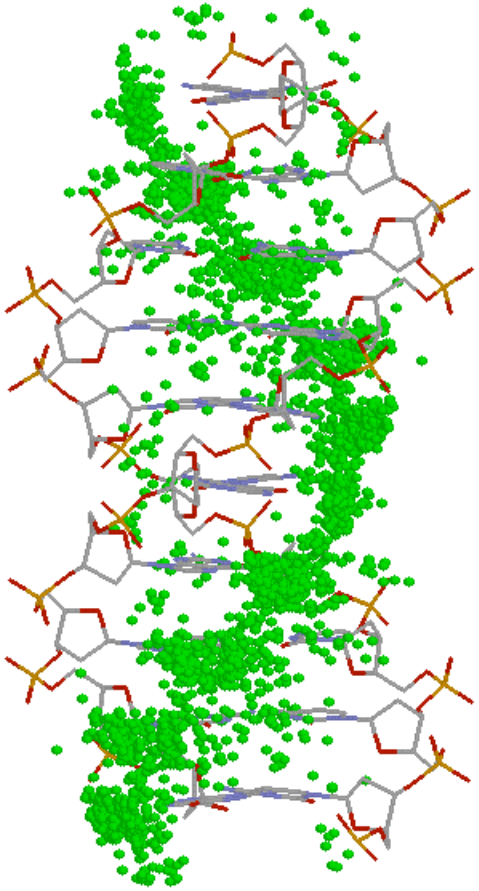


Protein·DNA

Cationic binding sites superimposed on ideal *B*-DNA mimic known sequence-specific build-up of metal ions in minor/major grooves.



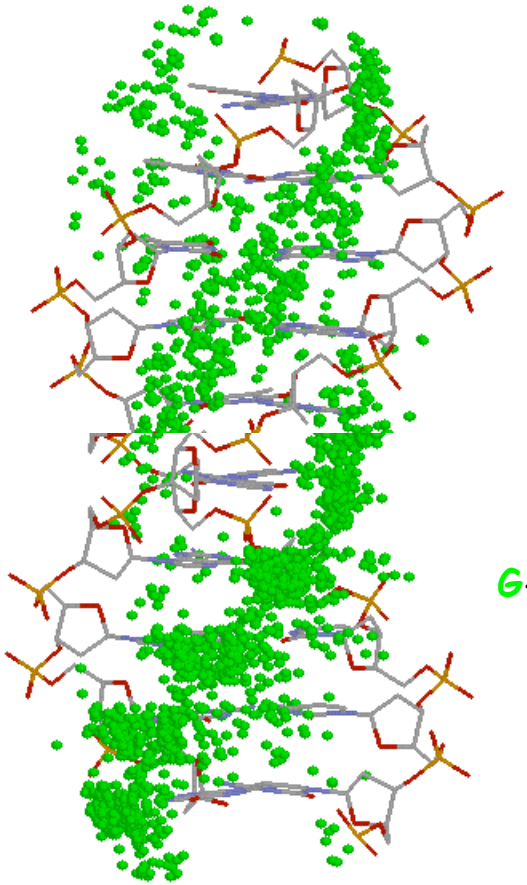
poly d*A*·poly dT



poly d*G*·poly d*C*

Because charged species build up on different edges of A·T vs. G·C base pairs, cations may accumulate on one face of the double helix for specific sequences, *e.g.*, poly dA_5G_5·poly dC_5T_5.

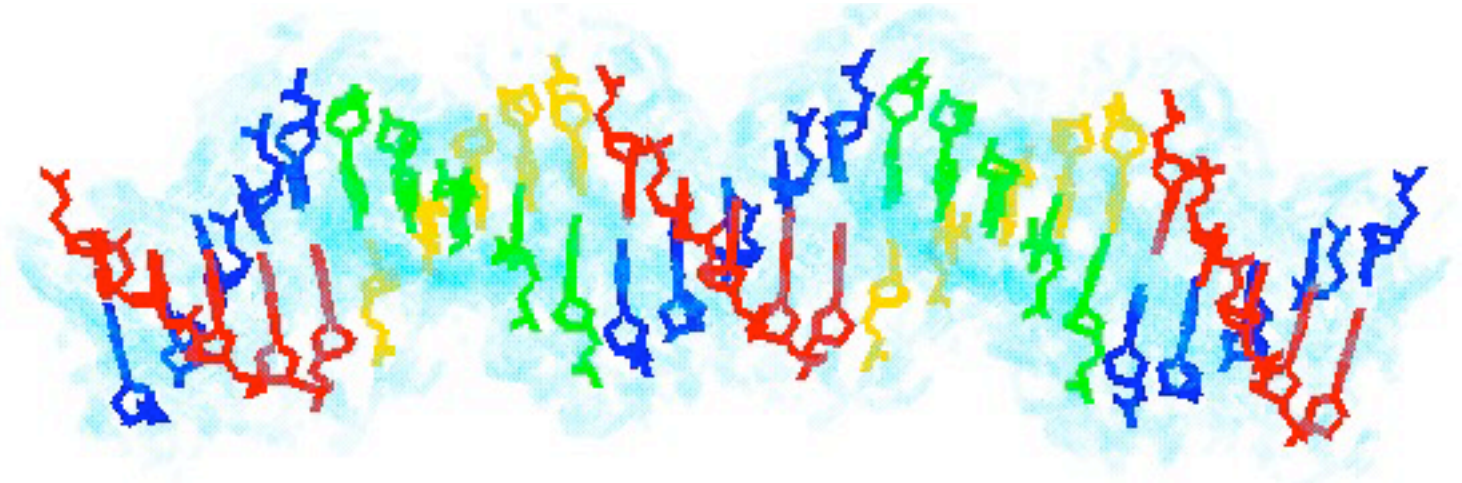
A·T minor-groove



G·C major-groove

poly dA_5G_5·poly dC_5T_5

DNA sequences with excess charge on one face of the double helix will adopt perturbed helical structures.



- The neutralization of phosphates by a localized excess of cationic charge curves the DNA.
- The curvature, in turn, facilitates the wrapping of DNA on proteins and the long-range recognition of specific sequences.

