NUCLEIC ACIDS

Are responsible for information storage and transfer and they direct the synthesis of proteins

High Accuracy in the information transfer

- Self-replication
- Self-repair

DNA (Deoxyribonucleic acid)

RNA (Ribonucleic acid)

DNAs have molecular weights between 10MD and 10000MD

Generally (but not always) the amount of DNA in a system increases with system complexity.

DNA is a double stranded linear helical molecule

RNA exists in several forms

- tRNA Transfer RNA 25kD
- mRNA messager RNA 100-150kD
- rRNA ribosomal RNA 40-1600kD

To store information we need an alphabet and a dictionary

Source	# of bases	Length	MW	Information content
Polyoma virus	4600	1.6 mm	3 MD	9×10^3
T2 Phage	185000	63 mm	122 MD	$3.7\ 10^5$
E Coli	3.4×10^6	1.2 mm	2.3 GD	$6.8\ 10^6$
Drosophila	6×10^7	2.1 cm	43 GD	$1.2\ 10^8$
Human cell	3×10^9	1 m	2 TD	$6\ 10^9$

The special aspect is that the **geometry** of the AT and GC pairs, measured from their point of attachment to the sugar moiety, is identical.

Two more features to be considered

1. Base pairing

2. Base stacking

In model compounds, other (than Watson&Crick) schemes for base pairing are found.

The preference between the W&C pairing appears to be an energetic one

Base-pair	K (M^{-1})	ΔH (Kcal/M)	-ΔS (cal/deg.M)	-ΔG (Kcal/M)
U-U	6.1	4.3	11.0	1.1
A-A	3.1	4.0	11.4	0.7
A-U	**100**	**6.2**	**11.8**	**2.7**
C-C	28	6.3	15	2.0
G-G	10^3-10^4	8.5-10	15	4.1
G-C	**10^4-10^5**	**10-11.5**	**15**	**5.5**

The problem of specificity of pairing

Thermodynamic data on base pairing do not show a very high specificity for AT and CG pairs. Other pairs are possible.

The specificity is much greater in macromolecules

Some pairs such as A-U can form more than one complex

Base stacking

Polyadenylic acid at low temperature shows a structure in which bases are stacked one on top of the other.

At high temperature this structure is destroyed

Macromolecules tend to aggregate in solution

Stacking is more favorable between

purine<-->purine

than between

pyrimidine<-->pyrimidine

Solvent plays a crucial role in base stacking

Physical origin of the stacking interaction

1. Induced dipole interactions between the pi-electron clouds of the stacked bases

2. Hydrophobic interactions

ROLE OF WATER AND IONS

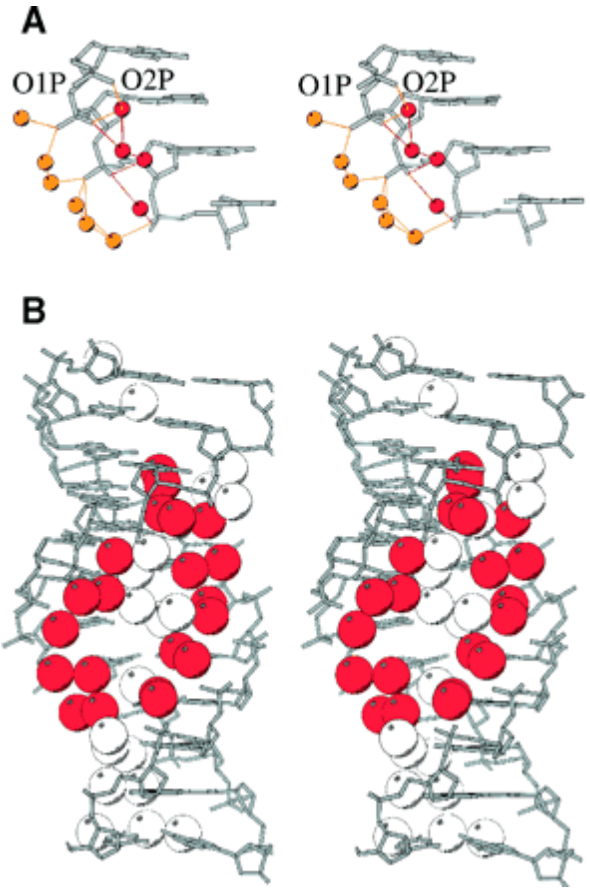
Some open problems

1. Helix-coil transitions in DNA: Dynamics cannot be predicted by current theories (We will see more later) Large number of conformations, loops etc.

2. Dynamic picture of DNA: Double helix
   Ion atmosphere

   High frequency dielectric relaxation

   Water spines

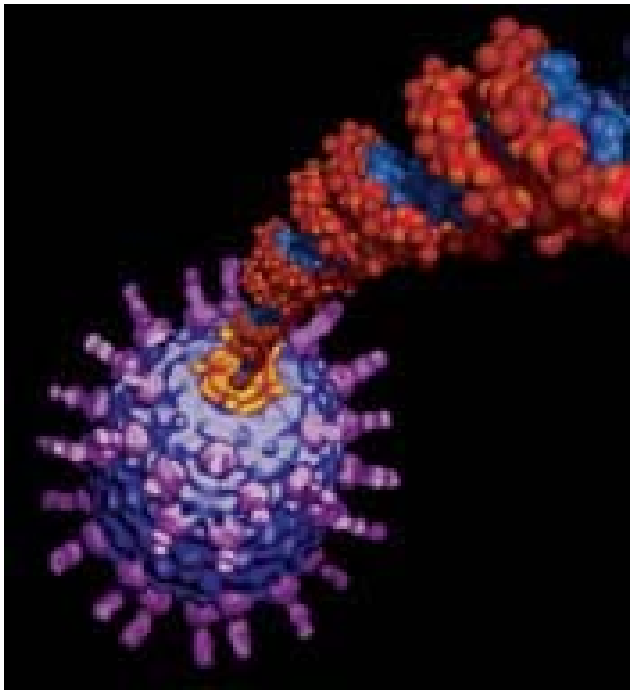


Stereos showing hydration patterns in G3C3. (**A**) Water molecules binding to phosphates. Those H-bonded to the outer O1P are yellow, while those bonded to the inner O2P are red. (**B**) Major groove hydration, with the double spine of water molecules in red, and other ordered water molecules in white.

