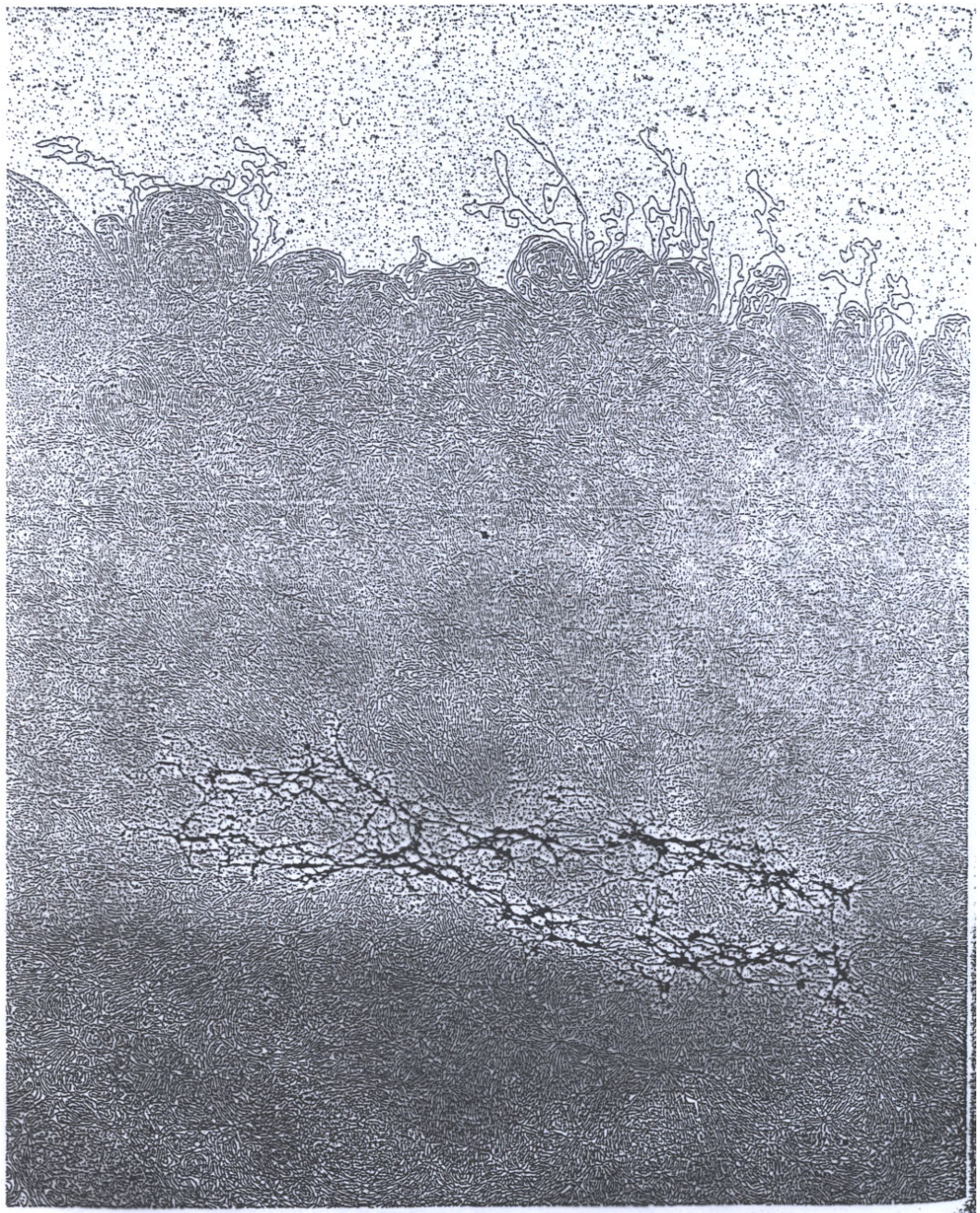
DNA



Electron micrograph of DNA spilling out of a single disrupted human chromosome. Only about half of the DNA in one chromosome is shown. (Courtesy of James Paulson and Ulrich Laemmli.)

Gene Expression - Transcription

- Genes are expressed as encoded proteins in a 2 step process: transcription + translation
- Central dogma of biology:

  DNA → RNA → protein
- Transcription: copy DNA strand making RNA (in nucleus)
  - Uses ~50 different transcription factors (proteins) and RNA polymerase
  - Locally opens DNA helix, assembles RNA strand following base pairing and re-closes DNA
  - Many different types of RNA are synthesized, including
    - Messenger (m- later translated into protein);
    - ribosomal (r- builds ribosomes used to synthesize proteins)(4types)
    - transfer (t- carries amino acids to growing protein) (32 types)
  - Three types of RNA polymerase – one for each of above RNA types
  - Most genes are split into segments (exons average 140 nucleotides code for amino acids in protein and introns, much longer – up to 500,000 nucleotides – code for a.a. not in protein
  - Introns must be cut and spliced out of mRNA

Translation Summary

- Translation involves the reading of mRNA nucleotide triplets (codons) to form tRNAs, specific for each amino acid, which then assemble these to form a polypeptide chain (the protein) - (this occurs in the cytoplasm)
- The specific steps of translation are
  - Initiation: mRNA is aligned on the ribosome and is read downstream (5' to 3') till the start codon AUG is found
  - Elongation: using the energy in GTP, coded tRNA-amino acid complexes are brought in and the amino acid is covalently attached via peptide bonds to the previous amino acid and the ribosome moves codon by codon along the mRNA repeating this process
  - Termination: when the ribosome reaches the stop codon (UAA< UAG or UGA), a protein release factor causes the polypeptide to be released from the ribosome and the ribosome splits into its subunits for later reassembly and use

Animation at http://www.johnkyrk.com/DNAtranslation.html

Post-translational Protein Modification

- As proteins are being synthesized in the cytoplasm, they are often (~80%) protected by chaperones to keep them from aggregating so they can fold properly or by chaperonins (~20%), hollow cylinders lined with hydrophobic amino acids into which the newly synthesized protein fits – both use ATP energy
- After some quality control checks, newly synthesized proteins can be altered by:
  - Co-valent attachment of sugar residues
  - Or phosphate groups
  - Or sulfate groups