Atomic-level representation of the naturally curved DNA sequence and surrounding ion cloud

---> AAAAACCCCCAAAAACCCCCAAAAA --->

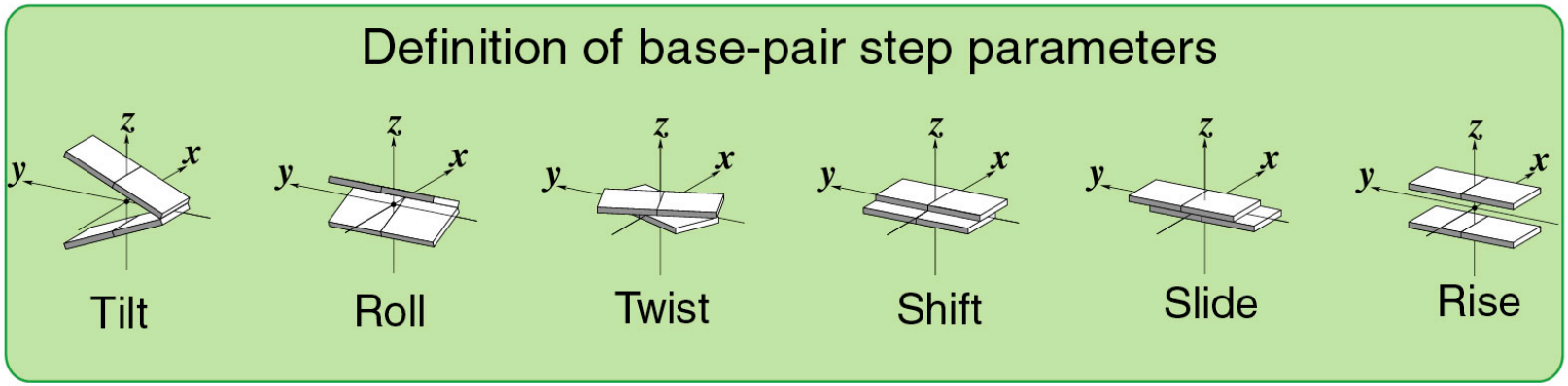
<--- TTTTTGGGGGTTTTTGGGGGTTTTT <---

What is the cost of DNA deformation on the nucleosome?

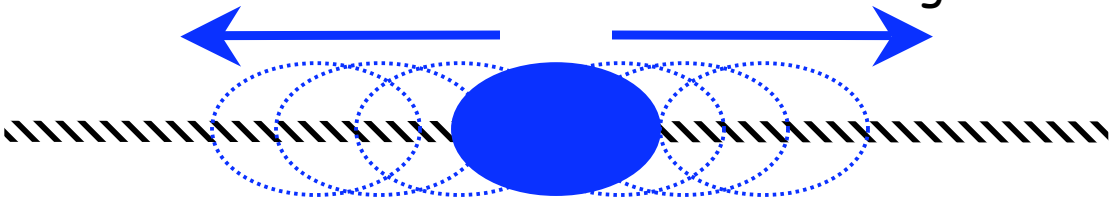
Knowledge-based potentials can be used to estimate the cost of (sequence-dependent) DNA deformation on the nucleosome.

The minimum-energy rest states of a given XZ dimer are equated to the mean values of the step parameters in other high-resolution (non-nucleosomal) structures.

$$\theta_1^{XZ} = \left\langle \theta_1^{XZ} \right\rangle$$



The cost of "threading" DNA on the nucleosome scaffold is computed as the sum of dimeric deformation energies V^{XZ}.



$$V^{XZ} = \sum_{i=1}^{6} \sum_{j=1}^{6} f_{ij}^{XZ} \left(\theta_i^{XZ} - \left\langle \theta_i^{XZ} \right\rangle \right) \left(\theta_j^{XZ} - \left\langle \theta_j^{XZ} \right\rangle \right)$$

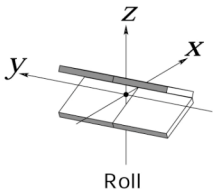
$$i, j = 1 - 6 \ (\text{Tilt}, \text{Roll}, \text{Twist}, \ \text{Shift}, \ \text{Slide}, \ \text{Rise})$$

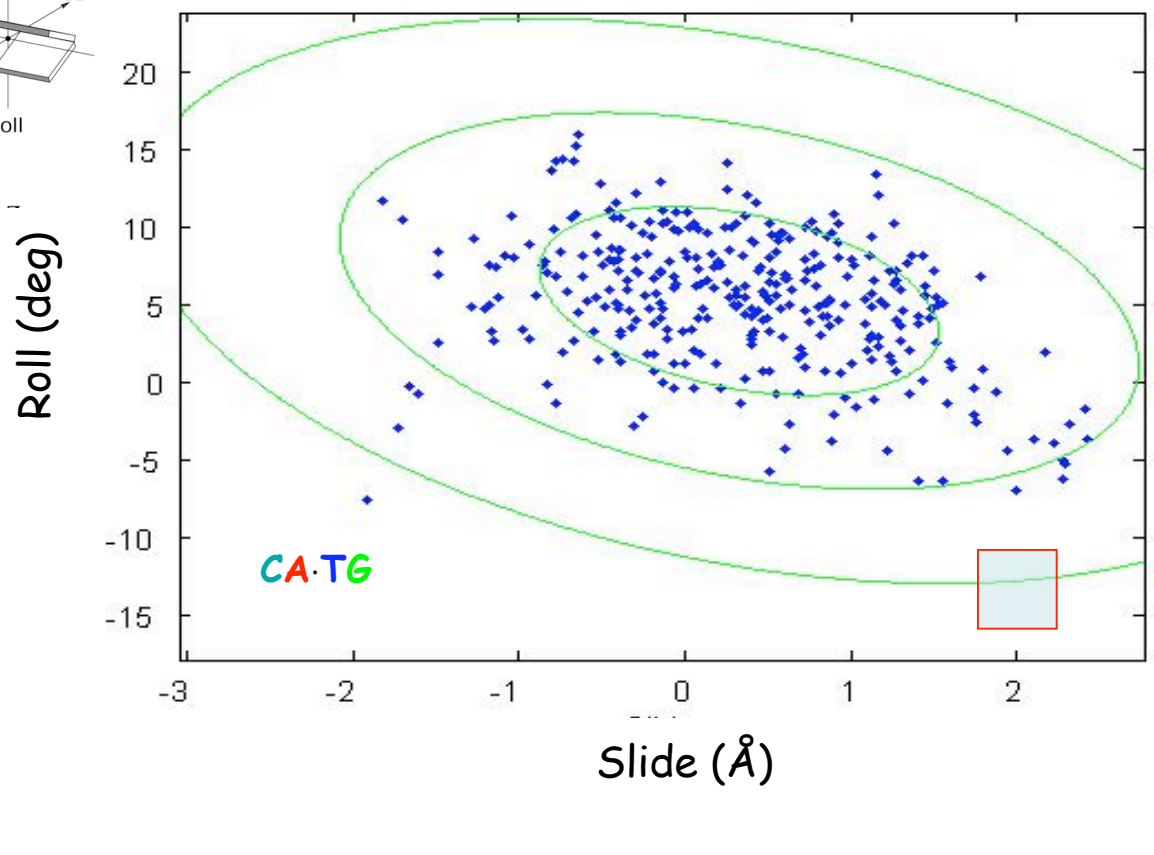
The rest state is defined by the mean sequence-dependent step parameters.