3. Collapsed form of DNA

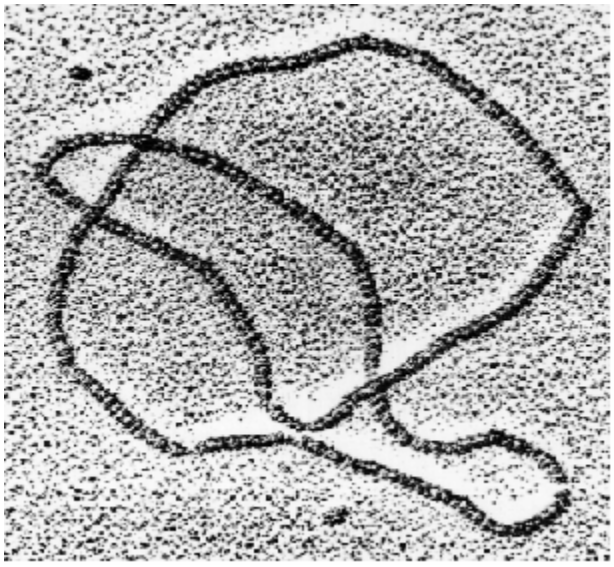
DNA inside viruses

DNA-Protein interactions



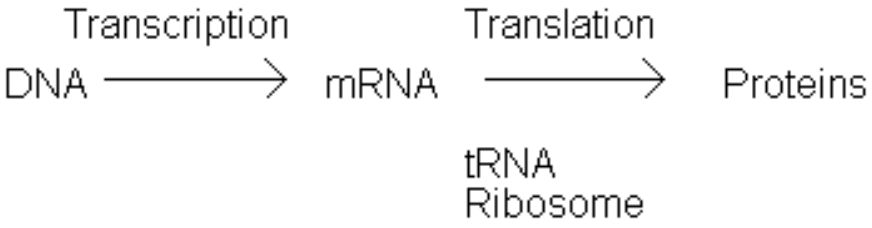
A powerful molecular motor (yellow) translocates the twisted strands of DNA (right) of the Bacillus subtilis bacteriophage ø29 into a protein capsid. By using optical tweezers to pull on the DNA while it is being packed, UC Berkeley and University of Minnesota researchers have measured the force generated by the motor and the packing pressure.
Sander Tans and Doug Smith, based on cryoelectron microscope capsid image from Marc Morais and

Michael Rossmann (Purdue Univ.) and DNA image fromPaul Thiessen (Chemical Graphics)



Nucleic Acids: Biochemical properties

The information contained in DNA is not used directly to control biosynthesis



In **Transcription**, a mRNA molecule is synthesized by RNA-polymerase according to the instructions given by the DNA template. RNA-polymerase is a complex enzyme made of many subunits. One subunit (s) recognizes start signals on the DNA. The binding of RNA-polymerase locally unfolds the DNA. There is also a specific signal to terminate transcription (r-factor).

The gene and the protein are collinear

Mutations are produced by changes in the base sequence of DNA.

The translation is produced by tRNA.

tRNA, unlike other nucleic acids, has a 3-D folded structure. It contains some unusual bases.

Many sequences for t-RNA are known

The sequences folded in a cloverleaf pattern with half of the residues base-paired. T-RNA has a L-shaped configuration.

The codon recognition works only in the anticodon part. The aminoacid does not seem to play an important role.

t-RNA can recognize more than one codon.

The pattern of degeneracy of the genetic code indicates that

XYU

XYC

are equivalent

The genetic code

The genetic code is the relationship between the sequence of the 4 bases in DNA and the sequence of the 20 aminoacids.

The code is universal. All the living organisms use the same code

- A sequence of 3 bases on the DNA (the codon) codes for one aminoacid
- The code is non overlapping
- The code is read sequentially, from a fixed starting point
- There are no commas between words
- The genetic code is degenerate (64 possibilities, 20 aminoacids)
- The genetic code is known

First Position		Second Position of Codon				Third Position
		T	**C**	**A**	**G**	
	T	TTT Phe [F] TTC Phe [F] TTA Leu [L] TTG Leu [L]	TCT Ser [S] TCC Ser [S] TCA Ser [S] TCG Ser [S]	TAT Tyr [Y] TAC Tyr [Y] TAA *Ter* [end] TAG *Ter* [end]	TGT Cys [C] TGC Cys [C] TGA *Ter* [end] TGG Trp [W]	**T** **C** **A** **G**
	C	CTT Leu [L] CTC Leu [L] CTA Leu [L] CTG Leu [L]	CCT Pro [P] CCC Pro [P] CCA Pro [P] CCG Pro [P]	CAT His [H] CAC His [H] CAA Gln [Q] CAG Gln [Q]	CGT Arg [R] CGC Arg [R] CGA Arg [R] CGG Arg [R]	**T** **C** **A** **G**
	A	ATT Ile [I] ATC Ile [I] ATA Ile [I] ATG Met [M]	ACT Thr [T] ACC Thr [T] ACA Thr [T] ACG Thr [T]	AAT Asn [N] AAC Asn [N] AAA Lys [K] AAG Lys [K]	AGT Ser [S] AGC Ser [S] AGA Arg [R] AGG Arg [R]	**T** **C** **A** **G**
	G	GTT Val [V] GTC Val [V] GTA Val [V] GTG Val [V]	GCT Ala [A] GCC Ala [A] GCA Ala [A] GCG Ala [A]	GAT Asp [D] GAC Asp [D] GAA Glu [E] GAG Glu [E]	GGT Gly [G] GGC Gly [G] GGA Gly [G] GGG Gly [G]	**T** **C** **A** **G**