Nucleic Acids

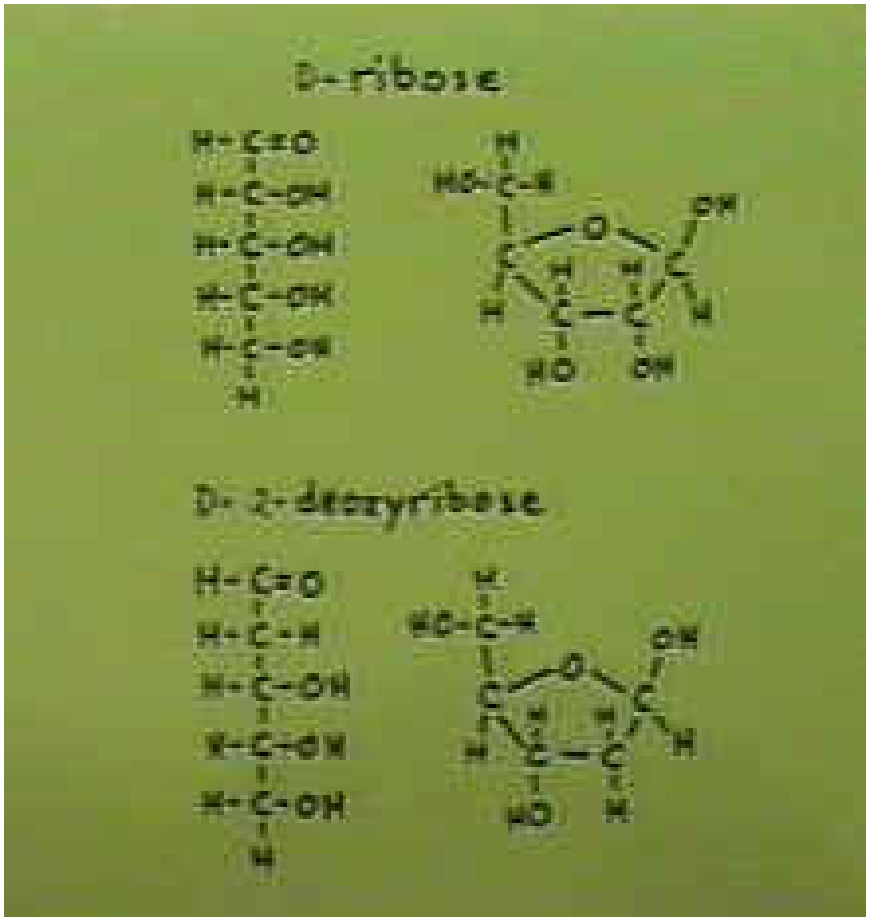
- DNA and RNA make up 5 – 15% of the dry weight of cells
- 4 different sub-units in DNA or RNA
- Each nucleotide = base + sugar + phosphoric acid

  sugar = D-ribose in RNA or 2-deoxy D-ribose in DNA

  base = 2 purines, A + G (with 2-rings) and 3 pyrimidines (C + T or U- in RNA) (with 1-ring)

  Base + sugar = nucleoside + phosphate group = nucleotide

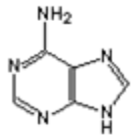
Nucleic Acid Structure



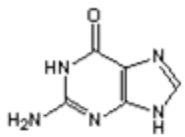
NUCLEIC ACID STRUCTURES

Amine Bases

Purines

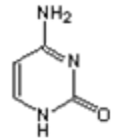


Adenine (A)

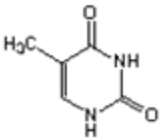


Guanine (G)

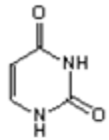
Pyrimidines



Cytosine(C)

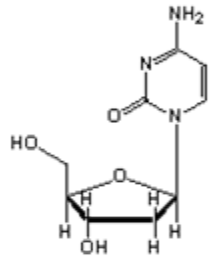


Thymine (T)

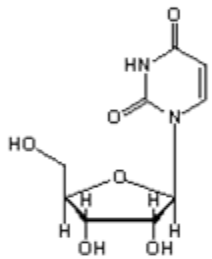


Uracil (U)

Nucleosides

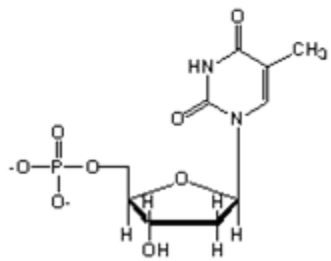


2'-Deoxycytidine

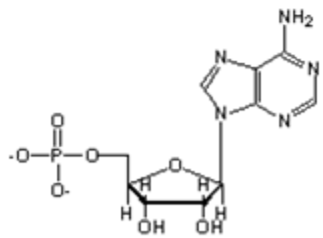


Uridine

Nucleotides



2'-Deoxythymidine 5'-monophosphate

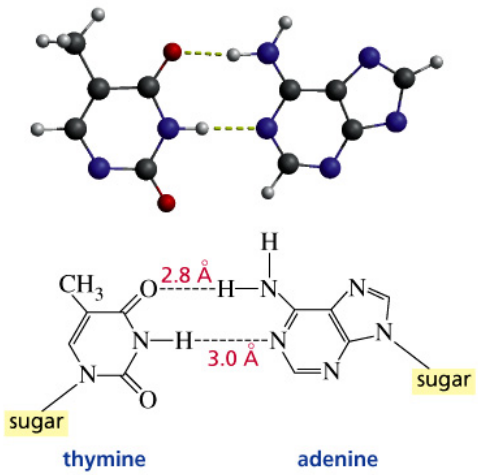


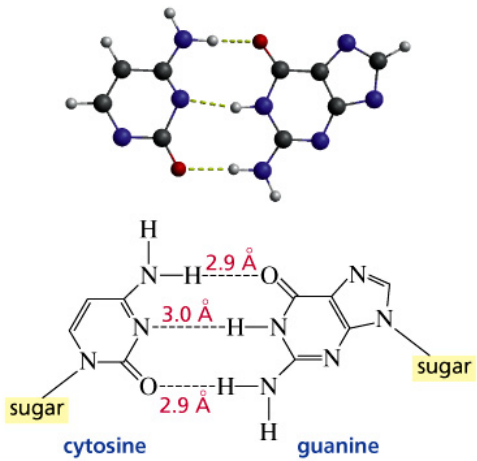
Adenosine 5'-monophsophate (5'-AMP)

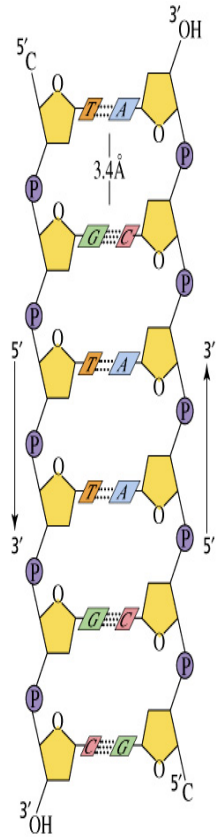
P. Kline 11/21/97

DNA Primary Structure

- Sequence of bases







Secondary Structure

- Watson-Crick helix is the standard configuration of DNA
- Base-base interactions stabilize the DNA
- Bases are planar – pyrimidines are exactly planar, purines are nearly so
- Principle of Base Pairing:

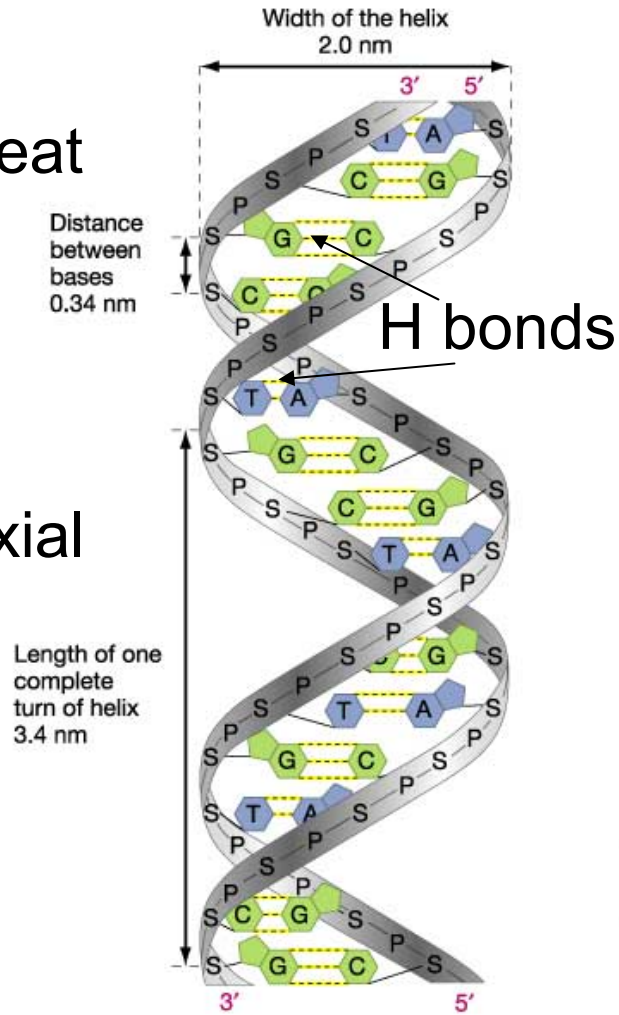
purine	pyrimidine
A	T (2 H bonds)
G	C (3 H bonds)

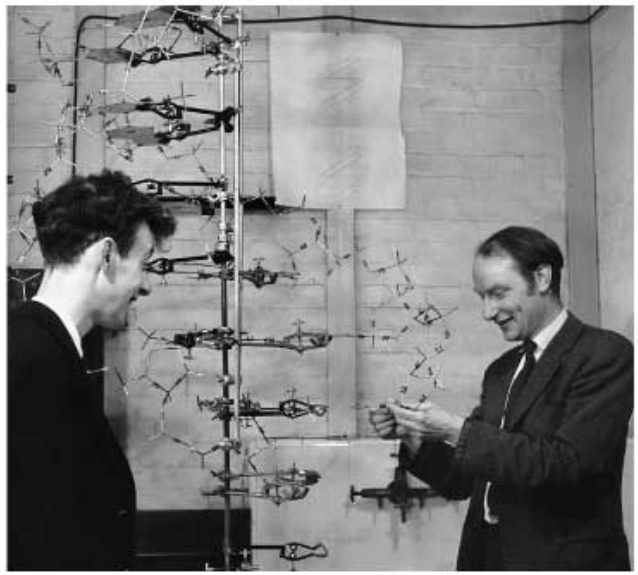
- Genetic info is all in either strand
- So, A+G = T+C = total bases in either strand and the ratio (A+T)/(G+C) is characteristic of a species – it does not depend on cell type or age