The deformed state adopts the step parameters of DNA on the nucleosome.

Different DNA sequences are aligned over the structural motif.

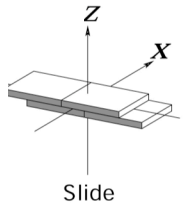
Pyrimidine-purine steps take up large deformations in Roll as well as Slide, including the extreme conformational distortions of nucleosomal DNA.‡





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

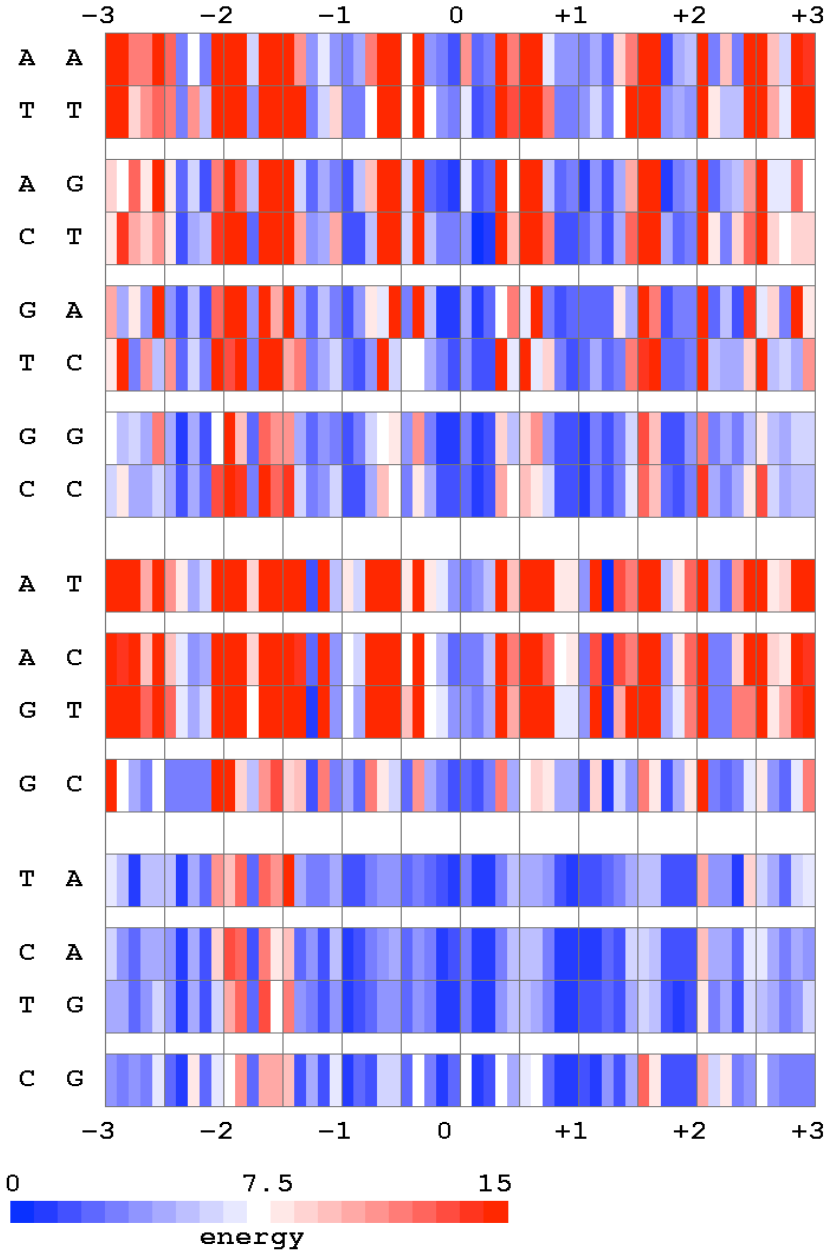
Nucleosomal DNA deformational extremes



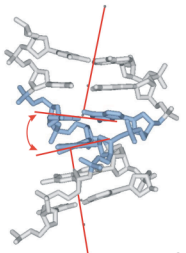
‡Scatter plots and derived potentials of 341 CA·TG dimer steps taken from 239 protein-DNA crystal complexes of 2.5 Å or better resolution filtered to exclude over-represented structures

"Cost" of threading base-pair steps at any position on the DNA pathway around the nucleosome core particle is lowest for pyrimidine-purine steps

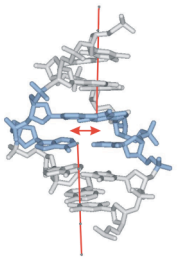
Pathway of the central 60 bp of DNA in contact with H3 and H4 proteins on either side of the nucleosomal dyad (PDB_ID pd0287).



Barriers occur at dimer steps where Roll and Slide exhibit large, concerted deformations.