An explanation of the Genetic Code: DNA is a two-stranded molecule. Each strand is a polynucleotide composed of **A** (adenosine), **T** (thymidine), **C** (cytidine), and **G** (guanosine) residues polymerized by "dehydration" synthesis in linear chains with specific sequences. Each strand has polarity, such that the 5'-hydroxyl (or 5'-phospho) group of the first nucleotide begins the strand and the 3'-hydroxyl group of the final nucleotide ends the strand; accordingly, we say that this strand runs 5' to 3' ("*Five prime to three prime*") . It is also essential to know that the two strands of DNA run *antiparallel* such that one strand runs 5' -> 3' while the other one runs 3' -> 5'. At each nucleotide residue along the double-stranded DNA molecule, the

nucleotides are complementary. That is, **A** forms two hydrogen-bonds with **T**; **C** forms three hydrogen bonds with **G**. In most cases the two-stranded, antiparallel, complementary DNA molecule folds to form a helical structure which resembles a spiral staircase. This is the reason why DNA has been referred to as the "Double Helix".

Why is the genetic code degenerate?

1. Minimize the effect of mutations

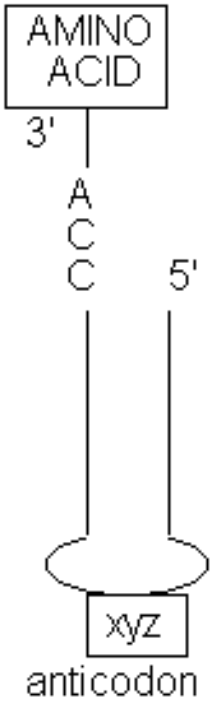
2. Permits to change base composition without altering the code
3. Was the genetic code always the same?

The translation mRNA to proteins requires an adapter molecule

The adapter tRNA contains

- aminoacid attachment site
- template recognition site

The joining of an amino-acid to tRNA to form the aminoacyl-tRNA is catalyzed by a specific enzyme. There are at least 20 different synthetase enzymes



Allowed pairing

Anticodon	**Codon**	
first base	third base	
C	G	specific
A	U	specific
U	A or G	one variation
G	U or C	one variation

I	U,C or A	two variations

Conclusions: The first two bases of a codon pair in the standard way (C-G and A-U)

The degeneracy of the code arises from imprecision in the pairing of the third base of the codon

Modeling the DNA helix

The structural parameters of the **B-DNA** helix in solution are summarized in the figure below:

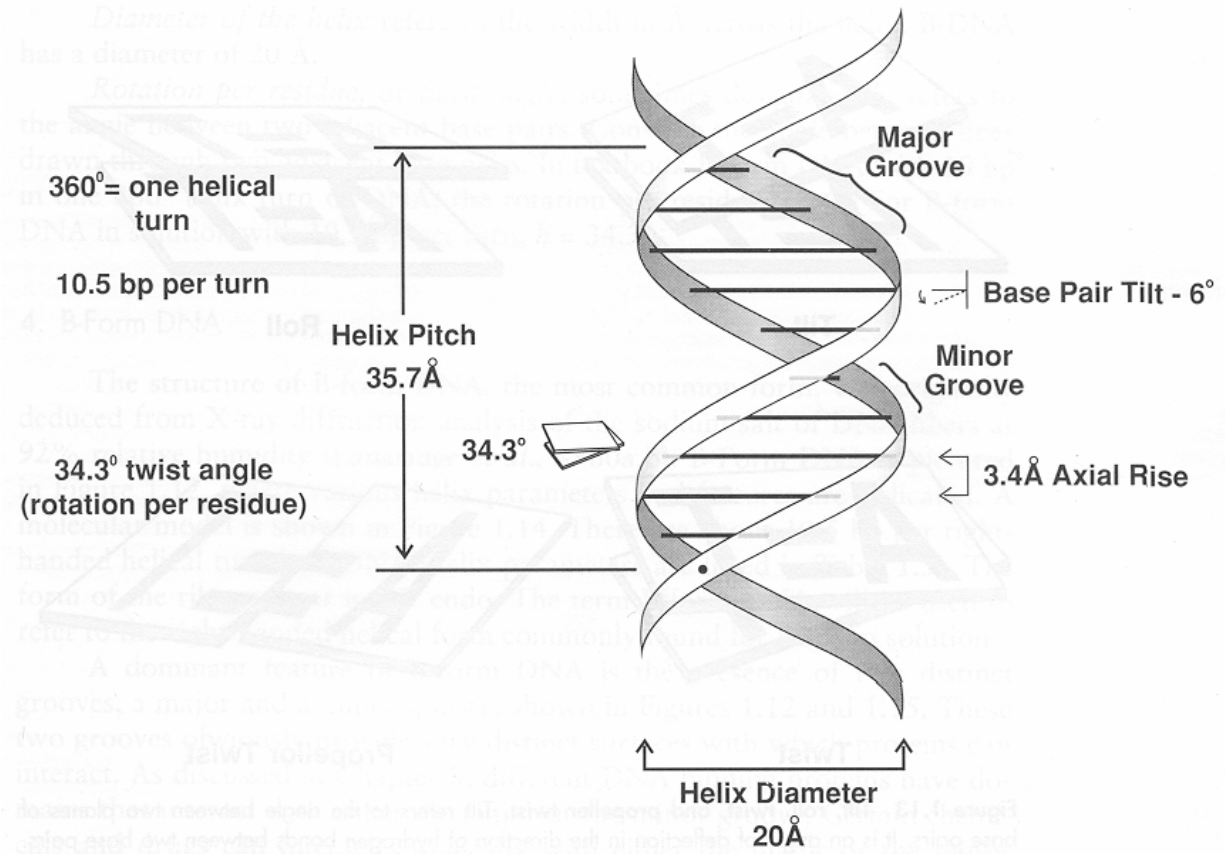


Figure 1.12 The DNA double helix in solution: structural parameters. The Watson–Crick double helix is composed of about 10.5 base pairs per helical turn. Since 360° constitutes one helical turn, there would be a 34.3° twist angle or rotation per residue between adjacent base pairs (see Table 1.4). The helix pitch or length per helical turn is 35.7 Å. The axial rise or distance between two planar base pairs is 3.4 Å. The base pair tilt or deviation from the horizontal plane of the bases is about -6°. The helix diameter or the width of the helix is about 20 Å. Note the positions of the minor groove and the major groove.

From chemistry we know:

- distance between adjacent sugars or phosphates in DNA chain: ~ 6 Å
- thickness of flat part of a DNA base: 3.3 Å

>How can we tuck the insoluble bases into the center of a DNA molecule where they can avoid water and at the same time get rid of the “holes” between successive bases (the exposure of their hydrophobic surfaces to water is energetically unfavorable!)?

One solution would be to form a skewed ladder in which the sugar-phosphate chains tilt to an angle of about 30° from the horizontal. One such a step is shown schematically in the top part of the figure below. However, using molecular models we would see that this leads to many unacceptable contacts between neighboring atoms.

The bases can stack onto each other just as well, without gaps, if they **twist** into the shape of a helix. The two chains climb from the horizontal at the same angle as before (~30°) but now they lay on the surface of a cylinder of diameter 18 Å.

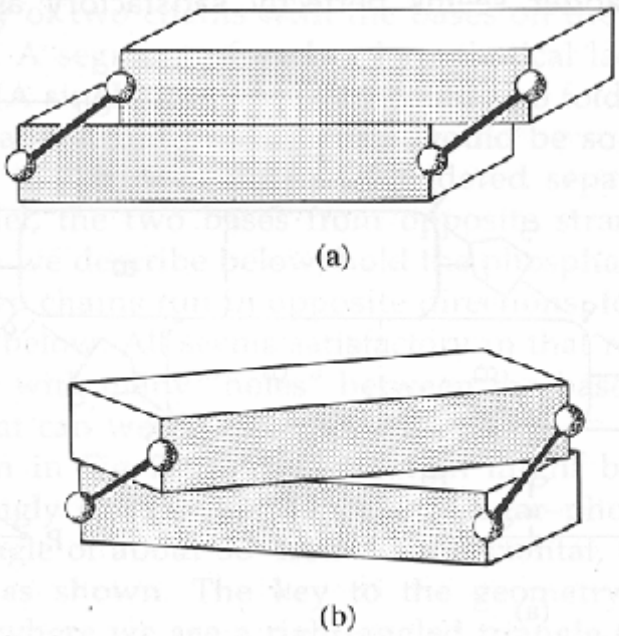


Figure 2.5 (a) Stacking of base-pairs as in the skewed ladder (b) Stacking of base-pairs by means of helical twist.

Schematic representations of a complete helical turn are shown below:

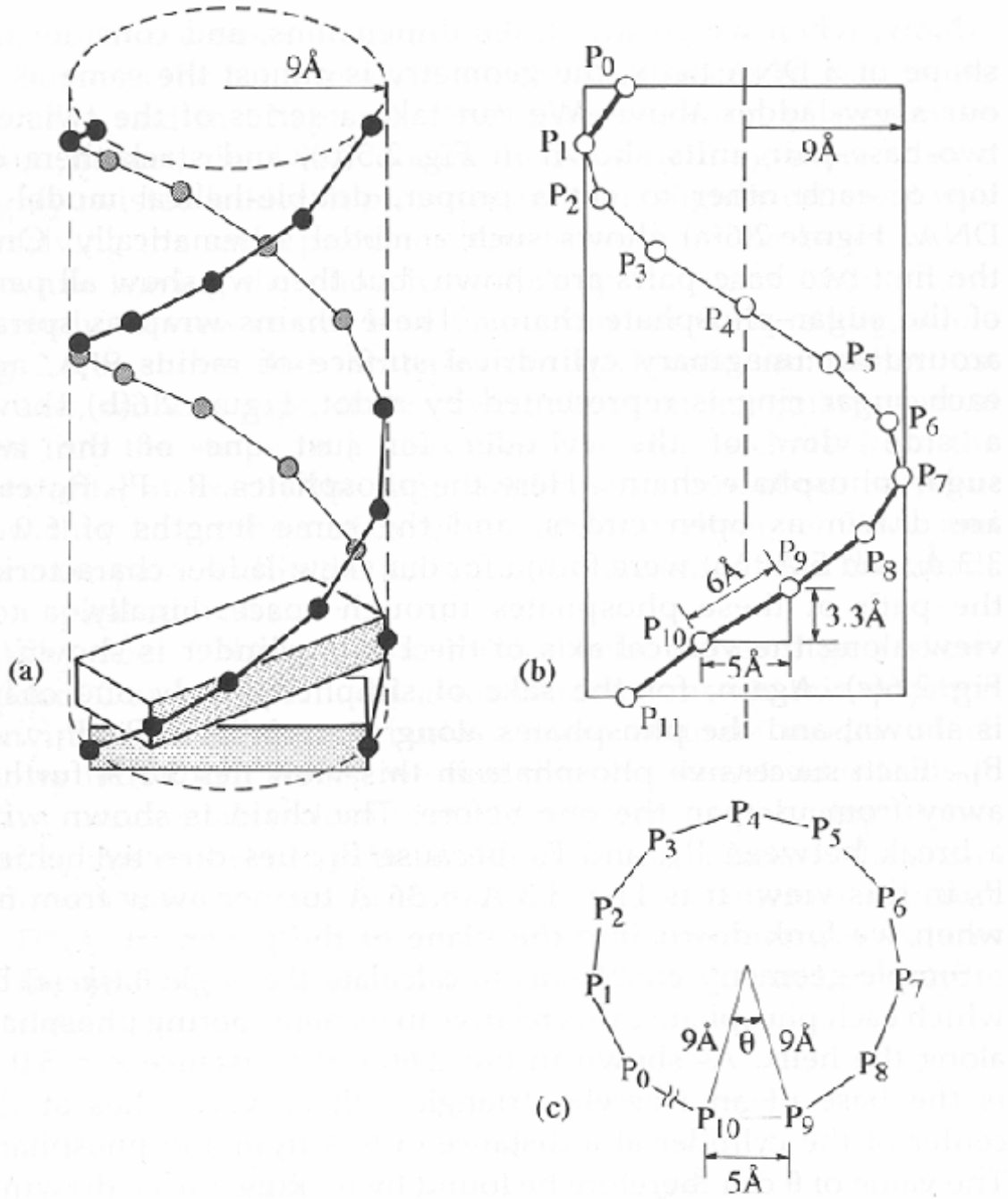


Figure 2.6 Sugar–phosphate chains wrapped helically around a cylinder: three views. In (a), sugar rings are drawn as shaded or filled circles, while phosphates are thin lines. In (b), phosphates are drawn as open circles, while sugars are thin lines. In (c), the view is down the long axis of the cylinder, looking along the dashed line in part (b).

From part (c) in the figure we can calculate the angle made by each phosphate-to-phosphate rotation:

θ = 2*Arcsin(2.5/9) = 32.3° (about 1/11 part of circle)

This simple calculation agrees pretty well with experiment.

Almost all DNA double helices have between 10 and 12 phosphates per turn of the helix, e.g. A-form DNA has 11 phosphates per turn, B-form DNA has 10, and Z-DNA has 12.

The sense of the rotation is right-handed in A- and B-DNA, left-handed in Z-DNA. The base pairs are mostly of a Watson-Crick kind, joining guanine (G) with cytosine (C) and adenine (A) with thymine (T).

The overall size of each base pair is about the same. Therefore it was sometimes thought that the base sequence of DNA could not influence its three-dimensional structure. This has been proven incorrect. E.g. in high resolution X-ray structures the base-pair twist angles range from 28° to 40° about a mean of 34°. The mean value is close to the angle we predicted in our modeling exercise, but the large variation was not predicted.

Looking closer it can be seen that a 32°-twist angle leaves four overhanging portions of the base surfaces unprotected from water:

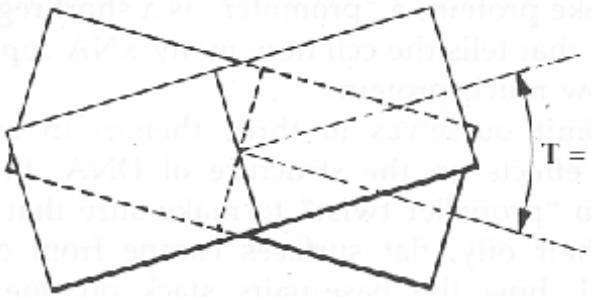


Figure 3.1 Two base-pairs with 32° of right-handed helical twist: the minor-groove edges are drawn with heavy shading, as in Fig. 3.5.

Rotating each of the bases *within the same strand* about its long axis improves the stacking and excludes water from a larger fraction of the surfaces of the two bases:

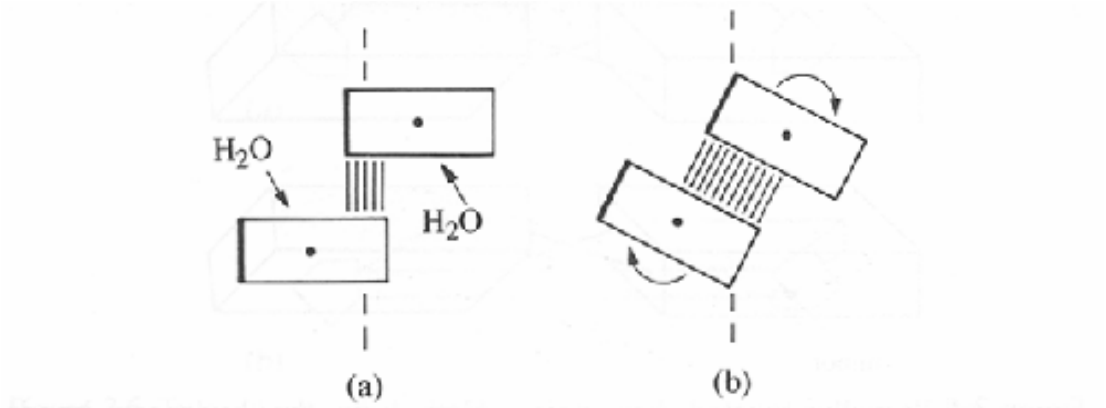


Figure 3.3 Propeller twist, as in (b), allows greater overlap of bases within the same strand and reduces the area of contact between the bases and water.