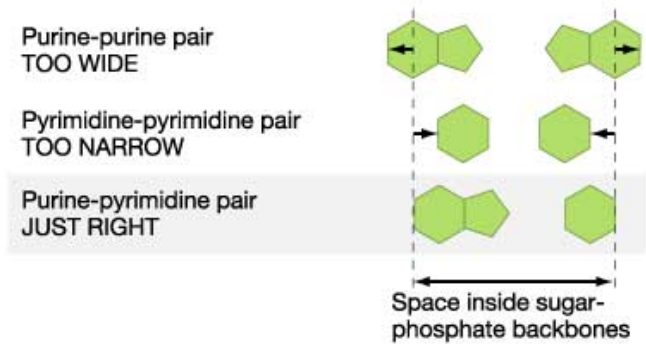
Watson-Crick (or B) DNA helix

Base repeat

Pitch = axial repeat

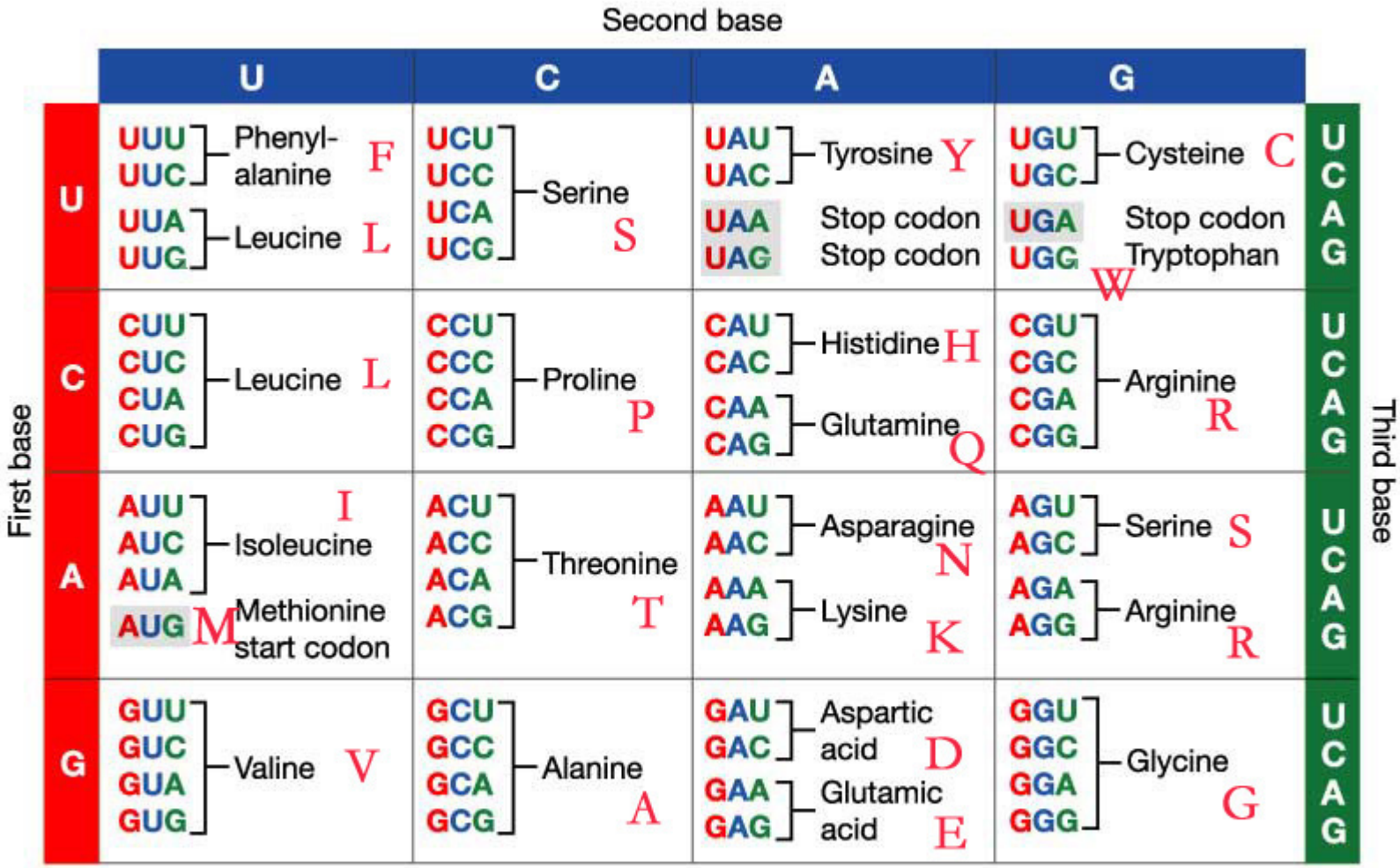






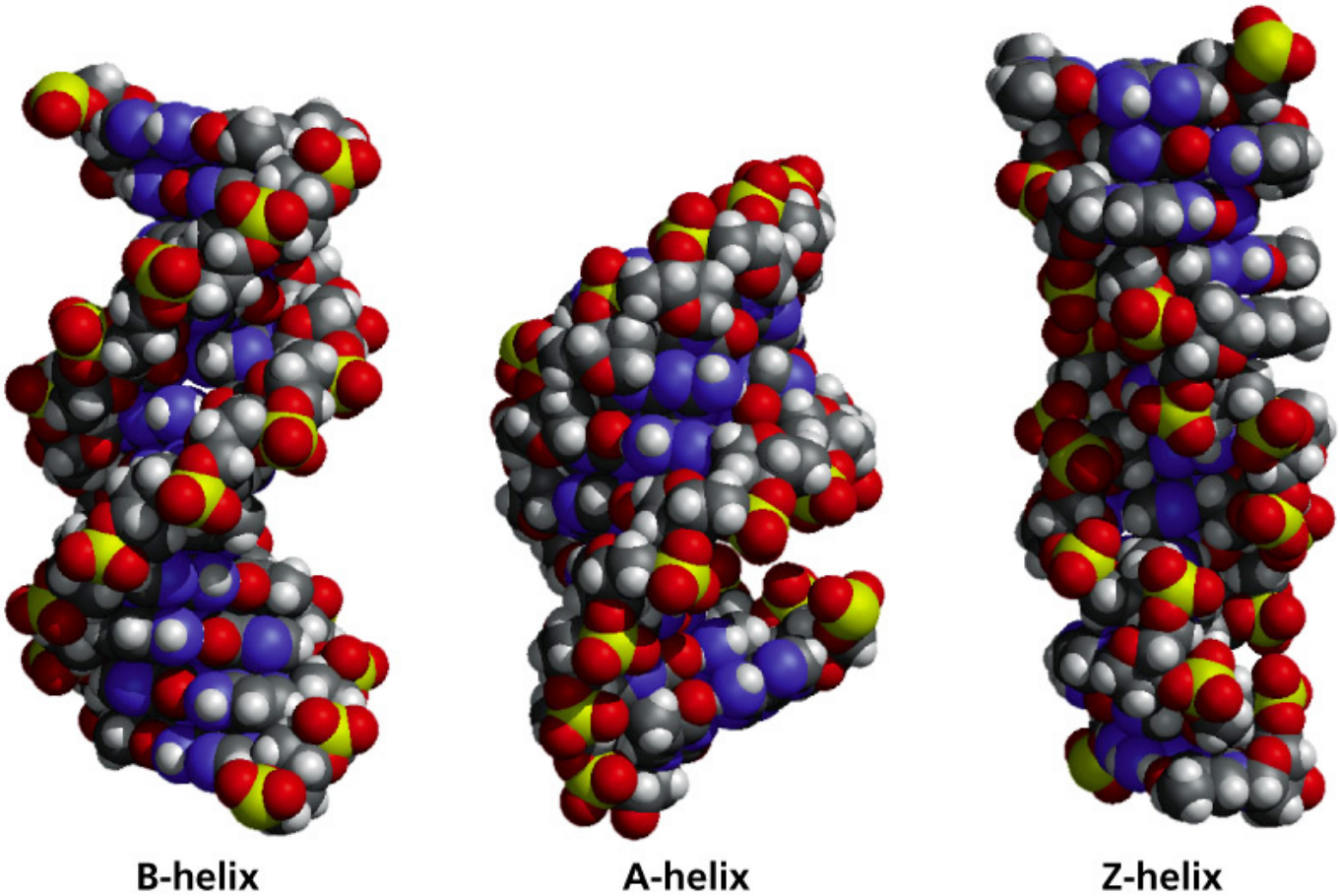
Genetic Code

- Codons = triplet of bases – $4^3 = 64$ (degenerate)