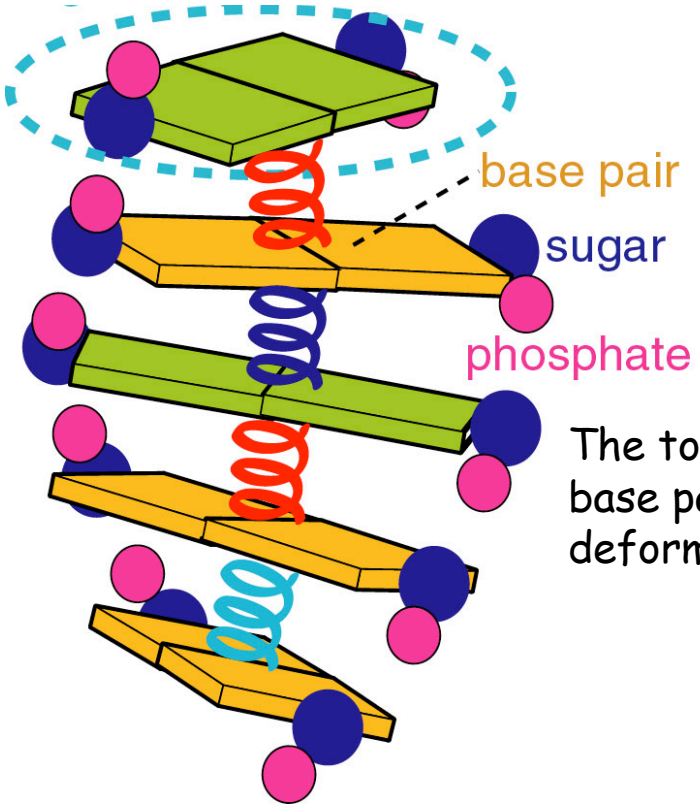
Roll < −20°



Slide > 2 Å


1 CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAA
2 TGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAA
3 GGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAAC
4 GAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACG
5 AGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGC
6 GAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCA
7 AATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCAC
8 ATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACG
9 TCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGT
10 CCCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTA
11 CCGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTAC
12 CGGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACG
13 GGTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGC
14 GTGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCG
15 TGCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGC
16 GCCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCT
17 CCGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTG
18 CGAGGCCGCTCAATTGGTCGTAGCAAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGT

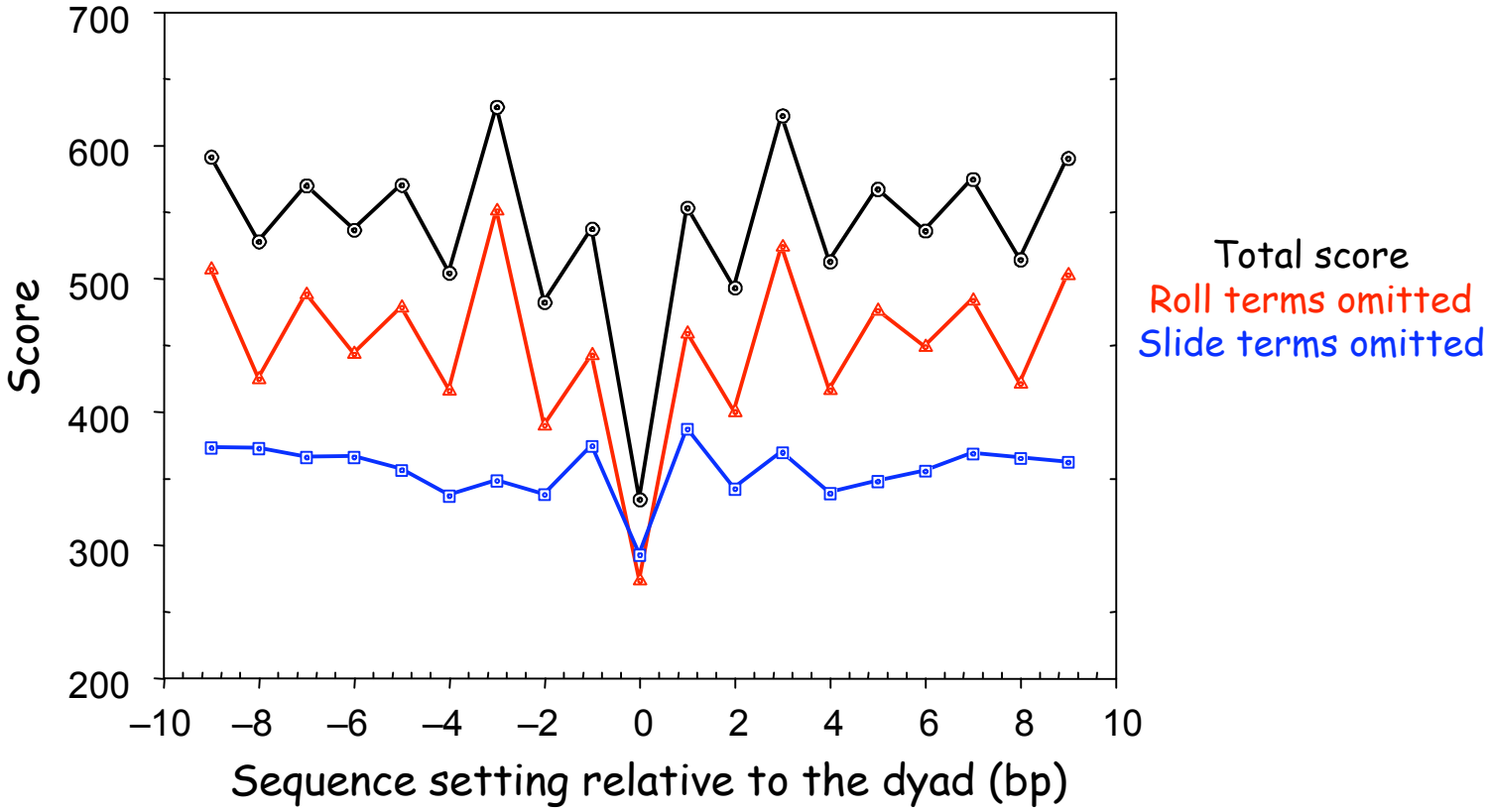
DNA threading score



The total cost Ψ of threading a DNA sequence of N base pairs on the nucleosome is the sum of the deformation scores of all N base-pair steps:

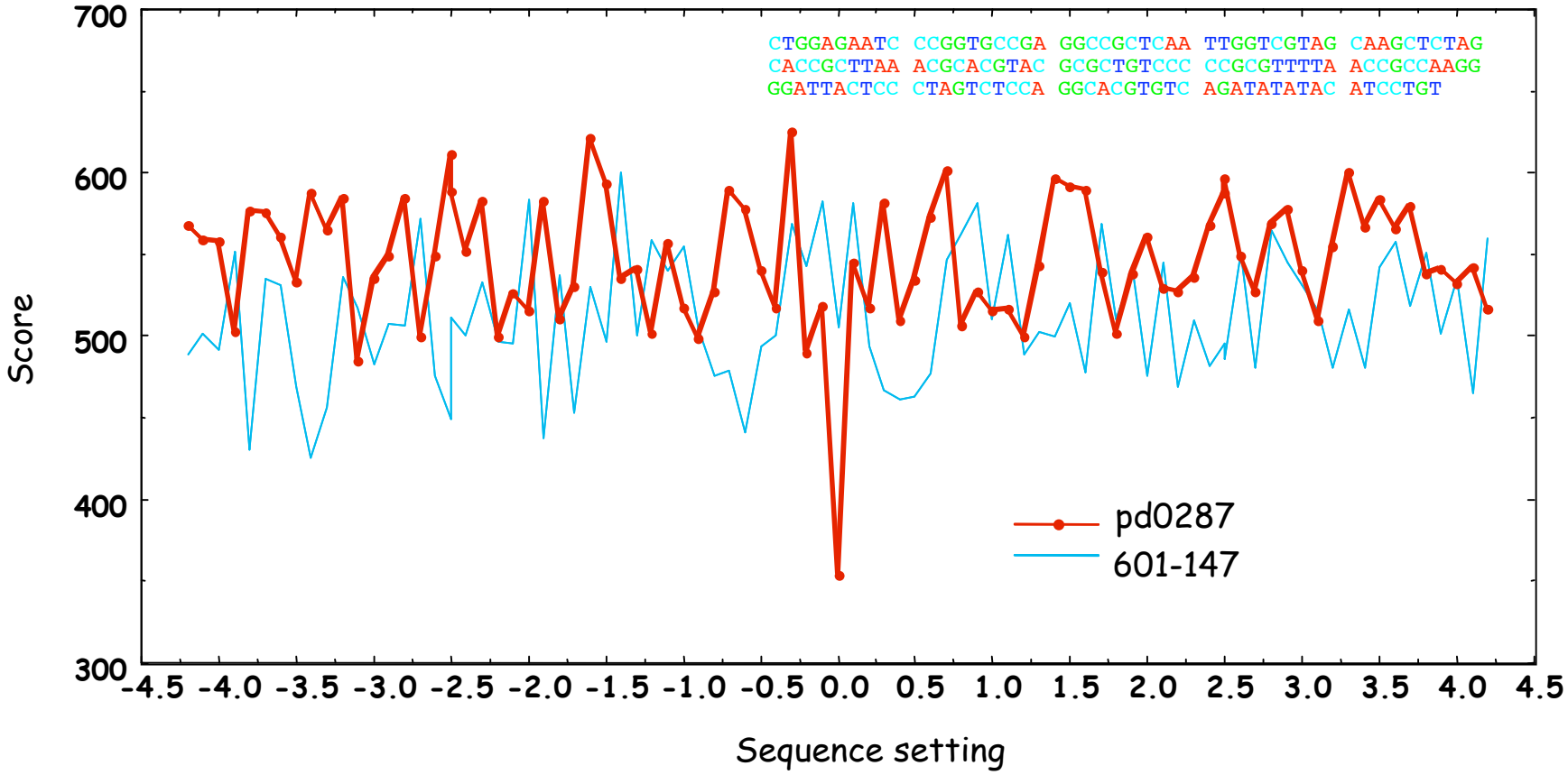
$$\Psi = \sum_{n=1}^{N} V_n$$

Energy of threading the crystallized nucleosome sequence on the currently best resolved core particle structure (NDB_ID: pd0287).



DNA structural template shortened to 129 bp by removing the straight, 9 bp pieces at either end of the superhelical pathway (129 = 147 – 9×2). This allows for 19 different settings of the 147 bp sequence on the structural template (19 = 1 + 9×2). The shift of one relative to the other is reported along the abscissa.

The threading energies of the currently best positioning sequence on the central 60 bp of the nucleosome scaffold are generally lower but the setting is more mobile than the energy and setting of the crystallized sequence.



Calculation assumes that folding pattern is fixed and independent of sequence;
601-147 sequence courtesy of Jonathan Widom