The two bases of the other strand exclude water by rotating in the opposite direction. This leads to the **propeller twist**. Typically 10° to 20° propeller twist is seen in DNA structures. The propeller twist distorts the hydrogen bonds that hold the two bases together.

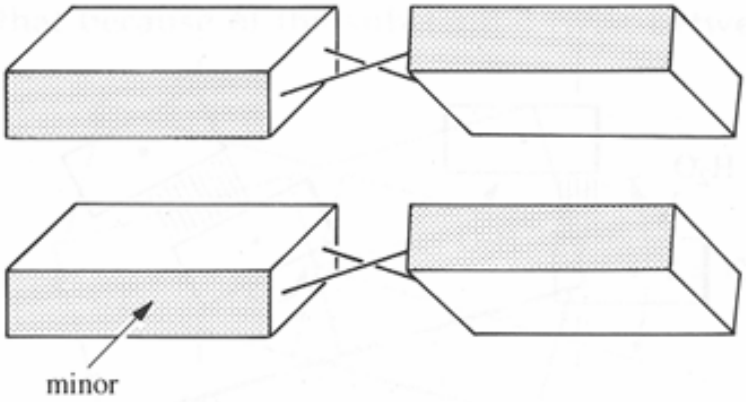


Figure 3.4 Propeller-twisted base-pairs. Note how the hydrogen bonds between bases are distorted by this motion, yet remain intact. The minor-groove edges of the bases are shaded.

Some DNA sequences have unusually high propeller twist, e.g. DNA regions with all adenines in one strand and all thymines in the other. A propeller twist of 20-30° in this sequence allows hydrogen bonds between adjacent base pairs, from adenine N6-H in one strand to thymine O4 in the other strand and in the neighboring base pair:

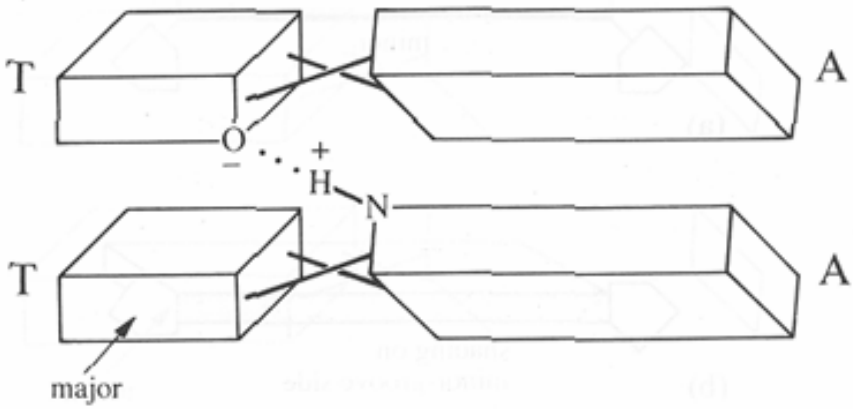


Figure 3.6 Propeller-twisted A–T pairs, showing an additional hydrogen bond between the base-pairs in the major groove (as proposed originally by Hillary Nelson).

Propeller twist is a property of a single base pair. The most important motions of a *base pair relative to its neighbor* are the **twist**, **roll** and **slide**. The positive sense of these three motions is shown schematically below. For simplicity, the propeller twist in each base pair is not shown.