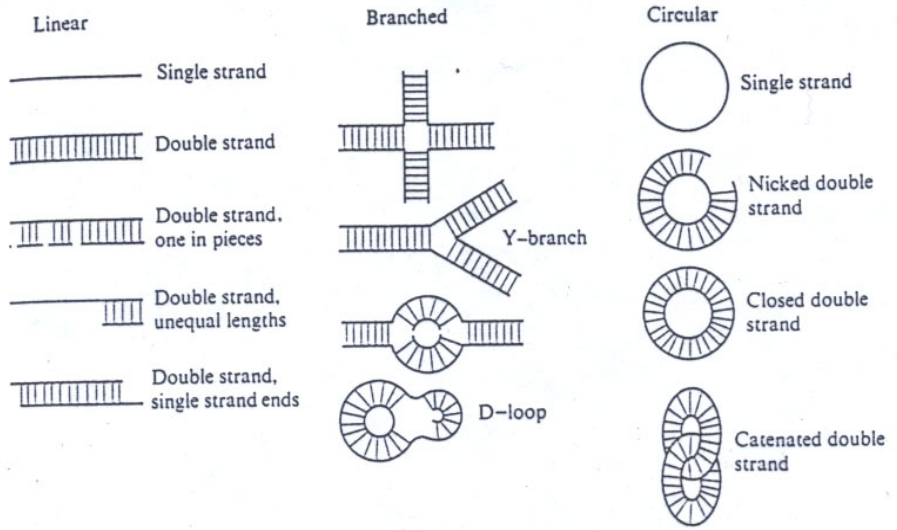
Other DNA Helices

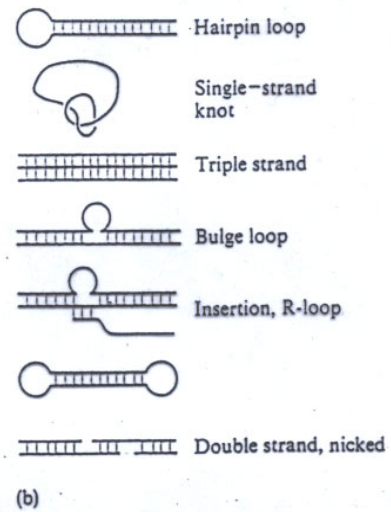
- Left-handed Z



Forms of DNA

- Sizes of Nucleic Acids
  - Shortest RNAs 75 – 80 residue t-RNAs
  - Largest RNAs 200,000 res.
  - Smallest DNA few 1000 res.
  - Largest DNA 10^8 res.





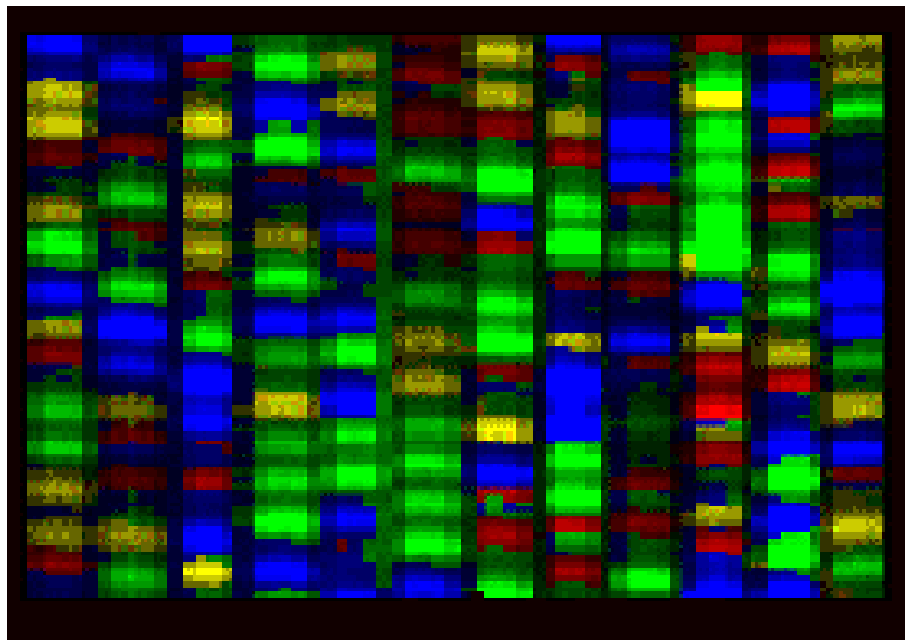
(b)

Figure 3-4

Some covalent nucleic acid chain structures. Regions with perfect base pairing are shown hatched. (a) Structures that have been observed in naturally occurring nucleic acids. (b) Structures that have been prepared synthetically.

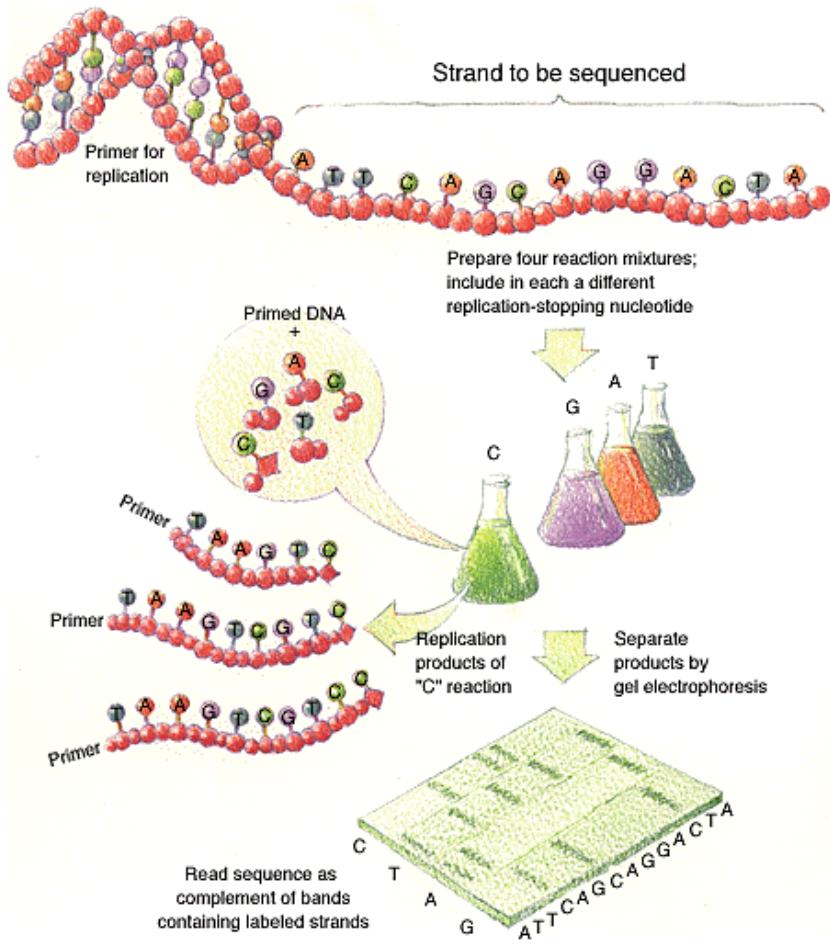
DNA sequencing

- Human Genome Project
  - Chromosomes (50 – 250 million bases) broken into short pieces
  - Fragments (500 bases) separated by gel electrophoresis



This image shows a portion of a fluorescence-based sequence gel. Each column of colored bars represents labeled DNA fragments which can be read as follows: blue = C, green = A, yellow = G, and red = T.

- In the much-automated Sanger sequencing method (based on Maxim & Gilbert ’77), the single-stranded DNA to be sequenced is "primed" for replication with a short complementary strand at one end.
- This preparation is then divided into four batches, and each is treated with a different replication-halting nucleotide (depicted here with a diamond shape), together with the four "usual" nucleotides. Each replication reaction then proceeds until a reaction-terminating nucleotide is incorporated into the growing strand, whereupon replication stops. Thus, the "C" reaction produces new strands that terminate at positions corresponding to the G's in the strand being sequenced. (Note that when long strands are being sequenced the concentration of the reaction-terminating nucleotide must be carefully chosen, so that a "normal" C is usually paired with a G; otherwise, replication would typically stop with the first or second G.)
- Gel electrophoresis -- one lane per reaction mixture -- is then used to separate the replication products, from which the sequence of the original single strand can be inferred.