Cyclization of short DNA fragments

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

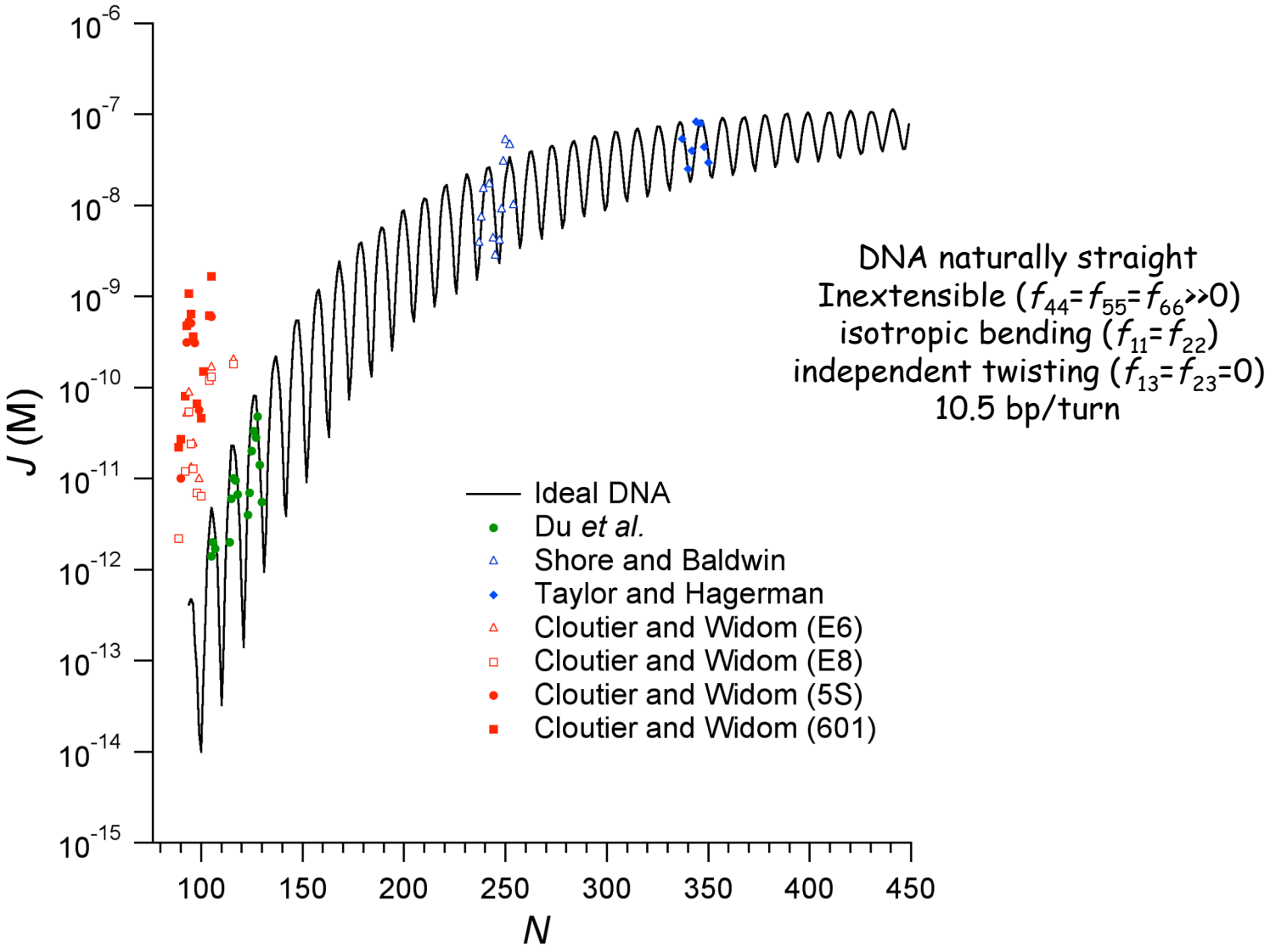
$$J = K_1 / K_2$$

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

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

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.

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.



The DNA molecules found experimentally 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.

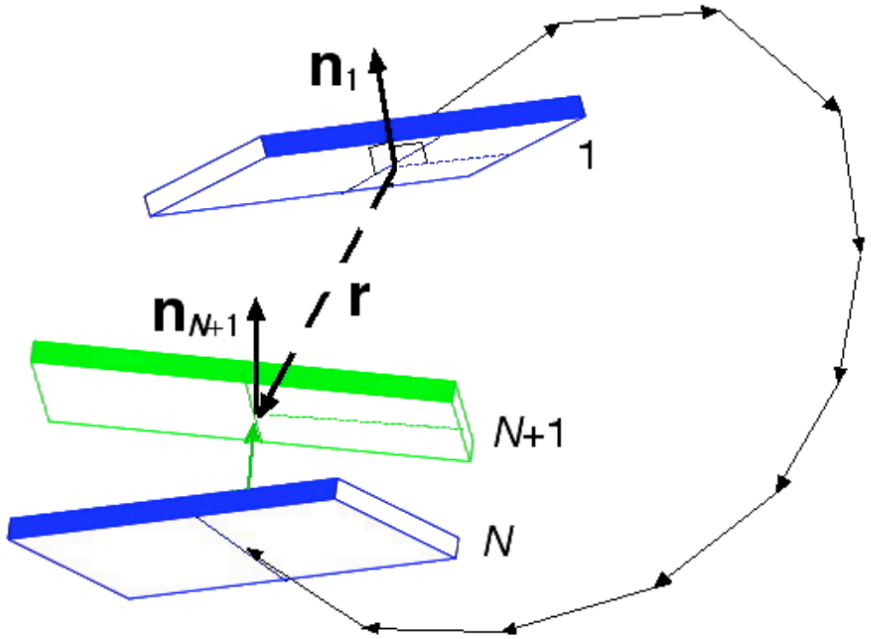
ggccgggtcg **TA**gC**A**agctc **TA**gC**A**ccgct
TAaacgC**A**cg **TA**cgcgc**TG**t c**TA**ccgcgtt
t**TA**accgc**CA** a**TA**ggat**TA**c t**TA**c**TA**gtct
c**TA**c

$J_{obs} = 1.07 \times 10^{-9} \pm 2.3 \times 10^{-11}$

$J_{calc} = 2.7 \times 10^{-12}$

Cloutier & Widom

The likelihood of DNA cyclization can be estimated from the number of simulated configurations of a linear molecule in which terminal residues are positioned so as to insure successful ligation.

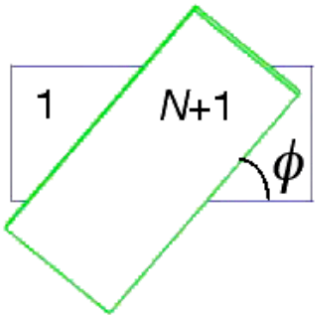


The hypothetical base pair N+1 overlaps the first base pair in a perfectly closed DNA of N base pairs.

$\mathbf{r} = \mathbf{0}$

$\cos\gamma = 1$

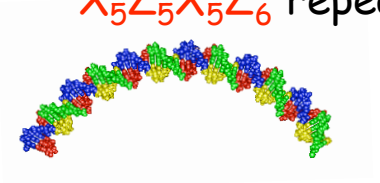
$\phi = 0$



Introduction of intrinsic curvature enhances the likelihood of cyclization but does not account for the observed oscillations in the J factor.