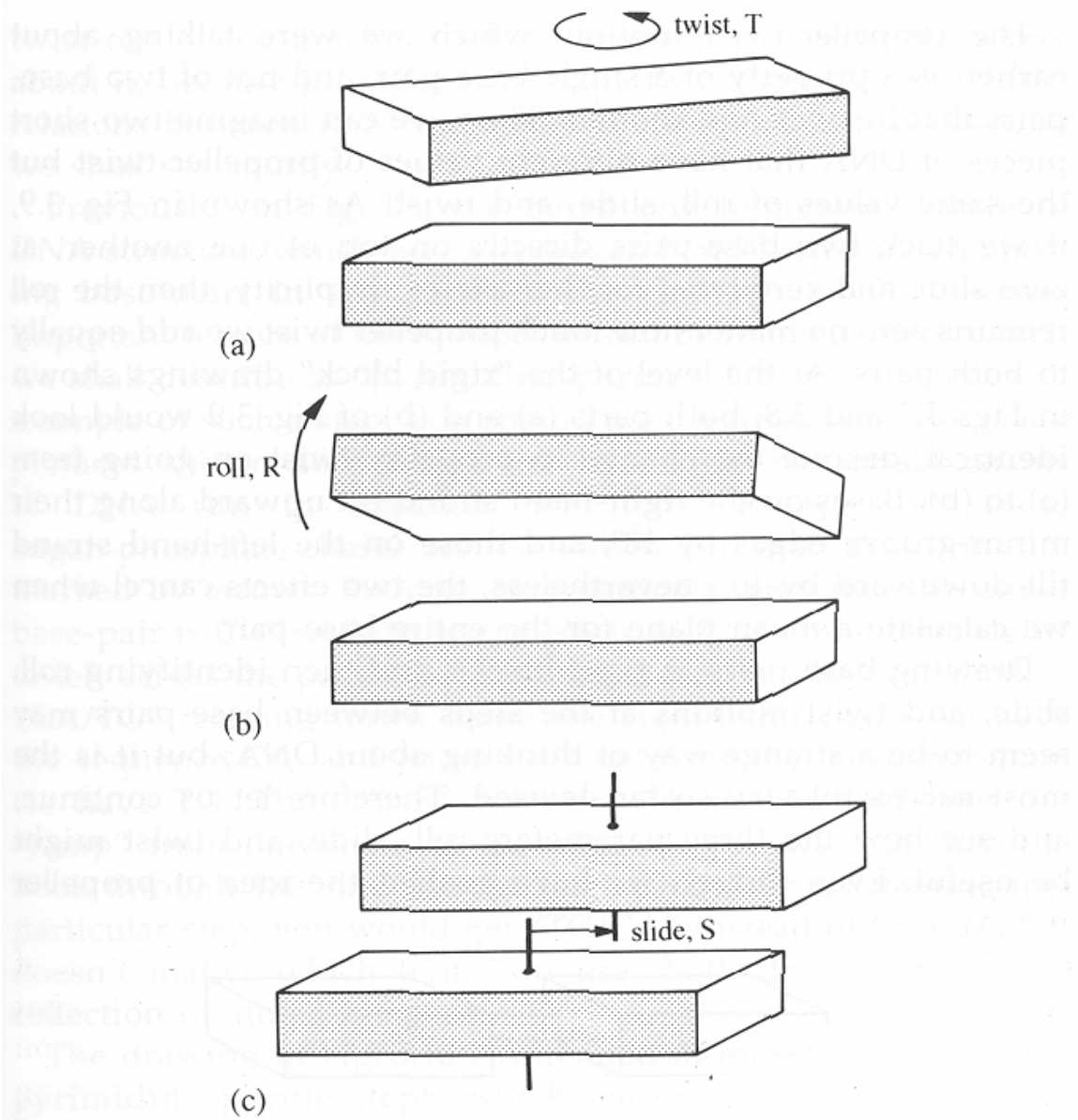


Figure 3.8 Twist, roll, and slide motions at a base-pair step. Each drawing defines the positive sense of twist, roll, or slide, as used in this book.

- The **twist** corresponds to the rotation about the local twist axis that runs vertically through, or near, the centers of any two neighboring base pairs. Typical values of base pair twist in real DNA molecules as determined from high-resolution structures range from 28° to 40° about a mean of 34°.
- The **roll** describes the rolling-open of base pairs along their long axes. Angles of roll vary from +20° to -10° in the usual DNA structures. The roll is positive if base pairs open up towards the minor groove side.
- The **slide** is a translational motion describing the relative sliding of neighboring base pairs along their long axes. Slide is defined as positive if the upper pair goes further to the right than the lower pair when looking at the minor groove edges. Typically slide values range from +2Å to -1Å in real DNA structures; the sugar-phosphate backbone does not allow more motion.

The values of roll, twist and slide are related to each other in ways that depend on the base composition of the step. Consider e.g. a pyrimidine-purine step, e.g. CA/TG (5'-CA-3' in one strand, 5'-TG-3' in the other strand running in opposite direction). Two different kinds of configurations are found for these steps in DNA structures. The large purine bases can slide away from each other by -1Å in order to avoid a steric clash (present as a result of the propeller twist in the base pairs!):

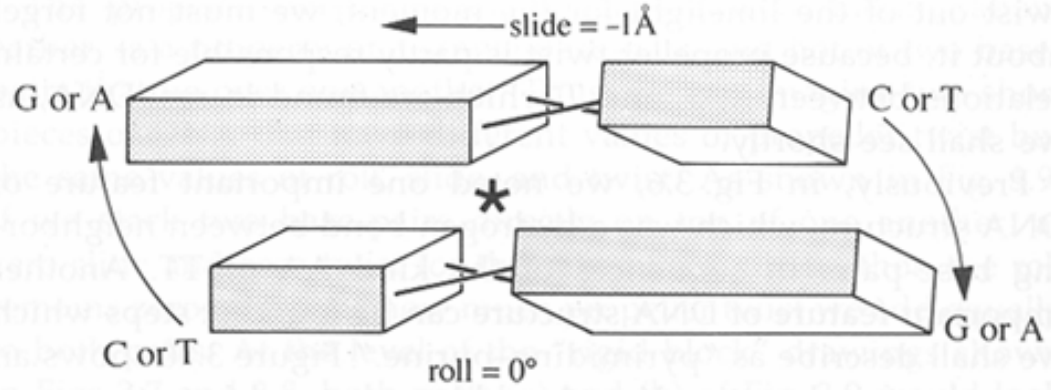


Figure 3.10 A pyrimidine–purine step with zero roll: negative slide is needed to avoid a steric clash at *, if the base-pairs have propeller twist.

The purine bases can also slide on top of one another by +2Å. To maintain a 20°- propeller twist in the base pairs, the roll angle between them becomes large and positive, +20°.