Human Genome

- With 3.1647×10^9 base pairs, human DNA could code for 10^7 proteins (ave M ~ 300 aa)
- But, lots of nonesense coding (50%)+ stop codes – average gene = 3000 bp, largest = dystrophin = 2.4×10^6 bp
- Number of genes (coded proteins) = 30-35,000 (<2% of DNA codes for proteins)
- Only 3x more proteins than in the fly
- 99.9% of genes are identical for all people
- Over 50% of genes have unknown function

DNA structure in chromosomes

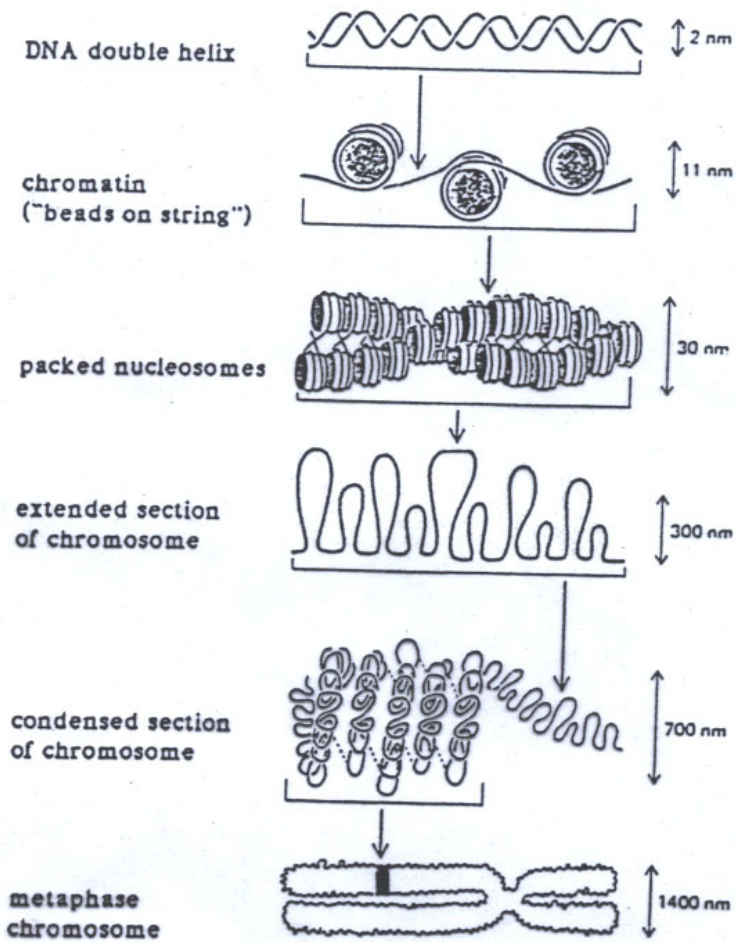


Figure 4.9 Packing of DNA into mammalian chromosomes. (Adapted from [2], cour
of Garland Publishing, Inc.)

Superhelical (Circular) DNA

or

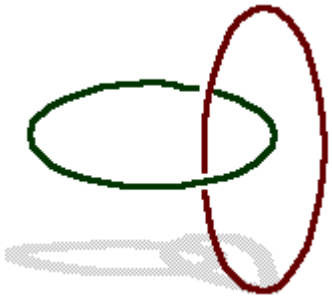
A Lesson in Topology

- Twist (Tw) of DNA refers to the number of turns of the double helix around its axis. So, since normally B-DNA has 10.5 bp/turn, linear B-DNA with N bp has a Tw = N/10.5
- If we twist the DNA we can decrease or increase the Tw, and unwind or overwind the DNA – to accommodate the twist, the free ends can simply rotate about one another
- If the DNA is constrained to form a closed circle then the structure can become more complex.

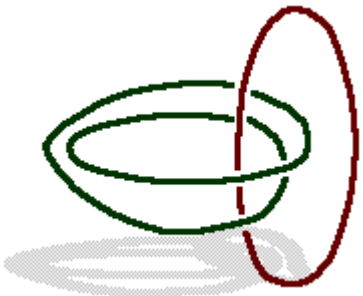
Writhe and Linking Number

- With circular DNA, a change in Tw will not be compensated by rotation of free ends and the (closed circular) ccDNA becomes strained – rubber band analogy – leads to writhing or supercoiling of the circle
- Wr refers to the coiling of the ds DNA axis
- Supercoils can be characterized by the superhelical density σ = Wr/Tw = # supercoils per turn of DNA
- Tw and Wr are related through the Linking Number (Lk) by Lk = Tw + Wr
- Lk is the number of times the two strands cross when confined to a plane
- Lk is fixed unless bonds are broken - topoisomerases

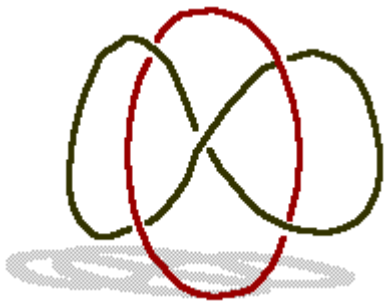
Linking Number



The linking number is 1.



- The linking number is 2.



- Is it hard to calculate the linking number?
- After changing the view of the problem shown in the last picture, Then you can get this picture.

Do you know the linking number now? The answer is 0.

Supercoiled DNA

- DNA is supercoiled to fit within cells
  - *E coli*: 4, 000 kbp, 1.36 cm vrs. Cell diameter of 10 μ
- Superhelical topology is described by Linking #
- L = T + W
  - T: Twists, number of times each strand crosses the other
  - W: Writhe, amount of supercoiling
    - negative supercoiling: left-handed turns
    - positive supercoiling : right-handed turns
- Superhelical density s = $\Delta L/L_0$
  - Typical s = -0.06