N	Intrinsic curvature					J_{obs}
	210	420	630	840	Straight	601-TA
95	8.42×10^{-6}	3.83×10^{-9}	1.52×10^{-10}	4.48×10^{-11}	4.75×10^{-12}	6.39×10^{-10}
100	3.21×10^{-10}	1.62×10^{-12}	3.94×10^{-14}	2.74×10^{-14}	3.25×10^{-14}	4.57×10^{-11}
105	3.77×10^{-5}	2.37×10^{-8}	1.18×10^{-9}	2.18×10^{-10}	2.31×10^{-11}	1.66×10^{-9}

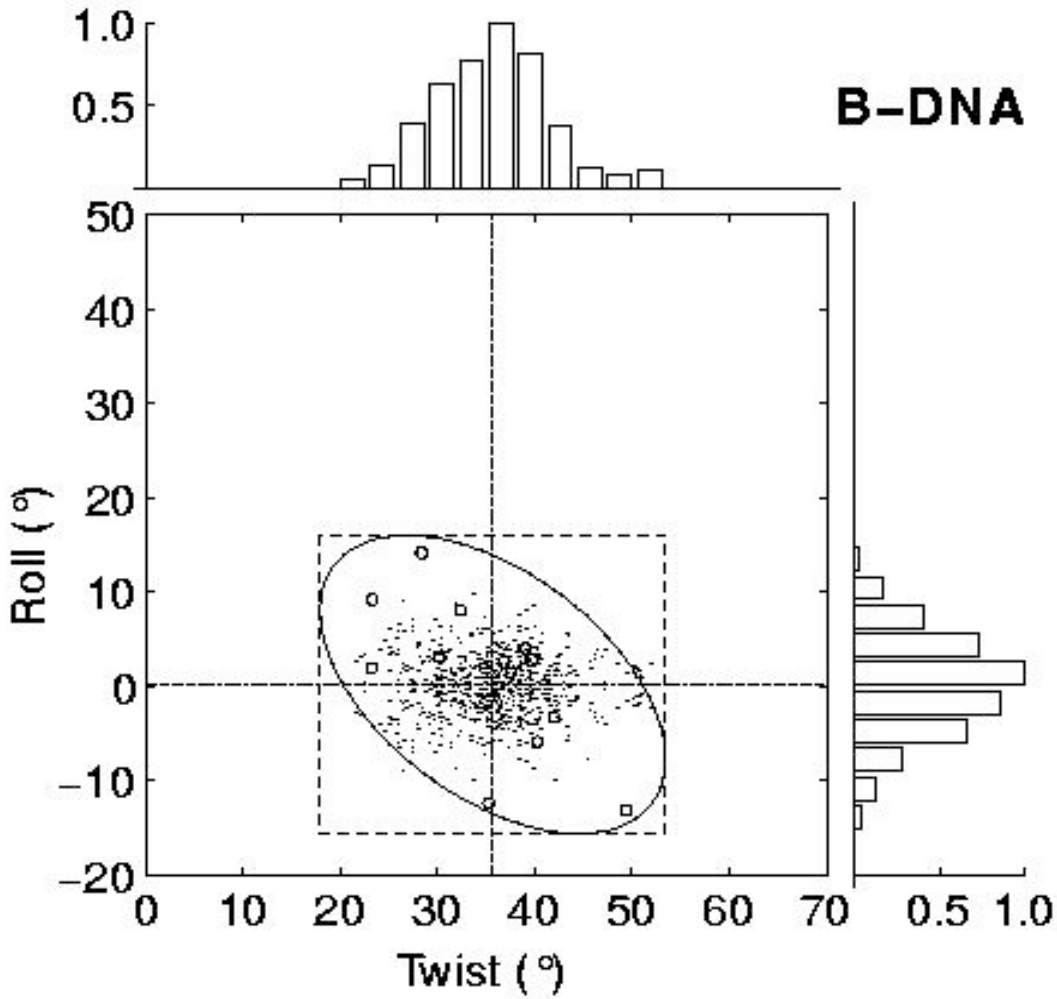
Intrinsic curvature measured in terms of the number of base pairs of an idealized $X_5Z_5X_5Z_6$ repeating sequence designed to form an O-ring of the chain length N.



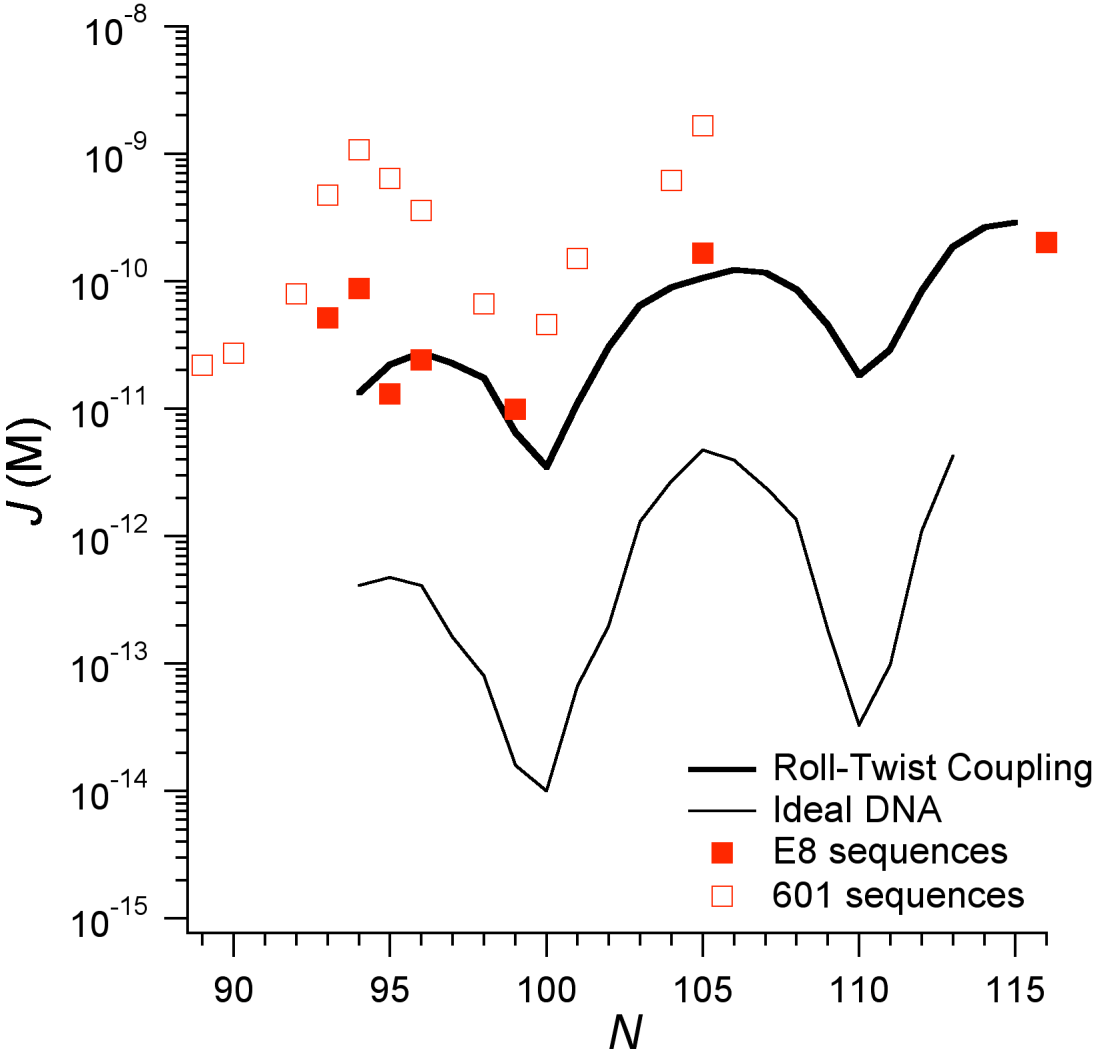
DNA model: $f_{44}=f_{55}=f_{66}>>0$, $f_{11}=f_{22}$, $f_{13}=f_{23}=0$, 10.5 bp/turn