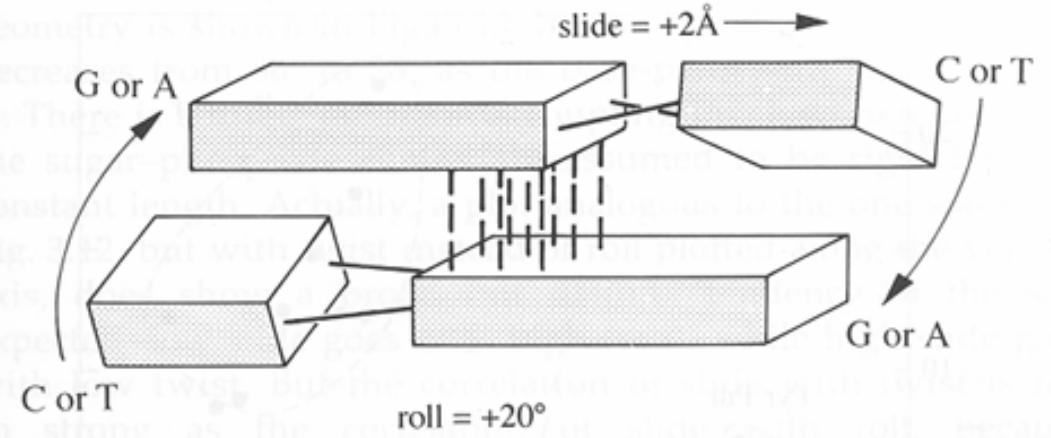


Figure 3.11 A pyrimidine–purine step in an alternative configuration, having positive roll and slide, due to the cross-chain stacking of purines.

Intermediate values of slide (between –1 and +2 Å) can be expected to be less stable because the large purines will neither avoid each other fully, nor stack fully on top of one another. The change of the roll angle as we change the slide (slide –1 Å gives roll=0° while slide +2 Å gives roll=20°) is a direct consequence of the propeller twist.

Plotting the roll and slide values for many base pair steps determined by X-ray crystallography shows a broad but sloping band going form low roll-low slide to high roll-high slide. A plot of the twist angle against the slide value also shows a broad band, going from low slide-high twist to high slide-low twist. The latter interdependence results from the connection of the base pairs to the fairly rigid sugar-phosphate backbone. Different values of roll (R), slide (S) and twist (T) generate different kinds of double helix. Let's look what happens when the same values of roll, slide and twist are repeated over 11 base pairs:

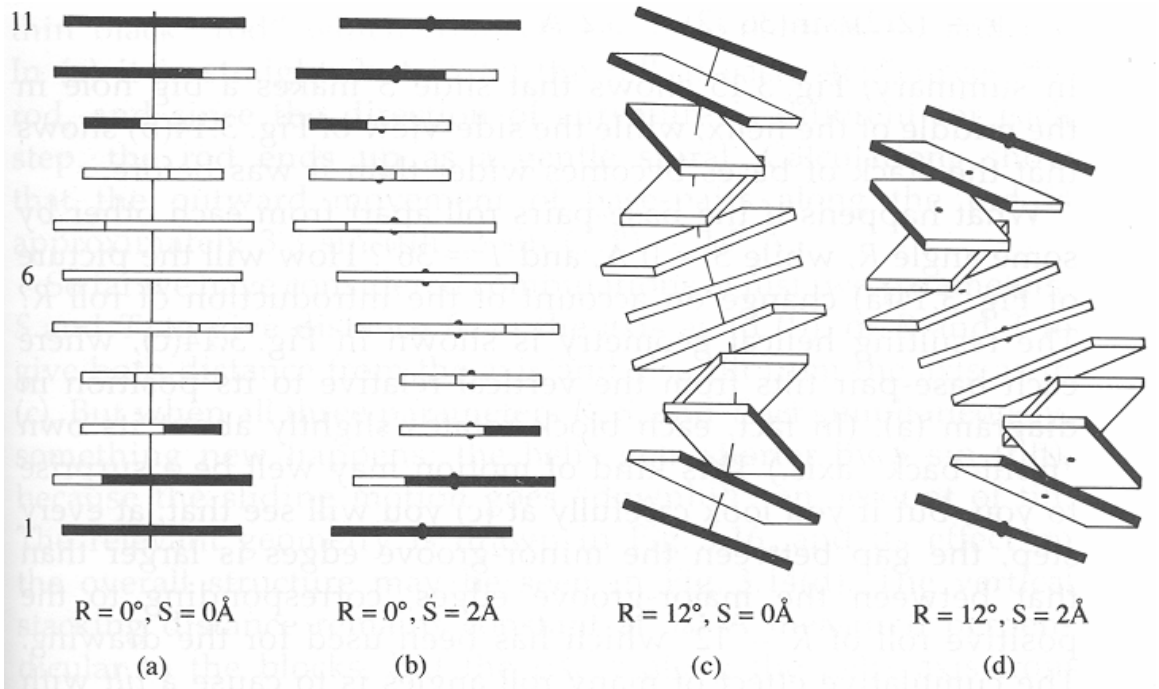


Figure 3.14 One complete helical turn of DNA having $T = 36°$, showing the effects of introducing uniform roll R or slide S at each step. Broadly, (a) corresponds to the "B" form of DNA, while (d) corresponds to the "A" form as shown in Fig. 2.7. Parts (b) and (c) correspond to structures intermediate between "B" and "A" which have, in fact, been seen recently in DNA crystals by X-ray diffraction.

(a) R = 0°, S = 0 Å and T = 36° >base pairs parallel to each other, exactly 10 steps needed to complete one helical turn
(b) R = 0°, S = +2 Å, T = 36° >the base pairs spiral outwards from a central helix axis (you can calculate the outward displacement of the base pairs), making a big hole in the center of the helix and a wider helix.
(c) R = 12°, S = 0 Å, T = 36° >causes a tilt with respect to the vertical axis, also the base pairs move out slightly form the axis
(d) R = 12°, S = +2 Å, T = 36° >helix gets shorter

(a) corresponds roughly to B-DNA, (d) to A-DNA. Clearly, the overall shape of a helix depends on the local parameters roll, slide and twist over a series of base pair steps.
Can we use this type of local information (R, S, T) to understand the geometry of DNA on a larger scale?