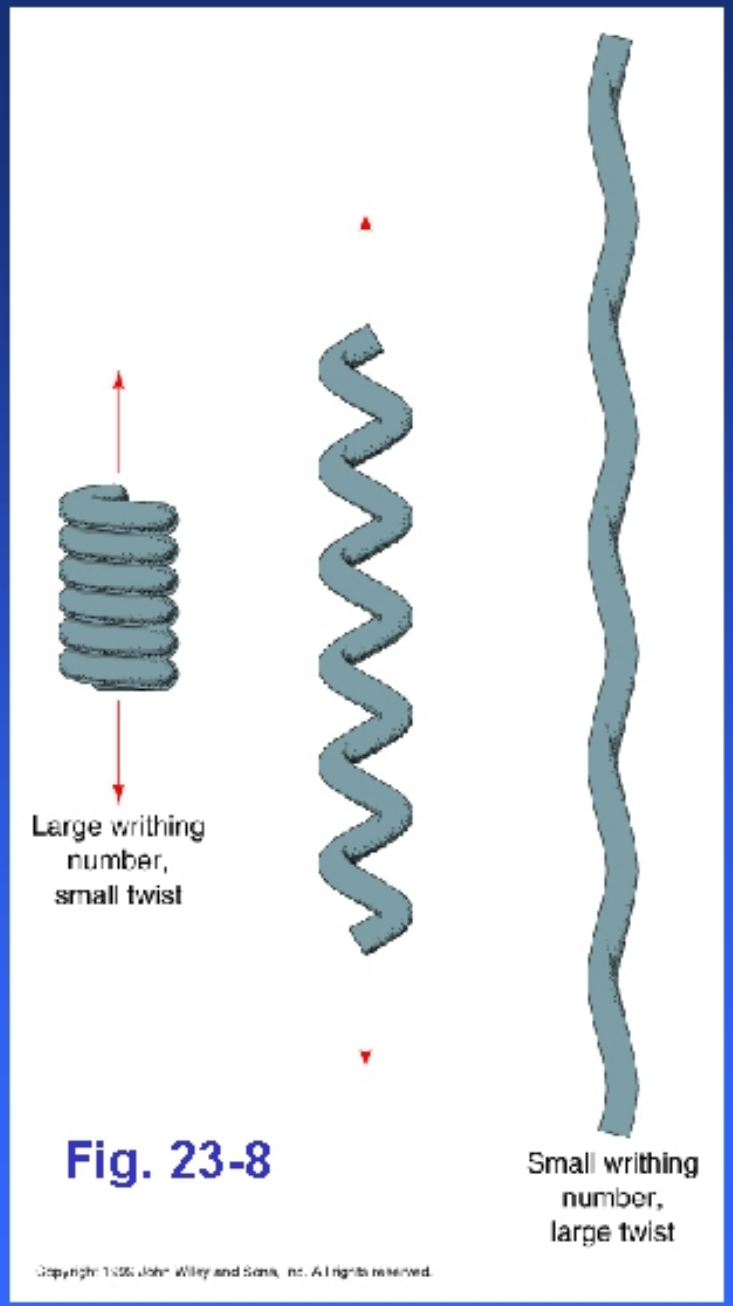


Fig. 23-8

Changes in Superhelix topology

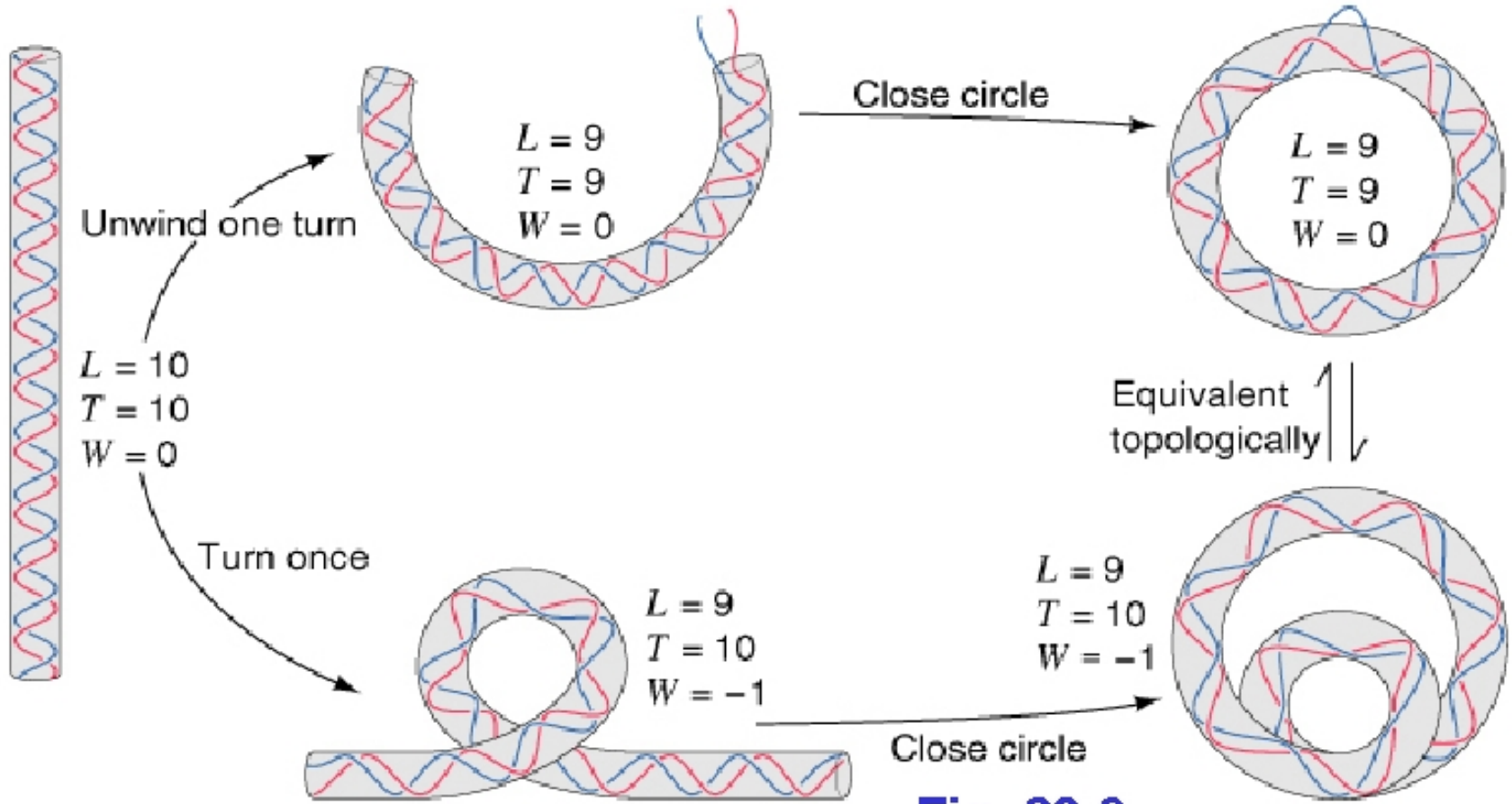
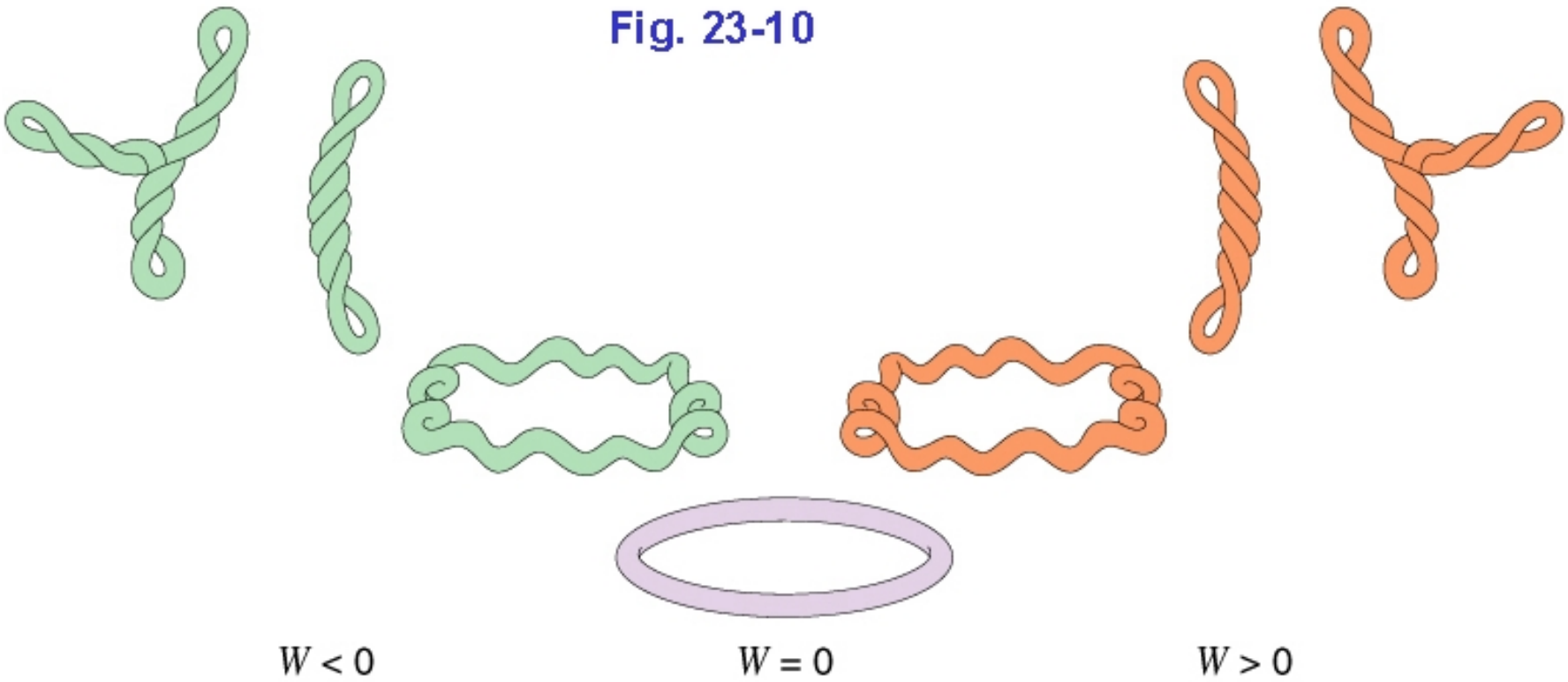