Roll and Twist angles are intrinsically ***coupled*** in DNA structures

Gaussian distribution of observed values forms basis of "knowledge-based" elastic potential.

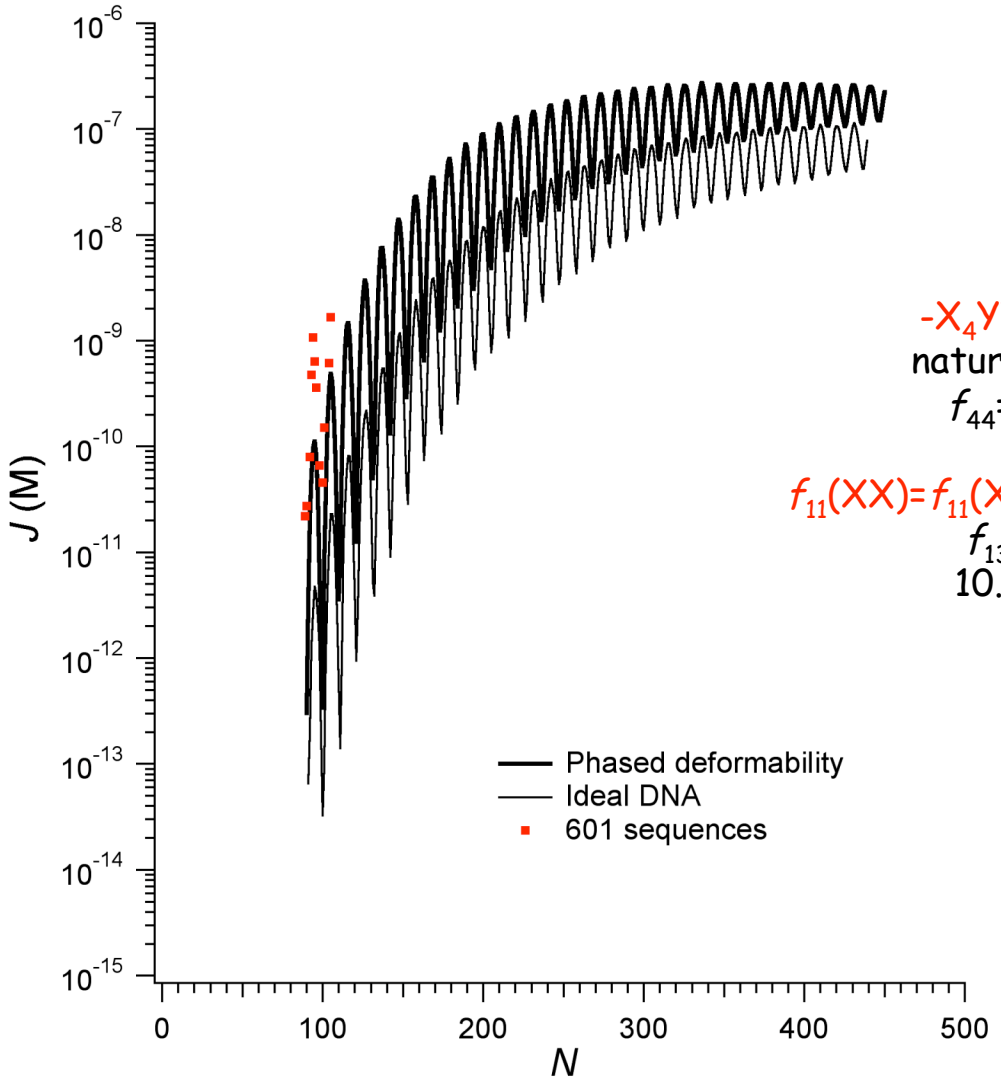


The introduction of phased Roll-Twist coupling in ideal DNA increases the *J* factor and concomitantly reduces the amplitude of oscillations in log *J* with *N*.



-$X_5Y_5X_5Y_6$- repeat
naturally straight
$f_{44}=f_{55}=f_{66}>>0$
$f_{11}=f_{22}$
$f_{13}=0$, $f_{23}\neq 0$
10.5 bp/turn

The softening of DNA with force constants consistent with the deformability and placement of TA steps in nucleosome-positioning sequences increases the ease of ring closure but does not dampen the observed oscillations in the *J* factor .



-X_4YYX_4- repeat
naturally straight
$f_{44}=f_{55}=f_{66}>>0$
$f_{11}=f_{22}$
$f_{11}(XX)=f_{11}(XY)=f_{11}(YX)=4f_{11}(YY)$
$f_{13}=0$, $f_{23}=0$
10.5 bp/turn

New findings and open questions

A sequence-dependent build-up of "ions" around the DNA base pairs appears to underlie its intrinsic structural properties.

Slide determines the pitch and positions DNA on the crystalline nucleosome.

Sequence-dependent properties of base-pair steps appear to facilitate ring closure of short DNA sequences.

Do other sequences fold in the same way as the single DNA sequence crystallized on the nucleosome?

How reliable is the sequence-dependent elastic representation of DNA dimeric structure?