Fig. 23-9

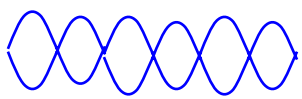
Progressive unwinding of negatively supercoiled DNA

Fig. 23-10



Example Study of Supercoiled DNA

- Example from my own work: Superhelical DNA (pBR322 plasmid DNA) = double stranded Watson-Crick DNA constrained to form a closed ring. Native supercoiled DNA is in dynamic equilibrium with energies from double helix, bending and twisting determining the tertiary conformation.
- Schematic:

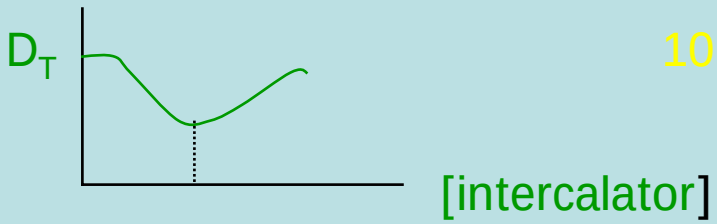


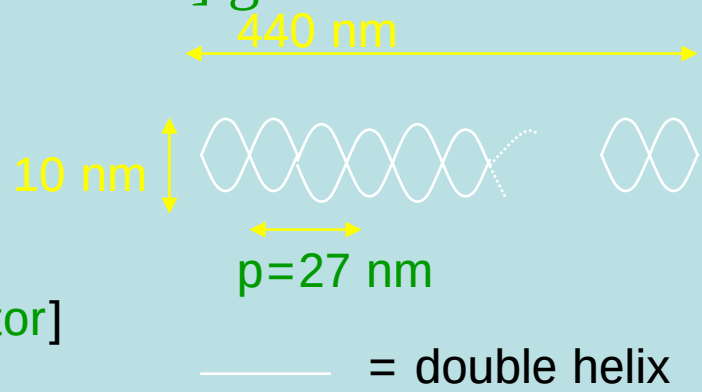
—— = double helix

Changing the Superhelical Density

- Addition of an intercalating agent (e.g. EtBr) causes the DNA to unwind (by a fixed θ/EtBr added)
- Two types of measurements made:
  - 1. q-dependence of $D_{apparent}$ at low concentrations
  - 2. titration measurements of D_T (low q) with intercalator
- From #1 we find best rod parameters: 10 x 440 nm rod

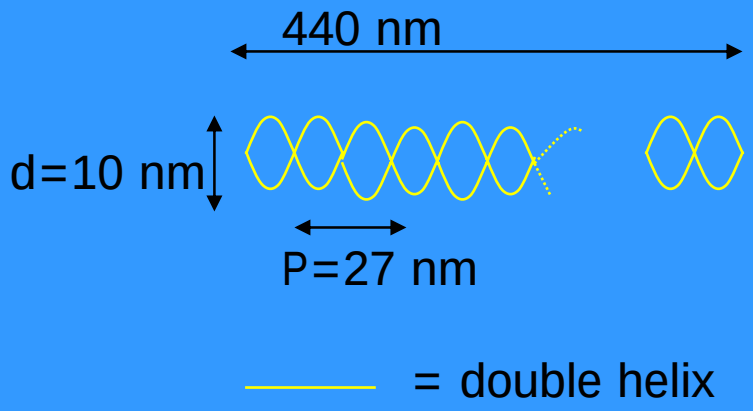
•Minimum in D_T vs [intercalator] gives number of superhelical turns





Conformational Changes III

SIMPLE PICTURE OF pBR322 in Solution



- Contour length of helix $C = \tau\sqrt{\pi^2 d^2 + p^2}$ = 1380 nm
- Independent value for C:

  C = Number base pairs x axial repeat distance =

  (4362) x (3.34A) = 1480 nm